Differentiated Output-Based Privacy-Preserving Average Consensus

Jieming Ke, *Student Member, IEEE*, Jimin Wang, *Member, IEEE*, and Ji-Feng Zhang, *Fellow, IEEE*

***Abstract*—This letter investigates the differentiated output-based privacy-preserving average consensus problem over digraphs. A new stochastic obfuscation algorithm is proposed to achieve better privacy-preserving effect. When the output messages for at least one out-neighbour are not leaked, the algorithm can be designed to achieve any pre-given consensus accuracy and privacy-preserving level simultaneously by properly selecting the private weights. Even if all the output messages are leaked, the algorithm can still ensure that each agent's initial state is protected to a certain extent. The mean square convergence of the algorithm is proved. The efficiency of the algorithm is verified by a numerical example.**

***Index Terms*—Privacy-preserving, cooperative control, multi-agent systems, average consensus, Fisher information.**

I. Introduction

COOPERATIVE and coordinated control of multi-agent systems has been a hot area of research. As the basis and core algorithm, average consensus algorithms play a more and more important role in many applications and services [1], [2]. In consensus algorithms, the initial state of each agent is an important fusion source and may contain sensitive information or personal privacy [3]. However, the explicit information interactions in classical distributed consensus algorithms ignore such privacy-preserving issues. Therefore, it is urgent to protect each agent's initial state in the classical distributed consensus algorithms.

At present, many researches have been done on the privacy-preserving issue in control systems [4]. Many methods have been proposed to achieve this goal. The existing approaches can be roughly divided into three categories: the structure technique, the deterministic transformation technique, and the stochastic obfuscation technique. The structure technique is based on the standard observability for state reconstruction and input observability for input re-identification. For example, by introducing a virtual state, the algorithm proposed in [5] decomposes each agent's state into two substates and obtains a secure structure against the potential passive attackers. However, the privacy of each agent cannot be protected when all its neighbors are directly connected to the potential passive attackers. For the deterministic transformation technique, the homomorphic encryption scheme is one of the frequently used methods. By using homomorphic encryption, the privacy issues of control systems have been studied, such as cooperative control [6], fast covariance intersection [7] and distributed optimization [8], [9], etc. The encryption method provides high-dimensional security, but leads to heavy load on computing and communication, which may be unapplicable to networks with limited energy and resource.

Roughly speaking, the stochastic obfuscation technique mainly protects the real state by adding noises [10]. Among others, the differential privacy mechanism has been widely utilized to privacy-preserving control and optimization. For example, differentially private average consensus algorithms [11], [12], [13] and differentially private distributed algorithms for stochastic aggregative games have been considered [14]. It is worth mentioning that in existing methods of the stochastic obfuscation technique, an agent sends the same output messages to all its out-neighbours. Then, a honest-but-curious neighbour can obtain all the output messages of the agent, and it is enough for an eavesdropper to eavesdrop one output channel for the output messages. The limitation motivates us to design a new stochastic obfuscation algorithm that sends differentiated output messages to different out-neighbours.

Motivated by the limitation of the approaches mentioned above, we propose a new differentiated output-based privacy-preserving average consensus protocol over digraphs in this letter. The privacy-preserving level of the initial state is quantified by Fisher information. For each agent, besides adding noise to its state in classical privacy-preserving consensus protocols, the random private weights of the algorithm are also designed. Our contributions are as follows.

1) A new privacy-preserving distributed consensus algorithm over digraphs is proposed, which can achieve a stronger privacy-preserving effect compared with the existing

Manuscript received 28 October 2022; revised 27 December 2022; accepted 20 January 2023. Date of publication 30 January 2023; date of current version 9 February 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFA0703800, and in part by the National Natural Science Foundation of China under Grant 62203045 and Grant T2293770. Recommended by Senior Editor M. Guay. *(Corresponding author: Jimin Wang.)*

Jieming Ke and Ji-Feng Zhang are with the Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: kejieming@amss.ac.cn; jif@iss.ac.cn).

Jimin Wang is with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China (e-mail: jimwang@ustb.edu.cn).

Digital Object Identifier 10.1109/LCSYS.2023.3240655

literature [10], [11], [12], [13]. Specifically, the algorithm can be designed to achieve any pre-given consensus accuracy and privacy-preserving level simultaneously if the output messages for at least one out-neighbour are not leaked.

2) The mean square convergence of the algorithm can be ensured, and simultaneously, the privacy of the agents' initial states can be preserved even if all input and output messages of the agent are leaked to the potential passive attackers.

The rest of this letter is organized as follows. In Section II, the problem formulation is given. The main results, including convergence analysis and privacy analysis, are presented in Sections III and IV respectively. In Section V, one example is provided to verify the efficiency of our protocol. Finally, a brief conclusion is given in Section VI.

II. Problem Formulation

A. Preliminaries

In this letter, we use $\mathbb{R}^n$ and $\mathbb{R}^{n\times n}$ to denote the sets of n-dimensional vectors and n-dimensional real square matrices, respectively. The notation $\otimes$ stands for Kronecker product. I_n represents n-dimensional identity matrix. $\mathbf{1}_n$ and $\mathbf{0}_n$ are n-dimensional column vectors with all elements being 1 and 0, respectively. The notation $\mathrm{diag}(x_1,\ldots,x_n)$ denotes the diagonal matrix with diagonal elements being $x_1,\ldots,x_n$. For a random variable $X\in\mathbb{R}$, $\mathbb{E}X$ and $\mathrm{Var}(X)$ denote the expectation and the variance of X, respectively. $\mathrm{Normal}(\mu,\sigma^2)$ denotes the normal distribution with mean μ and variance σ^2.

A digraph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ with N agents and M edges is introduced to describe the relationship between agents. $\mathcal{V}=\{1,\ldots,N\}$ is the agent set. $\mathcal{E}=\mathcal{V}\times\mathcal{V}$ is the edge set. An ordered pair $(i,j)\in\mathcal{E}$ represents that the agent j can receive messages from the agent i. Besides, define $\mathcal{N}_i^{\mathrm{in}}=\{j:(j,i)\in\mathcal{E}\}$ and $\mathcal{N}_i^{\mathrm{out}}=\{j:(i,j)\in\mathcal{E}\}$ as the set of in-neighbours and out-neighbours of the agent i, respectively. N_i represents the number of the elements in $\mathcal{N}_i^{\mathrm{out}}$. Moreover, a path in the digraph is an ordered sequence $(v_1,\ldots,v_k)$ with $(v_i,v_{i+1})\in\mathcal{E}$ for all $i=1,\ldots,k-1$. The graph $\mathcal{G}$ is called strongly connected if there is a path from i to j as well as from j to i for any agents i and j.

B. Problem Statement

Consider a set of N agents coupled by a communication graph $\mathcal{G}$. The dynamics of the agent i is described by

$$x_i(t)=x_i(t-1)+u_i(t-1), \tag{1}$$

where $x_i(t)\in\mathbb{R}$ is the agent i's state. The initial state $x_i(0)=\theta_i$. The control input $u_i(t)$ is designed to compute the average of the agents' initial states $\frac{1}{N}\sum_{i=1}^{N}\theta_i$. The input design relies on the message transmission in the communication graph $\mathcal{G}$. Below, we give the definition of the asymptotic unbiased mean square average consensus of the system (1).

Definition 1 (Asymptotic Unbiased Mean Square Average Consensus) [2]: For given initial states $\theta_1,\ldots,\theta_N$, the system (1) is said to achieve asymptotic unbiased mean square average consensus with ρ-accuracy, if there exists a random variable x^* with $\mathbb{E}x^*=\frac{1}{N}\sum_{i=1}^{N}\theta_i$ and $\mathrm{Var}(x^*)=\rho$ such that $\lim_{t\to\infty}\mathbb{E}(x_i(t)-x^*)^2=0$ for any agent $i\in\mathcal{V}$.

To achieve the asymptotic unbiased mean square average consensus, the following assumption about the digraph $\mathcal{G}$ is required.

Assumption 1: The digraph $\mathcal{G}$ is strongly connected.

In the average consensus protocol designed, the initial state θ_i is always considered as the sensitive information that the potential passive attackers are probably interested in. Generally, there are two types of potential passive attackers: honest-but-curious neighbours and eavesdroppers. Honest-but-curious neighbours probably compute the sensitive information from received messages. Eavesdroppers eavesdrop output messages in the network to access the sensitive information. In the traditional stochastic obfuscation methods [10], [11], [12], [13], the agent i sends the same output messages

$$y_i(t)=x_i(t)+d_i(t) \tag{2}$$

to all its out-neighbours. In that case, if output messages for an out-neighbour are leaked, then the potential passive attackers obtain all the output messages of the agent. To promote the privacy-preserving level, we try to design a new algorithm that sends differentiated output messages to different out-neighbours. The privacy-preserving level of the average consensus protocol is described by Fisher information.

Definition 2 (Privacy-Preserving Level) [15]: The privacy leakage of the sensitive information $\theta\in\mathbb{R}$ from the output message y is quantified by the Fisher information

$$\mathcal{I}_y(\theta)=\mathbb{E}\left[\frac{\partial\ln(p(y|\theta))}{\partial\theta}\right]^2=\int_{y\in\mathcal{Y}}p(y|\theta)\left[\frac{\partial\ln(p(y|\theta))}{\partial\theta}\right]^2\mathrm{d}y,$$

where $p(y|\theta)$ is the conditional probability density function of y given θ, and $\mathcal{Y}$ is the set of all possible y. Then, the privacy-preserving level can be quantified by $\mathcal{I}_y^{-1}(\theta)$.

Remark 1: Fisher information is appropriate to quantify the privacy leakage, because for any unbiased estimate of x denoted by $\hat{\theta}(y)$, it holds that $\mathbb{E}(\theta-\hat{\theta}(y))^2\geq\mathcal{I}_y^{-1}(\theta)$.

C. Algorithm Design

In the subsection, the privacy-based output messages $y_{ij}(t)$ and the control input $u_i(t)$ are designed to prevent the privacy leakage problem mentioned above.

We firstly introduce public balanced weights a_{ij} that satisfy $\sum_{j\in\mathcal{N}_i^{\mathrm{in}}}a_{ji}=\sum_{j\in\mathcal{N}_i^{\mathrm{out}}}a_{ij}=:\Delta_i$ for all agents i. The public balanced weights a_{ij} can be generated from the distributed imbalance-correcting algorithm in [16].

For the privacy-based output messages, we try to design differentiated output messages for different out-neighbours. To realize the differentiated output design, we introduce private weights w_{ij} for the agent i, which is generated by

$$\left[w_{ij_1},\quad\ldots,\quad w_{ij_{N_i}}\right]^{\top}\sim\mathrm{Normal}\left(\mu_i^w,\frac{(\sigma_i^w)^2}{\theta_i}\Sigma_{N_i}^w\right), \tag{3}$$

where $j_1,\ldots,j_{N_i}$ are the out-neighbours of the agent i, $\sigma_i^w>0$,

$$\mu_i^w=\frac{\Delta_i(M-NN_i)}{NN_i^2}\mathbf{1}_{N_i}=\left[\frac{\Delta_i}{N_i^2}\left(\frac{\sum_{j=1}^{N}N_j}{N}\right)-\frac{\Delta_i}{N_i}\right]\mathbf{1}_{N_i},$$

and $\Sigma_n^w=I_n-\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\top}$.

Remark 2: In (3), if $\theta_i = 0$, then it can be replaced with $\theta' \sim \text{Normal}(0, \varepsilon)$ for some small $\varepsilon > 0$ in the private weight design. For expression convenience, we assume that the sensitive information $\theta_i \neq 0$. Besides, for Σ_n^w, it holds that $\text{rank}(\Sigma_n^w) = n-1$ and $\Sigma_n^w \mathbf{1}_n = \mathbf{0}_n$.

Remark 3: If centralized information on M and N is available, μ_i^w can be directly obtained. Otherwise, μ_i^w can be obtained by using an average consensus algorithm [5] to calculate the average of N_i.

Based on the private weights w_{ij} and the public balanced weights a_{ij}, the output messages $y_{ij}(t)$ are designed as

$$\begin{cases} y_{ij}(t) = x_i(t) + q_i(t)\frac{w_{ij}}{a_{ij}}\theta_i + d_{ij}(t), \\ q_i(t) = \prod_{s=1}^{t}\Big(1 - \frac{\beta\Delta_i}{N_i(s+t_0)}\Big), \end{cases} \tag{4}$$

where $\beta > \frac{N_i}{2\Delta_i}$, $t_0 > \frac{\beta\Delta_i}{N_i}$ and $d_{ij}(t) \sim \text{Normal}(0, \sigma_{ij}^2)$. The noises $d_{ij}(t)$ are mutually independent for different agents i, out-neighbours j and time indexes t. The noises $d_{ij}(t)$ are also independent of the private weights w_{ij}.

Under the output messages (4), we design the stochastic approximation-based controller as follows

$$u_i(t-1) = \frac{\beta}{N_i(t+t_0)}\left[\sum_{k\in\mathcal{N}_i^{\text{in}}} a_{ki}(y_{ki}(t-1) - x_i(t-1))\right]. \tag{5}$$

Remark 4: If we remove the term $q_i(t)\frac{w_{ij}}{a_{ij}}\theta_i$ in (4), the proposed algorithm will degenerate to the traditional stochastic consensus algorithm in [1]. Section IV will further figure out that by introduce the term $q_i(t)\frac{w_{ij}}{a_{ij}}\theta_i$, the algorithm can achieve a better privacy-preserving level.

III. Convergence Analysis

This section will focus on the convergence analysis of the algorithm including the asymptotic unbiased mean square average consensus and the accuracy estimation.

Theorem 1: Apply the controller (5) and output message (4) with private weights w_{ij} generated from (3) to the multi-agent system (1). Suppose Assumption 1 holds. Then, the system (1) achieves asymptotic unbiased mean square average consensus with accuracy

$$\rho \leq \frac{\beta^2}{t_0 M^2}\sum_{(i,j)\in\mathcal{E}} a_{ij}^2\sigma_{ij}^2. \tag{6}$$

Proof: Set $z_{ij}(t) = x_i(t) + q_i(t)\frac{w_{ij}}{a_{ij}}\theta_i$. Then, by (1), (4) and (5), one can get

$$\begin{cases} z_{ij}(t) = z_{ij}(t-1) \\ \qquad + \frac{\beta}{N_i(t+t_0)}\Big[\sum_{k\in\mathcal{N}_i^{\text{in}}} a_{ki}\big(y_{ki}(t-1) - z_{ij}(t-1)\big)\Big], \\ y_{ij}(t) = z_{ij}(t) + d_{ij}(t). \end{cases}$$

The above dynamics is exactly the traditional stochastic approximation consensus algorithm of the virtual agents z_{ij} [1], whose communication graph is denoted as $\mathcal{L}(\mathcal{G})$. Note that each agent z_{ij} in $\mathcal{L}(\mathcal{G})$ represents an edge (i,j) in $\mathcal{G}$. And, there exists an edge in $\mathcal{L}(\mathcal{G})$ from the agent z_{i_1,j_1} to the agent z_{i_2,j_2} if and only if $j_1 = i_2$. Therefore, $\mathcal{L}(\mathcal{G})$ is exactly the line digraph of $\mathcal{G}$ [17]. Because of the strong connectedness of $\mathcal{G}$, the line digraph of $\mathcal{G}$ is also strongly connected [17]. Then, by [1, Th. 7], there exists a random variable x^* with $\text{Var}(x^*) < \infty$ such that $\lim_{t\to\infty}\mathbb{E}(z_{ij}(t) - x^*)^2 = 0$ for any $(i,j) \in \mathcal{E}$. Note that $\lim_{t\to\infty} z_{ij}(t) - x_i(t) = \lim_{t\to\infty} q_i(t)\frac{w_{ij}}{a_{ij}}\theta_i = 0$. Therefore, given any $(i,j) \in \mathcal{E}$, it holds that

$$\lim_{t\to\infty}\mathbb{E}\big(x_i(t) - x^*\big)^2 = 0. \tag{7}$$

Next, we prove $\mathbb{E}x^* = \frac{1}{N}\sum_{i=1}^{N}\theta_i$.

Denote $\zeta_{ij}(t) = x_i(t) + \frac{N_i}{\Delta_i}q_i(t)w_{ij}\theta_i$. Then, by (1) and (5), it holds that

$$\begin{aligned} &\sum_{(i,j)\in\mathcal{E}} \zeta_{ij}(t) \\ &= \sum_{(i,j)\in\mathcal{E}} \zeta_{ij}(t-1) + \frac{\beta}{t+t_0}\sum_{(k,i)\in\mathcal{E}} a_{ki}(y_{ki}(t-1) - x_i(t-1)) \\ &\quad - \frac{\beta}{t+t_0}\sum_{(i,j)\in\mathcal{E}} a_{ij}q_i(t-1)\frac{w_{ij}}{a_{ij}}\theta_i \\ &= \sum_{(i,j)\in\mathcal{E}} \zeta_{ij}(t-1) + \frac{\beta}{t+t_0}\sum_{(i,j)\in\mathcal{E}} a_{ij}d_{ij}(t-1), \end{aligned} \tag{8}$$

which implies

$$\mathbb{E}\left[\sum_{(i,j)\in\mathcal{E}} \zeta_{ij}(t)\right] = \mathbb{E}\left[\sum_{(i,j)\in\mathcal{E}} \zeta_{ij}(0)\right] = \frac{M}{N}\sum_{i=1}^{N}\theta_i. \tag{9}$$

Note that $\lim_{t\to\infty}\zeta_{ij}(t) - x_i(t) = \lim_{t\to\infty}\frac{N_i}{\Delta_i}q_i(t)w_{ij}\theta_i = 0$. Then, by (7) and (9), the asymptotic unbiased mean square average consensus of the system (1) can be verified.

For the accuracy estimation, by (8), it holds that

$$\begin{aligned} &\mathbb{E}\left[\sum_{(i,j)\in\mathcal{E}} \zeta_{ij}(t) - \frac{M}{N}\sum_{i=1}^{N}\theta_i\right]^2 \\ &= \mathbb{E}\left[\sum_{(i,j)\in\mathcal{E}} \zeta_{ij}(t-1) - \frac{M}{N}\sum_{i=1}^{N}\theta_i\right]^2 + \frac{\beta^2}{(t+t_0)^2}\sum_{(i,j)\in\mathcal{E}} a_{ij}^2\sigma_{ij}^2 \\ &= \mathbb{E}\left[\sum_{(i,j)\in\mathcal{E}} \zeta_{ij}(0) - \frac{M}{N}\sum_{i=1}^{N}\theta_i\right]^2 + \sum_{s=1}^{t}\frac{\beta^2}{(s+t_0)^2}\sum_{(i,j)\in\mathcal{E}} a_{ij}^2\sigma_{ij}^2 \\ &\leq \frac{\beta^2}{t_0}\sum_{(i,j)\in\mathcal{E}} a_{ij}^2\sigma_{ij}^2. \end{aligned}$$

Therefore, we have

$$\begin{aligned} \mathbb{E}\left(x^* - \frac{1}{N}\sum_{i=1}^{N}\theta_i\right)^2 &= \frac{1}{M^2}\mathbb{E}\left(Mx^* - \frac{M}{N}\sum_{i=1}^{N}\theta_i\right)^2 \\ &\leq \frac{1}{M^2}\left(\sqrt{\mathbb{E}\left(Mx^* - \sum_{(i,j)\in\mathcal{E}}\zeta_{ij}(t)\right)^2} + \sqrt{\frac{\beta^2}{t_0}\sum_{(i,j)\in\mathcal{E}} a_{ij}^2\sigma_{ij}^2}\right)^2. \end{aligned}$$

By $\zeta_{ij}(t)$ converging to $x_i(t)$ and $x_i(t)$ converging to x^* in the mean square sense, one can get $\mathbb{E}(Mx^* - \sum_{(i,j)\in\mathcal{E}}\zeta_{ij}(t))^2$

converges to 0 as t goes to infinity. Therefore, we have $\rho \leq \frac{\beta^2}{t_0 M^2} \sum_{(i,j)\in\mathcal{E}} a_{ij}^2 \sigma_{ij}^2$. ■

Remark 5: By Theorem 1, it is found that the key factor on the consensus accuracy is the variance of the noises $d_{ij}(t)$ rather than the private weights. σ_i^w in the private weight design does not appear in the accuracy estimate (6) in Theorem 1. Therefore, the agent i can increase σ_i^w for privacy-preserving purpose without losing the consensus accuracy.

IV. Privacy Analysis

In this section, we give the privacy analysis of the algorithm by Fisher information. Before that, some useful lemmas and corollaries are listed below.

Lemma 1: Given $\alpha \in \mathbb{R}$ and an invertible matrix $A \in \mathbb{R}^{n\times n}$ with $\alpha \mathbf{1}_n^\top A^{-1}\mathbf{1}_n \neq -1$, then

$$\mathbf{1}_n^\top \Big(A + \alpha \mathbf{1}_n^\top \mathbf{1}_n\Big)^{-1} \mathbf{1}_n = \frac{\mathbf{1}_n^\top A^{-1}\mathbf{1}_n}{1+\alpha \mathbf{1}_n^\top A^{-1}\mathbf{1}_n}.$$

Proof: By [18, Th. 1.1.17], it holds that

$$\begin{aligned}\mathbf{1}_n^\top \Big(A + \alpha \mathbf{1}_n^\top \mathbf{1}_n\Big)^{-1} \mathbf{1}_n &= \mathbf{1}_n^\top A^{-1}\mathbf{1}_n - \frac{\alpha\big(\mathbf{1}_n^\top A^{-1}\mathbf{1}_n\big)^2}{1+\alpha \mathbf{1}_n^\top A^{-1}\mathbf{1}_n} \\ &= \frac{\mathbf{1}_n^\top A^{-1}\mathbf{1}_n}{1+\alpha \mathbf{1}_n^\top A^{-1}\mathbf{1}_n}.\end{aligned}$$

The proof is thereby concluded. ■

Lemma 2: Given $\alpha > 0$, $a \in \mathbb{R}^m$ and a series of invertible matrices $A_1, \ldots, A_m \in \mathbb{R}^{n\times n}$ with $\alpha \sum_{i=1}^m \mathbf{1}_n^\top A_i^{-1}\mathbf{1}_n < 1$, then

$$\begin{aligned}&(a\otimes \mathbf{1}_n)^\top \Big[\mathrm{diag}(A_1,\ldots,A_m) - \alpha \mathbf{1}_{mn}\mathbf{1}_{mn}^\top\Big]^{-1}(a\otimes \mathbf{1}_n) \\ &\leq \frac{\sum_{i=1}^m a_i^2 \mathbf{1}_n^\top A_i^{-1}\mathbf{1}_n}{1-\alpha\sum_{i=1}^m \mathbf{1}_n^\top A_i^{-1}\mathbf{1}_n},\end{aligned}$$

where a_i is the i-th component of the vector a.

Proof: By Cauchy Inequality, it holds that

$$\left(\sum_{i=1}^m a_i^2 \mathbf{1}_n^\top A_i^{-1}\mathbf{1}_n\right)\left(\sum_{i=1}^m \mathbf{1}_n^\top A_i^{-1}\mathbf{1}_n\right) \geq \left(\sum_{i=1}^m a_i \mathbf{1}_n^\top A_i^{-1}\mathbf{1}_n\right)^2,$$

which together with [18, Th. 1.1.17] implies

$$\begin{aligned}&(a\otimes \mathbf{1}_n)^\top \Big[\mathrm{diag}(A_1,\ldots,A_m) - \alpha \mathbf{1}_{mn}\mathbf{1}_{mn}^\top\Big]^{-1}(a\otimes \mathbf{1}_n) \\ &= \sum_{i=1}^m a_i^2 \mathbf{1}_n^\top A_i^{-1}\mathbf{1}_n + \frac{\alpha\Big(\sum_{i=1}^m a_i \mathbf{1}_n^\top A_i^{-1}\mathbf{1}_n\Big)^2}{1-\alpha\sum_{i=1}^m \mathbf{1}_n^\top A_i^{-1}\mathbf{1}_n} \\ &\leq \frac{\sum_{i=1}^m a_i^2 \mathbf{1}_n^\top A_i^{-1}\mathbf{1}_n}{1-\alpha\sum_{i=1}^m \mathbf{1}_n^\top A_i^{-1}\mathbf{1}_n}.\end{aligned}$$

This proves the lemma. ■

Lemma 3: If positive real numbers α and γ satisfy $\alpha\gamma > 1$ and $\alpha < 1 + t_0$, then we have

$$\sum_{t=1}^\infty \prod_{s=1}^t \left(1 - \frac{\alpha}{s+t_0}\right)^\gamma < \frac{\alpha\gamma + t_0}{\alpha\gamma - 1}.$$

Proof: By [13, Lemma 3.2], it holds that $\prod_{s=1}^t (1-\frac{\alpha}{s+t_0}) \leq (\frac{1+t_0}{t+t_0})^\alpha$ for all $t \geq 1$. Therefore, by

$$\begin{aligned}\sum_{t=1}^\infty \frac{1}{(t+t_0)^{\alpha\gamma}} &= \frac{1}{(1+t_0)^{\alpha\gamma}} + \sum_{t=2}^\infty \frac{1}{(t+t_0)^{\alpha\gamma}} \\ &< \frac{\alpha\gamma + t_0}{(\alpha\gamma-1)(1+t_0)^{\alpha\gamma}},\end{aligned}$$

we have $\sum_{t=1}^\infty \prod_{s=1}^t (1-\frac{\alpha}{t+t_0})^\gamma < \frac{\alpha\gamma+t_0}{\alpha\gamma-1}$. ■

Lemma 4: Assume that the sensitive information $x \in \mathbb{R}$ is observed by $y = vx + d$, where $v \in \mathbb{R}^n$ and the noise $d \sim \mathrm{Normal}(\mu, \Sigma)$ with $\mu \in \mathbb{R}^n$ and invertible $\Sigma \in \mathbb{R}^{n\times n}$. Then,

1) $\mathcal{I}_y(x) = v^\top \Sigma^{-1} v$;
2) for any invertible $G \in \mathbb{R}^{n\times n}$ and given vector $b \in \mathbb{R}^n$, $\mathcal{I}_y(x) = \mathcal{I}_{Gy+b}(x)$.

Proof: For a), it holds that

$$p(y|x) = \frac{\exp\Big(-\frac{1}{2}(y-vx-\mu)^\top \Sigma^{-1}(y-vx-\mu)\Big)}{\sqrt{(2\pi)^n \det(\Sigma)}}.$$

Therefore, we have

$$\begin{aligned}\mathcal{I}_y(x) &= \mathbb{E}\Big(v^\top \Sigma^{-1}(y-vx-\mu)\Big)^2 \\ &= v^\top \Sigma^{-1}\mathbb{E}\Big[(y-vx-\mu)(y-vx-\mu)^\top\Big]\Sigma^{-1}v \\ &= v^\top \Sigma^{-1} v.\end{aligned}$$

For b), it holds that $Gy+b = Gx+Gd+b$, where $Gd+b \sim \mathrm{Normal}(\mu+b, G\Sigma G^\top)$. Then, by a), we have

$$\mathcal{I}_{Gy+b}(x) = v^\top G^\top \Big(G\Sigma G^\top\Big)^{-1} Gv = v^\top \Sigma^{-1} v = \mathcal{I}_y(x).$$

The proof is completed. ■

The privacy-preserving level is estimated in the following theorem. Note that the information leakage of the input messages affects the privacy-preserving level, and the input messages cannot be decided by the agent. Therefore, the theorem assumes that all input messages are leaked, which is the worst case for the input messages.

Theorem 2: Apply the controller (5) to the multi-agent system (1). If for the agent i, all input messages of the agent i and all the output messages for its out-neighbours $j_1, \ldots, j_m$ are leaked to the potential passive attackers, then,

$$\mathcal{I}^{-1}_{Y_{j_1,\ldots,j_m}(t)}(\theta_i) \geq \frac{1 - \frac{(\sigma_i^w)^2}{N_i}\sum_{s=1}^m \mathcal{I}_{ij_s}}{\sum_{s=1}^m \Big(a_{ij_s} + \frac{M-NN_i}{NN_i}\Big)^2 \mathcal{I}_{ij_s}},$$

where

$$Y_{j_1,\ldots,j_m}(t) = \Big[y_{ij_1}(t) \quad \cdots \quad y_{ij_1}(0) \quad \cdots \quad y_{ij_m}(t) \quad \cdots \quad y_{ij_m}(0)\Big]^\top,$$

and

$$\mathcal{I}_{ij_s} = \frac{2\beta\Delta_i + N_i t_0}{a_{ij_s}^2 \sigma_{ij_s}^2 (2\beta\Delta_i - N_i) + (\sigma_i^w)^2(2\beta\Delta_i + N_i t_0)}.$$

Proof: By (1) and (5), it holds that

$$x_i(t) = q_i(t)\theta_i + \sum_{s=1}^t \frac{q_i(t)}{q_i(s)} \frac{\beta}{N_i(s+t_0)} \left[\sum_{k\in\mathcal{N}_i^{\mathrm{in}}} a_{ki} y_{ki}(s-1)\right].$$

which implies

$$\begin{aligned} y_{ij}(t) &= q_i(t)\left(1+\frac{w_{ij}}{a_{ij}}\right)\theta_i + d_{ij}(t) \\ &\quad + \sum_{s=1}^{t}\frac{q_i(t)}{q_i(s)}\frac{\beta}{N_i(s+t_0)}\left[\sum_{k\in\mathcal{N}_i^{\text{in}}} a_{ki}y_{ki}(s-1)\right], \\ &= \frac{q_i(t)}{a_{ij}}\left(\left(a_{ij_1}+w_{ij_1}\right)\theta_i + \frac{a_{ij_1}d_{ij_1}(t)}{q_i(t)}\right) \\ &\quad + \sum_{s=1}^{t}\frac{q_i(t)}{q_i(s)}\frac{\beta}{N_i(s+t_0)}\left[\sum_{k\in\mathcal{N}_i^{\text{in}}} a_{ki}y_{ki}(s-1)\right], \end{aligned}$$

where $q_i(t)$ is defined as in (4). Therefore, there exist an invertible $G \in \mathbb{R}^{m(t+1)\times m(t+1)}$ and a $b \in \mathbb{R}^{m(t+1)}$ such that $Y_{j_1,\ldots,j_m}(t) = GY_q(t) + b$, where

$$\begin{aligned} Y_q(t) = \Big[&\left(a_{ij_1}+w_{ij_1}\right)\theta_i + \tfrac{a_{ij_1}d_{ij_1}(t)}{q_i(t)}, \ldots, \\ &\left(a_{ij_1}+w_{ij_1}\right)\theta_i + \tfrac{a_{ij_1}d_{ij_1}(0)}{q_i(t)}, \ldots, \\ &\left(a_{ij_m}+w_{ij_m}\right)\theta_i + \tfrac{a_{ij_m}d_{ij_m}(t)}{q_i(t)}, \ldots, \\ &\left(a_{ij_m}+w_{ij_m}\right)\theta_i + \tfrac{a_{ij_m}d_{ij_m}(0)}{q_i(0)}\Big]. \end{aligned}$$

Then, by Lemma 4, it holds that $\mathcal{I}_{Y_{j_1,\ldots,j_m}(t)}(\theta_i) = \mathcal{I}_{Y_q(t)}(\theta_i)$.

By (3), one can figure out that

$$Y_q(t) \sim \text{Normal}\left(\mu_q, \Sigma_q\right),$$

where

$$\mu_q = \left[\left(a_{ij_1}+\tfrac{M-NN_i}{NN_i}\right)\mathbf{1}_{t+1}^\top \quad \cdots \quad \left(a_{ij_1}+\tfrac{M-NN_i}{NN_i}\right)\mathbf{1}_{t+1}^\top\right]^\top,$$

and

$$\begin{aligned} \Sigma_q = \text{diag}\Bigg(&\frac{a_{ij_1}^2\sigma_{ij_1}^2}{q_i^2(t)}, \ldots, \frac{a_{ij_1}^2\sigma_{ij_1}^2}{q_i^2(0)}, \ \cdots \\ &\frac{a_{ij_m}^2\sigma_{ij_m}^2}{q_i^2(t)}, \ldots, \frac{a_{ij_m}^2\sigma_{ij_m}^2}{q_i^2(0)}\Bigg) \\ &+ (\sigma_i^w)^2 I_m \otimes \mathbf{1}_{t+1}\mathbf{1}_{t+1}^\top - \frac{(\sigma_i^w)^2}{N_i}\mathbf{1}_{m(t+1)}\mathbf{1}_{m(t+1)}^\top. \end{aligned}$$

By Lemmas 1, and 3, it holds that for $s = 1, \ldots, m$,

$$\begin{aligned} &\mathbf{1}_{t+1}^\top\left(\text{diag}\left(\frac{a_{ij_s}^2\sigma_{ij_s}^2}{q_i^2(t)}, \ldots, \frac{a_{ij_s}^2\sigma_{ij_s}^2}{q_i^2(0)}\right) + (\sigma_i^w)^2\mathbf{1}_{t+1}\mathbf{1}_{t+1}^\top\right)^{-1}\mathbf{1}_{t+1} \\ &= \frac{\sum_{l=0}^{t} q_i^2(l)}{a_{ij_s}^2\sigma_{ij_s}^2 + (\sigma_i^w)^2\left(\sum_{l=0}^{t} q_i^2(l)\right)} \\ &\le \frac{2\beta\Delta_i + N_i t_0}{a_{ij_s}^2\sigma_{ij_s}^2(2\beta\Delta_i - N_i) + (\sigma_i^w)^2(2\beta\Delta_i + N_i t_0)} = \mathcal{I}_{ij_s}. \end{aligned}$$

Then, by applying Lemma 2, we have

$$\begin{aligned} \mathcal{I}_{Y_{j_1,\ldots,j_m}(t)}(\theta_i) &= \mathcal{I}_{Y_q(t)}(\theta_i) = \mu_q^\top\Sigma_q^{-1}\mu_q \\ &\le \frac{\sum_{s=1}^{m}\left(a_{ij_s}+\frac{M-NN_i}{NN_i}\right)^2\mathcal{I}_{ij_s}}{1-\frac{(\sigma_i^w)^2}{N_i}\sum_{s=1}^{m}\mathcal{I}_{ij_s}}. \end{aligned}$$

This proves the theorem. ■

Remark 6: It is worth noticing that the lower bound of $\mathcal{I}^{-1}_{Y_{j_1,\ldots,j_m}(t)}(\theta_i)$ given in Theorem 2 increases as σ_i^w grows. Recall Remark 5. That means the agent i can promote the privacy-preserving level without losing the consensus accuracy by increasing σ_i^w.

The following theorem estimates the privacy-preserving level when σ_i^w is sufficiently large.

Theorem 3: Under the condition of Theorem 2, we have the following assertions:

a) if $m < N_i$, then $\lim_{\sigma_i^w\to\infty}\mathcal{I}^{-1}_{Y_{j_1,\ldots,j_m}(t)}(\theta_i) = \infty$;

b) if $m = N_i$, then

$$\begin{aligned} &\lim_{\sigma_i^w\to\infty}\mathcal{I}^{-1}_{Y_{j_1,\ldots,j_{N_i}}(t)}(\theta_i) \\ &\ge \frac{(2\beta\Delta_i - N_i)\sum_{s=1}^{N_i} a_{ij_s}^2\sigma_{ij_s}^2}{N_i(2\beta\Delta_i + N_i t_0)\sum_{s=1}^{N_i}\left(a_{ij_s}+\frac{M-NN_i}{NN_i}\right)^2}, \end{aligned}$$

Proof: When $m < N_i$, a) of the theorem can be verified by $\lim_{\sigma_i^w\to\infty}\mathcal{I}_{ij_s} = 0$, $\lim_{\sigma_i^w\to\infty}\frac{(\sigma_i^w)^2}{N_i}\sum_{s=1}^{m}\mathcal{I}_{ij_s} = \frac{m}{N_i} < 1$ and Theorem 2.

When $m = N_i$, by Theorem 2, it holds that

$$\begin{aligned} &\lim_{\sigma_i^w\to\infty}\mathcal{I}^{-1}_{Y_{j_1,\ldots,j_{N_i}}(t)}(\theta_i) \\ &\ge \lim_{\sigma_i^w\to\infty}\frac{1-\frac{(\sigma_i^w)^2}{N_i}\sum_{s=1}^{N_i}\mathcal{I}_{ij_s}}{\sum_{s=1}^{N_i}\left(a_{ij_s}+\frac{M-NN_i}{NN_i}\right)^2\mathcal{I}_{ij_s}} \\ &= \lim_{\sigma_i^w\to\infty}\frac{\sum_{s=1}^{N_i}\frac{(\sigma_i^w)^2 a_{ij_s}^2\sigma_{ij_s}^2(2\beta\Delta_i - N_i)}{a_{ij_s}^2\sigma_{ij_s}^2(2\beta\Delta_i - N_i) + (\sigma_i^w)^2(2\beta\Delta_i + N_i t_0)}}{N_i\sum_{s=1}^{N_i}\left(a_{ij_s}+\frac{M-NN_i}{NN_i}\right)^2(\sigma_i^w)^2\mathcal{I}_{ij_s}} \\ &= \frac{(2\beta\Delta_i - N_i)\sum_{s=1}^{N_i} a_{ij_s}^2\sigma_{ij_s}^2}{N_i(2\beta\Delta_i + N_i t_0)\sum_{s=1}^{N_i}\left(a_{ij_s}+\frac{M-NN_i}{NN_i}\right)^2}, \end{aligned}$$

which verifies b) of the theorem. ■

Remark 7: Theorems 1, and 3 imply that when the output messages for at least one out-neighbour are not leaked, we can decrease σ_{ij} to achieve any pre-given consensus accuracy, and increase σ_i^w to achieve any pre-given privacy-preserving level simultaneously. However, it is worth mentioning that differentiated output-based methods bring extra communication overhead.

V. Numerical Simulation

This section considers a discrete-time multi-agent system of four agents coupled by the communication digraph illustrated in Fig. 1. The public balanced weights a_{ij} are also given in Fig. 1.

In this example, the initial states of the agents are set to be $\theta_1 = 3$, $\theta_2 = 1$, $\theta_3 = -1$, and $\theta_4 = -3$. We employ the controller (5) with $\beta = 3$ and $t_0 = 9$. The variances of the noises in output messages (4) are set to be $\sigma_{12}^2 = \sigma_{23}^2 = \sigma_{34}^2 = \sigma_{41}^2 = 0.49$ and $\sigma_{13}^2 = 0.64$.

For the private weights generated by (3), when $(\sigma_1^w)^2 = 10000$, Fig. 2 shows that the states of the agents tend to the same value, and approach the average of the initial states.

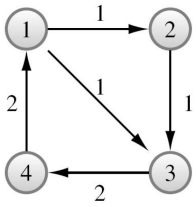

Fig. 1. Communication topology.

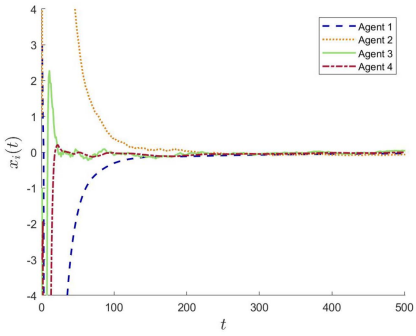

Fig. 2. The trajectories of the agents' states.

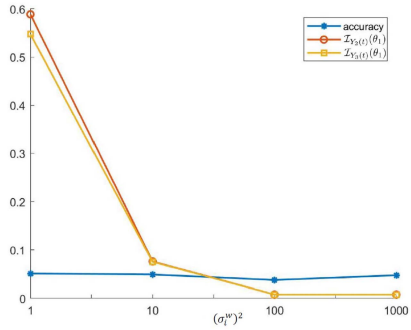

Fig. 3. The consensus accuracy and information leakages.

Besides, for the each case of $(\sigma_1^w)^2 = 1, 10, 100, 1000$, we repeat 50 experiments and compute the average of $(\frac{1}{4}\sum_{i=1}^4 x_i(1000) - \frac{1}{4}\sum_{i=1}^4 \theta_i)^2$ to approximate the consensus accuracy. The results are given in Fig. 3, which illustrates that the consensus accuracy does not increase as $(\sigma_1^w)^2$ grows. Meanwhile, the privacy leakage amounts $\mathcal{I}_{Y_2(t)}(\theta_1)$ and $\mathcal{I}_{Y_3(t)}(\theta_1)$ rapidly reach 0 as $(\sigma_1^w)^2$ grows. The experiment verifies that Agent 1 can increase σ_i^w to achieve any pre-given privacy-preserving level without losing the consensus accuracy if the output messages for Agents 2 and 3 are not both leaked. This cannot be achieved for traditional stochastic obfuscation algorithms [10], [11], [12], [13], which is consistent with theoretical analysis.

VI. Conclusion

In this letter, we have studied the differentiated output-based privacy-preserving average consensus over digraphs. A new privacy-preserving distributed consensus algorithm has been proposed, which can achieve a stronger privacy-preserving effect. Specifically, the algorithm can be designed to achieve any pre-given consensus accuracy and privacy-preserving level simultaneously achieved if the output messages for at least one out-neighbour are not leaked. The mean square convergence of the algorithm has been provided, and privacy protection of the initial state is quantified by Fisher information. Finally, a numerical example has been provided to verify the efficiency of the algorithm. In the future work, how to keep low communication overhead and high privacy-preserving level simultaneously can be considered.

References

[1] M. Huang and J. H. Manton, "Stochastic consensus seeking with noisy and directed inter-agent communication: Fixed and randomly varying topologies," *IEEE Trans. Autom. Control*, vol. 55, no. 1, pp. 235–241, Jan. 2010.

[2] T. Li. and J.-F. Zhang, "Mean square average-consensus under measurement noises and fixed topologies: Necessary and sufficient conditions," *Automatica*, vol. 45, no. 8, pp. 1929–1936, Aug. 2009.

[3] Y. Mo and R. M. Murray, "Privacy preserving average consensus," *IEEE Trans. Autom. Control*, vol. 62, no. 2, pp. 753–765, Feb. 2017.

[4] J. Zhang, J. Tan, and J. Wang, "Privacy security in control systems," *Sci. China Inform. Sci.*, vol. 64, Jul. 2021, Art. no. 176201.

[5] Y. Wang, "Privacy-preserving average consensus via state decomposition," *IEEE Trans. Autom. Control*, vol. 64, no. 11, pp. 4711–4716, Nov. 2019.

[6] M. S. Darup, A. Redder, and D. E. Quevedo, "Encrypted cooperative control based on structured feedback," *IEEE Control Syst. Lett.*, vol. 3, no. 1, pp. 37–42, Jan. 2019.

[7] M. Ristic, B. Noack, and U. D. Hanebeck, "Secure fast covariance intersection using partially homomorphic and order revealing encryption schemes," *IEEE Control Syst. Lett.*, vol. 5, no. 1, pp. 217–222, Jan. 2021.

[8] K. Tjell and R. Wisniewski, "Private aggregation with application to distributed optimization," *IEEE Control Syst. Lett.*, vol. 5, no. 5, pp. 1591–1596, Nov. 2021.

[9] X. Huo and M. Liu, "Privacy-preserving distributed multi-agent cooperative optimization-paradigm design and privacy analysis," *IEEE Control Syst. Lett.*, vol. 6, pp. 824–829, Jun. 2022.

[10] J. He, L. Cai, P. Cheng, J. Pan, and L. Shi, "Consensus-based data-privacy preserving data aggregation," *IEEE Trans. Autom. Control*, vol. 64, no. 12, pp. 5222–5229, Dec. 2019.

[11] E. Nozari, P. Tallapragada, and J. Cortés, "Differentially private average consensus: Obstructions, trade-offs, and optimal algorithm design," *Automatica*, vol. 81, pp. 221–231, Jul. 2017.

[12] X.-K. Liu, J.-F. Zhang, and J. Wang, "Differentially private consensus algorithm for continuous-time heterogeneous multi-agent systems," *Automatica*, vol. 122, Dec. 2020, Art. no. 109283.

[13] J. M. Wang, J. Ke, and J.-F. Zhang, "Differentially private bipartite consensus over signed networks with time-varying noises," 2022, *arXiv:2212.11479*.

[14] J. Wang, J.-F. Zhang, and X. He, "Differentially private distributed algorithms for stochastic aggregative games," *Automatica*, vol. 142, Aug. 2022, Art. no. 110440.

[15] F. Farokhi and H. Sandberg, "Ensuring privacy with constrained additive noise by minimizing Fisher information," *Automatica*, vol. 99, pp. 275–288, Jan. 2019.

[16] B. Gharesifard and J. Cortés, "Distributed strategies for generating weight-balanced and doubly stochastic digraphs," *Eur. J. Control*, vol. 18, no. 6., pp. 539–557, 2012.

[17] F. Harary and R. Z. Norman, "Some properties of line digraphs," *Rendiconti del Circolo Matematico di Palermo*, vol. 9, no. 2, pp. 161–168, May 1960.

[18] L. Guo, *Time-Varying Stochastic Systems, Stability and Adaptive Theory, Second Edition*. Beijing, China: Science Press, 2020.