Privacy-preserving Average Consensus: Privacy Analysis and Optimal Algorithm Design

Jianping He$^{1,2}$, Lin Cai$^2$, Chengcheng Zhao$^3$, Peng Cheng$^3$, and Xinping Guan$^1$

Abstract—Privacy-preserving average consensus aims to guarantee the privacy of initial states and asymptotic consensus on the exact average of the initial value. In existing work, it is achieved by adding and subtracting variance decaying and zero-sum random noises to the consensus process. However, there is lack of theoretical analysis to quantify the degree of the privacy protection. In this paper, we introduce the maximum disclosure probability that the other nodes can infer one node's initial state within a given small interval to quantify the privacy. We develop a novel privacy definition, named $(\epsilon,\delta)$-data-privacy, to depict the relationship between maximum disclosure probability and estimation accuracy. Then, we prove that the general privacy-preserving average consensus (GPAC) provides $(\epsilon,\delta)$-data-privacy, and provide the closed-form expression of the relationship between $\epsilon$ and $\delta$. Meanwhile, it is shown that the added noise with uniform distribution is optimal in terms of achieving the highest $(\epsilon,\delta)$-data-privacy. We also prove that when all information used in the consensus process is available, the privacy will be compromised. Finally, an optimal privacy-preserving average consensus (OPAC) algorithm is proposed to achieve the highest $(\epsilon,\delta)$-data-privacy and avoid the privacy compromise. Simulations are conducted to verify the results.

Index Terms—Average consensus, Data privacy, Optimal algorithm, Disclosure probability.

I. INTRODUCTION

Consensus has attracted extensive attention over the past decades, since it is an efficient algorithm for distributed computing and control. A consensus algorithm refers to the action that nodes in the network reach a global agreement regarding a certain opinion using their local neighbors' information only [1]. Due to the strong robustness and scalability, consensus has been applied in a variety of areas, e.g., coordination and cooperation [2], [3], distributed estimation and optimization [4], [5], sensor fusion [6], distributed energy management [7] and sensing scheduling [8], and time synchronization [9]–[11].

Average consensus is the most commonly adopted consensus algorithm, where the agreement reached by the algorithm equals the average of all nodes’ initial states. For traditional average consensus algorithms, each node will broadcast its real state to neighbor nodes during consensus process. Hence, under traditional average consensus algorithms, the state information of each node is disclosed to its neighbor nodes.

However, in some applications, the initial states of nodes are private information, which means that nodes do not want to release their real initial states to other nodes [16]. For example, consensus algorithm is adopted in social networks for a group of members to compute the common opinion on a subject [17]. In this application, each member may want to keep his personal opinion on the subject secret to other members. Also, in the multi-agent rendezvous problem [18], a group of nodes want to eventually rendezvous at a certain location, while the participants may not want to release their initial locations to others. This means that when the privacy is concerned, each node’s real state may not be available to the other nodes, and thus the traditional consensus algorithm becomes invalid.

Recently, researchers have investigated the privacy-preserving average consensus problem, which aims to guarantee that the privacy of initial state is preserved while average consensus can still be achieved [12]–[16]. The basic idea is to add random noise to the real state value during the communication to protect the privacy, and then carefully design the noise adding process such that average consensus is achieved. For example, Huang et al. [13] designed a differentially private iterative synchronous consensus algorithm by adding independent and exponentially decaying Laplacian noises to the consensus process. Their algorithm can guarantee differential privacy. As the algorithm may converge to a random value, the exact average consensus may not be guaranteed. Nozari et al. [14] pointed out and proved that it is impossible to achieve average consensus and differential privacy simultaneously. Hence, they design a novel linear Laplacian-based consensus algorithm, which guarantees that an unbiased estimate of the average consensus can be achieved almost surely with differential privacy. Manitara and Hadjicostis [15] proposed a privacy-preserving average consensus algorithm by adding correlated noises to the consensus process. The proposed algorithm guarantees the initial state of each node cannot be perfectly inferred by the other “malicious” nodes. A sufficient condition is provided under which the privacy of benign agents’ initial states are preserved. More recently, Mo and Murray in [16] well addressed the privacy-preserving average consensus problem by designing a novel PPAC algorithm, where exponentially decaying and zero-sum normal noises are added to traditional consensus process. They proved that PPAC algorithm achieves the exact average consensus in the mean-square sense, and also proved that the algorithm achieves minimum privacy breach in the sense of disclosed space. Braca et al. in [22] examined the interplay between learning and privacy over multi-agent consensus networks. They provided an analytical characterization of the interplay...
between learning and privacy for the consensus perturbing and preserving strategy, respectively.

However, there is lack of theoretical results to quantify the degree of the privacy protection and what is the relationship between the estimation accuracy and privacy. To fill this gap, in this paper, we provide theoretical privacy analysis for the GPAC algorithm (consider the general noise adding process) in the sense of the maximum disclosure probability that other nodes can infer one node’s initial state within a given small interval (a given estimation accuracy). A privacy definition, named $(\epsilon, \delta)$-data-privacy, which is first introduced in our previous work [21], is exploited to depict the maximum disclosure probability. This privacy definition reveals the relationship between privacy and estimation accuracy. We provide theoretical results to quantify the degree of the privacy preservation strategy, respectively.

To fill this gap, we design a novel OPAC algorithm to achieve average consensus as well as data-privacy. The main contributions of this paper are summarized as follows.

- We prove that the GPAC algorithm provides $(\epsilon, \delta)$-data-privacy, and obtain a closed-form expression of the relationship between the estimation accuracy and the privacy (the relationship between $\epsilon$ and $\delta$).
- We prove that for the added random noises, the uniform distribution is optimal in the sense that a PPAC algorithm can achieve the highest privacy when the mean and variance of noises are fixed.
- We prove that when all the information used in the consensus process are available for the estimation, the maximum disclosure probability will converge to one, i.e., the initial state of a node is perfectly inferred. This result reveals how the exact initial state can be inferred.
- We design a novel OPAC algorithm to achieve average consensus while guarantees the highest $(\epsilon, \delta)$-data-privacy. It is proved that OPAC algorithm converges to the exact average consensus, and avoids the privacy to be lost even if all the information used in the consensus process is available for the estimation.

The remainder of this paper is organized as follows. Section II introduces preliminary results and problem formulation. In Section III, we provide theoretical results on the degree of privacy protection. The OPAC algorithm is proposed in Section IV. Section V verifies the main results and conclusions are given in Section VI.

II. PRELIMINARIES AND PROBLEM FORMULATION

The network is abstracted as an undirected and connected graph, $G = (V, E)$, where $V$ is the set of nodes and $E$ is the set of the communication links (edges) between nodes. $(i, j) \in E$ if and only if (iff) nodes $i$ and $j$ can communicate with each other. Let $N_i$ be the neighbor set of node $i$, where $j \in N_i$ iff $(i, j) \in E$, i.e., $N_i = \{j | j \in V, (i, j) \in E, j \neq i\}$.

A. Average Consensus

Suppose that there are $n$ $(n \geq 3)$ nodes in the network (i.e., $|V| = n$), and each node $i$ has an initial scalar state $x_i(0)$, where $x_i(0) \in R$. For an average consensus algorithm, each node will communicate with its neighbor nodes and update its state based on the received information to obtain the average of all initial state’s values. Hence, the traditional average consensus algorithm is given as follows,

$$x_i(k + 1) = w_{ii}x_i(k) + \sum_{j \in N_i}w_{ij}x_j(k),$$  \hspace{1cm} (1)

for $\forall i \in V$, which can be written in the matrix form as

$$x(k + 1) = Wx(k),$$  \hspace{1cm} (2)

where $w_{ii}$ and $w_{ij}$ are weights, and $W$ is the weight matrix. It is well known from [19] that if, 1) $w_{ii} > 0$, and $w_{ij} > 0$ for $(i, j) \in E$ and $w_{ij} = 0$ for otherwise; and 2) $W$ is a doubly stochastic matrix, then average consensus can be achieved by (1), i.e.,

$$\lim_{k \to \infty} x_i(k) = \frac{\sum_{\ell=1}^{n}x_i(0)}{n} = \bar{x}. \hspace{1cm} (3)$$

When the privacy of nodes’ initial states are concerned, all nodes are unwilling to release its real state to the neighbor nodes at each iteration. It means that each $x_i(k)$ is unavailable in (1). To preserve the privacy of nodes’ initial states, a widely used approach is to add a random noise to the real state value when a node needs to communicate with its neighbor nodes at each iteration. We define a new state as

$$x_i^+ = x_i(k) + \theta_i(k), i \in V,$$  \hspace{1cm} (4)

where $\theta_i(k)$ is the added random noise for privacy preservation at iteration $k$. With the noise adding process, the update equation (1) is changed to,

$$x_i(k + 1) = w_{ii}x_i^+(k) + \sum_{j \in N_i}w_{ij}x_j^+(k)$$

$$= w_{ii}[x_i(k) + \theta_i(k)] + \sum_{j \in N_i}w_{ij}[x_j(k) + \theta_j(k)],$$  \hspace{1cm} (6)

for $\forall i \in V$. Therefore, a privacy-preserving average consensus algorithm is to design the added noises (including the distribution and the correlations among them), such that the goal of (1) is achieved under (5). Note that in (4), the noise $\theta_i(k)$ is a general random noise (where its distribution is not fixed), the algorithm (4)–(6) is thus named as the general privacy-preserving average consensus (GPAC) algorithm in the remainder part of this paper.

B. Privacy Definitions

Under (4), the broadcast information sequence of node $i$ is $x_i^+(0), x_i^+(1), ..., x_i^+(k)$, which will be received by its neighbor nodes. Hence, each neighbor node $j$ can infer/estimate the initial state $x_i(0)$ with the received information sequence from node $i$. Note that each of the information output, $x_i^+(k)$, equals the weighted sum of the received information in the previous round plus a noise. Based on the information output, node $j$ will take the probability over the space of all noises
where \( \{ \theta_i(k) \}_{k=0}^{\infty} \) (where the space is denoted by \( \Theta \)) to estimate the values of the added noises. It then will infer \( x_i(0) \) by using the difference between each information output and the estimated noises, i.e., \( \hat{x}_i(0) = x_i^+(k) - \theta_i(k) \), where \( \theta_i(k) \) is the estimation of random noise \( \theta_i^k \) (\( \theta_i^k = x_i^+(k) - x_i(0) \)). Under this estimation, we have

\[
\Pr \{ |\hat{x}_i(0) - x_i(0)| \leq \epsilon \} = \Pr \{ |\hat{\theta}_i(k) - \theta_i^k| \leq \epsilon \},
\]

where \( \epsilon \geq 0 \) is a small constant.

To investigate the relationship between the estimation accuracy and privacy, by referring to [21], we then introduce an privacy definition, named \((\epsilon, \delta)\)-data-privacy, where \( 0 \leq \epsilon \) and \( 0 \leq \delta \leq 1 \), as follows.

**Definition 2.1:** A GPAC algorithm provides \((\epsilon, \delta)\)-data-privacy, if and only if (iff),

\[
\delta = \max_{\theta_i(k) \in \Theta, k \geq 0} \Pr \{ |\hat{\theta}_i(k) - \theta_i^k| \leq \epsilon \},
\]

where \( \theta_i^k = x_i^+(k) - x_i(0) \) and \( \hat{\theta}_i(k) \) is the estimation of \( \theta_i^k \).

In the above definition, the estimation accuracy is denoted by parameter \( \epsilon \) and the privacy is expressed by parameter \( \delta \). From (8), it follows that \( \delta \) is the maximum probability that each neighbor node \( j \) can successfully estimate the initial state \( x_i(0) \) in a given interval \( [x_i(0) - \epsilon, x_i(0) + \epsilon] \) with the information output of node \( i \) only. \( \delta \) is thus named as the maximum disclosure probability.

**Definition 2.2:** Given an \( \epsilon \), if algorithms \( A_1 \) and \( A_2 \) provide \((\epsilon, \delta_1)\)-data-privacy and \((\epsilon, \delta_2)\)-data-privacy, respectively, where \( \delta_1 < \delta_2 \), then we say that \( A_1 \) achieves a higher \((\epsilon, \delta)\)-data-privacy than \( A_2 \).

### C. Problem Formulation

In this paper, we will investigate the privacy of the GPAC algorithm (4)–(6) based on the definition of \((\epsilon, \delta)\)-data-privacy, and then design an optimal privacy-preserving average consensus (OPAC) algorithm in terms of \((\epsilon, \delta)\)-data-privacy protection. In summary, we will consider the following four critical problems: i) how to quantify and analyze the privacy of the GPAC algorithm; ii) how will the distribution and correlation of the added random noises affect the privacy; iii) when and how will a node’s exact initial state be inferred by the other nodes; iv) how to achieve the optimal \((\epsilon, \delta)\)-data-privacy and the exact average consensus, and avoid the privacy of nodes’ initial states to be lost.

### III. Privacy Analysis of GPAC

Before presenting the main results, we first give the basic assumptions and the information set used for state estimation. Assume that the distribution and the correlation of the random variable \( \theta_i(k), k = 0, 1, \ldots \), and the update rule of the GPAC algorithm are available to all nodes. The full topology information and \( n \) are assumed to be unknown to any node, which means that each node cannot know the neighbor set of its neighbor nodes and the number of nodes in the whole network. The initial states of nodes are assumed to be independent from each other. For estimation, if there is no information of a variable, then the variable is viewed with domain \( R \). For simplicity, we assume that \( \theta_i(k) \) and \( \theta_j(k) \) are independently and identically distributed (i.i.d) \( \forall k \geq 0 \) and \( i \neq j \). Let \( X \) be the output of a random variable whose distribution is unknown and with domain \( R \). Without the knowledge of the distribution, according to the principle of maximum entropy, we should take the same probability over all the possible values of the random variable to estimate the values of \( X \). Therefore, given an estimation \( \hat{X} \), it is reasonable to assume that

\[
\Pr \{ |\hat{X} - X| \leq \epsilon \} \leq \int_{\nu_{\theta_i(0)} - \epsilon}^{\nu_{\theta_i(0)} + \epsilon} f_{\theta_i(0)}(y) dy.
\]

Then, we define two information sets of node \( i \) up to iteration \( k \) as follows,

\[
\mathcal{T}_1^i(k) = \{ x_i^+(0), \ldots, x_i^+(k) \},
\]

and

\[
\mathcal{T}_2^i(k) = \{ N_i, w_{ii}, w_{ij}, x_i^+(0), x_j^+(0), \ldots, x_i^+(k), x_j^+(k) | j \in N_i \}. \tag{11}
\]

The information set \( \mathcal{T}_1^i(k) \) only includes the states \( x_i^+(\ell), \ell = 0, 1, \ldots, k \), which are used for communication at iteration \( \ell \). Thus, its neighbor nodes can easily obtain \( \mathcal{T}_1^i(k) \) by storing the information received from node \( i \) at each iteration. The information set \( \mathcal{T}_2^i(k) \) includes all information used in consensus process (6) for node \( i \). Other nodes may obtain these information by an eavesdropping attack.

#### A. Privacy of the Algorithm

In this subsection, based on the definition of \((\epsilon, \delta)\)-data-privacy, we first analyze the privacy of the GPAC algorithm and reveal the relationship between the privacy and estimation accuracy, when \( \mathcal{T}_1^i(k) \) is available only.

**Theorem 3.1:** If \( \mathcal{T}_1^i(k) \) is the only information available to node \( j \) to estimate the value of \( x_i(0) \) at iteration \( k \), then

\[
\delta(k) = \max_{\theta_i(k) \in \Theta, k \in \mathbb{N}^+} \Pr \{ |\hat{\theta}_i(k) - \theta_i^k| \leq \epsilon | \mathcal{T}_1^i(k) \}
= \max_{\theta_i(0) \in \Theta} \Pr \{ |\hat{\theta}_i(0) - \theta_i^0| \leq \epsilon | \mathcal{T}_1^i(0) \} \tag{12}
= \max_{\theta_i(0) \in \Theta} \int_{\nu_{\theta_i(0)} - \epsilon}^{\nu_{\theta_i(0)} + \epsilon} f_{\theta_i(0)}(y) dy, \tag{13}
\]

i.e., the relationship between the privacy and the estimation accuracy always satisfies (13), and the maximum disclosure probability is not increased with iteration.

**Proof:** We first prove that, under \( \mathcal{T}_1^i(0), \) (13) holds. With \( \mathcal{T}_1^i(0) \), node \( j \) can estimate \( x_i(0) \) based on the fact that

\[
x_i^+(0) = x_i(0) + \theta_i(0) = x_i(0) + \theta_i^0,
\]

and the corresponding estimation \( \hat{x}_i(0) \) satisfies

\[
\hat{x}_i(0) = x_i^+(0) - \hat{\theta}_i(0).
\]
Then, for any estimation \( \hat{\theta}_i(0) \), we have

\[
\Pr \left\{ |\hat{\theta}_i(0) - \theta^0_i| \leq \epsilon |x_i^0(0) \right\} \\
= \Pr \left\{ \theta_i(0) \in [\hat{\theta}_i(0) - \epsilon, \hat{\theta}_i(0) + \epsilon] |x_i^0(0) \right\} \\
= \int_{\hat{\theta}_i(0) - \epsilon}^{\hat{\theta}_i(0) + \epsilon} f_{\theta_i(0)}(y)dy \\
\leq \max_{\theta_i(0) \in \Theta} \int_{\hat{\theta}_i(0) - \epsilon}^{\hat{\theta}_i(0) + \epsilon} f_{\theta_i(0)}(y)dy,
\]

which means that (13) holds under information \( I^0_i(0) \) at iteration \( k = 0 \).

Then, we prove that (13) holds under \( I^0_i(1) \). With \( I^0_i(1) \), node \( j \) can estimate \( x_i(0) \) by using the fact of both (14) and the following equation for estimation,

\[
\frac{x_i^+(1)}{w_{ij}} = \frac{x_i(1) + \theta_i(1)}{w_{ii}} \\
= x_i^+(0) + \sum_{l \in N_i} \frac{w_{il}}{w_{ii}} x_i^+(0) + \frac{1}{w_{ii}} \theta_i(1) \\
= x_i(0) + \theta_i(0) + \frac{1}{w_{ii}} \theta_i(1) + \sum_{l \in N_i} \frac{w_{il}}{w_{ii}} x_i^+(0).
\]

Using (14) only, we have

\[
\Pr \left\{ |\hat{\theta}_i(0) - \theta^1_i| \leq \epsilon |x_i^0(1) \right\} \\
= \int_{\hat{\theta}_i(0) - \epsilon}^{\hat{\theta}_i(0) + \epsilon} f_{\theta_i(0)}(y)dy \\
\leq \max_{\theta_i(0) \in \Theta} \int_{\hat{\theta}_i(0) - \epsilon}^{\hat{\theta}_i(0) + \epsilon} f_{\theta_i(0)}(y)dy.
\]

Then, we consider the estimation using (17) only. Let

\[
\theta_i^1 = \theta_i(0) + \frac{1}{w_{ii}} \theta_i(1) + \sum_{l \in N_i} \frac{w_{il}}{w_{ii}} x_i^+(0) \\
= \theta_i(0) + \frac{1}{w_{ii}} \theta_i(1) + \theta_i^1(1) \\
= \theta_i^0(0) + \theta_i^1(1).
\]

For any estimation \( \hat{\theta}_i(1) \) (the estimation of \( \theta_i^1(1) \)), we have

\[
\Pr \left\{ |\hat{\theta}_i(1) - \theta_i^1| \leq \epsilon |x_i^0(1) \right\} \\
\leq \Pr \left\{ \theta_i^1 - \hat{\theta}_i(1) \leq \epsilon |x_i^0(1), w_{ii}, \theta_i(0), \theta_i(1) \right\} \\
\leq \Pr \left\{ \theta_i^1 - \theta_i^0(1) - \hat{\theta}_i(1) + \theta_i^1(1) \leq \epsilon |x_i^0(1), w_{ii}, \theta_i^1(1) \right\} \\
\leq \Pr \left\{ |\theta_i^1(1) - \theta_i^0(1)| \leq \epsilon |x_i^0(1), w_{ii} \right\},
\]

where \( \theta_i^1(1) = \theta_i(1) - \theta_i^1(0) \) can be viewed as one of the estimation of \( \theta_i^1(1) \). Since the initial states of nodes are independent from each other and the topology information is not available for estimating/inferring, there is at least one variable included in \( \theta_i^1(1) \) which is unknown to the other nodes. Hence, \( \theta_i^1(1) \) is viewed as a random variable in (20) and its distribution is not available to the estimation. It follows that

\[
\Pr \left\{ |\hat{\theta}_i(1) - \theta_i^1| \leq \epsilon |x_i^0(1) \right\} \\
\leq \Pr \left\{ |\hat{\theta}_i(1) - \theta_i^1| \leq \epsilon |x_i^0(1), w_{ii}, \theta_i(0), \theta_i(1) \right\} \\
\leq \max_{z \in \Theta} \int_{z-\epsilon}^{z+\epsilon} f_{\theta_i(0)}(y)dy,
\]

where we have used the assumption (9). Meanwhile, note that one node can combine (14) and (17) together for estimation.

In this case, we have

\[
\Pr \left\{ \hat{x}_i(0) \in [x_i(0) - \epsilon, x_i(0) + \epsilon] |x_i^0(1) \right\} \\
\leq \max_{t_1, t_2 \in \Theta} \int_{t_1-\epsilon}^{t_1+\epsilon} \int_{t_2-\epsilon}^{t_2+\epsilon} f_{\theta_i(0)}(y)dy \\
\leq \max_{t_1, t_2 \in \Theta} \int_{t_1-\epsilon}^{t_1+\epsilon} \int_{t_2-\epsilon}^{t_2+\epsilon} f_{\theta_i(0)}(y)dy.
\]

From (18), (21) and (22), one concludes that (13) holds under information \( I^0_i(1) \) at iteration \( k = 1 \).

Following the similar analysis, we can prove that (13) holds under information set \( I^0_i(k) \) at any iteration \( k \). It means that \( \delta(k) \) is not an increased function of iteration \( k \), although there are more information of \( I^0_i(k) \) than \( I^0_i(0) \) for \( k > 0 \).

We thus have completed the proof.

From the above proof, it is observed that the privacy is not decreased with iteration when only the information set \( I^0_i(k) = \{ I^0_i(k) \mid k = 0, 1, \ldots, \infty \} \) is available for estimation. The main reason is that based on \( I^0_i(1) \), node \( j \) cannot know the neighbor set information of node \( i \), so that after one iteration there are unknown information embedded into \( x_i^+(k) \) for \( k \geq 1 \). Hence, after one iteration, using \( x_i^+(k) \) for \( k \geq 1 \) cannot improve the estimation accuracy. Also, one can see that the value of \( \delta \) does not depend on the estimation approaches. Hence, we state the following theorem.

**Theorem 3.2:** If \( I^0_i(1) \) is the only information available to the other nodes to estimate the value of \( x_i(0) \), the GPAC algorithm achieves \( (\epsilon, \delta) \)-data-privacy, where \( \epsilon \) and \( \delta \) satisfy

\[
\delta = \max_{\theta_i(0) \in \Theta} \int_{\hat{\theta}_i(0) - \epsilon}^{\hat{\theta}_i(0) + \epsilon} f_{\theta_i(0)}(y)dy
\]

and \( \lim_{\epsilon \to 0} \delta = 0 \).

**Remark 3.3:** It should be noticed that the results in the above two theorems are obtained under the assumption that the topology information is unknown to the nodes. If the assumption is relaxed, the above results could not be true for the GPAC algorithm in some cases. For example, if the topology information is available and \( N_i \subseteq N_j \), then \( x_i^0(1) + \sum_{l \in N_i} w_{il} x_i^+(0) = \frac{w_{ij}}{w_{ii}} x_j^0(1) \) in (17) are available to node \( j \). It leads to that the value of \( \theta_i(1) \) is released, which may decrease the uncertainty of \( \theta_i(0) \) due to the correlation between them. Then, \( f_{\theta_i(0)}(y) \) in (18) will have a smaller variance than \( f_{\theta_i(0)}(y) \), such that \( \delta \) increases w.r.t. \( k \) in this case. Therefore, (12) and (13) are no longer guaranteed.

From the above theorem, one sees that \( \delta \) depends only on \( f_{\theta_i(0)}(y) \) and \( \epsilon \) since the estimation \( \hat{\theta}_i(0) \) can be any value in
the domain of $\theta_i(0)$. Thus, $\delta$ is a function of $f_{\theta_i(0)}(y)$ and $\epsilon$, i.e., $\delta = \delta(f_{\theta_i(0)}(y), \epsilon)$. Based on Definition 2.2, a smaller $\delta$ can provide a higher $(\epsilon, \delta)$-data-privacy for any given $\epsilon$. Then, we aim to find the optimal distribution for $\theta_i(0)$ such that the algorithm can achieve the highest $(\epsilon, \delta)$-data-privacy.

**B. Optimal Noise Distribution**

In this subsection, we find an optimal distribution for the noise adding process in the sense of achieving the highest $(\epsilon, \delta)$-data-privacy for the GPAC algorithm. Note that a smaller $\epsilon$ means a higher accuracy estimation. It means that when $\epsilon$ becomes smaller, the value of $\delta$ is more important for the privacy preservation. Hence, we define the optimal distribution for privacy concerns as follows.

**Definition 3.4:** Let $f_{\theta_i(0)}^*(y)$ be the optimal distribution of $\theta_i(0)$, it means that for any given distribution $f_{\theta_i(0)}(y)$, there exists an $\epsilon$ such that $\delta(f_{\theta_i(0)}^*(y), \epsilon) < \delta(f_{\theta_i(0)}(y), \epsilon)$ holds for $\forall \epsilon \in (0, \epsilon_1]$.

To obtain the optimal distribution described in Definition 3.4, we define $\arg \min_{f_{\theta_i(0)}(y)} \delta = f_{\theta_i(0)}^*(y)$. Then, we formulate the following minimization problem,

$$\min_{f_{\theta_i(0)}(y)} \delta \quad \text{s.t.} \quad \mathbb{E}\{\theta_i(0)\} = 0,$$

$$\mathbb{V}\{\theta_i(0)\} = \sigma^2. \tag{24}$$

The solution of (24) is the optimal distribution for the added noises with a given mean and variance in terms of $(\epsilon, \delta)$-data-privacy for the GPAC algorithm.

**Theorem 3.5:** If $\theta_i(0)$ is the only information available to node $j$ to estimate the value of $x_i(0)$, then the optimal solution of problem (24) is that

$$f_{\theta_i(0)}^*(y) = \begin{cases} \frac{1}{2\sqrt{3}\sigma}, & \text{if } y \in [-\sqrt{3}\sigma, \sqrt{3}\sigma], \\ 0, & \text{otherwise}, \end{cases} \tag{25}$$

i.e., given the finite variance, the uniform distribution is optimal.

**Proof:** We prove this theorem by contradiction. Without loss of generality, we assume that $\sigma^2 = \frac{1}{2}$. Let $f_1(y)$ and $f_2(y)$ be the PDF of two random variables with mean 0 and variance $\sigma^2 = \frac{1}{2}$, and they follow a uniform and non-uniform distribution, respectively. Clearly, we have

$$f_1(y) = \begin{cases} \frac{1}{2}, & \text{if } y \in [-1, 1], \\ 0, & \text{otherwise}. \end{cases} \tag{26}$$

Suppose that the non-uniform distribution $f_2(y)$ is the optimal distribution. From Definition 3.4, there exists an $\epsilon_2$, such that

$$\max_{t \in R} \int_{t-\epsilon}^{t+\epsilon} f_1(y)dy > \max_{t \in R} \int_{t-\epsilon}^{t+\epsilon} f_2(y)dy, \tag{27}$$

holds for $\forall \epsilon \in (0, \epsilon_2]$. Since the above equation holds for arbitrarily small value of $\epsilon$, we infer that

$$\max_{y \in R} f_1(y) > \max_{y \in R} f_2(y).$$

Since $f_1(y)$ is a uniform distribution satisfying (26),

$$f_1(y) - f_2(y) > 0, \quad y \in [-1, 1].$$

It directly follows that

$$\int_{-1}^{1} f_1(y)dy - \int_{-1}^{1} f_2(y)dy > 0. \tag{28}$$

From the definition of a PDF, we have $\int_{-1}^{1} f_1(y)dy = 1$. Then, we infer from (28) that

$$\int_{-1}^{1} f_2(y)dy < 1. \tag{29}$$

Since both $f_1(y)$ and $f_2(y)$ have mean 0 and variance $\sigma^2 = \frac{1}{2}$, we have

$$\int_{-\infty}^{\infty} f_1(y) y^2 dy - \int_{-\infty}^{\infty} f_2(y) y^2 dy = 0, \tag{30}$$

which means that

$$\int_{-1}^{1} (f_1(y) - f_2(y)) y^2 dy = \left(\int_{-\infty}^{1} + \int_{1}^{\infty}\right) f_2(y) y^2 dy. \tag{31}$$

For the left hand side of (31), we have

$$\int_{-1}^{1} (f_1(y) - f_2(y)) y^2 dy \leq \int_{-1}^{1} (f_1(y) - f_2(y)) dy$$

$$= 1 - \int_{-1}^{1} f_2(y)dy. \tag{32}$$

For the right hand side of (31), since we have $\int_{-\infty}^{\infty} f_2(y)dy = 1$ and (29), it holds that

$$\left(\int_{-\infty}^{1} + \int_{1}^{\infty}\right) f_2(y) y^2 dy \leq 1 - \int_{-1}^{1} f_2(y)dy. \tag{33}$$

Combining (31), (32) and (33) renders a contradiction that

$$1 - \int_{-1}^{1} f_2(y)dy < \int_{-1}^{1} (f_1(y) - f_2(y)) y^2 dy$$

$$< 1 - \int_{-1}^{1} f_2(y)dy. \tag{34}$$

Hence, we cannot find a non-uniform distribution $f_2(y)$ such that the value of $\delta$ is smaller than that under uniform distribution $f_1(y)$. It means that, given the finite variance, the uniform distribution is the optimal solution of (24). Then, based on the definition of uniform distribution, it is not difficult to obtain (25).

We thus have completed the proof.

We have known that for the existing PPAC algorithm proposed in [16], the normal distribution noises is used in the noise adding process. It follows from Theorem 3.2 that the PPAC algorithm provides $(\epsilon, \delta)$-data-privacy with

$$\delta = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy.$$
If we use the uniform distribution noises to substitute the normal distribution noises, it can still provide \((\epsilon, \delta)\)-data-privacy, where \(\delta = \frac{\epsilon}{\sqrt{2\sigma}}\). Clearly, given a small \(\epsilon \ll \sigma^2\), we have
\[
\frac{\epsilon}{\sqrt{2\sigma}} < \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left(-\frac{y^2}{2\sigma^2}\right) dy,
\]
which means that the privacy of PPAC is enhanced.

C. Privacy Compromission

In this subsection, we reveal that for the GPAC algorithm, when \(I^1_i(k)\) (including more information than \(I^0_i(k)\), e.g., the topology information and information used in consensus process) is available to other nodes for estimation, the exact initial state of node \(i\) can be perfectly inferred, and thus the privacy of the initial state is compromised.

**Theorem 3.6:** If the information set \(I^1_i(k)\) of node \(i\) is available to the other nodes for estimation, then
\[
\delta(k) \geq \max_{\hat{\theta}_i(0) \in \Theta} \int_{\hat{\theta}_i(0) - \epsilon}^{\hat{\theta}_i(0) + \epsilon} f_{\hat{\theta}_i(0)|\theta_i(1),\ldots,\theta_i(k)}(y) dy, \forall k \geq 0,
\]...

where \(f_{\hat{\theta}_i(0)|\theta_i(1),\ldots,\theta_i(k)}(y)\) is the conditional PDF of \(\hat{\theta}_i(0)\) given conditions \(\theta_i(1), \ldots, \theta_i(k)\). Then, if \(\sum_{\ell=0}^{\infty} \theta_i(\ell) = 0\), we have \(\delta = 1\), i.e., \(x_i(0)\) is disclosed and the privacy is compromised.

**Proof:** Based on the information set \(I^1_i(k)\), the information of weights of states \(s(5)\) is available. That is the state sequence \(x_i(1), x_i(2), \ldots, x_i(k)\) of node \(i\) is released to other nodes. Then, with (4), one obtains the values of \(\hat{\theta}_i(1), \hat{\theta}_i(2), \ldots, \hat{\theta}_i(k)\). Thus, when \(k > 0\), all the adding noises and the states of node \(i\) are available to other nodes, except \(x_i(0)\) and \(\theta_i(0)\).

Then, under information set \(I^1_i(k)\), using (14), we have
\[
\Pr\left\{ |\hat{\theta}_i(0) - \theta_i(0)| \leq \epsilon \right\}_{I^1_i(k)} = \int_{\hat{\theta}_i(0) - \epsilon}^{\hat{\theta}_i(0) + \epsilon} f_{\hat{\theta}_i(0)|\theta_i(1),\ldots,\theta_i(k)}(y) dy.
\]

According the definition of \(\delta\), it follows that
\[
\delta(k) \geq \max_{\hat{\theta}_i(0) \in \Theta} \Pr\left\{ |\hat{\theta}_i(0) - \theta_i(0)| \leq \epsilon \right\}_{I^1_i(k)} \geq \max_{\hat{\theta}_i(0) \in \Theta} \int_{\hat{\theta}_i(0) - \epsilon}^{\hat{\theta}_i(0) + \epsilon} f_{\hat{\theta}_i(0)|\theta_i(1),\ldots,\theta_i(k)}(y) dy,
\]
which means that (35) holds.

When \(\sum_{\ell=0}^{\infty} \theta_i(\ell) = 0\), we have
\[
\theta_i(0) = -\sum_{\ell=1}^{\infty} \theta_i(\ell).
\]

Since \(\theta_i(1), \theta_i(2), \ldots, \theta_i(k)\) are available under \(I^1_i(k)\) for any integer \(k\), \(\theta_i(0)\) is inferred with (38) when \(k \to \infty\), i.e., \(\theta_i(0)\) is fixed and no longer a random variable given \(\theta_i(1), \theta_i(2), \ldots, \theta_i(\infty)\). It follows that
\[
\lim_{k \to \infty} \max_{\hat{\theta}_i(0) \in \Theta} \int_{\hat{\theta}_i(0) - \epsilon}^{\hat{\theta}_i(0) + \epsilon} f_{\hat{\theta}_i(0)|\theta_i(1),\ldots,\theta_i(k)}(y) dy = 1,
\]
which implies that \(\delta = 1\). Actually, when both \(x_i^+(0)\) and \(\theta_i(0)\) in (14) are disclosed, \(x_i(0)\) is disclosed.

We thus have completed the proof.

Consider the existing PPAC algorithms, e.g., [16], [21]. One obtains the correlation of the added noises satisfies
\[
\sum_{\ell=0}^{k} \theta_i(\ell) = \theta_i(0) + \sum_{\ell=1}^{k} [\phi^0 \nu_i(\ell) - 0^{\ell-1} 
u_i(\ell - 1) = \nu_i(0) - \phi^0 \nu_i(0) + \phi^1 \nu_i(1) - \phi^0 \nu_i(1) + \phi^2 \nu_i(2) - \ldots - \phi^{k-1} \nu_i(k - 1) + \phi^k \nu_i(k)]
\]
where \(\nu_i(k)\) is a random variable with fixed mean \((= 0)\) and variance \((= \sigma^2)\). Given \(\theta_i(1), \ldots, \theta_i(k)\), we obtain that \(\theta_i(0) = \phi_i(k) - \sum_{\ell=1}^{k} \theta_i(\ell)\), where \(\sum_{\ell=1}^{k} \theta_i(\ell)\) is known. Then, we have
\[
\Pr\left\{ |\hat{\theta}_i(0) - \theta_i(0)| \leq \epsilon \right\}_{I^1_i(k)} = \Pr\left\{ |\phi_i(k) - \phi_i(k)| \leq \epsilon \right\}_{I^1_i(k)} = \int_{\phi_i(k) - \epsilon}^{\phi_i(k) + \epsilon} f_{\phi_i(k)}(y) dy.
\]
and
\[
\max_{\hat{\theta}_i(0) \in \Theta} \int_{\hat{\theta}_i(0) - \epsilon}^{\hat{\theta}_i(0) + \epsilon} f_{\hat{\theta}_i(0)|\theta_i(1),\ldots,\theta_i(k)}(y) dy = \max_{\hat{\theta}_i(0) \in \Theta} \int_{\hat{\theta}_i(0) - \epsilon}^{\hat{\theta}_i(0) + \epsilon} f_{\hat{\theta}_i(0)}(y) dy,
\]
which satisfies (35). When \(k \to \infty\), we have
\[
\lim_{k \to \infty} \max_{\hat{\theta}_i(0) \in \Theta} \int_{\hat{\theta}_i(0) - \epsilon}^{\hat{\theta}_i(0) + \epsilon} f_{\hat{\theta}_i(0)}(y) dy = 1,
\]
since the variance of \(\phi\) satisfies \(\lim_{k \to \infty} \phi^2 \sigma^2 = 0\), and thus \(\delta = 1\). Therefore, it further verifies the result given in Theorem 3.6.

D. Further Discussion on Privacy

Differential privacy is a well-known and widely used privacy concept in computer and communication area [20], and it has been employed in control and network systems recently [23]. A differentially private algorithm promises that any two similar/close inputs will have approximately the same outputs, so that an adversary cannot infer from the data output with a high probability whether the data are associated with a single user or not. It has been proved by Nozari et al. in [14] that nodes in the network system cannot simultaneously converge to the average of their initial states and preserve differential privacy of their initial states. This motivated us to develop the definition of the \((\epsilon, \delta)\)-data-privacy. The proposed \((\epsilon, \delta)\)-data privacy can be used to reveal the relationship between the
the maximum data disclosure probability (δ) under a given estimation accuracy range (ε).

Consider the general noise adding mechanism that added a random noise to the initial data for data publishing. It is well known that when the adding noise is Laplacian noise, the mechanism ensures differential privacy, but if the noise is Gaussian or Uniform noise, the differential privacy cannot be guaranteed. Hence, the uniform noise is not good in the sense of differential privacy. However, in term of (ε, δ)-data-privacy, it is shown in this paper that both the Gaussian and Uniform noise are (ε, δ)-data-private, and using the Uniform noise can achieve the highest privacy. Clearly, the privacy of (ε, δ)-data-privacy is different from that of differential privacy. It is worth to investigate the relationship between these two kinds of privacy definition in theory, which beckons further investigation.

IV. OPAC ALGORITHM

In this section, we design an OPAC algorithm to achieve the highest (ε, δ)-data-privacy, and at the same time to avoid privacy to be compromised even if the information 𝑈 COVID1 (∞) of each node 𝑖 is available to other nodes.

A. Algorithm Design

From the privacy analysis in the above section, we note that the uniform distribution is optimal for the added noise in terms of achieving the highest (ε, δ)-data-privacy with δ = 1/√3σ (given variance σ). Hence, in each iteration of the OPAC algorithm, we will use uniform distribution noise. We also note that privacy is compromised when 𝑈 COVID1 (∞) is available. It is because that the nodes can use 𝑈 COVID1 (∞) to obtain the real values of θ̃(1), θ̃(2), ..., θ̃(∞), and then use the correlation ∑∞ k=0 θ̃(k) = 0 to infer θ̃(0), and thus the value of 𝑥̃(0) is revealed. To avoid the privacy compromission in this case, we introduce a secret continuous function 𝑅 𝑃 𝑃 𝑃 𝑃 : ⎣ ⎦ → ⎣ ⎦ for node 𝑖 with respect to its neighbor node 𝑗, Suppose that 𝑅 𝑃 𝑃 𝑃 𝑃 (𝑧) and 𝑅 𝑃 𝑃 𝑃 𝑃 𝑃 (𝑧) are only available to nodes 𝑖 and 𝑗, and 𝑅 𝑃 𝑃 𝑃 𝑃 𝑃 (𝑧) may or may not equal to 𝑅 𝑃 𝑃 𝑃 𝑃 𝑃 (𝑧). Then, the OPAC algorithm is described as follows.

B. Convergence and Privacy Analysis

In this subsection, we analyze the convergence and the privacy of the OPAC algorithm.

Theorem 4.1: Using the OPAC algorithm, we have (3) holds for ∀𝑖 ∈ 𝑉, i.e., an exact average consensus is achieved.

Proof: From Theorem 4.1 of [21], we know that if the added noises in (4) are bounded and decaying, and the sum of all nodes’ added noises equals zero, then average consensus can be achieved. In the following, we prove that the added noises used for the OPAC algorithm satisfy these conditions.

We first prove that the added noises are bounded and exponentially decaying. Clearly, θ̃(0) = ν̃(0) ∈ [−√3σ, √3σ] is bounded. Since each 𝑅 𝑃 𝑃 𝑃 𝑃 (𝑧) is continuous function, its value is bounded for any given 𝑧. Then, it follows from (41) that 𝑁 𝑃 𝑃 𝑃 𝑃 (0) is bounded. For 𝑘 ≥ 1, because 𝑁 𝑃 𝑃 𝑃 𝑃 (𝑘) is selected from interval [−√3σ, √3σ] and θ̃(𝑘) is generated by (42), it is not difficult to infer that each θ̃(𝑘) is bounded. Meanwhile, it follows from (42) that

$$\lim_{k \to \infty} |θ̃(k)| ≤ \lim_{k \to \infty} |g^k ν̃(k) - g^{k-1} ν̃(k-1)|$$

$$≤ \lim_{k \to \infty} |g^k \sqrt{3}σ + g^{k-1} \sqrt{3}σ| = 0,$$

which means that the noises are decaying and converge to 0.

Next, we prove that the sum of all nodes’ added noises equals to zero. Note that

$$\sum_{k=1}^{∞} θ̃(k) = \sum_{k=1}^{∞} θ̃(0) + \sum_{k=1}^{∞} θ̃(1) + \sum_{k=2}^{∞} θ̃(k)$$

$$= \sum_{i=1}^{n} ν̃(0) + \sum_{i=1}^{n} (g^i ν̃(1) - ν̃(i))$$

$$= \sum_{i=1}^{n} ν̃(0) - \sum_{i=1}^{n} ν̃(i),$$

where we have used the fact that $g^∞ ν̃(i) = 0$. Substituting (41) into the above equation yields that

$$\sum_{i=1}^{n} ν̃(i) = \sum_{i=1}^{n} ν̃(0) - \sum_{j \in N_i} (F̃(z̃_ij) - F̃(z̃_ji))$$

Since for each pair of $F̃(z̃_ij) - F̃(z̃_ji)$ using in node $i$, there exists a pair of $F̃(z̃_ij) - F̃(z̃_ji)$ with negative value using

Algorithm 1: OPAC Algorithm

1. Initialization: Each node $i$ selects a uniform distribution random variable $ν̃(0)$ from interval $[−\sqrt{3}σ, \sqrt{3}σ]$, and arbitrarily selects a constant sequence $z̃_ij ∈ R$ for $j ∈ N_i$.

2. Let $θ̃(0) = ν̃(0)$ and $x̃_i(0) = x_i(0) + θ̃(0)$. Then, each node $i$ transmits $x̃_i(0)$ and $z̃_ij$ to its neighbor node $j$.

3. Each node $i$ calculates $θ̃(0)$ by

$$θ̃(0) = ν̃(0) - \sum_{j \in N_i} [F̃(z̃_ij) - F̃(z̃_ji)], ∀i ∈ V.$$  (41)

4. Iteration: Each node updates its state with (5).

5. Each node generates a uniform distribution random variable $ν̃(k)$ from interval $[−\sqrt{3}σ, \sqrt{3}σ]$ for $k ≥ 1$.

6. Each node $i$ uses $θ̃(k)$ in (4) to get $x̃_i(k)$, where

$$θ̃(k) = \begin{cases} \rho g^k ν̃(1) - ν̃(0), & \text{if } k = 1; \\ g^k ν̃(k) - g^{k-1} ν̃(k-1), & \text{if } k ≥ 2, \end{cases}$$  (42)

where $ρ ∈ (0, 1)$ is a constant for all nodes.

7. Each node $i$ communicates with its neighbors with $x̃_i(k)$.

8. Let $k = k + 1$ and go to step 4.
in node \( j \), it follows that

\[
\sum_{i=1}^{n} \sum_{j \in N_i} [F_{ji}(z_{ji}) - F_{ij}(z_{ij})] = 0.
\]

Hence, we have \( \sum_{j=1}^{n} \sum_{k=0}^{\infty} \theta_i(k) = 0 \).

Thus, the proof is completed.

The following theorem can be obtained from Theorem 3.2 directly, since OPAC is one of the GPAC algorithm.

**Theorem 4.2:** If \( T_i^0 \) is the only information available to the other nodes to estimate the value of \( x_i(0) \), then the OPAC algorithm achieves \((\epsilon, \delta)\)-data-privacy, where \( \delta = \frac{\epsilon}{\sqrt{3\sigma}} \) and \( \lim_{\epsilon \to 0} \delta = 0 \).

Then, the following theorem shows that under \( T_i^1 \), the privacy compromise can be avoided by OPAC.

**Theorem 4.3:** Suppose that the information set \( T_i^j \) of node \( i \) is available to the other nodes and each node has at least two neighbors (i.e., \( |N_i| \geq 2 \) for all \( i \in V \)). Then, the privacy compromise can be avoided by OPAC.

**Proof:** It has been known that when \( T_i^1 \) of node \( i \) is available to other nodes, its neighbor node \( j \) can obtain the real values of \( \theta_i(1), \theta_i(2), \ldots, \theta_i(\infty) \). Then, the value of \( \sum_{k=1}^{\infty} \theta_i(k) \) is released. Note that

\[
\sum_{k=1}^{\infty} \theta_i(k) = (g^1 \nu_i(1) - \tilde{\nu}_i(0)) + \sum_{k=2}^{\infty} \theta_i(k)
\]

\[
= (g^1 \nu_i(1) - \tilde{\nu}_i(0)) + (g^{\infty} \nu_i(\infty) - g^1 \nu_i(1))
\]

\[
= \tilde{\nu}_i(0).
\]

It means that the value of \( \tilde{\nu}_i(0) \) is released and available to node \( j \). From (41), one sees that \( \tilde{\nu}_i(0) \neq \theta_i(0) \) and

\[
\tilde{\nu}_i(0) = \theta_i(0) - \sum_{j \in N_i} [F_{ji}(z_{ji}) - F_{ij}(z_{ij})].
\]

Since \( |N_i| \geq 2 \) and only \( F_{ji} \) and \( F_{ij} \) is known to node \( j \), there exists \( F_{ji}(z_{ji}) - F_{ij}(z_{ij}) \) for \( j \in N_i \). In (43) is not known by node \( j \). Meanwhile, \( F_{ji}(z_{ji}) - F_{ij}(z_{ij}) \) has domain \( R \), thus one infers that for any \( c \in [-\sqrt{3\sigma}, \sqrt{3\sigma}] \),

\[
\Pr\{\tilde{\theta}_i(0) = c|\tilde{\nu}_i(0)\} = \Pr\{\theta_i(0) = c\}.
\]

Hence, even if the value of \( \tilde{\nu}_i(0) \) is released, node \( j \) cannot increase the estimation accuracy of \( \theta_i(0) \) with (43). One thus concludes that based on the OPAC algorithm, the privacy compromise is avoided.

We thus have completed the proof.

If node \( i \) has only one neighbor node \( j \), node \( j \) can infer the value of \( F_{ji}(z_{ji}) - F_{ij}(z_{ij}) \). Then, from (43), node \( j \) can obtain the value of \( \theta_i(0) \) and \( x_i(0) \) when \( \tilde{\nu}_i(0) \) is known.

**Remark 4.4:** From the above two theorems, one sees that using OPAC algorithm, we have \( \delta = \frac{\epsilon}{\sqrt{3\sigma}} \), which is the optimal privacy that can be achieved from solving problem (24). Furthermore, \( \delta = \frac{\epsilon}{\sqrt{3\sigma}} \) can be guaranteed by OPAC algorithm under \( T_i^1(\infty) \). Thus, OPAC algorithm can achieve much higher \((\epsilon, \delta)\)-data-privacy than the existing PPAC.

\[1\]This is the main difference between OPAC and PPAC algorithm, and the main reason why OPAC can avoid privacy compromise.

V. **Performance Evaluation**

In this section, we conduct simulations to verify the obtained theoretical results and evaluate the performance of the proposed OPAC algorithm.

A. **Simulation Scenario**

Consider the network with 50 nodes which are randomly deployed in a 100m x 100m area, and the maximum communication range of each node is 30m. We consider the normal distribution and uniform distribution of the added noises, respectively, where the mean and variance of them are set 0 and \( \sigma^2 = 1 \). We set \( \sigma = 0.9 \). The initial states of the nodes are randomly selected from \([0, 10] \). The function, \( d(t) = \max_{i \in V} |x_i(t) - \bar{x}| \), is defined as the maximum difference between the nodes’ states and the average value.

B. **Verification**

Fig. 1(a) compares the convergence speed of the PPAC algorithm using normal and uniform distribution noises, in which the basic design is the same as PPAC proposed in [16]. It is observed that under the two different distributions, the PPAC algorithm has the same convergence speed. This justifies that the convergence speed only depends on the eigenvalues of the weighted matrix \( W \) and the value of \( \theta \) as proved in [16].

Fig. 1(b) compares the \((\epsilon, \delta)\)-data-privacy under \( T_i^0 \) with normal and uniform distribution noises. In simulation, we conduct 10,000 simulation runs. For each run, one node first generates a state \( \theta_i(0) \) randomly with the given distribution, and the other node generates 10,000 random numbers with the same distribution and use them as the estimation of \( \theta_i(0) \) (i.e., \( \tilde{\theta}_i(0) \)). Then, one get the probability of \( |\tilde{\theta}_i(0) - \theta_i(0)| \leq \epsilon \) in each run, and we use the maximum probability among these in all runs simulation as the value of \( \delta \). For the theoretical results, we use (23) to calculate the value of \( \delta \) under two different distributions. Clearly, one can observe from Fig. 1(b) that uniform distribution is much better than normal distribution in the sense of \((\epsilon, \delta)\)-data-privacy. It is also observed that \( \delta \) in simulation matches its value in theory.

Fig. 1(c) compares the \((\epsilon, \delta)\)-data-privacy under \( T_i^1 \) using normal and uniform distribution noises. The simulations here are conducted similarly as those in Fig. 1(b), except that when the iteration increases, the variance of the noises will be changed to \( \sigma^2 = \sigma^2 k \) since (40) will be used for estimation at iteration \( k \). We use (23) to calculate the value of \( \delta \), and the corresponding results are denoted as theoretical results. Both in simulation and theory, we set \( \epsilon = 0.2 \). As shown in Fig. 1(c), the maximum disclosure probability increases with iteration and will converge to 1, i.e., the privacy decays with iteration and will eventually be compromised.

C. **Evaluation**

In this subsection, we will evaluate the performance of the OPAC algorithm. Using the same setting as the above subsection, the OPAC algorithm can guarantee the similar privacy as the blue line shown in Fig. 1(b) under \( T_i^1 \). This is because
uniform distribution noise is used in OPAC and the secret function makes the subsequent \((k \geq 1)\) information cannot increase the disclosure probability. Therefore, the OPAC can guarantee much stronger privacy than the GPAC, since it can achieve the same data-privacy under \(I^0_i\) as the GPAC under \(I^1_i\).

Then, we test the convergence of the OPAC algorithm. Set \(F_{ij} = \frac{j+2i}{10}\). As shown in Fig. 2(a), we find that the nodes’ states will converge to the exact average with the OPAC, which means that an exact average consensus can be achieved by the proposed algorithm. Fig. 2(b) compares the convergence speed of the OPAC and PPAC, it is found that they almost have the same convergence speed. Hence, added secret function will not affect the convergence speed.

VI. CONCLUSIONS

In this paper, we investigated the privacy of the GPAC algorithm. We proposed a novel privacy definition, named \((\epsilon, \delta)\)-data-privacy, to depict the relationship between privacy and estimation accuracy, so that the degree of the privacy can be well quantified. We proved that the GPAC algorithm achieves \((\epsilon, \delta)\)-data-privacy, and obtained the closed-form expression of the relationship between \(\epsilon\) and \(\delta\). We also proved that the noise with uniform distribution guarantee a highest privacy when \(\epsilon\) is small enough. We revealed that the privacy will be lost when the information used in each consensus iteration is available to the other nodes. Then, to solve this problem and achieve highest \((\epsilon, \delta)\)-data-privacy, we proposed OPAC algorithm, followed by the convergence and privacy analysis. Lastly, simulations are conducted to demonstrate the efficiency of the proposed algorithm.

REFERENCES

[1] R. Olfati-Saber, J. A. Fax, and R. M. Murray, “Consensus and cooperation in networked multi-agent systems,” *Proceedings of the IEEE*, 95(1): 215–233, 2007.
[2] V. Blondel, J. M. Hendrickx, A. Olshevsky, and J. Tsitsiklis, “Convergence in multiagent coordination, consensus, and flocking,” in *Proc. of IEEE CDC*, 2005.
[3] W. Ren, B. Randal., and A. Ella, “Information consensus in multivehicle cooperative control: Collective group behavior through local interaction,” *IEEE Control Systems Magazine*, 27(2): 71–82, 2007.
[4] F. Pasqualetti, R. Carli, A. Bicchi, and F. Bullo, “Distributed estimation and detection under local information,” in *Proc. of IFAC*, 2010.
[5] G. Mateos, I. Schizas and G. Giannakis, “Distributed recursive least-squares for consensus-based in-network adaptive estimation,” *IEEE Trans. Signal Processing*, 57(11): 4583–4588, 2009.
[6] L. Xiao, S. Boyd, and S. Lall, “A scheme for robust distributed sensor fusion based on average consensus,” in *Proc. of ISIPSN*, 2005.
[7] C. Zhao, J. He, P. Cheng and J. Chen, “Consensus-based energy management in smart grid with transmission losses and directed communication,” *IEEE Trans. Smart Grid*, DOI: 10.1109/TSG.2015.2513772.
[8] J. He, L. Duan, F. Hou, P. Cheng, and J. Chen, “Multi-period scheduling for wireless sensor networks: A distributed consensus approach,” *IEEE Trans. Signal Processing*, 63(7): 1651–1663, 2015.
[9] L. Schenato and F. Fiorentin, “Average timesynch: A consensus-based protocol for clock synchronization in wireless sensor networks,” *Automatica*, 47(9): 1878–1886, 2011.
[10] R. Carli, and S. Zampieri, “Network clock synchronization based on the second order linear consensus algorithm,” IEEE Trans. Automat. Contr., 59(2): 409–422, 2014.

[11] J. He, P. Cheng, L. Shi, and J. Chen, “Time synchronization in WSNs: A maximum value based consensus approach,” IEEE Trans. Automat. Contr., 59(3): 660–674, 2014.

[12] J. Le Ny and G. Pappas, “Differentially private filtering,” IEEE Trans. Automat. Contr., 59(2): 341–354, 2014.

[13] Z. Huang, S. Mitra, and G. Dullerud, “Differentially private iterative synchronous consensus,” in Proc. ACM workshop on Privacy in the electronic society, 2012.

[14] E. Nozari, P. Tallapragada, and J. Cortes, “Differentially private average consensus: Obstructions, trade-offs, and optimal algorithm design,” arXiv preprint arXiv:1512.09039, 2016.

[15] N. Manitara and C. Hadjicostis, “Privacy-preserving asymptotic average consensus,” in Proc. of IEEE ECC, 2013.

[16] Y. Mo, and R. Murray, “Privacy preserving average consensus,” IEEE Trans. Automat Contr., accepted and to appear, 2016.

[17] M. DeGroot, “Reaching a consensus,” Journal of the American Statistical Association, 69(345), 118–121, 1974.

[18] J. Lin, A. S. Morse, and B. D. Anderson, “The multi-agent rendezvous problem,” in Proc. of IEEE CDC, 2003.

[19] A. Olshevsky and J. Tsitsiklis, “Convergence speed in distributed consensus and averaging,” SIAM Review, 53(4): 747–772, 2011.

[20] C. Dwork, “Differential privacy,” in Automata, languages and programming, Springer, 1-12, 2006.

[21] J. He, L. Cai, P. Cheng, M. Xing, J. Pan and L. Shi, “Private and accurate data aggregation against dishonest nodes,” https://arxiv.org/pdf/1609.06381v2.pdf, 2016.

[22] P. Braca, R. Lazzeretti, S. Marano, and V. Matta, “Learning with privacy in consensus + obfuscation,” IEEE Signal Processing Letters, 23(9): 1174-1178, 2016.

[23] J. Cortes, G. Dullerud, S. Han, J. Le Ny, S. Mitra, and G. Pappas, “Differential privacy in control and network systems,” in Proc. of IEEE CDC, 2016.