A Study of Privacy Preservation in Average Consensus Algorithm via Deterministic Obfuscation Signals

Navid Rezazadeh and Solmaz S. Kia, *Senior Member, IEEE*

***Abstract*—This article is a study on the use of additive obfuscation signals to keep the reference values of the agents in the continuous-time Laplacian average consensus algorithm private from eavesdroppers. Obfuscation signals are perturbations that agents add to their local dynamics and their transmitted-out messages to conceal their private reference values. An eavesdropper is an agent inside or outside the network that has access to some subset of the interagent communication messages, and its knowledge set also includes the network topology. Rather than focusing on using a zero-sum and vanishing additive signal, our work determines the necessary and sufficient conditions that define the set of admissible obfuscation signals that do not perturb the convergence point of the algorithm from the average of the reference values of the agents. Of theoretical interest, our results show that this class includes nonvanishing signals as well. Given this broader class of admissible obfuscation signals, we define a deterministic notion of privacy preservation. In this definition, privacy preservation for an agent means that neither the private reference value nor a finite set of values to which the private reference value of the agent belongs to can be obtained. Then, we evaluate the agents' privacy against eavesdroppers with different knowledge sets.**

***Index Terms*—Consensus algorithm, network systems, privacy preservation.**

I. Introduction

WE CONSIDER the *Laplacian average consensus* algorithm

$$\dot{x}^i(t) = -\sum_{j=1}^{N} \mathsf{a}_{ij}\,(x^i(t) - x^j(t)), \quad x^i(0) = \mathsf{r}^i \tag{1}$$

$i \in \mathcal{V} = \{1, \dots, N\}$, over a strongly connected and weight-balanced digraph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{A})$,[1] which drives x^i of each agent i to $\frac{1}{N}\sum_{j=1}^{N} \mathsf{r}^j$ as $t \to \infty$ [2]; r^i represents the *reference value* of agent $i \in \mathcal{V}$. This algorithm lacks privacy preservation because the reference value r^i is trivially revealed to all the in-neighbors of each agent $i \in \mathcal{V}$ and any external agent listening to the communication messages. Laplacian average consensus is a basic primitive algorithm that enables many other in-network distributed operations, e.g., sensor fusion [3] and distributed learning [4], [5]. Therefore, devising a privacy preservation augmentation for this algorithm is of importance in the literature. Our aim is to investigate whether in a network of $N \geq 3$ agents, the agents' reference value can be concealed from *eavesdroppers* by adding the obfuscation signals f^i and g^i to, respectively, the internal dynamics and the transmitted signal of each agent $i \in \mathcal{V}$, i.e.,

$$\dot{x}^i(t) = -\sum_{j=1}^{N} \mathsf{a}_{ij}\,\big(x^i(t) - y^j(t)\big) + f^i(t) \tag{2a}$$

$$y^i(t) = x^i(t) + g^i(t), \quad x^i(0) = \mathsf{r}^i \tag{2b}$$

while still guaranteeing that $x^i \to \frac{1}{N}\sum_{j=1}^{N} \mathsf{r}^j$ as $t \to \infty$.

Definition 1 (eavesdropper): An eavesdropper is an agent inside (internal agent) or outside (external agent) the network that stores and processes the accessible interagent communication messages $y^j(t)$, $t \in \mathbb{R}_{\geq 0}$, of all agents $j \in \mathcal{O} \subset \mathcal{V}$ in a network that implements (2) to obtain the private reference value of the other agents in the network, without interfering with the execution of algorithm (2). For an internal eavesdropper, $\mathcal{O}$ is the set of one-hop agents that communicate their outputs to the eavesdropper. For an external eavesdropper, $\mathcal{O}$ is the set of agents; it can intercept their outgoing messages. ■

The use of additive obfuscation signals has already been considered for privacy preservation for the average consensus algorithm, but there are some limiting assumptions. For example, adding a sequence of well-constructed vanishing stochastic obfuscation signals to the transmitted communication messages of the agents has been investigated in the literature for discrete-time implementation of (1) over connected undirected graphs [6], [7]. These results ensure the privacy preservation of the agents with respect to internal eavesdroppers that do not have access to at least one of the agent's transmitted-in signals. When deviation from the exact consensus value is tolerated, the authors in [8] and [9] have shown that the reference value of all the agents can be made private by perturbing the transmitted-out signal using a zero-mean randomly generated Gaussian or Laplacian noise.

Manuscript received 6 January 2023; revised 8 April 2023; accepted 17 May 2023. Date of publication 28 June 2023; date of current version 1 March 2024. This work was supported by NSF CAREER under Grant ECCS-1653838. A preliminary version of this work appeared in [1]. Recommended by Associate Editor Y. Wang. *(Corresponding author: Solmaz S. Kia.)*

The authors are with the Department of Mechanical and Aerospace Engineering, University of California Irvine, Irvine, CA 92697 USA (e-mail: nrezazad@uci.edu; solmaz@uci.edu).

Digital Object Identifier 10.1109/TCNS.2023.3290114

[1] See Section II for a brief description of the notation and the definitions.

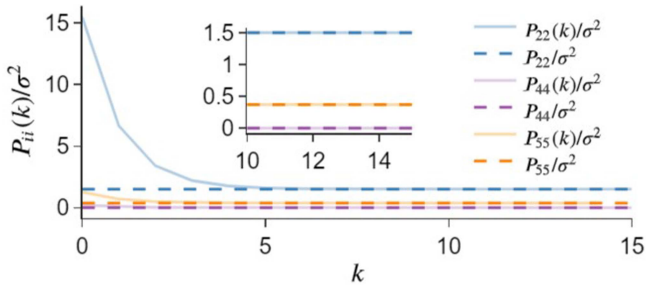

Fig. 1. Agent 1's (eavesdropper) maximum likelihood estimator's normalized error covariance when the method of [7] is used over graph of Fig. 2(b). Agent 4's privacy is not preserved because the estimator error covariance converges to zero. Even though agent 5's privacy is preserved due to the nonzero estimator error covariance, the privacy preservation might be limited because the estimator's error covariance is very small.

The deviation from the desired average is not quantified in [8], but is explained in [9] using the ϵ-differential privacy framework introduced in [10]. Additive obfuscation noises have also been used as a privacy preservation mechanism in other distributed algorithms such as distributed optimization [11]; distributed estimation [12], [13]; and distributed games [14]. Even though our focus in this article is on privacy preservation via additive obfuscation signals for alternative implementations of *Laplacian average consensus* algorithm (1), it is worth noting that the point-to-point gossip algorithm, secret massage passing, and encryption have also been used in the literature to provide privacy preservation for algorithm (1)'s discrete-time implementation [15], [16], [17], [18], [19], [20], [21], [22], [23]. Alternating communication signals using vanishing masking functions [24], as well as rewiring the graph and point-to-point communication to induce privacy preservation [25], [26], [27], [28], [29], [30], have also been explored in the literature. However, in practice, rewiring may be restrictive. The scheme of adding virtual nodes to the communication graph was also explored in [31].

The existing privacy preservation algorithms including our perliminary work [1] that use additive obfuscation signals are based on the assumption that the signals should be vanishing and zero-sum. These constraints are widely used as sufficient conditions to ensure the exact convergence of the algorithm. The existing results also use stochastic noise with shared parameters. These constrained choices can limit privacy guarantees. For example, the privacy guarantee in [7] is defined as a nonzero estimation covariance including relatively small nonzero values (see Fig. 1).

This article conducts a careful analysis of the use of *admissible* obfuscation signals for privacy preservation for the Laplacian average consensus algorithm (1) against internal and external eavesdroppers that know the network topology. We define the *admissible obfuscation signals* as integrable signals that do not perturb the algorithm's exact convergence. To make the study thorough, we add the obfuscation signals to the transmitted-out signals and also to the system dynamics, as shown in (2). A common trait of privacy preservation mechanisms that are intended not to perturb the exact convergence of the algorithm is that each uses a particular class of vanishing noises or perturbation functions [6], [7], [24]. One is then left to wonder whether stronger privacy preservation guarantees are achievable if a broader class of signals was considered. This article intends to answer this question. Thus, instead of using only a prespecified class of vanishing obfuscation signals, we investigate and obtain the necessary and sufficient conditions that define the set of admissible additive obfuscation signals that do not perturb the convergence point of the algorithm. A *theoretical* finding we arrive at is that the admissible obfuscation signals do not have to be necessarily vanishing. We conduct our study with respect to locally chosen signals from the admissible set. Understanding the nature of the admissible obfuscation signals is crucial in privacy preservation evaluations. It is rational to assume that the eavesdroppers are aware of the necessary conditions on such signals and use them to breach the privacy of the agents. We show that such knowledge enables the eavesdroppers that have access to all the transmitted-in and transmitted-out communication signals of a targeted agent to employ an observer to asymptotically reconstruct the targeted agent's reference value. Interestingly, we show that in this case, the privacy breach is inevitable even if the agent uses nonvanishing admissible obfuscation signals. Our analysis leads to the necessary and sufficient condition for an agent to stay private that at least one of its transmitted-in signals is not available to the eavesdropper. We define our notion of privacy preservation as follows.

Definition 2 (Privacy preservation): Consider an eavesdropper, as defined in Definition 1, that has access to $y^j(t)$, $t \in \mathbb{R}_{\geq 0}$, of all agents $j \in \mathcal{O} \subset \mathcal{V}$ in a network that implements (2) with locally chosen admissible perturbation signals (f^l, g^l), $l \in \mathcal{V}$. We say that the privacy of an agent $i \in \mathcal{V}$ is preserved if for any arbitrary large $\gamma \in \mathbb{R}_{>0}$, there exists a tuple $(x^{i'}(0) = \mathsf{r}^{i'}, f^{i'}(t), g^{i'}(t))$, with locally chosen admissible perturbations $(f^{i'}(t), g^{i'}(t))$ and $|\mathsf{r}^{i'} - \mathsf{r}^i| > \gamma$, such that $y^j(t) \equiv y^{j'}(t)$, $t \in \mathbb{R}_{\geq 0}$, for all $j \in \mathcal{O}$.

Adhering to this privacy definition means that the eavesdropper will be neither able to estimate the private reference value nor a finite set of values to which the private reference value of an agent belongs. This means that our approach to design admissible perturbation signals leads to a privacy preservation guarantee that is stronger than the privacy preservation in the stochastic approaches such as [7], where even though the exact reference value is concealed, an estimate with a quantifiable confidence interval on the reference value can be obtained; see Fig. 1 and Section V for more discussion.

II. Notation and Definitions

The Euclidean norm of vector $\mathbf{x} \in \mathbb{R}^n$ is $\|\mathbf{x}\| = \sqrt{\mathbf{x}^\top \mathbf{x}}$, and the (essential) supremum norm of a signal $f : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is $\|f\|_{\text{ess}} = (\text{ess})\sup\{\|f(t)\|, t \geq 0\}$. The set of measurable essentially bounded functions $f : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is denoted by $\mathcal{L}_n^\infty$. The set of measurable functions $f : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ that satisfy $\int_0^t \|f(\tau)\| d\tau < \infty$ is denoted by $\mathcal{L}_n^1$. For sets $\mathcal{A}$ and $\mathcal{B}$, the relative complement of $\mathcal{B}$ in $\mathcal{A}$ is $\mathcal{A} \backslash \mathcal{B} = \{x \in \mathcal{A} \,|\, x \notin \mathcal{B}\}$. To distinguish and emphasize that a variable in a network is local to an agent $i \in \mathcal{V} = \{1, \ldots, N\}$, we use superscripts, e.g., in (2), (f^i, g^i) are the local obfuscation signals of agent $i \in \mathcal{V}$.

If $p^i \in \mathbb{R}$ is a variable of agent $i \in \mathcal{V}$, the aggregated p^i's of the network is the vector $\mathbf{p} = [p^1, \dots, p^N]^\top \in \mathbb{R}^N$.

Graph theory: a weighted directed graph (digraph) is a triplet $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$, where $\mathcal{V} = \{1, \dots, N\}$ is the *node set*, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the *edge set* and $\mathbf{A} = [\mathsf{a}_{ij}] \in \mathbb{R}^{N \times N}$ is a weighted *adjacency* matrix with the property that $\mathsf{a}_{ij} > 0$ if $(i, j) \in \mathcal{E}$ and $\mathsf{a}_{ij} = 0$, otherwise. A weighted digraph is *undirected* if $\mathsf{a}_{ij} = \mathsf{a}_{ji}$ for all $i, j \in \mathcal{V}$. We follow [32] in definition of *in-neighbor* and *out-neighbor:* an edge from i to j, denoted by (i, j), means that agent i can read/obtain information from agent j; then, i is called an *in-neighbor* of j and j is called an *out-neighbor* of i. The set of the out-neighbors of an agent $i \in \mathcal{V}$ is $\mathcal{N}^i_{\text{out}}$, i.e., the set of agents that agent i has access to their information, $\mathcal{N}^i_{\text{in}}$ is the set of in-neighbors of agent i, i.e., the set of agents that have access to agent i's information. We define $\mathcal{N}^i_{\text{out}+i} = \mathcal{N}^i_{\text{out}} \cup \{i\}$ and $\mathcal{N}^i_{\text{in}+i} = \mathcal{N}^i_{\text{in}} \cup \{i\}$. A digraph is called *strongly connected* if for every pair of vertices, there is a directed path connecting them. We refer to a strongly connected and undirected graph as a *connected graph*. The *weighted out-degree* and *weighted in-degree* of a node i, are respectively, $\mathsf{d}^i_{\text{in}} = \sum_{j=1}^N \mathsf{a}_{ji}$ and $\mathsf{d}^i_{\text{out}} = \sum_{j=1}^N \mathsf{a}_{ij}$. We let $\mathsf{d}^{\max}_{\text{out}} = \max(\mathsf{d}^1_{\text{out}}, \dots, \mathsf{d}^N_{\text{out}})$. A digraph is *weight-balanced* if $\mathsf{d}^i_{\text{out}} = \mathsf{d}^i_{\text{in}}$ at each node $i \in \mathcal{V}$ (although they might be different across different nodes). The *(out-) Laplacian* matrix is $\mathbf{L} = [\ell_{ij}]$ is $\mathbf{L} = \mathbf{D}^{\text{out}} - \mathbf{A}$, where $\mathbf{D}^{\text{out}} = \text{Diag}(\mathsf{d}^1_{\text{out}}, \dots, \mathsf{d}^N_{\text{out}}) \in \mathbb{R}^{N \times N}$. Note that $\mathbf{L}\mathbf{1}_N = \mathbf{0}$. A digraph is weight-balanced iff $\mathbf{1}_N^\top \mathbf{L} = \mathbf{0}$. For a strongly connected and weight-balanced digraph, $\text{rank}(\mathbf{L}) = N - 1$, $\text{rank}(\mathbf{L} + \mathbf{L}^\top) = N - 1$, and $\mathbf{L}$ has one zero eigenvalue $\lambda_1 = 0$ and the rest of its eigenvalues have positive real parts. We let $\mathbf{R} \in \mathbb{R}^{N \times (N-1)}$ be a matrix whose columns are normalized orthogonal complement of $\mathbf{1}_N$. Then

$$\mathbf{T}^\top \mathbf{L} \mathbf{T} = \begin{bmatrix} 0 & \mathbf{0} \\ 0 & \mathbf{L}^+ \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} \frac{1}{\sqrt{N}} \mathbf{1}_N & \mathbf{R} \end{bmatrix}, \quad \mathbf{L}^+ = \mathbf{R}^\top \mathbf{L} \mathbf{R}. \tag{3}$$

For a strongly connected and weight-balanced digraph, $-\mathbf{L}^+$ is a Hurwitz matrix.

III. Admissible Obfuscations

We start our study by determining the space of the admissible obfuscation signals $f^i(t)$ and $g^i(t)$. Understanding the nature of these admissible signals for which the desired convergence point of the algorithm is preserved is crucial in evaluating the privacy guarantees of algorithm (2).

Definition 3 (Admissible obfuscation signals): We refer to the set of obfuscation signals $\{f^j \in \mathcal{L}_1^\infty, g^j \in \mathcal{L}_1^\infty\}_{j=1}^N$ in (2) that does not perturb the convergence of the algorithm, i.e., $\lim_{t\to\infty} x^i(t) = \frac{1}{N}\sum_{j=1}^N x^j(0) = \frac{1}{N}\sum_{j=1}^N \mathsf{r}^j$ for any $i \in \mathcal{V}$, as the admissible obfuscation signals.

The following theorem, whose proof is given in Appendix B, gives the necessary and sufficient conditions that define the set of the admissible obfuscation signals. We only require this set to be a subset of the integrable and essentially bounded signals so that the differential (2a) has a unique solution. Contrary to the common practice in the literature, the signals are not required to be vanishing.

Theorem 3.1 (The set of necessary and sufficient conditions on the admissible obfuscation signals): Consider algorithm (2) over a strongly connected and weight-balanced digraph with obfuscation signals $f^i, g^i \in \mathcal{L}_1^\infty$, $i \in \mathcal{V}$. Then, the trajectory $t \mapsto x^i(t)$ of all agents $i \in \mathcal{V}$ converges to $\frac{1}{N}\sum_{j=1}^N x^j(0) = \frac{1}{N}\sum_{j=1}^N \mathsf{r}^j$ as $t \to \infty$ if and only if

$$\lim_{t\to\infty} \int_0^t \sum_{k=1}^N \left(f^k(\tau) + \mathsf{d}^k_{\text{out}}\, g^k(\tau)\right) d\tau = 0 \tag{4a}$$

$$\lim_{t\to\infty} \int_0^t \mathrm{e}^{-\mathbf{L}^+(t-\tau)} \mathbf{R}^\top (\mathbf{f}(\tau) + \mathbf{A}\,\mathbf{g}(\tau))\, d\tau = \mathbf{0} \tag{4b}$$

for any $\mathbf{R}$ and $\mathbf{L}^+$ as defined in (3). ■

The necessary and sufficient conditions (4) show that the choice of admissible signals is highly coupled among the agents. However, the agents must choose the admissible signals privately and without cooperation with others. The next theorem, whose proof is given in Appendix B, offers a way for each agent $i \in \mathcal{V}$ to choose its own admissible signals (f^i, g^i) privately without revealing them explicitly to others.

Theorem 3.2 (Linear algebraic coupling): Consider algorithm (2) over a strongly connected and weight-balanced digraph. Let each agent $i \in \mathcal{V}$ choose its local obfuscation signals $f^i, g^i \in \mathcal{L}_1^\infty$ such that

$$\lim_{t\to\infty} \int_0^t \left(f^i(\tau) + \mathsf{d}^i_{\text{out}}\, g^i(\tau)\right) d\tau = \beta^i \tag{5}$$

where $\beta^i \in \mathbb{R}$. Then, the necessary and sufficient conditions to satisfy (4) are

$$\sum_{k=1}^N \beta^k = 0 \tag{6a}$$

$$\lim_{t\to\infty} \int_0^t \mathrm{e}^{-(t-\tau)} g^i(\tau)\, d\tau = \alpha \in \mathbb{R} \quad i \in \mathcal{V}. \tag{6b}$$

■

By enforcing condition (5), Theorem 3.2 shows that the coupling between the agents is a set of linear algebraic constraints. Now, if we set up the modified algorithm (2) in a way that each agent $i \in \mathcal{V}$ for example uses $\beta^i = 0$, and a common value $\alpha \in \mathbb{R}$, which can readily be $\alpha = 0$, each agent can choose its admissible obfuscation signals locally/privately according to (5) and (6b) and still guarantee convergence to the exact average consensus.

Definition 4 (Set of locally chosen admissible signals): For any given α and β^is satisfying (6a), $\mathcal{P}_{(\beta^i,\alpha)}$, $i \in \mathcal{V}$, denotes the set of integrable function tuples (f^i, g^i) satisfying (5) and (6).

Choosing signals that satisfy condition (5) is rather easy. However, condition (6b) appears to be more complex. The result below, whose proof is given in Appendix B, identifies three classes of signals that are guaranteed to satisfy condition (6b).

Lemma 3.1 (Signals that satisfy (6b)): For a given $\alpha \in \mathbb{R}$, let $g = g_1 + g_2 \in \mathcal{L}_1^\infty$ satisfy one of the conditions: (a) $\lim_{t\to\infty} g(t) = \alpha$; (b) $\lim_{t\to\infty} g_1(t) = \alpha$ and $\lim_{t\to\infty} \int_0^t g_2(\tau) d\tau = \bar{g} < \infty$; and (c) $\lim_{t\to\infty} g_1(t) = \alpha$ and

$\int_0^t \sigma(|g_2(\tau)|)d\tau < \infty$ for $t \in \mathbb{R}_{\geq 0}$, where σ is any class $\mathcal{K}_\infty$ function. Then, $\lim_{t\to\infty} \int_0^t \mathrm{e}^{-(t-\tau)} g(\tau)d\tau = \alpha$. ■

Of *theoretical interest*, Lemma 3.1 reveals a relaxation on the commonly seen condition in the literature, which requires the additive signal to be vanishing. Lemma 3.1 shows that admissible obfuscation signals $\{(f^j, g^j) \in \mathcal{P}_{(\beta^j,\alpha)}\}_{j=1}^N$ should not necessarily be vanishing signals even for $\alpha = 0$ and $\beta^i = 0$, $i \in \mathcal{V}$. For example, $g_1(t) = 0$ and $g_2(t) = \sin(\phi_0 + 2\pi(\frac{c}{2}t^2 + \omega_0 t))$, which is a waveform with linear chirp function [33], where ω_0 is the starting frequency at time $t = 0$, $c \in \mathbb{R}$ is the chirpiness constant, and ϕ_0 is the initial phase, satisfy condition (b) of Lemma 3.1 with $\alpha = 0$. When a nonzero α is used, the choices for nonvanishing g satisfying (6b) are much broader, e.g., according to condition (b) of Lemma 3.1, any function that asymptotically converges to α can be used.

Remark 3.1 (Admissible signals for discrete-time implementation): We can discretize the continuous-time Laplacian consensus algorithm (2) with time step $\delta \in (0, 1/\mathrm{d}_{\mathrm{out}}^{\max})$ [2] as

$$\mathbf{x}(k+1) = (\mathbf{I} - \delta\mathbf{L})\mathbf{x}(k) + \delta\mathbf{f}(k) + \delta\mathbf{A}\mathbf{g}(k). \tag{7}$$

Similar argument to those in the proof of Theorem 3.1 leads to the following conditions on the admissible obfuscation signals $(f^i(k), g^i(k))$, $i \in \mathcal{V}$:

$$\lim_{k\to\infty} \sum_{l=0}^{k} \sum_{i=1}^{N} (f^i(l) + \mathrm{d}_{\mathrm{out}}^i\, g^i(l)) = 0 \tag{8a}$$

$$\lim_{k\to\infty} \sum_{l=0}^{k-1} (\mathbf{I} - \delta\mathbf{L}^+)^l \mathbf{R}^\top (\mathbf{f}(k-1-l) + \mathbf{A}\,\mathbf{g}(k-1-l)) = \mathbf{0}. \tag{8b}$$

The algorithms in [6] and [7] are special cases of (7), which use $f^i(k) = g^i(k)$. They choose the obfuscation signals from a particular class of zero sum, which trivially satisfies (8a), and vanishing, which satisfies (8b), stochastic signals.

IV. Privacy Preservation Analysis

In light of Theorem 3.2, our proposed privacy preservation mechanism is as follows.

Definition 5 (Privacy preservation mechanism via additive obfuscation signals): Each agent $i \in \mathcal{V}$ implements (2), for which it chooses its own obfuscation signals (f^i, g^i) locally/privately from $\mathcal{P}_{(\beta^i,\alpha)}$.

α and β^is are the preset parameters of the modified algorithm (2). The complexity of choosing them is similar to choosing the algorithm parameters in [7], [18], [19], and [20]. As mentioned earlier, the straightforward choice is $\beta^i = 0$, $i \in \mathcal{V}$, and $\alpha = 0$.

To start the privacy preservation analysis, we first explicitly define the *knowledge set* of an eavesdropper that it uses to infer the private reference value of the other agents.[2] Without loss of generality, we assume that the internal eavesdropper is agent $1 \in \mathcal{V}$ and the external agent is agent ext.

Definition 6 (Knowledge set of an eavesdropper): The knowledge set of the internal eavesdropper agent 1 and external eavesdropper agent ext is

$$\mathcal{K}^a = \{\mathcal{Y}^a(\infty), \mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{A}), \text{ conditions (5) and (6)}, \alpha, \{\beta^i\}_{i=1}^N\} \tag{9}$$

$a \in \{1, \mathrm{ext}\}$. $\mathcal{Y}^1(t) = \{x^1(\tau), y^1(\tau), \{y^i(\tau)\}_{i\in\mathcal{N}_{\mathrm{out}}^1}\}_{\tau=0}^t$ is the set of signals available to agent 1. Let and $\mathcal{O} \subset \mathcal{V}$ be the set of agents that external eavesdropper ext has access to it. Thus, $\mathcal{Y}^{\mathrm{ext}}(t) = \{\{y^i(\tau)\}_{i\in\mathcal{O}}\}_{\tau=0}^t$. ■

The reader should notice that parameters α and β^i of every agent $i \in \mathcal{V}$ are part of the eavesdropper's knowledge set, indicating that the privacy preservation guarantees we provide do not depend on the lack of information on the value of these parameters.[3] For any given α and β^i, $\mathcal{P}_{(\beta^i,\alpha)}$ is an infinite set of integrable function tuples (f^i, g^i). Each agent $i \in \mathcal{V}$ decides locally/privately which (f^i, g^i) it chooses from $\mathcal{P}_{(\beta^i,\alpha)}$. Thus, the probability of an eavesdropper (internal or external) knowing what admissible pair of $(f^i, g^i) \in \mathcal{P}_{(\beta^i,\alpha)}$ agent i has chosen locally converges to zero as the cardinality of the set $\mathcal{P}_{(\beta^i,\alpha)}$ is infinity. This is in contrast to the stochastic methods such as [7], which instruct the agents to choose their obfuscation signals from a specific probability distribution, limiting the privacy guarantees; see Section V for more discussions.

Identifying the initial condition of the agents in the presence of unknown additive obfuscation signals may appear to be related to the classical concept of strong observability/detectability in control theory [35], [36]. However, the necessary conditions on the unknown admissible obfuscation signals (5) and (6) provide additional information to the eavesdropper. Such information is not being captured by the strong observability/detectability framework, rendering it inadequate for our study.

Consider the internal eavesdropper, agent 1, when it intends to obtain the initial condition of one of the agents $i \in \mathcal{V}$. The critical part of the knowledge set of an eavesdropper when it targets an agent is the signals that it has access to. Intuitively, when an eavesdropper agent does not have direct access to all the signals in $\{y^j(t)\}_{j\in\mathcal{N}_{\mathrm{out}+i}^i}$, a rational strategy appears to be that the eavesdropper agent estimates the states of the agents; it does not have access to their outputs. If those agents also have out-neighbors that their output signals are not available to the eavesdropper agent, then the eavesdropper agent should estimate the state of those agents as well, until the only inputs to the dynamics that it observes are the additive admissible perturbation signals. For example, in Fig. 2(a), to obtain the reference value of agent 6, agent 1 compensates for the lack of direct access to $y^7(t)$, which enters the dynamics of agent 6, by estimating the state of all the agents in subgraph $\mathcal{G}_3^1$. Our results below; however, show that this strategy is not effective. In fact, we show that an eavesdropper agent (internal or external) is able to uniquely identify the reference value of an agent $i \in \mathcal{V}$ if and only if it has direct access to $\{y^j(t)\}_{j\in\mathcal{N}_{\mathrm{out}+i}^i}$ for all $t \in \mathbb{R}_{\geq 0}$. To study privacy preservation

[2]We conduct our privacy preservation analysis from a single eavesdropper's point of view. The extension of our theoretical results to multiple collaborative and noncollaborative eavesdroppers is straightforward.

[3]The interested reader can find our extended results in [34] on privacy preservation against eavesdroppers that have various degrees of knowledge about α and β^is.

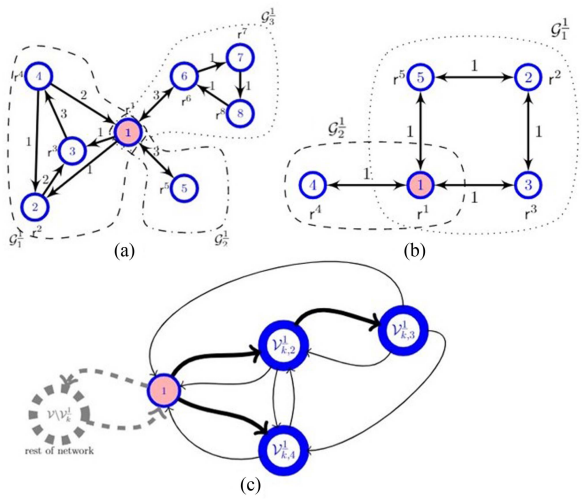

Fig. 2. (a) and (b) depict graphs $\mathcal{G}$ that node 1, the eavesdropper, is an articulation point of the undirected representation of $\mathcal{G}$. The islands of $\mathcal{G}$ induced by node 1 are highlighted by closed dashed curves. (c) kth induced island of node 1.

for agent $i \in \mathcal{V}$, we partition the graph into islands whose nodes are classified into different groups based on their information exchange by the eavesdropper and its out-neighbors (see Fig. 2). For that, note that removing eavesdropper agent 1 and its incident edges results in $\bar{n}^1 \geq 1$ disjoint strongly connected subgraphs $\bar{\mathcal{G}}_k^1 = (\bar{\mathcal{V}}_k^1, \bar{\mathcal{E}}_k^1) \subset \mathcal{G}(\mathcal{V}, \mathcal{E})$, $k \in \{1, \ldots, \bar{n}^1\}$. Adding agent 1 in subgraph $\bar{\mathcal{G}}_k^1$ and including its incident edges to this subgraph results in an *island graph* $\mathcal{G}_k^1 = (\mathcal{V}_k^1, \mathcal{E}_k^1) \subset \mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}_k^1 = \bar{\mathcal{V}}_k^1 \cup \{1\}$ and $\mathcal{E}_k^1 = \{(l, j) \in \mathcal{E} | \, l \in \mathcal{V}_k^1, \, j \in \mathcal{V}_k^1\}$. Every island of agent 1 is connected to the rest of the digraph $\mathcal{G}$ only through agent 1 [see Fig. 2(c)]. Thus, any information coming out of or going into any island of the eavesdropper goes through the eavesdropper. To simplify the notation, without loss of generality, carry out the subsequent study for agents in island $k = 1$, e.g., $\mathcal{G}_1^1$. Based on how each agent interacts with agent 1, we divide the agents of island $\mathcal{G}_1^1$ into the following three groups [see Fig. 2(c)]:

1) $\mathcal{V}_{1,2}^1 = \left\{ i \in \mathcal{V}_1^1 \,\middle|\, i \in \mathcal{N}_{\text{out}}^1, \, \mathcal{N}_{\text{out}}^i \not\subset \mathcal{N}_{\text{out}+1}^1 \right\}$.
2) $\mathcal{V}_{1,3}^1 = \left\{ i \in \mathcal{V}_1^1 \,\middle|\, i \notin \mathcal{N}_{\text{out}}^1 \right\}$.
3) $\mathcal{V}_{1,4}^1 = \left\{ i \in \mathcal{V}_1^1 \,\middle|\, i \in \mathcal{N}_{\text{out}}^1, \, \mathcal{N}_{\text{out}}^i \subseteq \mathcal{N}_{\text{out}+1}^1 \right\}$.

$\mathcal{V}_{1,4}^1$ is the set of agents, in which agent 1 has direct access to all their communication signals, while $\mathcal{V}_{1,2}^1$ and $\mathcal{V}_{1,3}^1$ are the sets of agents, in which some of interagent communication between them is not available to agent 1. Without loss of generality, in what follows, we assume that the agents in the network are labeled according to the ordered set $(1, \mathcal{V}_{1,2}^1, \mathcal{V}_{1,3}^1, \mathcal{V}_{1,4}^1, \mathcal{V} \backslash \mathcal{V}_1^1)$. We let the aggregated states and obfuscation signals of the agents in $\mathcal{V}_{1,l}^1$, $l \in \{2, 3, 4\}$, be $\mathbf{x}_l = [x^i]_{i \in \mathcal{V}_{1,l}^1}$, $\mathbf{g}_l = [g^i]_{i \in \mathcal{V}_{1,l}^1}$, and $\mathbf{f}_l = [f^i]_{i \in \mathcal{V}_{1,l}^1}$. Similarly, we let the aggregated states and obfuscation signals of the agents in $\mathcal{V} \backslash \mathcal{V}_1^1$ be $\mathbf{x}_5 = [x^i]_{i \in \mathcal{V} \backslash \mathcal{V}_1^1}$, $\mathbf{g}_5 = [g^i]_{i \in \mathcal{V} \backslash \mathcal{V}_1^1}$ and $\mathbf{f}_5 = [f^i]_{i \in \mathcal{V} \backslash \mathcal{V}_1^1}$. We partition $\mathbf{L}$, $\mathbf{A}$, and $\mathbf{D}^{\text{out}}$, respectively, to subblock matrices $\mathbf{L}_{ij}$'s, $\mathbf{A}_{ij}$'s, and $\mathbf{D}_{ij}^{\text{out}}$'s in a comparable manner to the partitioned aggregated state $(x^1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5)$ (see [34, Lemma 4.2]). By definition, $\mathbf{L}_{ij} = -\mathbf{A}_{ij}$, $i, j \in \{1, \ldots, 5\}$, $i \neq j$. With the right notation at hand, we present the following result that provides the privacy guarantee according to Definition 2 for the agents belonging to $\mathcal{V}_{1,2}^1$ and $\mathcal{V}_{1,3}^1$. Because every agent in $\mathcal{G}_1^1$ is connected to the rest of the agents in digraph $\mathcal{G}$ only through agent 1, all the out-neighbors and in-neighbors of agent 2 are necessarily in $\mathcal{G}_1^1$. The proof is given in Appendix B.

Lemma 4.1 (A case of indistinguishable admissible initial conditions for an internal eavesdropper): Let agent 1 be the internal eavesdropper whose knowledge set is as Definition 6. Let $\mathcal{G}_1^1 = (\mathcal{V}_1^1, \mathcal{E}_1^1)$ be an island of agent 1 that satisfies $\mathcal{V}_{1,2}^1 \neq \{\}$. Consider the modified static average consensus algorithm (2) over a strongly connected and weight-balanced digraph $\mathcal{G}$ where the agents are implementing $\{(x^i(0) = \mathsf{r}^i, f^i, g^i)\}_{i=1}^N$, with the locally chosen admissible obfuscation signals $(f^i, g^i) \in \mathcal{P}_{(\beta^i, \alpha)}$. Consider also an alternative execution of (2) with $\{(x^{i'}(0), f^{i'}, g^{i'})\}_{i=1}^N$ satisfying

$$\begin{aligned} &x^{1'}(0) = x^1(0), \;\; \mathbf{x}_4'(0) = \mathbf{x}_4(0), \;\; \mathbf{x}_5'(0) = \mathbf{x}_5(0) \\ &\mathbf{x}_2'(0) - \mathbf{x}_2(0) = -\mathbf{A}_{23}\mathbf{L}_{33}^{-1}(\mathbf{x}_3'(0) - \mathbf{x}_3(0)) \\ &x^{j'}(0) \in \mathbb{R}, \quad j \in \mathcal{V}_{1,3}^1 \end{aligned} \tag{10}$$

and

$$\begin{aligned} &f^{i'}(t) = f^i(t), \qquad\qquad i \in \mathcal{V} \setminus \mathcal{V}_{1,2}^1 \\ &f^{i'}(t) = f^i(t) - \left[\mathbf{A}_{23}\mathrm{e}^{-\mathbf{L}_{33}t}(\mathbf{x}_3'(0) - \mathbf{x}_3(0))\right]_{i-1}, \;\; i \in \mathcal{V}_{1,2}^1 \end{aligned} \tag{11}$$

and

$$\begin{aligned} &g^{i'}(t) = g^i(t), \qquad\qquad i \in \mathcal{V} \backslash \mathcal{V}_{1,2}^1 \\ &g^{i'}(t) = g^i(t) + \left[\mathrm{e}^{-\mathbf{D}_{22}t}(\mathbf{x}_2'(0) - \mathbf{x}_2(0)),\right]_{i-1}, \;\; i \in \mathcal{V}_{1,2}^1. \end{aligned} \tag{12}$$

Then

$$y^i(t) = y^{i'}(t), \quad t \in \mathbb{R}_{\geq 0}, \qquad i \in \mathcal{V} \backslash \mathcal{V}_{1,3}^1. \tag{13}$$

Moreover

$$\sum_{i=1}^N x^{i'}(0) = \sum_{i=1}^N x^i(0) = \sum_{i=1}^N \mathsf{r}^i \tag{14}$$

$$\lim_{t \to \infty} x^{i'}(t) = \frac{1}{N} \sum_{i=1}^N \mathsf{r}^i, \qquad i \in \mathcal{V} \tag{15}$$

$$\left(f^{i'}, g^{i'}\right) \in \mathcal{P}_{(\beta^i, \alpha)}, \qquad i \in \mathcal{V}. \tag{16}$$

■

Remark 4.1 (Lemma 4.1 leads to privacy preservation in accordance with Definition 2): Notice that by virtue of (16), $(f^{i'}, g^{i'})$, $i \in \mathcal{V}$, generated by (11) and (12), satisfies the locally chosen admissible obfuscation signals conditions (5) and (6) for the same α and β^i used to generate $\{f^i, g^i\}_{i=1}^N$. Next, notice

that according to (10) and $\mathbf{L}_{33}^{-1}$ being full rank and $\mathbf{A}_{23}$ having rows with at least one none zero entry, for any arbitrary $\gamma \in \mathbb{R}_{\geq}$, there always exists $x^{i'}(0)$ for $i \in (\mathcal{V}_{1,2}^1 \cup \mathcal{V}_{1,3}^1)$ that satisfies $|x^{i'}(0) - x^i(0)| > \gamma$ and $(f^{i'}, g^{i'}) \in \mathcal{P}_{(\beta^i,\alpha)}$, while signals received by the eavesdropper, as stated in (13), are identical for the execution of the algorithm using $\{(x^i(0) = \mathsf{r}^i, f^i, g^i)\}_{i=1}^N$ and $\{(x^{i'}(0), f^{i'}, g^{i'})\}_{i=1}^N$. This means that the privacy of all the agents in $(\mathcal{V}_{1,2}^1 \cup \mathcal{V}_{1,3}^1)$ is preserved in accordance with Definition 2. ■

We can develop similar results, as stated in the following corollary, for an external eavesdropper that does not have direct access to the output signal of some of the out-neighbors of agent $i \in \mathcal{V}$.

Corollary 4.1 (A case of indistinguishable admissible initial conditions for an external eavesdropper): Let agent ext be the external eavesdropper whose knowledge set is as Definition 6 where the eavesdropper has access only to $y^l(t)$, $l \in \mathcal{O} \subset \mathcal{V}$. Let $\bar{\mathcal{O}} = \{j \in \mathcal{V} | j \notin \mathcal{O} \text{ and } \exists i \in \mathcal{O} \text{ s.t. } i \in \mathcal{N}_{\text{in}}^j \subset \mathcal{O}\}$ be a nonempty set. Consider the modified static average consensus algorithm (2) over a strongly connected and weight-balanced digraph $\mathcal{G}$, where the agents are implementing $\{x^i(0) = \mathsf{r}^i, f^i, g^i\}_{i=1}^N$, with the locally chosen admissible obfuscation signals $(f^i, g^i) \in \mathcal{P}_{(\beta^i,\alpha)}$. For any $k \in \bar{\mathcal{O}}$, consider also an alternative execution of (2) with $\{x^{i'}(0), f^{i'}, g^{i'}\}_{i=1}^N$ satisfying

$$\begin{aligned} &x^{i'}(0) = x^i(0) \qquad i \in \mathcal{V}\backslash\mathcal{N}_{\text{in}+k}^k \\ &x^{i'}(0) - x^i(0) = -\frac{\mathsf{a}_{ik}}{d_{\text{out}}^k}(x^{k'}(0) - x^k(0)) \quad i \in \mathcal{N}_{\text{in}}^k \\ &x^{k'}(0) \in \mathbb{R} \end{aligned} \tag{17}$$

and

$$\begin{aligned} &f^{i'}(t) = f^i(t) \qquad i \in \mathcal{V}\backslash\mathcal{N}_{\text{in}+k}^k \\ &f^{i'}(t) = f^i(t) - \mathsf{a}_{ik}\mathrm{e}^{-d_{\text{out}}^k t}(x^{k'}(0) - x^k(0)) \quad i \in \mathcal{N}_{\text{in}}^k \end{aligned} \tag{18}$$

and

$$\begin{aligned} &g^{i'}(t) = g^i(t) \qquad i \in \mathcal{V}\backslash\mathcal{N}_{\text{in}+1}^k \\ &g^{i'}(t) = g^i(t) + \mathrm{e}^{-\mathsf{d}_{\text{out}}^i t}(x^{k'}(0) - x^k(0)) \quad i \in \mathcal{N}_{\text{in}}^k. \end{aligned} \tag{19}$$

Then

$$y^i(t) = y^{i'}(t), \quad t \in \mathbb{R}_{\geq 0}, \qquad i \in \mathcal{V}\backslash\{k\}. \tag{20}$$

Moreover,

$$\sum_{i=1}^N x^{i'}(0) = \sum_{i=1}^N x^i(0) = \sum_{i=1}^N \mathsf{r}^i \tag{21}$$

$$\lim_{t\to\infty} x^{i'}(t) = \frac{1}{N}\sum_{i=1}^N \mathsf{r}^i, \qquad i \in \mathcal{V} \tag{22}$$

$$\left(f^{i'}, g^{i'}\right) \in \mathcal{P}_{(\beta^i,\alpha)}, \qquad i \in \mathcal{V}. \tag{23}$$

■

The proof of Corollary 4.1 is given in Appendix B. A similar assertion to that of Remark 4.1 about privacy preservation compliance in accordance to Definition 2 can be made about the agents whose privacy is preserved by virtue of Corollary 4.1 with respect to an external eavesdropper.

Through Lemma 4.1 and Corollary 4.1, we have established that the privacy of agents when the eavesdropper, either internal or external, does not have access to at least one signal that is transmitted in to the agent is preserved. The next result, whose proof is given in Appendix B, shows that such a guarantee does not hold for agents whose incoming and outgoing signals are in the knowledge set of the eavesdropper.

Lemma 4.2 (Observer design for eavesdroppers with the knowledge set (9)): Consider the modified static average consensus algorithm (2) with a set of locally chosen admissible obfuscation signals $(f^i, g^i) \in \mathcal{P}_{(\beta^i,\alpha)}$, $i \in \mathcal{V}$, over a strongly connected and weight-balanced digraph $\mathcal{G}$. Let the knowledge set of the eavesdroppers be as in Definition 6. An external eavesdropper ext and internal eavesdropper agent 1 that has access to the output signals of agent $i \in \mathcal{V}$ and all its out-neighbors can employ, respectively, observer

$$\dot{\zeta} = \sum_{j=1}^N \mathsf{a}_{ij}\left(y^i - y^j\right), \quad \zeta(0) = -\beta^i - \alpha \tag{24a}$$

$$\dot{\eta} = -\eta + y^i, \qquad \eta(0) \in \mathbb{R} \tag{24b}$$

$$\nu^{\text{ext}}(t) = \zeta(t) + \eta(t) \tag{24c}$$

and observer

$$\dot{\psi} = \sum_{j=1}^N \mathsf{a}_{ij}\left(y^i - y^j\right), \quad \psi(0) = -\beta^i \tag{25a}$$

$$\nu^1(t) = \psi(t) + x^1(t) \tag{25b}$$

to asymptotically obtain r^i, $i \in \mathcal{V}$, i.e., $\nu^a \to \mathsf{r}^i$, $a \in \{\text{ext}, 1\}$ as $t \to \infty$. Moreover, at any time $t \in \mathbb{R}_{\geq 0}$, the estimation error of the observers, respectively, satisfies

$$\nu^{\text{ext}}(t) - \mathsf{r}^i = \eta(t) - x^i(t) + \int_0^t \left(f^i(\tau) + \mathsf{d}_{\text{out}}^i g^i(\tau)\right) d\tau - \beta^i - \alpha \tag{26a}$$

$$\eta(t) = \mathrm{e}^{-t}\eta_0 + \int_0^t \mathrm{e}^{-(t-\tau)}x^i(\tau)d\tau + \int_0^t \mathrm{e}^{-(t-\tau)}g^i(\tau)d\tau \tag{26b}$$

and

$$\nu^1(t) - \mathsf{r}^i = x^1(t) - x^i(t) + \int_0^t \left(f^i(\tau) + \mathsf{d}_{\text{out}}^i\, g^i(\tau)\right) d\tau \; - \beta^i. \tag{27}$$

■

The reader may have noticed the subtle difference between the computational cost of the observers for internal and external eavesdroppers. To construct observer (25), the internal eavesdropper uses its local state. To compensate for the lack of internal dynamics, the external eavesdropper is forced to employ a higher order observer (24) and invoke condition (6b), which the internal eavesdropper does not need.

Building on our results of eavesdropper observer design in Lemma 4.2 and indistinguishable reference values in Lemma 4.1 and Corollary 4.1, we establish the necessary and sufficient condition under which an eavesdropper with knowledge set (9) can discover the reference value of an agent $i \in \mathcal{V}$.

Theorem 4.1 (Privacy preservation using the modified average consensus algorithm (2) when the knowledge set of the eavesdroppers is given by Definition 6): Consider the modified static average consensus algorithm (2) with a set of locally chosen admissible obfuscation signals $\{f^i, g^i\}_{i=1}^N$ over a strongly connected and weight-balanced digraph $\mathcal{G}$. Let the knowledge set of the internal eavesdropper 1 and external agent ext be (9). Then, (a) agent 1 can reconstruct the exact initial value of agent $i \in \mathcal{V}\backslash\{1\}$ if and only if $i \in \mathcal{N}^1_{\text{out}}$ and $\mathcal{N}^i_{\text{out}} \subseteq \mathcal{N}^1_{\text{out}+1}$; and (b) the external agent ext can reconstruct the exact initial value of agent $i \in \mathcal{V}$ if and only if $\{\{y^j(\tau)\}_{j\in\mathcal{N}^i_{\text{out}+i}}\}_{\tau=0}^{\infty} \subseteq \mathcal{Y}^{\text{ext}}(\infty)$.

Proof: Proof of statement (a): Lemma 4.1 shows that if $i \notin \mathcal{N}^1_{\text{out}}$ or $\mathcal{N}^i_{\text{out}} \not\subset \mathcal{N}^1_{\text{out}+1}$, then agent 1 cannot uniquely identify the reference value r^i of agent i. Next, if $i \in \mathcal{N}^1_{\text{out}}$ and $\mathcal{N}^i_{\text{out}} \subseteq \mathcal{N}^1_{\text{out}+1}$, Lemma 4.2 guarantees that agent 1 can employ an observer to obtain the reference value of agent i. Next, suppose agent $i \in \mathcal{V}\backslash\{1\}$ satisfies $i \notin \mathcal{N}^1_{\text{out}}$ (respectively, $i \in \mathcal{N}^1_{\text{out}}$ and $\mathcal{N}^i_{\text{out}} \not\subset \mathcal{N}^1_{\text{out}+1}$). Without loss of generality, let $\mathcal{V}^1_1$ be the island of agent 1 that contains this agent i. Consequently, $i \in \mathcal{V}^1_{1,3}$ (respectively, $i \in \mathcal{V}^1_{1,2}$). Then, by virtue of Lemma 4.1, we know that there exists an infinite number of alternative admissible initial conditions and corresponding admissible obfuscation signals for any agents in $\mathcal{V}^1_{1,3} \cup \mathcal{V}^1_{1,2}$, for which the time histories of each signal transmitted to agent 1 are identical. Therefore, agent 1 cannot uniquely identify the initial condition of any agents in $\mathcal{V}^1_{1,3} \cup \mathcal{V}^1_{1,2}$. In light of Corollary 4.1 and Lemma 4.2, the proof of statement (b) is similar to that of statement (a) and is omitted for brevity. ■

Theorem 4.1 is of value from a transparency perspective. In using algorithm (2), the agents now know what other agents may discover their reference value. If privacy preservation is a must and it is important to not to deviate from the exact average, one also knows that the solution is to make sure that every agent has an exclusive out-neighbor that is not the out-neighbor of its out-neighbors. There are several classes of undirected graphs for which any two agents on the graph have an exclusive neighbor with respect to the other. Thus, by Theorem 4.1, the privacy of all the agents is preserved from any internal eavesdropper when they implement algorithm (2). Examples include cyclic bipartite undirected graphs, 4-regular ring lattice undirected graphs with $N > 5$, planar stacked prism graphs, directed ring graphs, and any biconnected undirected graph that does not contain a cycle with three edges (see [37] for the formal definition of these graph topologies). Theorem 4.1 also presents an opportunity to make agents private with respect to a particular or all the other agents by rewiring the graph so that the conditions of the theorem are satisfied.

Remark 4.2 (An eavesdropper cannot estimate the reference value of a private agent or a finite set of values to which the private reference value of an agent belongs): For any given admissible β^i and α, the cardinality of $\mathcal{P}_{(\beta_i,\alpha)}$ is infinity. Each agent chooses its own (f^i, g^i) privately from the infinite set $\mathcal{P}_{(\beta_i,\alpha)}$. On the other hand, given any $\lambda \in \mathbb{R}_{>0}$, according to the discussion in Remark 4.1, for any agent $i \in (\mathcal{V}^1_{k,3} \cup \mathcal{V}^1_{k,2})$, $k \in \{1, \ldots, \bar{n}^1\}$, there exists an infinite set of possible alternatives $\{(x^i_\ell(0), (f^i_\ell, g^i_\ell) \in \mathcal{P}_{(\beta_i,\alpha)})\}_{\ell=1}^{\infty}$, satisfying $\sum_{j=1}^N x^j_\ell(0) = \sum_{j=1}^N x^j(0)$ such that

$$\lambda < |x^i(0) - x^i_1(0)| < |x^i(0) - x^i_2(0)| < \cdots < |x^i(0) - x^i_\ell(0)| \tag{28}$$

$\ell \to \infty$ for which $y^j(t) \equiv y^j_\ell(t)$ for all $j \in \mathcal{N}^1_{\text{out}+1}$. Now, let $x'(0)$ be the estimate of eavesdropper on the reference value $x^i(0) = \mathsf{r}^i$, $i \in (\mathcal{V}^1_{k,3} \cup \mathcal{V}^1_{k,2})$. Given (28), agent 1 cannot have more confidence on $x'(0) = x^i(0)$ over $x'(0) = x^i_\ell(0)$. Therefore, the eavesdropper will not be able to estimate neither the state nor a finite set of values to which the initial value of an agent belongs. Similar argument can be made about privacy preservation with respect to external eavesdropper. It is important to note that Lemma 4.1 and consequently the stronger notion of the privacy that is established here are due to the use of obfuscation signals (f^i, g^i) and would have not been possible if we had used $f^i = g^i$, where f^i comes from a specific class of functions as in existing literature [7]. ■

We close our study by pointing out that even though agent 1 cannot obtain the initial condition of the individual agents in $\mathcal{V}^1_{k,2} \neq \{\}$ and $\mathcal{V}^1_{k,3}$, $k \in \{1, \ldots, \bar{n}^1\}$, it can obtain the average of initial conditions of those agents. Without loss of generality, we demonstrate our results for $k = 1$.

Proposition 4.1 (Island anonymity): Consider the dynamic consensus algorithm (2) over a strongly connected and weight-balanced digraph $\mathcal{G}$ in which $\mathcal{V}^1_{1,2} \neq \{\}$. Let $n_{2,3} = |\mathcal{V}^1_{1,2} \cup \mathcal{V}^1_{1,3}|$ and $\mathsf{d}^{1,1}_{\text{out}} = \sum_{j\in(\mathcal{V}^1_{1,2}\cup\mathcal{V}^1_{1,4})} \mathsf{a}_{1j}$ be the out-degree of agent 1 in subgraph $\mathcal{G}^1_1$. Then, eavesdropper 1 with the knowledge set (9) can employ the observer

$$\dot{\zeta}_i = \sum_{j=1}^N \mathsf{a}_{ij}(y^i - y^j), \quad \zeta_i(0) = -\beta^i, \quad i \in \mathcal{V}^1_{1,4}$$

$$\dot{\eta} = -\sum_{j\in\left(\mathcal{V}^1_{1,2}\cup\mathcal{V}^1_{1,4}\right)} a_{1j}\left(y^1 - y^j\right), \quad \eta(0) = -\sum_{j\in\mathcal{V}^1_1\backslash\{1\}} \beta^i$$

$$\mu(t) = \frac{\eta(t) - \sum_{i\in\mathcal{V}^1_{1,4}} \zeta_i}{n_{2,3}} + x^1(t)$$

to have $\lim_{t\to\infty} \mu(t) = \frac{1}{n_{2,3}} \sum_{j\in\left(\mathcal{V}^1_{1,2}\cup\mathcal{V}^1_{1,3}\right)} \mathsf{r}^j$.

The proof is given in Appendix B.

V. Privacy Preservation Discussion

The deterministic and stochastic approaches to privacy preservation withhold different definitions of a private agent. In our deterministic setup, privacy is preserved when an eavesdropper, despite its knowledge set, ends up in an underdetermined system of equations when it wants to obtain the reference value of an agent. Therefore, the eavesdropper is left with infinite guesses

of a private agent's reference value, which it cannot favor any of them more than the other (see Remark 4.2). However, the stochastic privacy of an agent is preserved when the eavesdropper's estimate of the reference value yields a nonzero uncertainty. For example, in [7], a maximum likelihood estimator is used by the eavesdropper to estimate the reference value of the other agents. It is shown that the variance of $P(k)$ of this estimator converges to a constant matrix P. The privacy statement determines that the agents' privacy whose corresponding component in P converges to zero is breached. More specifically, given a vector ζ, a space of the agents' initial condition $\zeta^\top \mathbf{x}(0)$ is disclosed to the eavesdropper if $\zeta^\top P\zeta = 0$ and if $\zeta^\top P\zeta > 0$, it is interpreted as conserving the privacy of the subspace. In this setting, for an agent whose corresponding component of P is nonzero, the eavesdropper does not know the agent's exact reference value, but it has an estimate of it. Hence, we tend to favor the deterministic notion of privacy over stochastic as the deterministic approach reveals less information. Fig. 1 is the replicate of the result of an example study over the graph in Fig. 2(b) in [7], which shows the evolution of the covariance of the maximum likelihood estimator of the eavesdropper. As expected, P_{44} converges to zero but P_{22} and P_{55} not. Even though P_{22} and P_{55} are nonzero, they are pretty small, indicating that the eavesdropper can have a good estimate of the reference values of these agents. In contrast, our privacy preservation ensures that for agents whose privacy is preserved, the eavesdropper neither can obtain the reference value nor establish an estimate.

Consider the network given in Fig. 2(b). To demonstrate over results, consider the following three implementations of the modified continuous-time Laplacian average consensus algorithm (2) with the reference values and the additive obfuscation signals as follows:

Case (1): $\mathbf{r} = [-3, 5, 1, -2, 10]^\top$

$$\mathbf{f}(t) = \left[-3, -2\cos\left(\frac{t}{t^2+1}\right), \frac{t^5}{t^5+1}, \tan\left(\frac{\pi}{4}\tanh(t)\right), -2\tanh(t)\right]^\top$$

$$\mathbf{g}(t) = \left[1 + 0.23\mathrm{e}^{-t}, \cos\left(10\pi\frac{t^2}{t^5+1}\right), \left(1 + \mathrm{e}^{-t}\sin(10t)\right)\tanh(t), 1 + \mathrm{e}^{-(t-1)^2}, \log\left(\mathrm{e} - \mathrm{e}^{0.1t}(1+\sin(t))\right)\right]^\top.$$

Case (2): $\mathbf{r} = [-3, 15, -4, -2, 5]^\top$

$$\mathbf{f}(t) = \left[-3, -2\cos\left(\frac{t}{t^2+1}\right), -10\mathrm{e}^{-2t} + \frac{t^5}{t^5+1}, \tan\left(\frac{\pi}{4}\tanh(t)\right), -10\mathrm{e}^{-2t} - 2\tanh(t)\right]^\top$$

$$\mathbf{g}(t) = \left[1 + 0.23\mathrm{e}^{-t}, \cos\left(10\pi\frac{t^2}{t^5+1}\right), 5\mathrm{e}^{-2t} + \left(1 + \mathrm{e}^{-t}\sin(10t)\right)\tanh(t), 1 + \mathrm{e}^{-(t-1)^2}, 5\mathrm{e}^{-2t} + \log\left(\mathrm{e} - \mathrm{e}^{0.1t}(1+\sin(t))\right)\right]^\top.$$

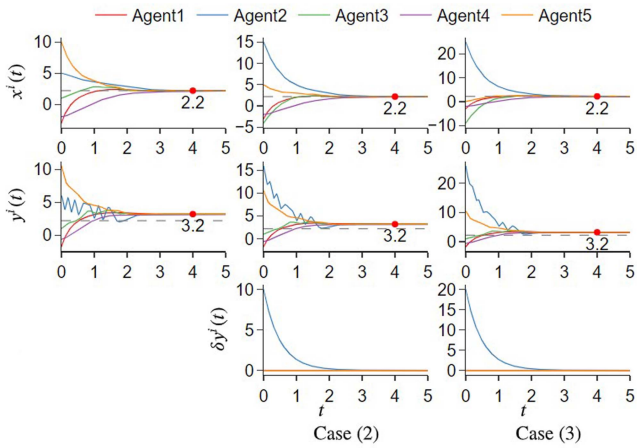

Fig. 3. Consensus results for three different cases.

Case (3): $\mathbf{r} = [-3, 25, -9, -2, 0]^\top$

$$\mathbf{f}(t) = \left[-3, -2\cos\left(\frac{t}{t^2+1}\right), -20\mathrm{e}^{-2t} + \frac{t^5}{t^5+1}, \tan\left(\frac{\pi}{4}\tanh(t)\right), -20\mathrm{e}^{-2t} - 2\tanh(t)\right]^\top$$

$$\mathbf{g}(t) = \left[1 + 0.23\mathrm{e}^{-t}, \cos\left(10\pi\frac{t^2}{t^5+1}\right), 10\mathrm{e}^{-2t} + \left(1 + \mathrm{e}^{-t}\sin(10t)\right)\tanh(t), 1 + \mathrm{e}^{-(t-1)^2}, 10\mathrm{e}^{-2t} + \log\left(\mathrm{e} - \mathrm{e}^{0.1t}(1+\sin(t))\right)\right]^\top.$$

Let Case (1) correspond to the actual operation case, and let the other two cases be admissible alternative ones. Here, all the admissible obfuscation signals are smooth, uniformly continuous, and nonvanishing. They satisfy (5), (6a), and (6b) with $\alpha = 1$ and $\beta^i = 0$, $i \in \mathcal{V} = \{1, 2, 3, 4, 5\}$. The plots in the top row of Fig. 3 confirms the convergence of the algorithm to the exact average, as guaranteed by Theorem 2. The plots in the second row of Fig. 3 show that the transmitted-out signal y^i of each agent $i \in \mathcal{V}$ satisfies $\lim_{t\to\infty} y^i(t) = \frac{1}{N}\sum_{j=1}^{N} \mathrm{r}^j + \alpha$. Let $\delta y^i(t)$ be the communication signals difference between Case (j), $j \in \{2, 3\}$ and Case (1). As seen in the two bottom plots in Fig. 3, only $\delta y^2(t)$ is nonzero. This means that agent 1, in all three cases, receives exactly the same transmission messages from its neighbors, agents 3–5. This result, as predicted by Theorem 2, shows that agent 1, the eavesdropper, cannot tell whether r^2 is equal to 5 of Case (1), 15 of Case (2), or 25 Case (3). Moreover, agent 1 is not able to say which one of these cases is more probable. A similar statement can be made about agents 3 and 5 whose privacy is guaranteed in our framework. While the privacy of agents 2, 3, and 5 is preserved, according to Lemma 4.2, agent 1 can employ an observer of form (25) to asymptotically estimate the reference value of agent 4. The response of this estimator is shown in Fig. 4. Here, to make a comparison study with respect to [7], we used the undirected graph of

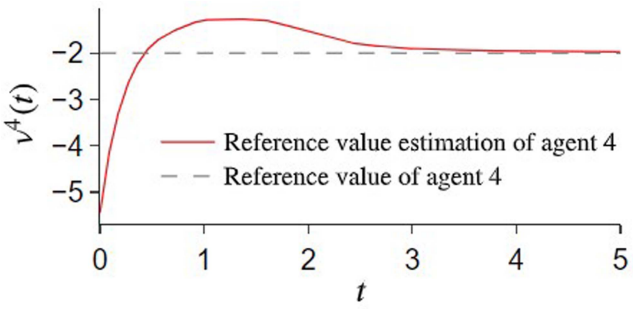

Fig. 4. Privacy breach of agent 4 in all three cases of Fig. 2(b).

Fig. 2(b). See [34] for a numerical simulation study using a directed graph.

VI. Conclusion

In this article, we studied the extent of privacy preservation that additive obfuscation signals can induce on the popular average consensus algorithm, when such signals are used to conceal the reference value of the agents from eavesdroppers. In the literature, a common trait of privacy preservation mechanisms that do not perturb the exact convergence of the algorithm is that each uses a particular class of vanishing noises or perturbation functions [6], [7], [24]. Our intent was to study whether stronger privacy preservation guarantees would be achievable if a broader class of signals was considered. To conduct a thorough study, we added obfuscation signals f^i and g^i to both the state equation and transmitted-out signal of each agent $i \in \mathcal{V}$ [see (2)]. Previous work [6] and [7] used $f^i = g^i$. Our study showed that privacy preservation still cannot be provided for agents, whose transmitted-out and all transmitted-in communication messages are available to the eavesdropper. However, using a broader class of admissible obfuscation signals and nonidentical obfuscation signals f^i and g^i for each agent $i \in \mathcal{V}$, we established a privacy preservation analysis framework that led to a stronger notion of privacy in the sense that an eavesdropper will be able to estimate neither the state nor a finite set of values to which the initial value of an agent belongs. We characterized the class of admissible obfuscation signals with a set of functionals with parameters α and β^i, $i \in \mathcal{V}$. The existing results [6], [7], [24], which use zero-sum and vanishing signals, are in fact using $\alpha = \beta^i = 0$. Thus, $\alpha = \beta^i = 0$ are known to any eavesdropper, internal or external. One can imagine that not every external eavesdropper is fully informed. Our extended results in [34] on privacy preservation against eavesdroppers that have various degrees of knowledge about α and β^is reveal, for example, that if β^i corresponding to the locally chosen admissible obfuscation signals of an agent $i \in \mathcal{V}$ is not known to an eavesdropper, the privacy of the agent i is preserved even if the eavesdropper knows all the transmitted input and output signals of agent i and the parameter α. Finally, notice that our study of the use of two obfuscation signals (f^i, g^i) as opposed to $f^i = g^i$ leads to the *theoretical* finding that the obfuscation signals can be nonvanishing. From a practical perspective, this finding can be interesting with respect to, for example, a naive eavesdropper without processing power. The use of nonvanishing signals conceals the final convergence value from this naive eavesdropper. The more interesting case; however, is the case of signals like chirp $g^i(t) = \sin(\phi_0 + 2\pi(\frac{c}{2}t^2 + \omega_0 t))$ that we identified through the results of Lemma 3.1. These signals lose uniform continuity when $t \to \infty$ and, thus, may be considered an impractical choice for an obfuscation signal. However, in practice, the consensus algorithms are terminated in finite time by tolerating some convergence error. As our example showed, the effect of this type of nonvanishing signals on $x^i(t)$ diminishes with time. Therefore, it is possible that we can have an acceptable convergence error but still be able to fully disguise the final convergence point even if the external eavesdropper knows α and β^i. A formal study of the use of the chirp-type signals for privacy preservation when the consensus algorithm is terminated in finite time is left as our future work.

Appendix A
Auxiliary Results

To provide proofs for our lemmas and theorems, we rely on a set of auxiliary results, which are given in this appendix.

Lemma 7.1 (Auxiliary result 1): Let $\mathbf{L}$ be the Laplacian matrix of a strongly connected and weight-balanced digraph. Recall $\mathbf{L}^+ = \mathbf{R}^\top \mathbf{L} \mathbf{R}$ from (3). Let $\mathbf{g}(t) = [g_1(t), \ldots, g_n(t)]^\top \in \mathcal{L}_n^\infty$. Then

$$\lim_{t\to\infty} \int_0^t \mathrm{e}^{-\mathbf{L}^+(t-\tau)} \mathbf{R}^\top \mathbf{L}\, \mathbf{g}(\tau) d\tau = \mathbf{0} \tag{29}$$

is guaranteed to hold if and only if

$$\lim_{t\to\infty} \int_0^t \mathrm{e}^{-(t-\tau)} g^i(\tau)\, d\tau = \alpha \in \mathbb{R}, \quad i \in \{1, \ldots, N\}. \tag{30}$$

Proof: Let

$$\dot{\boldsymbol{\zeta}} = -\mathbf{L}^+ \boldsymbol{\zeta} + \mathbf{R}^\top \mathbf{L} \mathbf{g}(t), \qquad \boldsymbol{\zeta}(0) \in \mathbb{R}^{N-1} \tag{31}$$

$$\dot{\boldsymbol{\eta}} = -\boldsymbol{\eta} + \mathbf{R}^\top \mathbf{L} \mathbf{g}(t), \qquad \boldsymbol{\eta}(0) \in \mathbb{R}^{N-1}. \tag{32}$$

The trajectories $t \mapsto \boldsymbol{\zeta}$ and $t \mapsto \boldsymbol{\eta}$ of these two dynamics for $t \in \mathbb{R}_{\geq 0}$ are given by

$$\boldsymbol{\zeta}(t) = \mathrm{e}^{-\mathbf{L}^+ t} \boldsymbol{\zeta}(0) + \int_0^t \mathrm{e}^{-\mathbf{L}^+(t-\tau)} \mathbf{R}^\top \mathbf{L} \mathbf{g}(\tau) d\tau \tag{33}$$

$$\boldsymbol{\eta}(t) = \mathrm{e}^{-t} \boldsymbol{\eta}(0) + \mathbf{R}^\top \mathbf{L} \int_0^t \mathrm{e}^{-(t-\tau)}\, \mathbf{g}(\tau) d\tau. \tag{34}$$

Let $\mathbf{e} = \boldsymbol{\zeta} - \boldsymbol{\eta}$. Then, the error dynamics between (31) and (32) is given by

$$\dot{\mathbf{e}} = -\mathbf{e} + (\mathbf{I} - \mathbf{L}^+) \boldsymbol{\zeta} \tag{35}$$

or equivalently

$$\dot{\mathbf{e}} = -\mathbf{L}^+ \mathbf{e} + (\mathbf{L}^+ + \mathbf{I}) \boldsymbol{\eta}. \tag{36}$$

Let (29) hold. Since $-\mathbf{L}^+$ is a Hurwitz matrix, we have $\lim_{t\to\infty} \boldsymbol{\zeta}(t) = 0$. Moreover, since $\mathbf{g}$ is essentially bounded, the trajectories of $\boldsymbol{\zeta}$ are guaranteed to be bounded. Therefore, considering error dynamics (35), by invoking the input-to-state

stability results [38], we have the guarantees that $\lim_{t\to\infty}\mathbf{e}(t)=\mathbf{0}$ and, consequently, $\lim_{t\to\infty}\boldsymbol{\eta}(t)=\mathbf{0}$. As such, from (34), we obtain

$$\mathbf{R}^\top\mathbf{L}\lim_{t\to\infty}\int_0^t \mathrm{e}^{-(t-\tau)}\mathbf{g}(\tau)d\tau=\mathbf{0}. \tag{37}$$

The null space of $\mathbf{R}^\top\mathbf{L}\in\mathbb{R}^{(N-1)\times N}$ is spanned by $\mathbf{1}_N$; therefore, $\lim_{t\to\infty}\int_0^t \mathrm{e}^{-(t-\tau)}\mathbf{g}(\tau)d\tau=\alpha\mathbf{1}_N,\quad \alpha\in\mathbb{R}$, which validates (30). Now, let (30) hold. Then, using (34), we obtain $\lim_{t\to\infty}\boldsymbol{\eta}(t)=\mathbf{0}$. Since $\mathbf{g}$ is essentially bounded, the trajectories of $\boldsymbol{\zeta}$ are guaranteed to be bounded. Thereby, considering error dynamics (36), by invoking the input-to-state stability results [38], we have the guarantees that $\lim_{t\to\infty}\mathbf{e}(t)=\mathbf{0}$ and, consequently, $\lim_{t\to\infty}\boldsymbol{\eta}(t)=\mathbf{0}$. Since $-\mathbf{L}^+$ is a Hurwitz matrix, we obtain (29) from (33). ■

Lemma 7.2 (Auxiliary result 2): Let $\mathbf{u}:\mathbb{R}_{\geq 0}\to\mathbb{R}^n$ be an essentially bounded signal and $\mathbf{E}\in\mathbb{R}^{n\times n}$ be a Hurwitz matrix.

(a) If $\lim_{t\to\infty}\mathbf{u}(t)=\bar{\mathbf{u}}\in\mathbb{R}^n$, and $\mathbf{E}\in\mathbb{R}^{n\times n}$, then

$$\lim_{t\to\infty}\int_0^t \mathrm{e}^{\mathbf{E}\,(t-\tau)}\mathbf{u}(\tau)d\tau=-\mathbf{E}^{-1}\,\bar{\mathbf{u}}. \tag{38}$$

(b) If $\lim_{t\to\infty}\int_0^t\mathbf{u}(\tau)d\tau=\bar{\mathbf{u}}\in\mathbb{R}^n$, then

$$\lim_{t\to\infty}\int_0^t \mathrm{e}^{\mathbf{E}\,(t-\tau)}\mathbf{u}(\tau)d\tau=\mathbf{0}. \tag{39}$$

Proof: To prove statement (a) we proceed as follows. Let $\boldsymbol{\mu}(t)=\mathbf{u}(t)-\bar{\mathbf{u}}$. Next, consider $\dot{\boldsymbol{\zeta}}=\mathbf{E}\,\boldsymbol{\zeta}+\boldsymbol{\mu}$ and $\boldsymbol{\zeta}(0)\in\mathbb{R}^n$, which gives $\boldsymbol{\zeta}(t)=\mathrm{e}^{\mathbf{E}\,t}\boldsymbol{\zeta}(0)+\int_0^t\mathrm{e}^{\mathbf{E}(t-\tau)}\boldsymbol{\mu}(\tau)d\tau, t\geq 0$. Since $\mathbf{E}$ is Hurwitz and $\boldsymbol{\mu}$ is an essentially bounded and vanishing signal, by virtue of the input-to-state stable (ISS) results for linear systems [38], we have $\lim_{t\to\infty}\boldsymbol{\zeta}(t)=0$. Consequently, $\lim_{t\to\infty}\int_0^t\mathrm{e}^{\mathbf{E}\,(t-\tau)}\boldsymbol{\mu}(\tau)d\tau=\mathbf{0}$, which guarantees (38).

To prove statement (b), we proceed as follows. Consider $\dot{\boldsymbol{\zeta}}=\mathbf{u},\ \dot{\boldsymbol{\eta}}=\mathbf{E}\boldsymbol{\eta}+\mathbf{u},\quad\boldsymbol{\zeta}(0)=\mathbf{0}$, and $\boldsymbol{\eta}(0)\in\mathbb{R}^n$, which result in $\boldsymbol{\zeta}(t)=\int_0^t\mathbf{u}(\tau)d\tau$ and

$$\boldsymbol{\eta}(t)=\mathrm{e}^{\mathbf{E}\,t}\boldsymbol{\eta}(0)+\int_0^t\mathrm{e}^{\mathbf{E}\,(t-\tau)}\mathbf{u}(\tau)d\tau. \tag{40}$$

Given the conditions on $\mathbf{u}$ both $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}$ are essentially bounded signals (recall that $\mathbf{E}$ is Hurwitz). Let $\mathbf{e}=\boldsymbol{\eta}-\boldsymbol{\zeta}$. Then, we can write $\dot{\mathbf{e}}=\mathbf{E}\,\mathbf{e}+\mathbf{E}\,\boldsymbol{\zeta}$ and $\mathbf{e}(0)=\boldsymbol{\eta}(0)\in\mathbb{R}^n$. Since $\boldsymbol{\zeta}$ is essentially bounded and satisfies $\lim_{t\to\infty}\mathbf{E}\boldsymbol{\zeta}(t)=\mathbf{E}\bar{\mathbf{u}}$, reasoning similar to that of the proof of statement (a), we can conclude $\lim_{t\to\infty}\mathbf{e}(t)=-\bar{\mathbf{u}}$. As a result, $\lim_{t\to\infty}\boldsymbol{\eta}(t)=\mathbf{0}$. Consequently, from (40), we obtain (39). ■

Lemma 7.3 (Auxiliary result 3): Let $\mathcal{G}$ be a strongly connected and weight-balanced digraph. Then, every island of any agent i is strongly connected and weight-balanced.

Proof: Without loss of generality, we prove our argument by showing that the island $\mathcal{G}_1^1$ of agent 1 is strongly connected and weight-balanced. By construction, we know that there is a directed path from every agent to every other agent in $\mathcal{G}_1^1$; therefore, $\mathcal{G}_1^1$ is strongly connected. Next, we show that $\mathcal{G}_1^1$ is weight-balanced. Let $\mathcal{V}_2=\mathcal{V}_1^1\backslash\{1\}$ and $\mathcal{V}_3=\mathcal{V}\backslash\mathcal{V}_2$. Let the nodes of $\mathcal{G}$ be labeled in accordance to $(1,\mathcal{V}_2,\mathcal{V}_3)$, respectively, and partition the graph Laplacian $\mathbf{L}$ accordingly as

$$\mathbf{L}=\begin{bmatrix}\mathsf{d}_{\text{out}}^1 & -\mathbf{A}_{12} & -\mathbf{A}_{13}\\ -\mathbf{A}_{21} & \mathbf{L}_{22} & \mathbf{0}\\ -\mathbf{A}_{31} & \mathbf{0} & \mathbf{L}_{33}\end{bmatrix}.$$

Since $\mathcal{G}$ is strongly connected and weight-balanced, we have $\mathbf{L}\mathbf{1}_N=\mathbf{0}$ and $\mathbf{1}_N^\top\mathbf{L}=\mathbf{0}$, which guarantee that

$$\mathbf{1}_{|\mathcal{V}_1^1|}^\top\begin{bmatrix}-\mathbf{A}_{12}\\ \mathbf{L}_{22}\end{bmatrix}=\mathbf{0},\qquad\begin{bmatrix}-\mathbf{A}_{21} & \mathbf{L}_{22}\end{bmatrix}\mathbf{1}_{|\mathcal{V}_1^1|}=\mathbf{0}. \tag{41}$$

Therefore

$$\mathbf{1}_{|\mathcal{V}_1^1|}^\top\begin{bmatrix}-\mathbf{A}_{12}\\ \mathbf{L}_{22}\end{bmatrix}\mathbf{1}_{|\mathcal{V}_1^1|}=0,\qquad\mathbf{1}_{|\mathcal{V}_1^1|}^\top\begin{bmatrix}-\mathbf{A}_{21} & \mathbf{L}_{22}\end{bmatrix}\mathbf{1}_{|\mathcal{V}_1^1|}=0$$

which we can use to conclude that $\mathtt{sum}(\mathbf{A}_{12}^\top)=\mathtt{sum}(\mathbf{A}_{21})$.

Let the Laplacian matrix of $\mathcal{G}_1^1$ be $\mathbf{L}_1^1$. Partitioning this matrix according to order node set $(1,\mathcal{V}_2)$, we obtain $\mathbf{L}_1^1=\begin{bmatrix}\mathsf{d}_{\text{out}}^{1,1} & -\mathbf{A}_{12}\\ -\mathbf{A}_{21} & \mathbf{L}_{22}\end{bmatrix}$, where $\mathsf{d}_{\text{out}}^{1,1}=\sum_{j\in\mathcal{V}_2}\mathsf{a}_{1j}=\mathtt{sum}(\mathbf{A}_{12}^\top)$. To establish that $\mathcal{G}_1^1$ is weight-balanced digraph, we next show that $\mathbf{1}_{|\mathcal{V}_1^1|}^\top\mathbf{L}_1^1=\mathbf{0}$. From $\mathbf{1}_N^\top\mathbf{L}=\mathbf{0}$, it follows that: $\mathbf{1}_{|\mathcal{V}_1^1|}^\top\begin{bmatrix}-\mathbf{A}_{12}\\ \mathbf{L}_{22}\end{bmatrix}=\mathbf{0}$. Therefore, to prove $\mathcal{G}_1^1$ is weight-balanced, we need to show that $\mathsf{d}_{\text{out}}^{1,1}+\mathtt{sum}(-\mathbf{A}_{21})=0$, which follows immediately from $\mathsf{d}_{\text{out}}^{1,1}=\mathtt{sum}(\mathbf{A}_{12}^\top)$ and $\mathtt{sum}(\mathbf{A}_{12}^\top)=\mathtt{sum}(\mathbf{A}_{21})$. ■

APPENDIX B
PROOF OF OUR MAIN RESULTS

Proof of Theorem 3.1: We prove first necessity. We write the algorithm (2) in compact form

$$\dot{\mathbf{x}}=-\mathbf{L}\,\mathbf{x}-\mathbf{L}\,\mathbf{g}+\mathbf{f}+\mathbf{D}^{\text{out}}\,\mathbf{g}=-\mathbf{L}\,\mathbf{x}+\mathbf{f}+\mathbf{A}\,\mathbf{g}. \tag{42}$$

Left multiplying both sides of (42) by $\mathbf{1}_N^\top$ gives $\sum_{j=1}^N\dot{x}^j(t)=\sum_{j=1}^N(f^i(t)+\mathsf{d}_{\text{out}}^i\,g^i(t))$, which results in $\sum_{j=1}^N x^j(t)=\sum_{j=1}^N x^j(0)+\int_0^t\sum_{j=1}^N(f^i(\tau)+\mathsf{d}_{\text{out}}^i\,g^i(\tau))\,d\tau$. Because $x^i(0)=\mathsf{r}^i$, to ensure $\lim_{t\to\infty}x^i(t)=\frac{1}{N}\sum_{j=1}^N\mathsf{r}^j,\ i\in\mathcal{V}$, we necessarily need (4b).

Next, we apply the change of variable

$$\mathbf{p}=\begin{bmatrix}p_1\\ \mathbf{p}_{2:N}\end{bmatrix}=\mathbf{T}\,\mathbf{x} \tag{43}$$

where $\mathbf{T}$ is defined in (3), to write (42) in the equivalent form

$$\dot{p}_1=\frac{1}{\sqrt{N}}\sum_{i=1}^N\left(f^i+\mathsf{d}_{\text{out}}^i\,g^i\right) \tag{44a}$$

$$\dot{\mathbf{p}}_{2:N}=-\mathbf{L}^+\,\mathbf{p}_{2:N}+\mathbf{R}^\top(\mathbf{f}+\mathbf{A}\,\mathbf{g}). \tag{44b}$$

The solution of (44) is

$$p_1(t)=\frac{1}{\sqrt{N}}\sum_{i=1}^N x^i(0) \tag{45a}$$

$$+\frac{1}{\sqrt{N}}\int_0^t\sum_{i=1}^N\left(f^i(\tau)+\mathsf{d}^i_{\text{out}}\,g^i(\tau)\right)d\tau$$

$$\mathbf{p}_{2:N}(t)=\mathrm{e}^{-\mathbf{L}^+t}\,\mathbf{p}_{2:N}(0)+\int_0^t\mathrm{e}^{-\mathbf{L}^+(t-\tau)}\mathbf{R}^\top(\mathbf{f}(\tau)+\mathbf{A}\,\mathbf{g}(\tau))\,d\tau. \quad (45b)$$

Given (4a), (45a) results in $\lim_{t\to\infty}p_1(t)=\frac{1}{\sqrt{N}}\sum_{i=1}^N x^i(0)=\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathsf{r}^i$. Consequently, given (43), to ensure $\lim_{t\to\infty}x^i(t)=\frac{1}{N}\sum_{j=1}^N\mathsf{r}^j$, $i\in\mathcal{V}$, we need

$$\lim_{t\to\infty}\mathbf{p}_{2:N}(t)=\mathbf{0}. \quad (46)$$

Because for a strongly connected and weight-balanced digraph, $-\mathbf{L}^+$ is a Hurwitz matrix, $\lim_{t\to\infty}\mathrm{e}^{-\mathbf{L}^+t}\mathbf{p}_{2:N}(0)=\mathbf{0}$. Then, the necessary condition for (46) is (4b).

The sufficiency proof follows from noting that under (4), the trajectories of (45) satisfy $\lim_{t\to\infty}p_1(t)=\frac{1}{\sqrt{N}}\sum_{i=1}^N x^i(0)$ and $\lim_{t\to\infty}\mathbf{p}_{2:N}(t)=\mathbf{0}$. Then, given (43) and $x^i(0)=\mathsf{r}^i$, we obtain $\lim_{t\to\infty}x^i(t)=\frac{1}{N}\sum_{j=1}^N\mathsf{r}^j$, $i\in\mathcal{V}$. ■

Proof of Theorem 3.2: Given (5), it is straightforward to see that (6a) is necessary and sufficient for (4a). Next, we observe that using (5), we can write $\lim_{t\to\infty}\int_0^t\mathbf{R}^\top(\mathbf{f}(\tau)+\mathbf{D}^{\text{out}}\,\mathbf{g}(\tau))d\tau=\mathbf{R}^\top\begin{bmatrix}\beta^1&\cdots&\beta^N\end{bmatrix}^\top$. Then, it follows from statement (b) of Lemma 7.2 that $\lim_{t\to\infty}\int_0^t\mathrm{e}^{-\mathbf{L}^+(t-\tau)}\mathbf{R}^\top(\mathbf{f}(\tau)+\mathbf{D}^{\text{out}}\,\mathbf{g}(\tau))d\tau=\mathbf{0}$. As a result, given $\mathbf{f}+\mathbf{A}\,\mathbf{g}=\mathbf{f}+\mathbf{D}^{\text{out}}\,\mathbf{g}-\mathbf{L}\,\mathbf{g}$, we obtain

$$\lim_{t\to\infty}\int_0^t\mathrm{e}^{-\mathbf{L}^+(t-\tau)}\mathbf{R}^\top(\mathbf{f}(\tau)+\mathbf{A}\,\mathbf{g}(\tau))\,d\tau=-\lim_{t\to\infty}\int_0^t\mathrm{e}^{-\mathbf{L}^+(t-\tau)}\mathbf{R}^\top\mathbf{L}\,\mathbf{g}(\tau)d\tau. \quad (47)$$

Given (47), by virtue of Lemma 7.1, (4b) holds if and only if (6b) holds. ■

Proof of Lemma 3.1: When condition (a) holds, the proof of the statement follows from statement (a) of Lemma 7.2. When condition (b) is satisfied, the proof follows from statements (a) and (b) of Lemma 7.2, which, respectively, give $\lim_{t\to\infty}\int_0^t\mathrm{e}^{-(t-\tau)}g_1(\tau)d\tau=\alpha$ and $\lim_{t\to\infty}\int_0^t\mathrm{e}^{-(t-\tau)}g_2(\tau)d\tau=0$. When condition (c) is satisfied, the proof follows from statement (a) of Lemma 7.2, which gives $\lim_{t\to\infty}\int_0^t\mathrm{e}^{-(t-\tau)}g_1(\tau)d\tau=\alpha$, and noting that $\int_0^t\mathrm{e}^{-(t-\tau)}g_2(\tau)d\tau$ is the zero state response of system $\dot\zeta=-\zeta+g_2$. Since $g_2(t)$ is essentially bounded, this system is ISS, and as a result, it is also integral ISS [38]. Then, $\int_0^t\mathrm{e}^{-(t-\tau)}g_2(\tau)d\tau=0$ follows from [38, Lemma 3.1]. ■

Proof of Lemma 4.1: Let the error variables of the two execution of (2) described in the statement be $\delta x^i(t)=x^{i'}(t)-x^i(t)$, $\delta y^i(t)=y^{i'}(t)-y^i(t)$, $\delta g^i(t)=g^{i'}(t)-g^i(t)$, and $\delta f^i(t)=f^{i'}(t)-f^i(t)$, $i\in\mathcal{V}$. Consequently

$$\delta x^1(0)=0,\quad \delta\mathbf{x}_4=\mathbf{0},\quad \delta\mathbf{x}_5(0)=\mathbf{0} \quad (48a)$$

$$\delta x^i(0)\in\mathbb{R},\qquad i\in\left(\mathcal{V}^1_{1,2}\cup\mathcal{V}^1_{1,3}\right) \quad (48b)$$

$$\delta\mathbf{x}_2(0)=-\mathbf{A}_{23}\mathbf{L}_{33}^{-1}\delta\mathbf{x}_3(0) \quad (48c)$$

and

$$\delta g^1(t)\equiv 0,\ \delta f^1(t)\equiv 0 \quad (49a)$$

$$\delta\mathbf{g}_l(t)\equiv\mathbf{0},\ \delta\mathbf{f}_l(t)\equiv\mathbf{0},\qquad l\in\{3,4,5\} \quad (49b)$$

$$\delta\mathbf{g}_2(t)=-\mathrm{e}^{-\mathbf{D}^{\text{out}}_{22}t}\delta\mathbf{x}_2(0),\ \delta\mathbf{f}_2(t)=-\mathbf{A}_{23}\mathrm{e}^{-\mathbf{L}_{33}t}\delta\mathbf{x}_3(0). \quad (49c)$$

Given the interagent interactions across the network based on agent grouping in accordance to the definition of the island $\mathcal{G}^1_1$ (see Fig. 2), the error dynamics pertained to the modified static average consensus algorithm (2) reads as

$$\begin{bmatrix}\delta\dot x^1\\ \delta\dot{\mathbf{x}}_2\\ \delta\dot{\mathbf{x}}_3\\ \delta\dot{\mathbf{x}}_4\\ \delta\dot{\mathbf{x}}_5\end{bmatrix}=-\underbrace{\begin{bmatrix}\mathsf{d}^1_{\text{out}}&-\mathbf{A}_{12}&\mathbf{0}&-\mathbf{A}_{14}&-\mathbf{A}_{15}\\ -\mathbf{A}_{21}&\mathbf{L}_{22}&-\mathbf{A}_{23}&-\mathbf{A}_{24}&\mathbf{0}\\ -\mathbf{A}_{31}&-\mathbf{A}_{32}&\mathbf{L}_{33}&-\mathbf{A}_{34}&\mathbf{0}\\ -\mathbf{A}_{41}&-\mathbf{A}_{42}&\mathbf{0}&\mathbf{L}_{44}&\mathbf{0}\\ -\mathbf{A}_{51}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{L}_{55}\end{bmatrix}}_{\mathbf{L}}\begin{bmatrix}\delta x^1\\ \delta\mathbf{x}_2\\ \delta\mathbf{x}_3\\ \delta\mathbf{x}_4\\ \delta\mathbf{x}_5\end{bmatrix}$$

$$+\underbrace{\begin{bmatrix}0&\mathbf{A}_{12}&\mathbf{0}&\mathbf{A}_{14}&\mathbf{A}_{15}\\ \mathbf{A}_{21}&\mathbf{A}_{22}&\mathbf{A}_{23}&\mathbf{A}_{24}&\mathbf{0}\\ \mathbf{A}_{31}&\mathbf{A}_{32}&\mathbf{A}_{33}&\mathbf{A}_{34}&\mathbf{0}\\ \mathbf{A}_{41}&\mathbf{A}_{42}&\mathbf{0}&\mathbf{A}_{44}&\mathbf{0}\\ \mathbf{A}_{51}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{A}_{55}\end{bmatrix}}_{\mathbf{A}}\begin{bmatrix}\delta g^1\\ \delta\mathbf{g}_2\\ \delta\mathbf{g}_3\\ \delta\mathbf{g}_4\\ \delta\mathbf{g}_5\end{bmatrix}+\begin{bmatrix}\delta f^1\\ \delta\mathbf{f}_2\\ \delta\mathbf{f}_3\\ \delta\mathbf{f}_4\\ \delta\mathbf{f}_5\end{bmatrix}. \quad (50)$$

Since, for a strongly connected and weight-balanced digraph, we have $\operatorname{rank}(\mathbf{L})=N-1$ and $-(\mathbf{L}+\mathbf{L}^\top)\le 0$, the subblock matrices $-\mathbf{L}_{33}$ and $-\mathbf{L}_{44}$ and $-\mathbf{L}_{55}$ satisfy $-(\mathbf{L}_{ii}+\mathbf{L}_{ii}^\top)<0, i\in\{1,\dots,5\}$. Thereby, they are invertible and Hurwitz matrices.

To establish (21), we show that $\mathbf{1}_N^\top\delta\mathbf{x}(0)=\mathbf{0}_N$. For this, note that taking into account (48), we can write

$$\delta\mathbf{x}(0)=\underbrace{\begin{bmatrix}0&0&0&0&0\\ 0&0&-\mathbf{A}_{23}&0&0\\ 0&0&\mathbf{L}_{33}&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0\end{bmatrix}}_{\mathbf{B}}\begin{bmatrix}0\\ \mathbf{L}_{33}^{-1}\delta\mathbf{x}_3(0)\\ \mathbf{L}_{33}^{-1}\delta\mathbf{x}_3(0)\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}. \quad (51)$$

Comparing $\mathbf{B}$ with the block partitioned $\mathbf{L}$ in (50), it is evident that $\mathbf{1}^\top\mathbf{B}=\mathbf{0}$ follows from $\mathbf{1}^\top\mathbf{L}=\mathbf{0}$. Consequently, we can deduce from (51) that $\mathbf{1}^\top\delta\mathbf{x}(0)=0$. Next, given (21), we validate (15) by invoking Theorem 3.2 and showing that the perturbation signals $(f^{i'},g^{i'})$, $i\in\mathcal{V}$, satisfy the sufficient conditions in (6). For $i\in\mathcal{V}\backslash\mathcal{V}^1_{1,2}$, the sufficient conditions in (6) are trivially satisfied. To show (6a) for $i\in\mathcal{V}^1_{1,2}$, we proceed as follows. First note that since (f^i,g^i), $i\in\mathcal{V}^1_{1,2}$, are admissible signals, they necessarily satisfy (6a). Next, note that using (10), we can write $\int_0^t(-\mathbf{A}_{23}\mathrm{e}^{-\mathbf{L}_{33}\tau}\delta\mathbf{x}_3(0)+\mathbf{D}^{\text{out}}_{22}\mathrm{e}^{-\mathbf{D}^{\text{out}}_{22}\tau}\delta\mathbf{x}_2(0))d\tau=\mathbf{A}_{23}\mathbf{L}_{33}^{-1}\mathrm{e}^{-\mathbf{L}_{33}t}\delta\mathbf{x}_3(0)-\mathrm{e}^{-\mathbf{D}^{\text{out}}_{22}\tau}\delta\mathbf{x}_2(0)$. Let $\mathfrak{B}_2=[\{\beta^i\}_{i\in\mathcal{V}^1_{1,2}}]$. Then, in light of the aforementioned observations and the fact that $-\mathbf{L}_{33}$ and $-\mathbf{D}^{\text{out}}_{22}$ are Hurwitz matrices, we can write $\lim_{t\to\infty}\int_0^t(\mathbf{f}'_2(\tau)+\mathbf{D}^{\text{out}}_{22}\mathbf{g}'_2(\tau))d\tau=\mathfrak{B}_2+$

$\lim_{t\to\infty}(\mathbf{A}_{23}\mathbf{L}_{33}^{-1}\mathrm{e}^{-\mathbf{L}_{33}t}\delta\mathbf{x}_3(0)-\mathrm{e}^{-\mathbf{D}_{22}^{\mathrm{out}}\tau}\delta\mathbf{x}_2(0))=\mathfrak{B}_2$, which shows that $(f^{i'},g^{i'})$, $i\in\mathcal{V}_{1,2}^1$, also satisfy the sufficient condition (6a). Establishing that $g^{i'}$, $i\in\mathcal{V}_{1,2}^1$, satisfies that the sufficient condition (6b) follows from the admissibility of g^i, $i\in\mathcal{V}_{1,2}^1$, which ensures that it satisfies (6b), and direct calculations as shown in the following: $\lim_{t\to\infty}\int_0^t \mathrm{e}^{-(t-\tau)}g^{i'}(\tau)\,d\tau=\alpha+\lim_{t\to\infty}\int_0^t \mathrm{e}^{-(t-\tau)}\mathrm{e}^{-\mathrm{d}_{\mathrm{out}}^i\tau}\delta x^i(0)\,d\tau=\alpha$. Here, we used the fact that for a strongly connected digraph, we have $\mathrm{d}_{\mathrm{out}}^i\geq 1$.

Next, we assume that (13) or equivalently

$$\delta y^1(t)=\delta x^1(t)+\delta g^1(t)\equiv\mathbf{0},\qquad t\in\mathbb{R}_{\geq 0}\tag{52a}$$
$$\delta\mathbf{y}_2(t)=\delta\mathbf{x}_2(t)+\delta\mathbf{g}_2(t)\equiv\mathbf{0},\qquad t\in\mathbb{R}_{\geq 0}\tag{52b}$$
$$\delta\mathbf{y}_4(t)=\delta\mathbf{x}_4(t)+\delta\mathbf{g}_4(t)\equiv\mathbf{0},\qquad t\in\mathbb{R}_{\geq 0}\tag{52c}$$
$$\delta\mathbf{y}_5(t)=\delta\mathbf{x}_5(t)+\delta\mathbf{g}_5(t)\equiv\mathbf{0},\qquad t\in\mathbb{R}_{\geq 0}\tag{52d}$$

hold. Then, for the given initial conditions (48), we identify the perturbation signals that make the error dynamics (50) render such an output. As we show in the following, these perturbation signals are exactly the same as (49). Then, the proof is established by the fact that given a set of initial conditions and integrable external signals, the solution of any linear ordinary differential equation is unique. That is, if we implement the identified inputs, the error dynamics is guaranteed to satisfy (52). If (52) holds, then the error dynamics (50) reads as

$$\delta\dot{x}^1=-\mathrm{d}_{\mathrm{out}}^1\delta x^1+\delta f^1\tag{53a}$$
$$\delta\dot{\mathbf{x}}_2=-\mathbf{D}_{22}^{\mathrm{out}}\delta\mathbf{x}_2+\mathbf{A}_{23}\delta\mathbf{x}_3+\mathbf{A}_{23}\delta\mathbf{g}_3+\delta\mathbf{f}_2\tag{53b}$$
$$\delta\dot{\mathbf{x}}_3=-\mathbf{L}_{33}\delta\mathbf{x}_3+\mathbf{A}_{33}\delta\mathbf{g}_3+\delta\mathbf{f}_3\tag{53c}$$
$$\delta\dot{\mathbf{x}}_4=-\mathbf{D}_{44}^{\mathrm{out}}\delta\mathbf{x}_4+\delta\mathbf{f}_4\tag{53d}$$
$$\delta\dot{\mathbf{x}}_5=-\mathbf{D}_{55}^{\mathrm{out}}\delta\mathbf{x}_5+\delta\mathbf{f}_5.\tag{53e}$$

Here, we used $\mathbf{L}_{ii}=\mathbf{D}_{ii}^{\mathrm{out}}-\mathbf{A}_{ii}$, $i\in\{1,2,4,5\}$. Next, we choose the perturbation signals according to (49). Then, for the given initial conditions (48), from (53), we obtain

$$\delta\dot{x}^1=-\mathrm{d}_{\mathrm{out}}^1\delta x^1,\quad\Rightarrow\delta x^1(t)=0\Rightarrow\delta y^1(t)\equiv 0\tag{54a}$$
$$\delta\dot{\mathbf{x}}_3=-\mathbf{L}_{33}\,\delta\mathbf{x}_3,\quad\Rightarrow\delta\mathbf{x}_3(t)=\mathrm{e}^{-\mathbf{L}_{33}t}\delta\mathbf{x}_3(0)\tag{54b}$$
$$\delta\dot{\mathbf{x}}_4=-\mathbf{D}_{44}^{\mathrm{out}}\delta\mathbf{x}_4,\quad\Rightarrow\delta\mathbf{x}_4(t)\equiv\mathbf{0},\Rightarrow\delta\mathbf{y}_4(t)\equiv\mathbf{0}\tag{54c}$$
$$\delta\dot{\mathbf{x}}_5=-\mathbf{D}_{55}^{\mathrm{out}}\delta\mathbf{x}_5,\quad\Rightarrow\delta\mathbf{x}_5(t)\equiv\mathbf{0},\Rightarrow\delta\mathbf{y}_5(t)\equiv\mathbf{0}\tag{54d}$$

for $t\in\mathbb{R}_{\geq 0}$. Substituting for $\mathbf{x}_3$ and $\delta\mathbf{f}_2$ into (53b), we obtain

$$\begin{aligned}\delta\dot{\mathbf{x}}_2&=-\mathbf{D}_{22}^{\mathrm{out}}\delta\mathbf{x}_2+\mathbf{A}_{23}\mathrm{e}^{-\mathbf{L}_{33}t}\delta\mathbf{x}_3(0)-\mathbf{A}_{23}\mathrm{e}^{-\mathbf{L}_{33}t}\delta\mathbf{x}_3(0)\\&=-\mathbf{D}_{22}^{\mathrm{out}}\delta\mathbf{x}_2,\Rightarrow\delta\mathbf{x}_2(t)=\mathrm{e}^{-\mathbf{D}_{22}^{\mathrm{out}}t}\delta\mathbf{x}_2(0)\end{aligned}\tag{55}$$

for $t\in\mathbb{R}_{\geq 0}$. Finally using $\delta\mathbf{g}_2$ in (49c), we obtain

$$\begin{aligned}\delta\mathbf{y}_2(t)&=\delta\mathbf{x}_2+\delta\mathbf{g}_2\\&=\mathrm{e}^{-\mathbf{D}_{22}^{\mathrm{out}}t}\delta\mathbf{x}_2(0)-\mathrm{e}^{-\mathbf{D}_{22}^{\mathrm{out}}t}\delta\mathbf{x}_2(0)\equiv\mathbf{0}\end{aligned}\tag{56}$$

for $t\in\mathbb{R}_{\geq 0}$. ■

Proof of Corollary 4.1: The proof can be deduced from the proof of Lemma 4.1. The proof trivially follows from (11), (12), and (17) through singling out an agent in $\mathcal{V}_{1,3}^1$ and finding all of its in-neighbors in $\mathcal{V}_{1,2}^1$. ■

Proof of Lemma 4.2: For an internal eavesdropper, given (2) and (25), we can write $\dot{\psi}+\dot{x}^i=f^i+\mathrm{d}_{\mathrm{out}}^i\,g^i$, which, because of $x^i(0)=\mathsf{r}^i$ and $\zeta(0)=-\beta^i$, gives $\psi(t)=-x^i(t)+\mathsf{r}^i+\int_0^t(f^i(\tau)+\mathrm{d}_{\mathrm{out}}^i\,g^i(\tau))d\tau-\beta^i$, $t\in\mathbb{R}_{\geq 0}$. Then, using (2b) and (25b), we obtain (27) as the estimation error. Subsequently, because of (5) and since $\lim_{t\to\infty}(x^1(t)-x^i(t))=0$, from (27), we obtain $\lim_{t\to\infty}\nu(t)=\mathsf{r}^i$. For an external eavesdropper, given (2) and (24a), we can write $\dot{\zeta}+\dot{x}^i=f^i+\mathrm{d}_{\mathrm{out}}^i\,g^i$, which, given $x^i(0)=\mathsf{r}^i$ and $\zeta(0)=-\beta^i-\alpha$, for $t\in\mathbb{R}_{\geq 0}$, gives

$$\zeta(t)=-x^i(t)+\mathsf{r}^i+\int_0^t(f^i(\tau)+\mathrm{d}_{\mathrm{out}}^i\,g^i(\tau))d\tau\;-\beta^i-\alpha.\tag{57}$$

On the other hand, using (2b), $t\mapsto\eta(t)$ is obtained from (26b). Then, tracking error (26a) is readily deduced from (24c) and (57). Next, given (5) and (6b) and also $\lim_{t\to\infty}\mathrm{e}^{-t}\eta_0=0$, we obtain $\lim_{t\to\infty}\nu(t)=\mathsf{r}^i+\lim_{t\to\infty}(-x^i(t)+\int_0^t\mathrm{e}^{-(t-\tau)}x^i(\tau)d\tau)$.

Subsequently, since $\lim_{t\to\infty}x^i(t)=\frac{1}{N}\sum_{j=1}^N\mathsf{r}^j$, we can conclude our proof by invoking Lemma 7.2 that guarantees $\lim_{t\to\infty}\int_0^t\mathrm{e}^{-(t-\tau)}x^i(\tau)d\tau=\lim_{t\to\infty}x^i(t)=\frac{1}{N}\sum_{j=1}^N\mathsf{r}^j$. ■

Proof of Proposition 4.1: Consider the aggregate dynamics of η and $\mathbf{x}_i$, $i\in\{2,3,4\}$, which reads as

$$\begin{bmatrix}\dot{\eta}\\\dot{\mathbf{x}}_2\\\dot{\mathbf{x}}_3\\\dot{\mathbf{x}}_4\end{bmatrix}=-\underbrace{\begin{bmatrix}\mathrm{d}_{\mathrm{out}}^{1,1}&-\mathbf{A}_{12}&\mathbf{0}&-\mathbf{A}_{14}\\-\mathbf{A}_{21}&\mathbf{D}_{22}^{\mathrm{out}}&-\mathbf{A}_{23}&-\mathbf{A}_{24}\\-\mathbf{A}_{31}&-\mathbf{A}_{32}&\mathbf{D}_{33}^{\mathrm{out}}&\mathbf{0}\\-\mathbf{A}_{41}&-\mathbf{A}_{42}&\mathbf{0}&\mathbf{D}_{44}^{\mathrm{out}}\end{bmatrix}}_{\mathbf{L}_1^1}\begin{bmatrix}y^1\\\mathbf{y}_2\\\mathbf{y}_3\\\mathbf{y}_4\end{bmatrix}+\begin{bmatrix}0\\\mathbf{f}_2+\mathbf{D}_{22}^{\mathrm{out}}\mathbf{g}_2\\\mathbf{f}_3+\mathbf{D}_{33}^{\mathrm{out}}\mathbf{g}_3\\\mathbf{f}_4+\mathbf{D}_{44}^{\mathrm{out}}\mathbf{g}_4\end{bmatrix}.$$

Notice that $\mathbf{L}_1^1$ is the Laplacian matrix of graph $\mathcal{G}_1^1$. By virtue of Lemma 7.3, we know that $\mathcal{G}_1^1$ is a strongly connected and weight-balanced digraph. Consequently, left multiplying both sides of equation above with $\mathbf{1}_{|\mathcal{V}_1^1|}^\top$ gives $\dot{\eta}+\sum_{j\in\mathcal{V}_1^1\setminus\{1\}}x^i=\sum_{j\in\mathcal{V}_1^1\setminus\{1\}}(f^j(t)+\mathrm{d}_{\mathrm{out}}^j\,g^j(t))$. Thereby, given $\eta(0)=-\sum_{j\in\mathcal{V}_1^1\setminus\{1\}}\beta^i$ and $x^i(0)=\mathsf{r}^i$, we obtain $\eta(t)=\sum_{j\in\mathcal{V}_1^1\setminus\{1\}}\mathsf{r}^j-\sum_{j\in\mathcal{V}_1^1\setminus\{1\}}x^j(t)+\sum_{j\in\mathcal{V}_1^1\setminus\{1\}}\int_0^t(f^j(\tau)+\mathrm{d}_{\mathrm{out}}^j\,g^j(\tau))d\tau-\sum_{j\in\mathcal{V}_1^1\setminus\{1\}}\beta^i$. On the other hand, following the proof of Theorem 4.2, we can conclude that $\sum_{i\in\mathcal{V}_{1,4}^1}\zeta_i(t)=\sum_{i\in\mathcal{V}_{1,4}^1}\mathsf{r}^i-\sum_{i\in\mathcal{V}_{1,4}^1}x^i(t)+\sum_{i\in\mathcal{V}_{1,4}^1}\int_0^t(f^i(\tau)+\mathrm{d}_{\mathrm{out}}^i\,g^i(\tau))d\tau-\sum_{i\in\mathcal{V}_{1,4}^1}\beta^i$. Therefore, we can write $n_{2,3}\,\mu(t)=\sum_{j\in(\mathcal{V}_{1,2}^1\cup\mathcal{V}_{1,3}^1)}\mathsf{r}^i\sum_{j\in(\mathcal{V}_{1,2}^1\cup\mathcal{V}_{1,3}^1)}x^i(t)-\sum_{j\in(\mathcal{V}_{1,2}^1\cup\mathcal{V}_{1,3}^1)}\beta^i+\sum_{j\in(\mathcal{V}_{1,2}^1\cup\mathcal{V}_{1,3}^1)}\int_0^t(f^j(\tau)+\mathrm{d}_{\mathrm{out}}^j\,g^j(\tau))d\tau+n_{2,3}\,x^1(t)$.

The proof then follows from the necessary condition (5) on the perturbation signals and the fact that $\lim_{t\to\infty} n_{2,3}\, x^1(t) - \sum_{j\in(\mathcal{V}^1_{1,2}\cup\mathcal{V}^1_{1,3})} x^i(t) = 0$ (recall that $\lim_{t\to\infty} x^i(t) = \lim_{t\to\infty} x^j(t)$, $\forall i, j \in \mathcal{V}$). ■

REFERENCES

[1] N. Rezazadeh and S. S. Kia, "Privacy preservation in a continuous-time static average consensus algorithm over directed graphs," in *Proc. Amer. Control Conf.*, 2018. pp. 5890–5895.

[2] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[3] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. 4th Int. Symp. Inf. Process. Sensor Netw.*, 2005, pp. 63–70.

[4] L. Georgopoulos and M. Hasler, "Distributed machine learning in networks by consensus," *Neurocomputing*, vol. 124, pp. 2–12, 2014.

[5] Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, and H. Yu, "Federated learning," *Synth. Lectures Artif. Intell. Mach. Learn.*, vol. 13, no. 3, pp. 1–207, 2019.

[6] N. E. Manitara and C. N. Hadjicostis, "Privacy-preserving asymptotic average consensus," in *Proc. Eur. Control Conf.*, 2013, pp. 760–765.

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

[8] M. Kefayati, M. S. Talebi, B. H. Khalaj, and H. R. Rabiee, "Secure consensus averaging in sensor networks using random offsets," in *Proc. IEEE Int. Conf. Telecommun.*, 2007, pp. 556–560.

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

[10] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3/4, pp. 211–407, 2014.

[11] Z. Huang, S. Mitra, and N. Vaidya, "Differentially private distributed optimization," in *Proc. Int. Conf. Distrib. Comput. Netw.*, 2015, Art. no. 4.

[12] J. L. Ny and G. Pappas, "Differentially private Kalman filtering," in *Proc. Allerton Conf. Commun., Control Comput.*, 2012, pp. 1618–1625.

[13] J. L. Ny and G. J. Pappas, "Differential private filtering," *IEEE Trans. Autom. Control*, vol. 59, no. 2, pp. 341–354, Feb. 2014.

[14] S. Gade, A. Winnicki, and S. Bose, "On privatizing equilibrium computation in aggregate games over networks," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 3272–3277, 2020.

[15] N. Gupta, J. Katz, and N. Chopra, "Privacy in distributed average consensus," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 9515–9520, 2017.

[16] L. Yu, W. Yu, and Y. Lv, "Multi-dimensional privacy-preserving average consensus in wireless sensor networks," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 3, pp. 1104–1108, Mar. 2022.

[17] W. Fang, M. Zamani, and Z. Chen, "Secure and privacy preserving consensus for second-order systems based on Paillier encryption," *Syst. Control Lett.*, vol. 148, 2021, Art. no. 104869.

[18] C. N. Hadjicostis and A. D. Domínguez-García, "Privacy-preserving distributed averaging via homomorphically encrypted ratio consensus," *IEEE Trans. Autom. Control*, vol. 65, no. 9, pp. 3887–3894, Sep. 2020.

[19] M. Ruan, H. Gao, and Y. Wang, "Secure and privacy-preserving consensus," *IEEE Trans. Autom. Control*, vol. 64, no. 10, pp. 4035–4049, Oct. 2019.

[20] N. Gupta and N. Chopra, "Model-based encryption: Privacy of states in networked control systems," in *Proc. 56th Allerton Conf. Commun., Control Comput.*, 2018, pp. 242–248.

[21] H. Gao, C. Zhang, M. Ahmad, and Y. Wang, "Privacy-preserving average consensus on directed graphs using push-sum," in *Proc. IEEE Conf. Commun. Netw. Secur.*, 2018, pp. 1–9.

[22] Y. Liu, J. Wu, I. R. Manchester, and G. Shi, "Gossip algorithms that preserve privacy for distributed computation—Part I: The algorithms and convergence conditions," in *Proc. IEEE Conf. Decis. Control*, 2018, pp. 4499–4504.

[23] Y. Liu, J. Wu, I. R. Manchester, and G. Shi, "Gossip algorithms that preserve privacy for distributed computation—Part II: Performance against eavesdroppers," in *Proc. IEEE Conf. Decis. Control*, 2018, pp. 5346–5351.

[24] C. Altafini, "A system-theoretic framework for privacy preservation in continuous-time multiagent dynamics," *Automatica*, vol. 122, 2020, Art. no. 109253.

[25] Y. Wang, J. Lu, W. Zheng, and K. Shi, "Privacy-preserving consensus for multi-agent systems via node decomposition strategy," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no. 8, pp. 3474–3484, Aug. 2021.

[26] Y. Xiong and Z. Li, "Privacy preserving discrete-time average consensus by injecting edge-based perturbations," in *Proc. 40th Chin. Control Conf.*, 2021, pp. 5413–5418.

[27] D. I. Ridgley, R. A. Freeman, and K. M. Lynch, "Simple, private, and accurate distributed averaging," in *Proc. 57th Allerton Conf. Commun., Control Comput.*, 2019, pp. 446–452.

[28] Y. Xiong and Z. Li, "Privacy preserving average consensus by adding edge-based perturbation signals," in *Proc. IEEE Int. Conf. Control Technol. Appl.*, 2020, pp. 712–717.

[29] S. Zhang, T. O. Timoudas, and M. A. Dahleh, "Consensus with preserved privacy against neighbor collusion," *Control Theory Technol.*, vol. 18, no. 4, pp. 409–418, 2020.

[30] I. L. D. Ridgley, R. A. Freeman, and K. M. Lynch, "Private and hot-pluggable distributed averaging," *IEEE Control Syst. Lett.*, vol. 4, no. 4, pp. 988–993, Oct. 2020.

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

[32] F. Bullo, J. Cortés, and S. Martínez, *Distributed Control of Robotic Networks* (Applied Mathematics Series). Princeton, NJ, USA: Princeton Univ. Press, 2009.

[33] P. Flandrin, *Explorations in Time-Frequency Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2018.

[34] N. Rezazadeh and S. S. Kia, "Privacy preservation in continuous-time average consensus algorithm via deterministic additive obfuscation signals," 2021, *arXiv:1904.05286*.

[35] M. Hou and R. J. Patton, "Input observability and input reconstruction," *Automatica*, vol. 34, no. 6, pp. 789–794, 1998.

[36] M. L. J. Hautus, "Strong detectability and observers," *Linear Algebra Appl.*, vol. 50, pp. 353–368, 1983.

[37] C. Read and R. Wilson, *An Atlas of Graphs*. London, U.K.: Oxford Univ. Press, 2005.

[38] S. N. Dashkovskiy, D. V. Efimov, and E. D. Sontag, "Input to state stability and allied system properties," *Autom. Remote Control*, vol. 72, no. 8, pp. 1579–1614, 2011.

Navid Rezazadeh received the B.Sc. degree in mechanical engineering from the Sharif University of Technology, Tehran, Iran, in 2013, and the M.Sc. and Ph.D. degrees in mechanical and aerospace engineering from the University of California, Irvine, CA, USA, in 2017 and 2022, respectively.

He is currently a Research Assistant with the Department of Mechanical and Aerospace Engineering, University of California. His research interests include privacy, distributed optimization, distributed decision making, and machine learning.

Dr. Rezazadeh was a recipient of University of California Irvine Doctoral and Holmes fellowship for his graduate studies.

Solmaz S. Kia (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in aerospace engineering from the Sharif University of Technology, Tehran, Iran, in 2001 and 2004, respectively, and the Ph.D. degree in mechanical and aerospace engineering from the University of California, Irvine (UCI), Irvine, CA, USA, in 2009.

She was a Senior Research Engineer with SySense Inc., El Segundo, CA, from 2009 to 2010. She held postdoctoral positions with the Department of Mechanical and Aerospace Engineering, University of California, San Diego, CA, and UCI. She is currently an Associate Professor with the Department of Mechanical and Aerospace Engineering, UCI. Her main research interests include distributed optimization/coordination/estimation, nonlinear control theory, and probabilistic robotics.

Dr. Kia was a recipient of the University of California President's Postdoctoral Fellowship in 2012–2014, the National Science Foundation CAREER Award in 2017, and the Best Control System Magazine (CSM) Paper Award in 2021. She is an Associate Editor for *Automatica*, IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS, and IEEE OPEN JOURNAL OF CONTROL SYSTEMS.