Resilience to Malicious Activity in Distributed Optimization for Cyberphysical Systems

Michal Yemini, Angelia Nedić, Stephanie Gil and Andrea Goldsmith

***Abstract*— Enhancing resilience in distributed networks in the face of malicious agents is in its infancy and many of its key theoretical results and applications have not been explored. This work develops a new algorithmic and analytical framework for achieving resilience to malicious agents in distributed optimization problems where a legitimate agent's dynamic is influenced by the values it receives from neighboring agents and its own self-serving target function. We show that by utilizing stochastic values of trust between agents it is possible to recover convergence to the system's global optimal point even in the presence of malicious agents. Additionally, we provide expected convergence rate guarantees in the form of an upper bound on the expected distance to the optimal value. Finally, we present numerical results that validate the analytical convergence guarantees we present in this paper even when the malicious agents are the majority of agents in the network.**

I. Introduction

Distributed optimization is at the core of various multi-agent tasks including distributed control and estimation, multi-robot tasks such as mapping, and many learning tasks such as Federated Learning [1]–[3]. Owing to a long history and much attention in the research community, the theory for distributed optimization has matured, and several important results provide rigorous performance guarantees in the form of convergence and convergence rate for different function types, underlying graph topologies, and noise [4]–[8]. However, an understanding of how these results hold in the face of malicious activity is largely unclear.

With the growing prevalence of multi-agent and cyber-physical systems, and their reliance on distributed optimization methods for correct functioning in the real world, it becomes critical that the vulnerability of these methods is well understood. In particular, malicious agents can greatly interfere with the result of a distributed optimization scheme, driving the convergence to a non-optimal result or preventing convergence altogether by either not sharing key information or by manipulating key information such as the shared gradients critical for the correct functioning of the distributed optimization scheme [9]–[11]. Note that while stochastic optimization methods have long given treatment to the problem of noise for these systems [12], [13], a malicious agent has the ability to inject intentionally biased or manipulated information which can lead to greater potential damage for these systems. As a result, recent works have increasingly turned attention to the investigation of robust and resilient versions of distributed optimization methods in the face of malicious intent and/or severe (potentially biased) noise [9]–[11], [14], [15]. These approaches can be coarsely divided into two categories, those that use the transmitted data between nodes to infer the presence of anomalies (for example see [10], [16]), and those that exploit additional side information from the network or the physicality of the underlying cyberphysical system to provide additional channels of resilience [17]–[19].

We are interested in investigating the class of problems where the physicality of the system plays an important role in achieving new possibilities of resilience for these systems. Indeed the physicality of cyberphysical systems has been shown to provide many new channels of verification and establishing *inter-agent trust* through watermarking [20], wireless signal characteristics [19], [21], side information [22], and camera or lidar data cross-validation [23].

We capitalize on this observation which motivates us to focus on a class of problems where there is some extra information in the system that can be exploited. We abstract this information as a value α_{ij} that indicates the likelihood with which an agent i can trust data received from another agent j. We show that under mild assumptions, when this information is available, several powerful results for distributed optimization can be recovered such as 1) ***convergence to the true optimal point*** in the case of minimization over the sum of strongly convex functions, and 2) ***characterization of convergence rate*** that depends on the network topology, the amount of trust observations acquired, and the number of legitimate and malicious agents in the system.

A. Related Work

In the absence of malicious agents, the legitimate agents can construct iterates converging to an optimal point $x^\star_{\mathcal{L}}$ by using either their gradients, or sub-gradients when their objective functions are not differentiable. Each agent i updates its data value by considering the data values of its neighbors, and its self-serving gradient direction of its objective function f_i or the directions obtained from its neighbors. Convergence to an optimal point $x^\star_{\mathcal{L}}$ can be achieved for constrained multi-agent problem (1) in [4], [6], [24]–[30] and with limited gradient information [31]. Also, a zero-order method has been proposed in [32]. Some works, such as [4], assume that the weight matrices, which dictate how agents incorporate the data they receive from their neighbors, are doubly-stochastic.

Michal Yemini and Andrea J. Goldsmith are with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: myemini@princeton.edu; goldsmith@princeton.edu). Angelia Nedić is with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85281 USA (e-mail: Angelia.Nedich@asu.edu). Stephanie Gil is with the Computer Science Department, School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02139 USA (e-mail: sgil@seas.harvard.edu).

However, works such as [27] overcome this assumption by performing additional weighted averaging steps. The convergence rate of the method (2) is at best $O\left(\frac{1}{T}\right)$ where T is the algorithm running time, see [29].

To harm the system, a malicious agent can send falsified data to their legitimate neighbors. If the legitimate agents are unaware of their malicious neighbors, then the malicious agents will succeed in controlling the system [9], [10], [33]–[35]. To combat the harmful effect of an attack, the approach taken in [33], [35] requires the existence of a set of *connected* legitimate agents (trusting each other) such that all other agents are connected to at least one trusted agent, which is unrealistic especially for robotic and ad-hoc networks with sporadic communications. The approaches in [9], [10], [34], [36] rely on the agent data values to detect and discard the malicious inputs and have an upper bound on the number of tolerable malicious agents (with a star network having the largest number - a half of the number of agents in the network) [34]. When the number of malicious agents exceeds the tolerable number, the attack succeeds and malicious agents evades detection. In contrast with the existing works, our proposed method provides a significantly stronger resilience to malicious activity by exploiting the physical aspect of the problem, i.e., the wireless medium. Thus, each legitimate agent can learn trustworthy neighbors while optimizing the system objective.

B. Paper Organization

The reset of this paper is organized as follows: Section II presents the system model and problem formulation. Section III presents our learning mechanism for detecting malicious agents. Section IV proposes our algorithm for resilient distributed optimization, and Section V provides analytical guarantees for its convergence. Finally, Section VI presents numerical results that to validate our analytical results, and Section VII concludes the paper.

II. Problem Formulation

We consider a multi-agent system of n agents communicating over a network, which is represented by an undirected graph, $\mathbb{G} = (\mathbb{V}, \mathbb{E})$. The node set $\mathbb{V} = \{1, \ldots, n\}$ represents the agents and the edge set $\mathbb{E} \subset \mathbb{V} \times \mathbb{V}$ represents the set of communication links, with $\{i, j\} \in \mathbb{E}$ indicating that agents i and j are connected. We study the case where an unknown subset of the agents is malicious and the trustworthy agents are learning which neighbors they can trust. Thus, $\mathbb{V} = \mathcal{L} \cup \mathcal{M}$ where $\mathcal{L}$ is the set of legitimate agents that execute computational tasks and share their data truthfully, and $\mathcal{M}$ denotes the set of agents that are not truthful. *The sets $\mathcal{L}$ and $\mathcal{M}$ are just modeling artifacts, and none of the legitimate agents knows if it has malicious neighbors or not, at any time.* Throughout the paper, we will use the subscripts $\mathcal{L}$ and $\mathcal{M}$ to denote the various quantities related to legitimate and malicious agents, respectively.

We are interested in a general distributed optimization problem, where the legitimate agents aim at optimizing a common objective whereas the malicious agents try to impair the legitimate agent by malicious injections of harmful data. The aim of the legitimate agents is to minimize distributively the sum of their objective functions in the constraint set $\mathcal{X} \subset \mathbb{R}^d$, i.e.:

$$x^\star_{\mathcal{L}} = \arg\min_{x \in \mathcal{X}} f_{\mathcal{L}}(x), \text{ with } f_{\mathcal{L}}(x) = \frac{1}{|\mathcal{L}|} \sum_{i \in \mathcal{L}} f_i(x). \quad (1)$$

By choosing a local update rule and exchanging some information with their neighbors, the legitimate agents want to determine an optimal solution $x^\star$ to Eq. (1). In contrast, the malicious agents aim to either lead the legitimate agents to a common non-optimal value x such that $f_{\mathcal{L}}(x) > f_{\mathcal{L}}(x^\star)$, or prevent the convergence of an optimization method employed by the legitimate agents.

A. Notation

We denote by $\|x\| \triangleq \sqrt{x^T x}$ the ℓ_2 vector norm. We let $\Pi_{\mathcal{X}}(x)$ be the projection of the point x onto the set $\mathcal{X}$, i.e.,

$$\Pi_{\mathcal{X}}(x) = \min_{y \in \mathcal{X}} \|y - x\|.$$

Finally, we denote by $\mathbf{E}(\cdot)$ the expectation operator.

B. The update rule of agents

We propose distributed methods with a significantly stronger resilience in comparison with [10], whereby this is enabled by each legitimate agent learning which neighbors it can trust while optimizing a system objective.

The update rule of legitimate agents. Each legitimate agent i updates $x_i(t)$ by considering the values $x_j(t)$ of its neighbors and *the gradient of its own objective function f_i* similarly to the iterates described in [4][1]. When malicious agents are present, this method takes the following form: for every legitimate agent i,

$$\begin{aligned} c_i(t) &= w_{ii}(t) x_i(t) + \sum_{j \in \mathcal{N}_i} w_{ij}(t) x_j(t), \\ y_i(t) &= c_i(t) - \gamma(t) \nabla f_i(c_i(t)), \\ x_i(t+1) &= \Pi_{\mathcal{X}}\left(y_i(t)\right), \end{aligned} \quad (2)$$

where $\gamma_i(t) > 0$ is the stepsize of agent i at time t and $\mathcal{N}_i$ is the set of neighbors of agent i in the communication graph. The set $\mathcal{N}_i$ is composed of both legitimate and malicious agents. Finally, the weights $w_{ij}(t), j \in \mathcal{N}_i \cup \{i\}$, are nonnegative and sum to 1.

The update rule for the malicious agents. Malicious agents $i \in \mathcal{M}$ choose values arbitrarily in the set $\mathcal{X}$. We assume that their actions are not known, and thus we do not model them.

C. Trust values

We employ a probabilistic framework of trustworthiness where we assume the availability of stochastic *observations of trust* between communicating agents. This information is abstracted in the form of a random variable α_{ij} defined below.

[1]Note, however, that [4] does not include projection on the set $\mathcal{X}$.

Definition II.1 (α_{ij})**.** For every $i \in \mathcal{L}$ and $j \in \mathcal{N}_i$, the random variable $\alpha_{ij} \in [0,1]$ represents the probability that agent j is a trustworthy neighbor of agent i. We assume the availability of such observations $\alpha_{ij}(t)$ throughout the paper.

This model of inter-agent trust observations has been used extensively in prior works [21], [37]. The focus of the current work is not the derivation of the α_{ij} values themselves, but rather on the derivation of a theoretical framework for achieving resilient distributed optimization using this model. Indeed, we show that much stronger results of convergence are achievable by properly exploiting this information in the network. We refer to [21] for an example of such an α_{ij} value. Intuitively, a random realization $\alpha_{ij}(t)$ of α_{ij} contains useful trust information regarding the legitimacy of a transmission.

We assume that a value of $\alpha_{ij}(t) > 0.5$ indicates a legitimate transmission and $\alpha_{ij}(t) < 0.5$ indicates a malicious transmission in a stochastic sense (misclassifications are possible). Note that $\alpha_{ij}(t) = 0.5$ means that the observation is completely ambiguous and contains no useful trust information for the transmission at time t.

We use the following assumptions throughout the paper:

Assumption 1. *(i)* [Sufficiently connected graph] *The subgraph $\mathbb{G}_{\mathcal{L}}$ induced by the legitimate agents is connected.*
(ii) [Homogeneity of trust variables] *There are scalars $E_{\mathcal{L}} > 0$ and $E_{\mathcal{M}} < 0$ such that*

$$E_{\mathcal{L}} \triangleq \mathbf{E}(\alpha_{ij}(t)) - 0.5 \qquad \text{for all } i \in \mathcal{L},\ j \in \mathcal{N}_i \cap \mathcal{M},$$

$$E_{\mathcal{M}} \triangleq \mathbf{E}(\alpha_{ij}(t)) - 0.5 \qquad \text{for all } i \in \mathcal{L},\ j \in \mathcal{N}_i \cap \mathcal{L}.$$

(iii) [Independence of trust observations] *The observations $\alpha_{ij}(t)$ are independent for all t and all pairs of agents i and j, with $i \in \mathcal{L}$, $j \in \mathcal{N}_i$. Moreover, for any $i \in \mathcal{L}$ and $j \in \mathcal{N}_i$, the observation sequence $\{\alpha_{ij}(t)\}$ is identically distributed. We note that these are standard assumptions when using the probabilistic trust framework employed here [21], [37].*

D. Assumptions on the objective functions and initial points

Assumption 2. *We assume that $\mathcal{X} \subset \mathbb{R}^d$ is a compact and convex and that there exists a known value $\eta > 0$ such that*

$$\|x\| \leq \eta, \quad \forall x \in \mathcal{X}, \tag{3}$$

We note that this η value is arbitrary and simply bounds the malicious agents inputs away from infinity.

Assumption 3. *For all legitimate agents $i \in \mathcal{L}$, the function f_i is μ-strongly convex and has L-Lipschitz continuous gradients, i.e., $\|\nabla f_i(x) - \nabla f_i(y)\| \leq L\|x-y\|$, for all $x \in \mathbb{R}^d$.*

Corollary 1. *When $\mathcal{X}$ is compact, Assumption 3 implies that there is a scalar G such that $\|\nabla f_i(x)\| \leq G, \forall x \in \mathcal{X}, i \in \mathcal{L}$.*

Assumption 4. *Let $\gamma(t) \geq 0$ be monotonically nonincreasing. We assume that $\sum_{t=1}^{\infty} \gamma(t) = \infty$ and $\sum_{t=1}^{\infty} \gamma^2(t) < \infty$.*

E. Objectives

The objective of this work is to arrive at strong convergence results for the distributed optimization problem in Eq. (1) in the presence of malicious agents $\mathcal{M}$. We wish to achieve this by carefully exploiting the availability of trust values $\alpha_{ij}(t)$ in the network. Specifically we aim to achieve the following:

(1) **Objective 1** - We wish to construct a sequence $\{w_{ij}(t)\}_{t=1,\ldots}$ of weights to be used in the iterates (2) for weighting the influence of neighboring nodes in computing each legitimate agent's update. Specifically, we wish to construct this sequence such that it converges to *the nominal weights* $\overline{w}_{ij}$ almost surely (a.s.), where $\overline{w}_{ij} = 0$ for all malicious edges where $i \in \mathcal{L}$ and $j \in \mathcal{M}$.
(2) **Objective 2** - Utilizing the proposed weights $\{w_{ij}(t)\}_{t=1,\ldots}$, we aim to show that the iterates given by (2) converge (in some sense) to the true optimal point $x^{\star}_{\mathcal{L}} \in \mathcal{X}$ under Assumptions 1-4.
(3) **Objective 3** - We aim to establish an upper bound on the expected value of $\|x_i(t) - x^{\star}_{\mathcal{L}}\|^2$, for all $i \in \mathcal{L}$, as a function of the time t, for the iterates $x_i(t)$ produced by the method.

III. Learning the sets of trusted neighbors

In this section we establish a few key characteristics of the stochastic observations $\alpha_{ij}(t)$ that result from the model described in Section II-C and that we will subsequently use in our analysis of the convergence of the iterates in Eq. (2). We consider the sum over a history of $\alpha_{ij}(t)$ values that we denote by $\beta_{ij}(t)$:

$$\beta_{ij}(t) = \sum_{k=0}^{t-1} (\alpha_{ij}(k) - 0.5) \text{ for } t \geq 1, i \in \mathcal{L}, j \in \mathcal{N}_i, \tag{4}$$

and define $\beta_{ij}(0) = 0$. Intuitively, following the discussion on α_{ij}'s immediately after Definition II.1, the values $\beta_{ij}(t)$ will tend towards positive values for legitimate agent transmissions $i \in \mathcal{L}$ and $j \in \mathcal{N}_i \cap \mathcal{L}$, and will tend towards negative values for malicious agent transmissions where $i \in \mathcal{L}$ and $j \in \mathcal{N}_i \cap \mathcal{M}$. We restate an important result shown in [37] regarding the exponential decay rate of misclassifications given a sum over the history of stochastic observation values that we will use extensively in the forthcoming analysis.

Lemma 1 (Lemma 2 [37])**.** *Consider the random variables $\beta_{ij}(t)$ as defined in Eq.* (4)*. Then, for every $t \geq 0$ and every $i \in \mathcal{L}$, $j \in \mathcal{N}_i \cap \mathcal{L}$,*

$$\Pr(\beta_{ij}(t) < 0) \leq \max\{\exp(-2tE^2_{\mathcal{L}}), \mathbb{1}_{\{E_{\mathcal{L}} < 0\}}\}.$$

Additionally, for every $t \geq 0$ and every $i \in \mathcal{L}, j \in \mathcal{N}_i \cap \mathcal{M}$,

$$\Pr(\beta_{ij}(t) \geq 0) \leq \max\{\exp(-2tE^2_{\mathcal{M}}), \mathbb{1}_{\{E_{\mathcal{M}} > 0\}}\}.$$

In other words, the probability of misclassifying malicious agents as legitimate, or vice versa, decays exponentially in the accrued number of observations t.

Corollary 2. *There exists a random finite time T_f such that*

$$\beta_{ij}(t) \geq 0 \text{ for all } t \geq T_f \text{ and all} i \in \mathcal{L}, j \in \mathcal{N}_i \cap \mathcal{L},$$
$$\beta_{ij}(t) < 0 \text{ for all } t \geq T_f \text{ and all} i \in \mathcal{L}, j \in \mathcal{N}_i \cap \mathcal{M}, \tag{5}$$

and there exists $i \in \mathcal{L}$ such that

$$\beta_{ij}(T_f - 1) < 0 \text{ for some } j \in \mathcal{N}_i \cap \mathcal{L}, \text{ or}$$
$$\beta_{ij}(T_f - 1) \geq 0 \text{ for some } j \in \mathcal{N}_i \cap \mathcal{M}. \quad (6)$$

Proof. It follows directly from [37, Proposition 1]. □

Define

$$D_{\mathcal{L}} \triangleq \sum_{i \in \mathcal{L}} |\mathcal{N}_i \cap \mathcal{L}| \quad \text{and} \quad D_{\mathcal{M}} \triangleq \sum_{i \in \mathcal{L}} |\mathcal{N}_i \cap \mathcal{M}|.$$

Additionally, we define upper bounds on the misclassification probabilities as

$$p_c(k) \triangleq \mathbb{1}_{\{k \geq 0\}} \left[D_{\mathcal{L}} e^{-2kE_{\mathcal{L}}^2} + D_{\mathcal{M}} e^{-2kE_{\mathcal{M}}^2} \right], \text{ and}$$
$$p_e(k) \triangleq D_{\mathcal{L}} \frac{\exp(-2kE_{\mathcal{L}}^2)}{1 - \exp(-2E_{\mathcal{L}}^2)} + D_{\mathcal{M}} \frac{\exp(-2kE_{\mathcal{M}}^2)}{1 - \exp(-2E_{\mathcal{M}}^2)}.$$

Using these quantities, we obtain some useful bounds on the probabilities of the events $(T_f = k)$ and $(T_f > k - 1)$ for any k, as follows.

Lemma 2. *For every $k \geq 0$*

$$\Pr(T_f = k) \leq \min\{p_c(k-1), 1\}, \quad (7)$$
$$\Pr(T_f > k - 1) \leq \min\{p_e(k-1), 1\}. \quad (8)$$

We present the proof of Lemma 2 in Appendix I.

IV. The Algorithm

This section presents Algorithm 1 which incorporates the agent's learning of inter-agent trust values into the dynamic (2) through the choice of the time-dependent weights $w_{ij}(t)$. These weights depend on a parameter T_0 that captures the number of trust measurements a legitimate agent collects before trusting one of its neighbors.

A. The weight matrix sequence

We define a time dependent *trusted neighborhood* for agent $i \in \mathcal{L}$ as:

$$\mathcal{N}_i(t) \triangleq \{j \in \mathcal{N}_i : \beta_{ij}(t) \geq 0\}. \quad (9)$$

This is the subset of neighbors that legitimate agent i classifies as its legitimate neighbors at time t. For all $t \geq 0$, let

$$d_i(t) \triangleq \max\{|\mathcal{N}_i(t)| + 1\} \geq 2 \qquad \text{for all } i \in \mathcal{L}.$$

At each time t, *every agent i sends the value $d_i(t)$ to its neighbors $j \in \mathcal{N}_i$ in addition to the value $x_i(t)$*. Alternatively, we can assume that agent i sends $d_i(t)$ to its neighbors only when the value $d_i(t)$ changes.

Legitimate agents are the most susceptible to make classification errors regarding the trustworthiness of their neighbors when they have a small sample size of trust value observations. Thus, we delay the updating of legitimate agents' values until time $T_0 \geq 0$. Up to that time the legitimate agents only collect observations of trust values.

Algorithm 1 The protocol of agent $i \in \mathcal{L}$.

Inputs: T, T_0, $\mathcal{N}_i$, $x_i(0)$, $\nabla f_i(\cdot)$, $\gamma(\cdot)$
Outputs: $x_i(T)$.
Set $\beta_{ij}(t) = 0$ for all $j \in \mathcal{N}_i$
for $t = 0, \ldots, T-1$ **do**
  Set $\mathcal{N}_i(t) = \{j \in \mathcal{N}_i : \beta_{ij}(t) \geq 0\}$
  Set $d_i(t) = \max\{|\mathcal{N}_i(t)| + 1\}$
  Send $x_i(t)$ and $d_i(t)$ to neighbors
  for $j \in \mathcal{N}_i$ **do**
    Receive $x_j(t)$ and $d_j(t)$
    Extract $\alpha_{ij}(t)$
    Set $\beta_{ij}(t+1) = \sum_{k=0}^{t} (\alpha_{ij}(k) - 0.5)$
    Set $w_{ij}(t) = \frac{\mathbb{1}_{\{t \geq T_0\}} \mathbb{1}_{\{\beta_{ij}(t) \geq 0\}}}{2 \max\{d_i(t), d_j(t)\}}$
  end for
  Set $w_{ii}(t) = 1 - \sum_{m \in \mathcal{N}_i} w_{im}(t)$
  Set $x_i(t+1)$ according to the dynamic (11)
end for

We define the weight matrix $W(t)$ by choosing its entries $w_{ij}(t)$ as follows: for every $i \in \mathcal{L}$, $j \in \mathcal{N}_i$,

$$w_{ij}(t) = \begin{cases} \frac{\mathbb{1}_{\{t \geq T_0\}}}{2 \cdot \max\{d_i(t), d_j(t)\}} & \text{if } j \in \mathcal{N}_i(t), \\ 0 & \text{if } j \notin \mathcal{N}_i(t) \cup \{i\}, \\ 1 - \sum_{m \in \mathcal{N}_i} w_{im}(t) & \text{if } j = i. \end{cases} \quad (10)$$

Using the weights (10) and letting the stepsize $\gamma(-k) = 0$, $\forall k \in \mathbb{N}_+$, the dynamic in Eq. (2) is equivalent to the following dynamic where agents *only consider the data values received from their trusted neighbors at time t, i.e., $\mathcal{N}_i(t)$,* when computing their own value updates:

$$c_i(t) = w_{ii}(t) x_i(t) + \sum_{j \in \mathcal{N}_i(t) \cap \mathcal{L}} w_{ij}(t) x_j(t) + \sum_{j \in \mathcal{N}_i(t) \cap \mathcal{M}} w_{ij}(t) x_j(t),$$
$$y_i(t) = c_i(t) - \gamma(t - T_0) \nabla f_i(c_i(t)),$$
$$x_i(t+1) = \Pi_{\mathcal{X}}\left(y_i(t)\right). \quad (11)$$

The dependence of the weights $w_{ij}(t)$ on the trust observation history $\beta_{ij}(t)$ comes in through the choice of time-dependent and random trusted neighborhood $\mathcal{N}_i(t)$ (cf. Eqn. 9). Consequently, some entries of the matrix $W(t)$ are also random, as seen from Eq. (10). The gradients $\nabla f_i(x_i(t))$ are stochastic due to the randomness of $x_i(t)$, however, they are not unbiased as typically assumed in stochastic approximation methods, including [38], thus we cannot readily rely on prior analysis for stochastic approximation methods. However, as we show in our subsequent analysis, the variance of $\|\nabla f_i(x_i(t))\|$ decays sufficiently fast and allows convergence to the optimal point even in the presence of malicious agents.

V. Analytical Results

This section analyzes the convergence characteristics of Algorithm 1. Our main result states that by incorporating trust values through $\beta_{ij}(t)$ for all $i \in \mathcal{L}$ and $j \in \mathcal{N}_i$, Algorithm 1 *converges (in the mean-squared) sense to the*

optimal value of the optimization problem in Eq. (1) even in the presence of malicious agents.

The two forthcoming auxiliary lemmas help us establish our main result given in Theorem 1.

Let us denote $d_{i,\mathcal{L}} \triangleq |\mathcal{N}_i \cap \mathcal{L}| + 1$. Next, we define the doubly stochastic matrix $\overline{W}_{\mathcal{L}} \in [0,1]^{|\mathcal{L}|\times|\mathcal{L}|}$ with the entries $[\overline{W}_{\mathcal{L}}]_{i,j}$, for every $i, j \in \mathcal{L}$:

$$[\overline{W}_{\mathcal{L}}]_{i,j} = \begin{cases} \frac{1}{2\cdot\max\{d_{i,\mathcal{L}}, d_{j,\mathcal{L}}\}} & \text{if}\, , j \in \mathcal{N}_i, \\ 0 & \text{if}\, , j \notin \mathcal{N}_i(t) \cup \{i\}, \\ 1 - \sum\limits_{m\in\mathcal{N}_i} w_{im}(t) & \text{if}\, j = i. \end{cases} \quad (12)$$

Note that $\overline{W}_{\mathcal{L}}$ is the *nominal* weight matrix, or what the weight matrix would be in the absence of malicious agents. Let $\sigma_2(A)$ be the second largest singular value of A, and denote $\rho_{\mathcal{L}} = \max_{k\geq 1} \sigma_2\left(\overline{W}_{\mathcal{L}}^{k}\right)$. Since $\mathbb{G}$ is connected and $\overline{W}_{\mathcal{L}}$ is doubly stochastic, $\rho_{\mathcal{L}} < 1$ is equal to the second largest eigenvalue modulus of $\overline{W}_{\mathcal{L}}$. By [39, Lemma 2.2] $\rho_{\mathcal{L}}$ can be upper bounded by $(1 - 1/(71|\mathcal{L}|^2))$, [40] improves the constant of this bound to 4.

Lemma 3. *Let $r \in \{1, 2\}$, $i \in \mathcal{L}$, and $t \geq 0$. Then,*

$$\mathbf{E}\left[\left(\sum_{j\in\mathcal{N}_i\cap\mathcal{L}} |w_{ij}(k) - \overline{w}_{ij}|\right)^r\right] \leq p_c(k),$$

$$\mathbf{E}\left[\left(\sum_{j\in\mathcal{N}_i\cap\mathcal{M}} w_{ij}(k)\right)^r\right] \leq \frac{p_c(k)}{2^r},$$

$$\mathbf{E}\left[|w_{ii}(k) - \overline{w}_{ii}|^r\right] \leq \frac{p_c(k)}{2^r}.$$

Proof. First, note that, $w_{ii}(t) \geq 0.5$ for every $i \in \mathcal{L}$, thus

$$|w_{ii}(t) - \overline{w}_{ii}| \leq 0.5, \text{ and } \sum_{j\in\mathcal{N}_i\cap\mathcal{M}} w_{ij}(k) \leq 0.5.$$

Additionally, $\sum_{j\in\mathcal{N}_i\cap\mathcal{L}} \|w_{ij}(k) - \overline{w}_{ij}\| > 0$ only if $W_{\mathcal{L}}(t) \neq \overline{W}_{\mathcal{L}}$. It follows by the triangle inequality that

$$\sum_{j\in\mathcal{N}_i\cap\mathcal{L}} |w_{ij}(k) - \overline{w}_{ij}| \leq \sum_{j\in\mathcal{N}_i\cap\mathcal{L}} [w_{ij}(k) + \overline{w}_{ij}] \leq 1.$$

The event $w_{ii}(t) - \overline{w}_{ii} \neq 0$ depends on $d_j(t)$, $j \in \mathcal{N}_i(t)$. Now, for all i and $j \in \mathcal{N}_i \cap \mathcal{L}$, we have

$$\mathbf{E}\left[\left(\sum_{j\in\mathcal{N}_i\cap\mathcal{L}} |w_{ij}(k) - \overline{w}_{ij}|\right)^r\right] \leq \mathbf{E}\left[1 \cdot \mathbb{1}_{\{W_{\mathcal{L}}(t)\neq\overline{W}_{\mathcal{L}}\}}\right] \leq p_c(k).$$

The rest of the proof follows similarly. □

Next, we present an auxiliary lemma to upper bound the expected distance between an agents' value and the average agents' values at time t.

Denote, $t/2 \triangleq \lfloor \frac{t}{2} \rfloor$,

$$\overline{x}_{\mathcal{L}}(t) \triangleq \frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}} x_i(t),$$

and

$$\delta_{\mathcal{M}}(t, T_0) \triangleq 2\eta\rho_{\mathcal{L}}^{t-T_0} + \frac{(2\eta\sqrt{p_c(T_0)} + G\gamma(0))\rho_{\mathcal{L}}^{(t-T_0)/2}}{1-\rho_{\mathcal{L}}} + \frac{2(\eta\sqrt{p_c((t+T_0)/2)} + G\gamma((t-T_0)/2))}{1-\rho_{\mathcal{L}}}. \quad (13)$$

Lemma 4. *For every $t \geq 0$*

$$\frac{1}{|\mathcal{L}|}\sum_{j\in\mathcal{L}} \mathbf{E}\|x_i(t) - \overline{x}_{\mathcal{L}}(t)\| \leq \delta_{\mathcal{M}}(t, T_0), \text{ and}$$

$$\frac{1}{|\mathcal{L}|}\sum_{j\in\mathcal{L}} \mathbf{E}\|x_i(t) - \overline{x}_{\mathcal{L}}(t)\|^2 \leq \widetilde{g}^2(t, T_0).$$

Proof. Denote $\phi_i(t) \triangleq \Pi_{\mathcal{X}}[y_i(t)] - c_i(t)$. The matrices $W_{\mathcal{L}}(t) \triangleq (w_{ij}(t))_{i,j\in\mathcal{L}}$ are random, vary with time, and can be sub-stochastic. Thus we cannot naively apply Lemma 5 in Appendix II for $\Delta_i(t) = \phi_i(t)$. Instead, we substitute

$$\Delta_i(t) = c_i(t) - \overline{c}_i(t) + \phi_i(t)$$

where $\overline{c}_i(t) \triangleq \overline{w}_{ii}x_i(t) + \sum_{j\in\mathcal{N}_i\cap\mathcal{L}} \overline{w}_{ij}x_j(t)$.
Thus, we have that

$$x_i(t+1) = \overline{w}_{ii}x_i(t) + \sum_{j\in\mathcal{N}_i\cap\mathcal{L}} \overline{w}_{ij}x_j(t) + \Delta_i(t).$$

By the Cauchy–Schwarz inequality for the ℓ_2 inner product, and by the Cauchy–Schwarz inequality *for expectations*

$$\begin{aligned}\mathbf{E}\left[\|\Delta_i(t)\|^2\right] &= \mathbf{E}\left[\|c_i(t) - \overline{c}_i(t) + \phi(t)\|^2\right] \\ &\leq \mathbf{E}\left[\|c_i(t) - \overline{c}_i(t)\|^2\right] + \mathbf{E}\left[\|\phi_i(t)\|^2\right] \\ &\quad + 2\sqrt{\mathbf{E}\left[\|c_i(t) - \overline{c}_i(t)\|^2\right]}\sqrt{\mathbf{E}\left[\|\phi_i(t)\|^2\right]}.\end{aligned}$$

Additionally, by Lemma 3 and Assumption 2,

$$\begin{aligned}&\mathbf{E}\left[\|c_i(t) - \overline{c}_i(t)\|^2\right] \\ &= \mathbf{E}\Bigg[\|[w_{ii}(t) - \overline{w}_{ii}]x_i(t) + \sum_{j\in\mathcal{N}_i\cap\mathcal{L}} [w_{ij}(t) - \overline{w}_{ij}]x_j(t) \\ &\qquad + \sum_{j\in\mathcal{N}_i\cap\mathcal{M}} w_{ij}(t)x_j(t)\|^2\Bigg] \\ &\leq 4\eta^2 p_c(t).\end{aligned}$$

Additionally, by Assumption 3 and the non-expansiveness property of the projection, since $c_i(t) \in \mathcal{X}$ for all i and all t, it follows that

$$\|\phi_i(t)\| \leq \|y_i(t) - c_i(t)\| \leq G\gamma(t - T_0).$$

Hence,

$$\begin{aligned}&\mathbf{E}\left[\|\Delta_i(t)\|^2\right] \\ &= \mathbf{E}\left[\|c_i(t) - \overline{c}_i(t) + \phi_i(t)\|^2\right] \\ &\leq 4\eta^2 p_c(t) + G^2\gamma^2(t - T_0) + 4G\eta\sqrt{p_c(t)}\,\gamma(t - T_0) \\ &= \left(2\eta\sqrt{p_c(t)} + G\gamma(t - T_0)\right)^2. \quad (14)\end{aligned}$$

We substitute $\delta(t) = \widetilde{\delta}(t + T_0) = 2\eta\sqrt{p_c(t + T_0)} + G\gamma(t)$,

$W(t) = \overline{W}_{\mathcal{L}}$, and $\rho = \rho_{\mathcal{L}}$ in Lemma 5 and use the transformation $t \to t - T_0$ to conclude the proof. □

Before presenting the main result of this paper we denote the special function $\overline{h}(T)$ which has the form

$$\overline{h}(T) \triangleq \frac{G^2 T}{\mu} + \frac{2G^2 T}{\mu(1-\rho_{\mathcal{L}})} + \frac{8(\mu+L)G^2}{\mu^2(1-\rho_{\mathcal{L}})^2} \ln\left(\frac{T+2}{2}\right) + \frac{2\eta G}{1-\rho_{\mathcal{L}}} + \frac{2(\mu+L)(\mu\eta+2G)^2}{\mu^2(1-\rho_{\mathcal{L}})^2} + \frac{2G^2+4G\eta(\mu+L)}{\mu(1-\rho_{\mathcal{L}})^3} + \frac{G^2(\mu+L)}{\mu^2(1-\rho_{\mathcal{L}})^4},$$

Note that this function grows linearly in T and is comprised of the first term which captures error rate for the centralized gradient descend optimization (see [41]) without malicious agents, and the following terms that include $\rho_{\mathcal{L}}$, capture the contribution from distributing the optimization over a decentralized network (without malicious agents) that is characterized by the second largest eigenvalue modulus of $\overline{W}_{\mathcal{L}}$.

Theorem 1. *Algorithm 1 converges to the optimal point $x^*_{\mathcal{L}}$ in the mean-squared sense for every collection $x_i(0) \in \mathcal{X}$, $i \in \mathcal{L}$, of initial points i.e.,*

$$\lim_{t\to\infty} \mathbf{E}\left[\|x_i(t) - x^\star_{\mathcal{L}}\|^2\right] = 0, \ \forall\, i \in \mathcal{L}, \tag{15}$$

whenever $\sum_{t=0}^{\infty} \gamma(t) = \infty$ and $\sum_{t=0}^{\infty} \gamma^2(t) < \infty$.

Moreover, let $\gamma(t) = \frac{2}{\mu(t+2)}$. Then, for every $T_0 \in \mathbb{N}$ and $T \geq T_0$ there exists a function $C_{\mathcal{M}}(T_0)$ that decreases exponentially with T_0 and is independent of T such that for any collection $x_i(0) \in \mathcal{X}$, $i \in \mathcal{L}$, and for all $T \geq T_0$,

$$\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}} \mathbf{E}\left[\|x_i(T) - x^\star_{\mathcal{L}}\|^2\right] \leq \min\left\{4\eta^2, \frac{4\overline{h}(T-T_0) + \mu C_{\mathcal{M}}(T_0)}{\mu(T-T_0)(T-T_0+1)}\right\}. \tag{16}$$

Intuitively, the $C_{\mathcal{M}}(T_0)$ term above represents the error term contributed by the presence of malicious agents in the distributed network. It can be seen that for large enough T the entire term on the right of the inequality (16) decays on the order of $O\left(\frac{1}{T}\right)$.

Before proceeding to prove this theorem, we point out that unlike the analysis for stochastic gradient models such as [38], in our model $w_{ij}(t)$ and $x_j(t)$ are correlated. This follows by the statistical dependence of $w_{ij}(t)$ and $w_{ij}(t-1)$. Thus, we cannot use the standard analysis which requires that $\mathbf{E}[w_{ij}(t)x_j(t)] = \mathbf{E}[w_{ij}(t)]\mathbf{E}[x_j(t)]$. Finally, we observe that the nonnegativity of the variance of random variables (15) and the sandwich theorem imply that $\lim_{t\to\infty} \mathbf{E}[\|x_i(t) - x^\star_{\mathcal{L}}\|] = 0$, $\forall\, i \in \mathcal{L}$. This result also holds since convergence in expectation in the rth moment implies convergence in expectation in the sth moment whenever $0 < s < r$.

Proof. By the non-expansiveness property of projections, for every $t \geq T_0$ and $i \in \mathcal{L}$,

$$\|x_i(t+1) - x^\star_{\mathcal{L}}\|^2 = \|\Pi_{\mathcal{X}}\left(y_i(t)\right) - x^\star_{\mathcal{L}}\|^2 \leq \|y_i(t) - x^\star_{\mathcal{L}}\|^2.$$

Denote

$$\overline{c}_i(t) \triangleq \overline{w}_{ii}x_i(t) + \sum_{j\in\mathcal{N}_i\cap\mathcal{L}} \overline{w}_{ij}x_j(t), \text{ and}$$
$$\overline{g}_i(t) \triangleq \overline{c}_i(t) - \gamma(t-T_0)\nabla f_i(\overline{c}_i(t)) - x^\star_{\mathcal{L}}.$$

Recall that $y_i(t) = c_i(t) - \gamma(t-T_0)\nabla f_i(c_i(t))$, then by the Cauchy-Schwarz inequality for the inner product on ℓ_2

$$\begin{aligned}
&\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}} \mathbf{E}\left[\|x_i(t+1) - x^\star_{\mathcal{L}}\|^2\right] \\
&= \frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}} \mathbf{E}\Big[\|\overline{c}_i(t) - \gamma(t-T_0)\nabla f_i(\overline{c}_i(t)) - x^\star_{\mathcal{L}} \\
&\quad + c_i(t) - \overline{c}_i(t) + \gamma(t-T_0)[\nabla f_i(\overline{c}_i(t)) - \nabla f_i(c_i(t))]\|^2\Big] \\
&\leq \underbrace{\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}} \mathbf{E}\left[\|\overline{g}_i(t)\|^2\right]}_{\text{(I)}} + \underbrace{\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}} \mathbf{E}\left[\|c_i(t) - \overline{c}_i(t)\|^2\right]}_{\text{(II)}} \\
&+ \underbrace{\frac{\gamma^2(t)}{|\mathcal{L}|}\sum_{i\in\mathcal{L}} \mathbf{E}\left[\|\nabla f_i(\overline{c}_i(t)) - \nabla f_i(c_i(t))\|^2\right]}_{\text{(III)}} \\
&+ \underbrace{\frac{2}{|\mathcal{L}|}\sum_{i\in\mathcal{L}} \mathbf{E}\left[\|\overline{g}_i(t)\| \cdot \|c_i(t) - \overline{c}_i(t)\|\right]}_{\text{(IV)}} \\
&+ \underbrace{\frac{2\gamma(t-T_0)}{|\mathcal{L}|}\sum_{i\in\mathcal{L}} \mathbf{E}\left[\|\overline{g}_i(t)\| \cdot \|\nabla f_i(\overline{c}_i(t)) - \nabla f_i(c_i(t))\|\right]}_{\text{(V)}} \\
&+ \underbrace{\frac{2\gamma(t-T_0)}{|\mathcal{L}|}\sum_{i\in\mathcal{L}} \mathbf{E}\Big[\|c_i(t) - \overline{c}_i(t)\| \cdot \|\nabla f_i(\overline{c}_i(t)) - \nabla f_i(c_i(t))\|\Big]}_{\text{(VI)}}.
\end{aligned}$$

By (3), Lemma 3, and the Cauchy-Schwarz inequality, it follows that

$$\begin{aligned}
\text{(II)} &= \mathbf{E}\Big[\|[w_{ii}(t) - \overline{w}_{ii}]x_i(t) + \sum_{j\in\mathcal{N}_i\cap\mathcal{L}} [w_{ij}(t) - \overline{w}_{ij}]x_j(t) \\
&\qquad + \sum_{j\in\mathcal{N}_i\cap\mathcal{M}} w_{ij}(t)x_j(t)\|^2\Big] \\
&\leq 4\eta^2 p_c(t),
\end{aligned} \tag{17}$$

Since ∇f_i are L-Lipschitz continuous

$$\text{(III)} \leq L^2\mathbf{E}\left[\|\overline{c}_i(t) - c_i(t)\|^2\right] \leq 4L^2\eta^2 p_c(t). \tag{18}$$

Now, by Lemma 3, $\mathbf{E}\left[\|c_i(t) - \overline{c}_i(t)\|\right] \leq 2\eta p_c(t)$. Thus,

$$\text{(IV)} \leq (2\eta + \gamma(t-T_0)G)\mathbf{E}\left[\|\ \overline{c}_i(t) - c_i(t)\|\right]$$

$$\leq 2\eta\left(2\eta+\gamma(t-T_0)G\right)p_c(t), \tag{19}$$

and by the L-Lipschitz continuity of ∇f_i

$$\text{(V)} \leq 2L\eta\left(2\eta+\gamma(t-T_0)G\right)p_c(t),$$
$$\text{(VI)} \leq L\mathbf{E}\left[\|c_i(t)-\overline{c}_i(t)\|^2\right] \leq 4L\eta^2 p_c(t). \tag{20}$$

Define $h_{\mathcal{M}}(t,T_0)$ as

$$4\eta^2 p_c(t)\left[2(L+1)+\gamma^2(t-T_0)L^2+\frac{\gamma(t-T_0)G(L+1)}{2\eta}\right],$$

and recall that $\overline{x}_{\mathcal{L}}(t)\triangleq\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}}x_i(t)$. Therefore,

$$\begin{aligned}
&\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}}\mathbf{E}\left[\|x_i(t+1)-x^\star_{\mathcal{L}}\|^2\right]\\
&\leq \text{(I)}+h_{\mathcal{M}}(t,T_0)\\
&\leq h_{\mathcal{M}}(t,T_0)+\frac{(1-\mu\gamma(t-T_0))}{|\mathcal{L}|}\sum_{i\in\mathcal{L}}\mathbf{E}\left[\|x_i(t)-x^\star_{\mathcal{L}}\|^2\right]\\
&+\gamma^2(t-T_0)G^2+\frac{2\gamma(t-T_0)}{|\mathcal{L}|}\sum_{i\in\mathcal{L}}\Bigg[G\mathbf{E}\left[\|x_i(t)-\overline{x}_{\mathcal{L}}(t)\|\right]\\
&\qquad+\frac{\mu+L}{2}\mathbf{E}\left[\|x_i(t)-\overline{x}_{\mathcal{L}}(t)\|^2\right]\Bigg],
\end{aligned}$$

where the last inequality follows from Assumption 3, the convexity of $\|\cdot\|^2$ and the double stochasticity of $\overline{W}_{\mathcal{L}}$.

Recall (13) and denote

$$\widetilde{h}_{\mathcal{M}}(t,T_0)\triangleq\gamma(t-T_0)G^2+2G\delta_{\mathcal{M}}(t,T_0)+(\mu+L)\delta^2_{\mathcal{M}}(t,T_0).$$

Here, the term $\widetilde{h}_{\mathcal{M}}(t,T_0)$ is affected by the distributed nature of our optimization process and the presence of the malicious agents. We utilize Lemma 4 to conclude that

$$\begin{aligned}
&\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}}\mathbf{E}\left[\|x_i(t+1)-x^\star_{\mathcal{L}}\|^2\right]\leq\gamma(t-T_0)\widetilde{h}_{\mathcal{M}}(t,T_0)\\
&+h_{\mathcal{M}}(t,T_0)+\frac{(1-\mu\gamma(t-T_0))}{|\mathcal{L}|}\sum_{i\in\mathcal{L}}\mathbf{E}\left[\|x_i(t)-x^\star_{\mathcal{L}}\|^2\right].
\end{aligned}$$

Thus, since $|\mathcal{L}|<\infty$

$$\lim_{t\to\infty}\mathbf{E}\left[\|x_i(t)-x^\star_{\mathcal{L}}\|^2\right]=0,\ \forall\, i\in\mathcal{L}, \tag{21}$$

whenever $\sum_{t=0}^{\infty}\gamma(t)=\infty$ and $\sum_{t=0}^{\infty}\gamma^2(t)<\infty$.

To prove the second part of the theorem, we let $\gamma(t)=\frac{2}{\mu(t+2)}$ as proposed in [41]. It follows that

$$\begin{aligned}
&\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}}\mathbf{E}\left[\|x_i(t+1)-x^\star_{\mathcal{L}}\|^2\right]\leq\frac{2\widetilde{h}_{\mathcal{M}}(t,T_0)}{\mu(t-T_0+2)}\\
&+h_{\mathcal{M}}(t,T_0)+\frac{t}{t-T_0+2}\cdot\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}}\mathbf{E}\left[\|x_i(t)-x^\star_{\mathcal{L}}\|^2\right].
\end{aligned} \tag{22}$$

Multiplying both sides by $(t-T_0+1)(t-T_0+2)$ and summing over the set $t\in\{T_0,T_0+1,\ldots,T-1\}$ yield the upper bound

$$\begin{aligned}
\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}}\mathbf{E}[\|x_i(T)-x^\star_{\mathcal{L}}\|^2]&\leq\frac{2\sum_{t=T_0}^{T-1}(t-T_0+1)\widetilde{h}_{\mathcal{M}}(t,T_0)}{\mu(T-T_0)(T-T_0+1)}\\
&+\frac{\sum_{t=T_0}^{T-1}(t-T_0+1)(t-T_0+2)h_{\mathcal{M}}(t,T_0)}{(T-T_0)(T-T_0+1)}.
\end{aligned}$$

Using the identity $\sqrt{a+b}\leq\sqrt{a}+\sqrt{b}$, we deduce that

$$\sqrt{p_c(t)}\leq\sqrt{D_{\mathcal{L}}}e^{-tE^2_{\mathcal{L}}}+\sqrt{D_{\mathcal{M}}}e^{-tE^2_{\mathcal{M}}}. \tag{23}$$

In addition, we utilize the identities for $|v|\in(0,1)$

$$\sum_{t=0}^{\infty}tv^t=\frac{v}{(1-v)^2}\ \text{and}\ \sum_{t=0}^{\infty}t^2v^t=\frac{2v}{(1-v)^3}.$$

Denote

$$\begin{aligned}
\widetilde{C}_1(T_0,E,D)&\triangleq\frac{16\eta e^{-T_0E^2}\sqrt{D}}{1-\rho_{\mathcal{L}}}\Bigg[\frac{G}{(1-e^{-E^2})^2}+\\
&\frac{G+(\eta+\frac{4}{\mu})(\mu+L)}{(1-\rho_{\mathcal{L}})^2}+\frac{(\mu+L)(G+2\mu\sqrt{D}e^{-T_0E^2})}{\mu(1-\rho_{\mathcal{L}})^3}\Bigg],
\end{aligned}$$

and

$$C_1(T_0)\triangleq\widetilde{C}_1(T_0,E_{\mathcal{L}},D_{\mathcal{L}})+\widetilde{C}_1(T_0,E_{\mathcal{M}},D_{\mathcal{M}}).$$

Further algebra yields that

$$\sum_{t=T_0}^{T-1}(t-T_0+1)\widetilde{h}_{\mathcal{M}}(t,T_0)\leq 2\overline{h}(t-T_0)+C_1(T_0),$$

here, the added term $C_1(T_0)$ captures the influence of the malicious agents on the term (I). Additionally, denote

$$\begin{aligned}
\widetilde{C}_2(T_0,E,D)&\triangleq\frac{4\eta(L+1)De^{-2T_0E^2}}{\left(1-e^{-2E^2}\right)}\cdot\\
&\left[\frac{4\eta e^{-2E^2}}{\left(1-e^{-2E^2}\right)^2}+\frac{(6\eta+\frac{G}{\mu})e^{-2E^2}}{1-e^{-2E^2}}+4\eta+\frac{4\eta L}{\mu^2}+\frac{G}{\mu}\right],
\end{aligned}$$

and

$$C_2(T_0)\triangleq\widetilde{C}_2(T_0,E_{\mathcal{L}},D_{\mathcal{L}})+\widetilde{C}_2(T_0,E_{\mathcal{M}},D_{\mathcal{M}}).$$

Then,

$$\begin{aligned}
&\sum_{t=T_0}^{T-1}(t-T_0+1)(t-T_0+2)h_{\mathcal{M}}(t,T_0)\\
&=8\eta^2(L+1)\sum_{t=0}^{T-T_0-1}(t+1)(t+2)p_c(t+T_0)\\
&+4\eta^2L^2\sum_{t=0}^{T-T_0-1}(t+1)(t+2)\gamma^2(t)p_c(t+T_0)\\
&+2\eta G(L+1)\sum_{t=0}^{T-T_0-1}(t+1)(t+2)\gamma(t)p_c(t+T_0)\\
&\leq C_2(T_0).
\end{aligned}$$

Letting $C_{\mathcal{M}}(T_0)=C_1(T_0)+C_2(T_0)$ concludes the proof. □

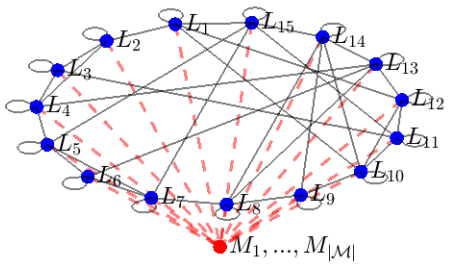

Fig. 1. Undirected graph $\mathbb{G}$. Two agents are neighbors if they are connected by an edge. Legitimate and malicious agents are depicted by blue and red nodes, respectively. Edges between legitimate agents are depicted by black solid lines. Edges between legitimate and malicious agents are depicted by red dashed lines.

Thus we have shown that indeed we are able to recover convergence *to the optimal value* of the original distributed optimization problem given in Eq. (1) even in the presence of malicious agents, and further, we have established an upper bound on the expected value of $\|x_i(t) - x^\star_{\mathcal{L}}\|^2$, for all $i \in \mathcal{L}$, as a function of the time t as given by Eq. (16) in Theorem 1.

VI. Numerical Results

This section presents numerical results that validate our convergence of Algo. 1. As a benchmark, we compare our results to that of [10] which is an adaptation of the W-MSR algorithm [42] for consensus to the case of distributed optimization. Following the notations in [10], we denote by F the maximal number of highest values and lowest values that each legitimate agent discards, overall a legitimate agent may ignore no more than $2F$ values.

We consider a distributed network with $|\mathcal{L}| = 15$ legitimate agents and $|\mathcal{M}| = \{15, 30\}$ malicious agents. To maximize the malicious agents' impact we assume that every malicious agent is connected to all the legitimate agents. The legitimate agents connectivity is captured by Fig. 1. The legitimate agents values are one-dimensional and lie in the interval $[-\eta, \eta]$, where $\eta = 50$. The legitimate agents aim to minimize the function $\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}}(x-u_i)^2$, where $(u_i)_{i=1}^{15} = (115.7, 163.3, -81.7, 127.2, -63.7, 58.4, -3.1, 62.9, 54.5, 144.9, -121.1, 9.3, -2.6, -124.5, 131)$. In this case, $x^\star_{\mathcal{L}} \approx 31.367$ where the approximation is to the third digit on the right. The initial values of the legitimate agents are chosen randomly in the set $[-\eta, \eta]$. Additionally, we choose $\gamma(t) = \frac{10}{t+2} \cdot \mathbb{1}_{\{t\geq 0\}}$. To maximize the harmful impact of malicious agent on our analytical results, choose the malicious agents' values to be equal to -50, i.e., $-\eta$, at all times. Finally, we have $\mathbf{E}[\alpha_{ij}] = 0.55$ if $j \in \mathcal{N}_i \cap \mathcal{L}$, and $\mathbf{E}[\alpha_{ij}] = 0.45$ if $j \in \mathcal{N}_i \cap \mathcal{M}$. The random variable α_{ij} is uniformly distributed on the interval $\left[\mathbf{E}[\alpha_{ij}] - \frac{\ell}{2}, \mathbf{E}[\alpha_{ij}] + \frac{\ell}{2}\right]$. We consider the values ℓ: $0.6, 0.8$, in both scenarios $|E_{\mathcal{L}}| = |E_{\mathcal{M}}| = 0.05$, however, the variance of the trust values when $\ell = 0.8$ are higher. We remark that the legitimate agents are ignorant regarding the values $\mathbf{E}[\alpha_{ij}]$ and ℓ. We average the results across 100 system realizations. Denote $\overline{e}(t) \triangleq \frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}}|x_i(t) - x^\star_{\mathcal{L}}|$.

Figs. 2 and 3 captures the average value of the distance of each legitimate agent from the optimal point $x^\star_{\mathcal{L}}$ normalized by the average of this initial distance, i.e., the average value of $\frac{\overline{e}(t)}{\overline{e}(0)}$, for each time t. We can see that the W-MSR algorithm fail to converge to the optimal solution. This occurs due to the high number of malicious agents which are higher than the tolerance threshold in [10] which is upper bounded by 2 following our argument in [37]. Additionally, the W-MSR algorithm is not guaranteed to converge to the optimal value $x^\star_{\mathcal{L}}$ but to a value in the convex hull of $\Pi_{[-\eta,\eta]}(u_i)$, $i \in \mathcal{L}$. In our case this convex hull is exactly the interval $[-\eta, \eta]$, thus the W-MSR algorithm cannot guarantee the reduction of the distance to the optimal value with respect to the interval $[\eta, \eta]$. In contrary, Algo. 1 provides resilience to malicious activity and can tolerate even $15 = |\mathcal{L}|$ and $30 = 2|\mathcal{L}|$ malicious agents, as evident in Figs. 2 and 3. Furthermore, we can see from Figs. 2 and 3 that Algo. 1 is robust to small values of $E_{\mathcal{L}}$ and $E_{\mathcal{M}}$. Finally, Figs. 2 and 3 shows that the variance of the trust values has more impact on T_0 values that are smaller than 50, this occurs

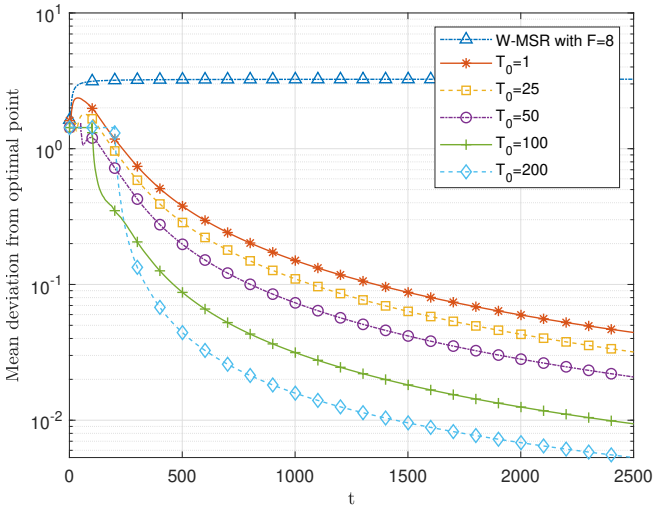

Fig. 2. Average of $\frac{\overline{e}(t)}{\overline{e}(0)}$ of as a function of t for $|\mathcal{M}| = 15$, $\ell = 0.8$.

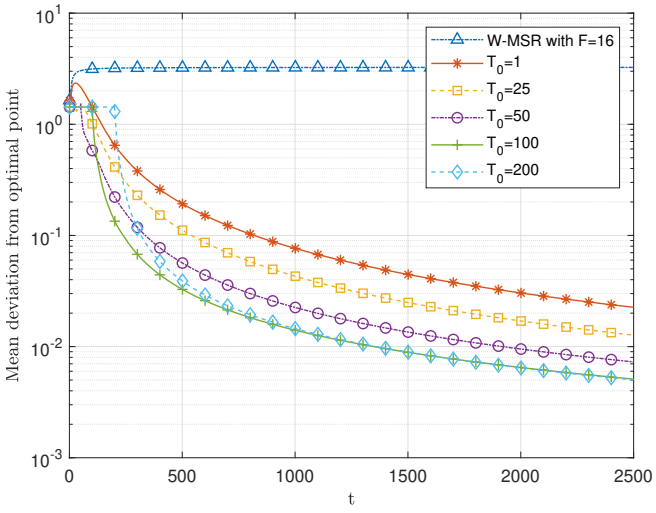

Fig. 3. $\frac{\overline{e}(t)}{\overline{e}(0)}$ as a function of t for $|\mathcal{M}| = 30$, $\ell = 0.6$.

since the higher variance of the trust values increases the misclassification errors. Since the probability of the these errors decreases with T_0, they are less impactful when T_0 is 100 or higher.

VII. Conclusions

This work studies the problem of resilient distributed optimization in the presence of malicious activity with an emphasis on cyberphysical systems. We consider the case where additional information in the form of stochastic inter-agent trust values are available. Under this model, we propose a mechanism for exploiting these trust values where legitimate agents learn to distinguish between their legitimate and malicious neighbors. We incorporate this mechanism to arrive at resilient distributed optimization where strong performance guarantees can be recovered. Specifically, we prove that our algorithm converges to the optimal solution of the nominal distributed optimization system with no malicious agents, and we present an upper bound on the expected distance of the agents' iterates from the optimal solution. Finally, we present numerical results that demonstrate the performance of our proposed distributed optimization framework.

Appendix I
Proof of Lemma 2

First, note that by the decision rule at time t we are guaranteed to make a classification mistake at time 0 since we initialize $\beta_{ij} = 0$, which means that at time $t = 0$ all the agents are classified as malicious by their neighbors.

Denote

$$\mathcal{E}(k) = \bigcup_{\substack{i\in\mathcal{L},\\ j\in\mathcal{N}_i\cap\mathcal{L}}} \{\beta_{ij}(k) < 0\} \quad \bigcup_{\substack{i\in\mathcal{L},\\ j\in\mathcal{N}_i\cap\mathcal{M}}} \{\beta_{ij}(k) < 0\}.$$

By Lemma 1, for all $t \geq 1$ we have that

$$\begin{aligned}
&\Pr(T_f = k)\\
&\overset{(a)}{\leq} \Pr\left(\mathcal{E}(k-1)\right)\\
&\overset{(b)}{\leq} \sum_{\substack{i\in\mathcal{L},\\ j\in\mathcal{N}_i\cap\mathcal{L}}} \Pr(\beta_{ij}(k-1) < 0) + \sum_{\substack{i\in\mathcal{L},\\ j\in\mathcal{N}_i\cap\mathcal{M}}} \Pr(\beta_{ij}(k-1) \geq 0)\\
&\overset{(c)}{\leq} \sum_{i\in\mathcal{L}} |\mathcal{N}_i\cap\mathcal{L}| \exp(-2(k-1)^+E_{\mathcal{L}}^2)\\
&\quad + \sum_{i\in\mathcal{L}} |\mathcal{N}_i\cap\mathcal{M}| \exp(-2(k-1)^+E_{\mathcal{M}}^2), \qquad (24)
\end{aligned}$$

where (a) follows from the definition of T_f in Corollary 2 which implies that if $T_f = k$ there must be a misclassification error in the legitimacy of agents at time $k-1$, (b) follows from the union bound, and (c) follows from Lemma 1.

Furthermore,

$$\begin{aligned}
&\Pr(T_f > t-1)\\
&= \Pr\left(\bigcup_{k\geq t} \mathcal{E}(k)\right)\\
&\leq \sum_{k=t}^{\infty} \Pr\left(\mathcal{E}(k)\right)\\
&\leq \sum_{k=t}^{\infty}\sum_{i\in\mathcal{L}} |\mathcal{N}_i\cap\mathcal{L}| \exp(-2(k-1)^+E_{\mathcal{L}}^2)\\
&\quad + \sum_{k=t}^{\infty}\sum_{i\in\mathcal{L}} |\mathcal{N}_i\cap\mathcal{M}| \exp(-2(k-1)^+E_{\mathcal{M}}^2)\\
&= D_{\mathcal{L}} \frac{\exp(-2(t-1)E_{\mathcal{L}}^2)}{1-\exp(-2E_{\mathcal{L}}^2)} + D_{\mathcal{M}} \frac{\exp(-2(t-1)E_{\mathcal{M}}^2)}{1-\exp(-2E_{\mathcal{M}}^2)}. \qquad (25)
\end{aligned}$$

Note that $\Pr(T_f > t-1)$ vanishes as t tends to infinity.

Appendix II

Here we extend [43, Lemma 11] to the case of d-dimensional vectors and random perturbations. Note that a naive implementation of [43, Lemma 11] for each dimension in $\{1,\ldots,d\}$ scale of the resulted upper bound by $\sqrt{d}$. The upper bound we next derive eliminates this scaling.

Let $A \in \mathbb{R}^{d\times|\mathcal{L}|}$ and denote by $[A]_j$ the jth column of A. The the Frobenius norm of the matrix A is defined as

$$\|A\|_{\mathrm{F}} \triangleq \sqrt{\sum_{i=1}^{d}\sum_{j\in\mathcal{L}} |a_{ij}|^2} = \sqrt{\sum_{j\in\mathcal{L}} \|[A]_j\|^2} = \sqrt{\sum_{i=1}^{d} \|[A^T]_i\|^2},$$

where A^T denotes the transpose matrix of the matrix A. Additionally, we denote by $\mathbf{1}$ the all-ones column vector with $|\mathcal{L}|$ entries.

Lemma 5. *Let $\mathcal{X} \subset \mathbb{R}^d$ be a compact and convex set. Additionally, let $W(t) \triangleq (w_{ij}(t))_{i,j\in\mathcal{L}}$ be deterministic doubly stochastic matrices such that $\sigma_2(W(t)) \leq \rho$ for all $t \geq 0$. Furthermore, let $\Delta_i(t) \in \mathbb{R}^{d\times 1}$ be random vectors, and let $X(t) \in \mathbb{R}^{d\times|\mathcal{L}|}$ be defined by the following dynamic*

$$X(t+1) = X(t)W^T(t) + \Delta(t), \qquad (26)$$

where $\Delta(t) = (\Delta_1,\ldots,\Delta_{|\mathcal{L}|})$.

Let us assume that there exists a non-increasing sequence $\delta(t)$ such that

$$\mathbf{E}(\|\Delta_i(t)\|^2) \leq \delta^2(t), \ \ \forall\, i \in \mathcal{L},$$

and let

$$\overline{X}(t) \triangleq \frac{X(t)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T = \frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}}[X(t)]_i\mathbf{1}^T.$$

Then,

$$\frac{\sum_{j\in\mathcal{L}} \mathbf{E}\|[X(t)]_j - \overline{X}(t)\|}{|\mathcal{L}|} \leq 2\eta\rho^t + \frac{\delta(0)\rho^{t/2}}{1-\rho} + \frac{2\delta(t)}{1-\rho},$$

and

$$\frac{\sum_{j\in\mathcal{L}} \mathbf{E}\|[X(t)]_j - \overline{X}(t)\|^2}{|\mathcal{L}|} \leq \left[2\eta\rho^t + \frac{\delta(0)\rho^{t/2}}{1-\rho} + \frac{2\delta(t)}{1-\rho}\right]^2$$

Proof. First, we utilize the upper bound $\frac{1}{2}(\|[A]_i\|^2 +$

$\|[A]_j\|) \geq \|[A]_i\| \cdot \|[A]_j\|$ to deduce that

$$\sum_{j\in\mathcal{L}} \|[A]_j\| \leq \sqrt{|\mathcal{L}|} \cdot \|A\|_{\mathrm{F}}. \tag{27}$$

Furthermore,

$$\sum_{j\in\mathcal{L}} \|[A]_j\|^2 = \|A\|_{\mathrm{F}}^2. \tag{28}$$

Thus, we focus our efforts on upper bounding $\|X(t) - \overline{X}(t)\|_{\mathrm{F}}$ and $\|X(t) - \overline{X}(t)\|_{\mathrm{F}}^2$.

Observe that

$$\begin{aligned}\left\|\Delta(t) - \frac{\Delta(t)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right\|_{\mathrm{F}} &= \sqrt{\sum_{j\in\mathcal{L}}\left\|[\Delta(t)]_j - \frac{1}{|\mathcal{L}|}\sum_{k\in\mathcal{L}}[\Delta(t)]_k\right\|^2} \\ &\leq \sqrt{\sum_{j\in\mathcal{L}}\|[\Delta(t)]_j\|^2} \\ &= \|\Delta(t)\|_{\mathrm{F}}. \end{aligned} \tag{29}$$

Additionally,

$$\begin{aligned}&\left\|\left(X(t) - \frac{X(t)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right)W^T(t)\right\|_{\mathrm{F}} \\ &= \left\|W(t)\left(X(t) - \frac{X(t)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right)^T\right\|_{\mathrm{F}} \\ &\leq \sqrt{\sum_{i=1}^{d}\left\|W(t)\left([X^T(t)]_i - \mathbf{1}\frac{\mathbf{1}^T[X^T(t)]_i}{|\mathcal{L}|}\right)\right\|^2} \\ &\leq \sqrt{\sum_{i=1}^{d}\rho^2\left\|\left([X^T(t)]_i - \mathbf{1}\frac{\mathbf{1}^T[X^T(t)]_i}{|\mathcal{L}|}\right)\right\|^2} \\ &= \rho\left\|X(t) - \frac{X(t)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right\|_{\mathrm{F}}. \end{aligned} \tag{30}$$

It follows that

$$\begin{aligned}&\|X(t+1) - \overline{X}(t+1)\|_{\mathrm{F}} \\ &= \left\|X(t+1) - \frac{X(t+1)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right\|_{\mathrm{F}} \\ &= \left\|X(t)W^T(t) + \Delta(t) - \frac{[X(t)+\Delta(t)]\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right\|_{\mathrm{F}} \\ &\overset{(a)}{\leq} \left\|X(t)W^T(t) - \frac{X(t)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right\|_{\mathrm{F}} + \left\|\Delta(t) - \frac{\Delta(t)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right\|_{\mathrm{F}} \\ &\overset{(b)}{\leq} \left\|X(t)W^T(t) - \frac{X(t)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right\|_{\mathrm{F}} + \|\Delta(t)\|_{\mathrm{F}} \\ &\overset{(c)}{=} \left\|\left(X(t) - \frac{X(t)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right)W^T(t)\right\|_{\mathrm{F}} + \|\Delta(t)\|_{\mathrm{F}} \\ &\overset{(d)}{\leq} \rho\left\|X(t) - \frac{X(t)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right\|_{\mathrm{F}} + \|\Delta(t)\|_{\mathrm{F}}, \end{aligned} \tag{31}$$

where (a) follows from the triangle inequality, (b) follows from (29), (c) follows from the double stochasticity of $W(t)$, and (d) follows from (30).

Thus,

$$\begin{aligned}&\|X(t) - \overline{X}(t)\| \\ &\leq \rho^t\left\|X(0) - \frac{X(0)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right\| + \sum_{k=0}^{t-1}\rho^{t-1-k}\|\Delta(k)\|. \end{aligned} \tag{32}$$

Now, since $\mathbf{E}(\|\Delta_i(t)\|^2) \leq \delta^2(t)$ for all $i \in \mathcal{L}$, and by the non-negativity of the variance of $\|\Delta(k)\|_{\mathrm{F}}$

$$\begin{aligned}\mathbf{E}(\|\Delta(k)\|_{\mathrm{F}}) &\leq \sqrt{\mathbf{E}\left(\|\Delta(k)\|_{\mathrm{F}}^2\right)} \\ &= \sqrt{\sum_{j\in\mathcal{L}}\mathbf{E}\left(\|[\Delta(k)]_j\|^2\right)} \\ &\leq \sqrt{|\mathcal{L}|\delta^2(k)} \\ &= \sqrt{|\mathcal{L}|}\delta(k). \end{aligned} \tag{33}$$

then $\mathbf{E}(\|\Delta_i(t)\|) \leq \delta(t)$ for all $i \in \mathcal{L}$.

If follows from (3)

$$\begin{aligned}&\left\|X(0) - \frac{X(0)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right\|_{\mathrm{F}} \\ &= \sqrt{\sum_{j\in\mathcal{L}}\left\|\left[X(0) - \frac{X(0)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right]_j\right\|^2} \\ &\leq \sqrt{\sum_{j\in\mathcal{L}}\left(\|[X(0)]_j\| + \left\|\left[\frac{X(0)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right]_j\right\|\right)^2} \\ &\leq \sqrt{\sum_{j\in\mathcal{L}}(2\eta)^2} \\ &= 2\eta\sqrt{|\mathcal{L}|}. \end{aligned} \tag{34}$$

It follows that

$$\begin{aligned}&\frac{\mathbf{E}\|X(t) - \overline{X}(t)\|_{\mathrm{F}}}{|\mathcal{L}|} \\ &\leq \frac{2\eta}{\sqrt{|\mathcal{L}|}}\rho^t + \frac{1}{\sqrt{|L|}}\sum_{k=0}^{t-1}\rho^{t-1-k}\delta(k) \\ &\leq \frac{2\eta}{\sqrt{|\mathcal{L}|}}\rho^t + \frac{\delta(0)}{\sqrt{|L|}}\cdot\frac{\rho^{t/2}}{1-\rho} + \frac{\delta(t/2)}{\sqrt{|L|}(1-\rho)}. \end{aligned} \tag{35}$$

Similarly,

$$\begin{aligned}&\|X(t) - \overline{X}(t)\|_{\mathrm{F}}^2 \\ &\leq \rho^{2t}\left\|X(0) - \frac{X(0)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right\|^2 \\ &+ 2\rho^t\left\|X(0) - \frac{X(0)\mathbf{1}}{|\mathcal{L}|}\mathbf{1}^T\right\|\sum_{k=0}^{t-1}\rho^{t-1-k}\|\Delta(k)\|_{\mathrm{F}} \\ &+ \sum_{k_1=0}^{t-1}\sum_{k_2=0}^{t-1}\rho^{t-1-k_1}\rho^{t-1-k_2}\|\Delta(k_1)\|_{\mathrm{F}}\cdot\|\Delta(k_2)\|_{\mathrm{F}} \\ &\leq \rho^{2t}4\eta^2|\mathcal{L}| + 4\rho^t\eta\sqrt{|\mathcal{L}|}\sum_{k=1}^{t-1}\rho^{t-1-k}\|\Delta(k)\|_{\mathrm{F}} \\ &+ \sum_{k_1=0}^{t-1}\sum_{k_2=0}^{t-1}\rho^{t-1-k_1}\rho^{t-1-k_2}\|\Delta(k_1)\|_{\mathrm{F}}\cdot\|\Delta(k_2)\|_{\mathrm{F}}. \end{aligned} \tag{36}$$

Additionally, by the Cauchy–Schwarz inequality *for expec-*

tations

$$\mathbf{E}\left(\|\Delta(k_1)\|_{\mathrm{F}} \cdot \|\Delta(k_2)\|_{\mathrm{F}}\right) \leq \sqrt{\mathbf{E}\left(\|\Delta(k_1)\|_{\mathrm{F}}^2\right)\mathbf{E}\left(\|\Delta(k_2)\|_{\mathrm{F}}^2\right)} \leq |\mathcal{L}|\delta(k_1)\delta(k_2). \quad (37)$$

Therefore,

$$\begin{aligned}&\frac{\mathbf{E}\left(\|X(t)-\overline{X}(t)\|_{\mathrm{F}}^2\right)}{|\mathcal{L}|}\\ &\leq 4\eta^2\rho^{2t} + 4\rho^t\eta\left[\frac{\delta(0)\rho^{t/2}}{1-\rho}+\frac{\delta(t/2)}{(1-\rho)}\right]\\ &\quad+\left[\frac{\delta(0)\rho^{t/2}}{1-\rho}+\frac{\delta(t/2)}{(1-\rho)}\right]^2\\ &= \left[2\eta\rho^t+\frac{\delta(0)\rho^{t/2}}{1-\rho}+\frac{\delta(t/2)}{(1-\rho)}\right]^2. \quad (38)\end{aligned}$$

We conclude the proof by using the upper bounds (27) and (28). □

REFERENCES

[1] T. Halsted, O. Shorinwa, J. Yu, and M. Schwager, "A survey of distributed optimization methods for multi-robot systems," 2021. [Online]. Available: https://arxiv.org/pdf/2103.12840.pdf

[2] M. Zhu and S. Martínez, *Distributed optimization-based control of multi-agent networks in complex environments*. Springer, 2015.

[3] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, 2020.

[4] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Automat. Contr.*, vol. 54, no. 1, pp. 48–61, 2009.

[5] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

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

[7] G. Lan and Y. Zhou, "Asynchronous decentralized accelerated stochastic gradient descent," *IEEE J. Sel. Areas Inf. Theory*, vol. 2, no. 2, pp. 802–811, 2021.

[8] G. Lan, Y. Ouyang, and Y. Zhou, "Graph topology invariant gradient and sampling complexity for decentralized and stochastic optimization," 2021, https://arxiv.org/pdf/2101.00143.pdf.

[9] N. Ravi, A. Scaglione, and A. Nedić, "A case of distributed optimization in adversarial environment," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2019, pp. 5252–5256.

[10] S. Sundaram and B. Gharesifard, "Distributed optimization under adversarial nodes," *IEEE Trans. Automat. Contr.*, vol. 64, no. 3, pp. 1063–1076, 2019.

[11] A.-Y. Lu and G.-H. Yang, "Distributed secure state estimation in the presence of malicious agents," *IEEE Trans. Automat. Contr.*, vol. 66, no. 6, pp. 2875–2882, 2021.

[12] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Automat. Contr.*, vol. 31, no. 9, pp. 803–812, 1986.

[13] A. Nedić and A. Olshevsky, "Stochastic gradient-push for strongly convex functions on time-varying directed graphs," *IEEE Trans. Automat. Contr.*, vol. 61, no. 12, pp. 3936–3947, 2016.

[14] M. Zhu and S. Martínez, "On distributed constrained formation control in operator–vehicle adversarial networks," *Automatica*, vol. 49, no. 12, pp. 3571–3582, 2013.

[15] K. Saulnier, D. Saldana, A. Prorok, G. J. Pappas, and V. Kumar, "Resilient flocking for mobile robot teams," *IEEE Robotics and Automation letters*, vol. 2, no. 2, pp. 1039–1046, 2017.

[16] F. Pasqualetti, A. Bicchi, and F. Bullo, "Consensus computation in unreliable networks: A system theoretic approach," *IEEE Trans. Automat. Contr.*, vol. 57, no. 1, pp. 90–104, 2012.

[17] A. A. Cárdenas, T. Roosta, G. Taban, and S. Sastry, *Cyber Security: Basic Defenses and Attack Trends*, 2008, pp. 73–101.

[18] A. A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *2008 The 28th International Conference on Distributed Computing Systems Workshops*, 2008, pp. 495–500.

[19] J. Xiong and K. Jamieson, "Securearray: Improving wifi security with fine-grained physical-layer information," in *Proceedings of the 19th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom, 2013.

[20] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 93–109, 2015.

[21] S. Gil, S. Kumar, M. Mazumder, D. Katabi, and D. Rus, "Guaranteeing spoof-resilient multi-robot networks," *Autonomous Robots*, vol. 41, pp. 1383–1400, 2017.

[22] C. E. Shannon, "Channels with side information at the transmitter," *IBM Journal of Research and Development*, vol. 2, no. 4, pp. 289–293, 1958.

[23] C. Pippin and H. Christensen, "Trust modeling in multi-robot patrolling," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 59–66.

[24] A. Nedić, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. Automat. Contr.*, vol. 55, no. 4, pp. 922–938, 2010.

[25] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *J. Optim. Theory Appl.*, vol. 147, no. 3, pp. 516–545, 2010.

[26] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE Trans. Automat. Contr.*, vol. 57, no. 3, pp. 592–606, 2012.

[27] K. I. Tsianos, S. Lawlor, and M. G. Rabbat, "Push-sum distributed dual averaging for convex optimization," in *IEEE Conf. Decision Control (CDC)*, 2012, pp. 5453–5458.

[28] ——, "Consensus-based distributed optimization: Practical issues and applications in large-scale machine learning," in *Allert. Conf. Commun. Control Comput.*, 2012, pp. 1543–1550.

[29] K. I. Tsianos and M. G. Rabbat, "Distributed strongly convex optimization," in *Allert. Conf. Commun. Control Comput.*, 2012, pp. 593–600.

[30] A. Nedić and A. Olshevsky, "Stochastic gradient-push for strongly convex functions on time-varying directed graphs," *IEEE Trans. Automat. Contr.*, vol. 61, no. 12, pp. 3936–3947, 2016.

[31] S. Magnússon, C. Enyioha, N. Li, C. Fischione, and V. Tarokh, "Convergence of limited communication gradient methods," *IEEE Trans. Automat. Contr.*, vol. 63, no. 5, pp. 1356–1371, 2018.

[32] Y. Tang, J. Zhang, and N. Li, "Distributed zero-order algorithms for nonconvex multiagent optimization," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 1, pp. 269–281, 2021.

[33] W. Abbas, Y. Vorobeychik, and X. Koutsoukos, "Resilient consensus protocol in the presence of trusted nodes," in *International Symposium on Resilient Control Systems (ISRCS)*, 2014, pp. 1–7.

[34] B. Turan, C. A. Uribe, H.-T. Wai, and M. Alizadeh, "Resilient primal-dual optimization algorithms for distributed resource allocation," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 1, pp. 282–294, 2021.

[35] C. Zhao, J. He, and Q.-G. Wang, "Resilient distributed optimization algorithm against adversarial attacks," *IEEE Trans. Automat. Contr.*, vol. 65, no. 10, pp. 4308–4315, 2020.

[36] T. Ding, Q. Xu, S. Zhu, and X. Guan, "A convergence-preserving data integrity attack on distributed optimization using local information," in *IEEE Conf. Decision Control (CDC)*, 2020, pp. 3598–3603.

[37] M. Yemini, A. Nedić, A. J. Goldsmith, and S. Gil, "Characterizing trust and resilience in distributed consensus for cyberphysical systems," *IEEE Trans. Robot.*, vol. 38, no. 1, pp. 71–91, 2022.

[38] M. O. Sayin, N. D. Vanli, S. S. Kozat, and T. Başar, "Stochastic subgradient algorithms for strongly convex optimization over distributed networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 4, no. 4, pp. 248–260, 2017.

[39] A. Olshevsky, "Linear time average consensus and distributed optimization on fixed graphs," *SIAM Journal on Control and Optimization*, vol. 55, no. 6, pp. 3990–4014, 2017.

[40] B. Charron-Bost, "Geometric bounds for convergence rates of averaging algorithms," 2020. [Online]. Available: https://arxiv.org/pdf/2007.04837.pdf

[41] S. Lacoste-Julien, M. Schmidt, and F. R. Bach, "A simpler approach to obtaining an O(1/t) convergence rate for the projected stochastic subgradient method," 2012. [Online]. Available: https://arxiv.org/pdf/1212.2002.pdf

[42] H. J. LeBlanc, H. Zhang, X. Koutsoukos, and S. Sundaram, "Resilient asymptotic consensus in robust networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 4, pp. 766–781, 2013.

[43] A. Nedić, A. Olshevsky, and M. G. Rabbat, "Network topology and communication - computation tradeoffs in decentralized optimization," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 953–976, 2018.