

# Deterministic Privacy Preservation in Static Average Consensus Problem

Amir-Salar Esteki and Solmaz S. Kia<sup>✉</sup>, *Senior Member, IEEE*

**Abstract**—In this letter, we consider the problem of privacy preservation in the static average consensus problem. This problem normally is solved by proposing privacy preservation augmentations for the popular first order Laplacian-based algorithm. These mechanisms however come with computational overhead, may need coordination among the agents to choose their parameters and also alter the transient response of the algorithm. In this letter, we show that an alternative iterative algorithm that is proposed in the literature in the context of dynamic average consensus problem has intrinsic privacy preservation and can be used as a privacy preserving algorithm that yields the same performance behavior as the well-known Laplacian consensus algorithm but without the overheads that come with the existing privacy preservation methods.

**Index Terms**—Privacy preservation, static average consensus, networked systems.

## I. INTRODUCTION

THIS letter considers the problem of privacy preservation in the in-network average consensus (PrivCon) problem, where the objective is to enable a group of  $\mathcal{V} = \{1, \dots, N\}$  communicating agents interacting over a strongly connected and weight-balanced digraph  $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{A})$ ,<sup>1</sup> see Fig. 1, to calculate  $\mathbf{r}^{\text{avg}} = \frac{1}{N} \sum_{j=1}^N \mathbf{r}^j$  using local interactions but without disclosing their local *reference value*  $\mathbf{r}^i$ ,  $i \in \mathcal{V}$ , to each other. The privacy preservation for average consensus is normally formalized as concealing the reference value of each agent from a malicious agent that is defined as follows.

**Definition 1 (Malicious Agent):** A malicious agent in the network is an agent that wants to obtain the reference value of the other agents without perturbing/interrupting the execution of the average consensus algorithm. The knowledge set of this malicious agent consists of (a) the network topology  $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{A})$ , (b) its own local states and reference input, (c) the received signals from its out-neighbors, and (d) the agreement state  $x^i(k)$  of each agent  $i \in \mathcal{V}$  converges asymptotically to  $\mathbf{r}^{\text{avg}}$ .

Manuscript received September 14, 2020; revised November 21, 2020; accepted December 7, 2020. Date of publication December 22, 2020; date of current version March 26, 2021. This work was supported by NSF under Award ECCS-1653838. Recommended by Senior Editor C. Seatzu. (Corresponding author: Solmaz S. Kia.)

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

Digital Object Identifier 10.1109/LCSYS.2020.3046432

<sup>1</sup>For graph theoretic definitions used hereafter see Section II.

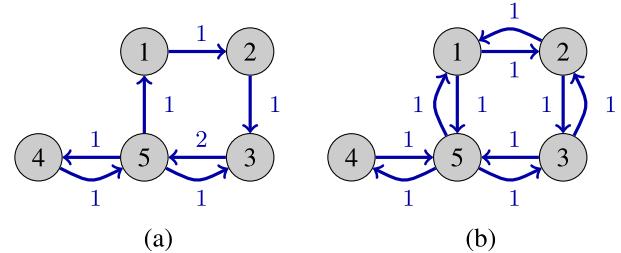


Fig. 1. The network in plot (a) is an example of a strongly connected and weight-balanced digraph, while the network in plot (b) is an example of a connected undirected graph. An edge  $i \xrightarrow{a_{ij}} j$  from an agent  $i$  to agent  $j$  means that agent  $i$  can obtain information from agent  $j$ ;  $a_{ij} > 0$  is the corresponding adjacency matrix element.

The solutions to the PrivCon problem in the literature mainly center around designing privacy preservation augmentation mechanisms for the popular average consensus algorithm

$$x^i(k+1) = x^i(k) - \Delta \sum_{j=1}^N \mathbf{a}_{ij}(x^j(k) - x^i(k)), \\ x^i(0) = \mathbf{r}^i, \quad i \in \mathcal{V}, \quad (1)$$

which with proper choice of stepsize  $\Delta$  guarantees  $x^i \rightarrow \mathbf{r}^{\text{avg}}$ ,  $i \in \mathcal{V}$ , as  $k \rightarrow \infty$  [1]. In a weighted digraph,  $\mathbf{a}_{ij} > 0$  if there is an edge from agent  $i$  to agent  $j$ , i.e., agent  $i$  can obtain information from agent  $j$ , otherwise  $\mathbf{a}_{ij} = 0$ . Thus, algorithm (1) requires each agent  $i \in \mathcal{V}$  to share its reference value  $\mathbf{r}^i$  with its in-neighbors (agents that receive messages from  $i$ ) in the first step of the algorithm, resulting in a trivial breach of privacy. Standard observability analysis [2] shows also that when the adjacency matrix  $\mathbf{A} = [\mathbf{a}_{ij}]$  of the network topology is known to an agent  $i \in \mathcal{V}$ , agent  $i$  may obtain the initial condition of algorithm (1), and consequently, the reference value of all or some of the other agents of the network; see [3] for details.

A popular approach explored to induce privacy preservation to algorithm (1) is the encryption where either a trusted third-party generates the public-key [4] or agents share their keys in which the network is restricted to a point-to-point or tagged undirected communication framework [5], [6]. Also, [5] requires extra communication channels for key generation purposes. Moreover, in [5], [6], the network should

be connected at the first step of communication and therefore, it does not have robustness to possible switching in the topology. Differential privacy [7]–[9] and additive obfuscation noise methods [10]–[12] are other popular techniques to conceal the *exact* reference value of the agents, but the malicious agent can obtain an estimate on the reference value using a stochastic estimator. In addition, final convergence is perturbed in [7]–[10] and convergence rate is altered in [11]. Moreover, these classes of privacy preserving measures cause noisy transient response for the algorithm, which results in excessive energy expenditure if agents use local controls to track the consensus trajectory as a reference input. Lastly, the guarantees established in all the methods discussed above are only for connected undirected graphs. Interested reader can also find privacy preservation methods for the continuous-time implementation of algorithm (1) in [13] and [14].

In this letter, we focus on a solution for PrivCon problem that does not perturb the final convergence value. We address the PrivCon problem by proposing to use an alternative algorithm (see algorithm (2)) that yields similar transient and convergence performance to that of algorithm (1). We analyze the privacy preservation of this algorithm carefully and show that this algorithm intrinsically yields exactly the same privacy preservation guarantees as in [10]–[12] in terms of which agents' privacy can be preserved. More precisely, we show that similar to the results in [10]–[12] the privacy of any agent that has at least one out-neighbor that is not the out-neighbor of the malicious agent is preserved. In comparison to the privacy preservation by use of the additive obfuscation noise methods [10]–[12], however, algorithm (2) offers a deterministic and stronger sense of privacy preservation, i.e., it meets the privacy preservation definition given below.

**Definition 2 (Privacy preservation):** Privacy of an agent is preserved from a malicious agent if the malicious agent cannot obtain any estimate on the reference value of the agent.

To show that the privacy of an agent  $i$  is preserved in the sense of Definition 2, we show that there exists other values of  $r^i$ , *arbitrarily* different from the true one, that give the same trajectory for the information the malicious agent receives from its out-neighbors. Therefore, the malicious agent cannot distinguish between the true reference value and the arbitrary alternatives. This, in essence, is a stronger notion of privacy preservation than that of the  $\epsilon$ -differential privacy [15] or that of [11] where even though the malicious agent cannot obtain the exact reference value of the agents but it obtains an estimate on the reference value.

In our demonstration study in Section IV, we compare in particular our suggested PrivCon problem solution to the additive obfuscation noise method of [11]. Reference [11] is a prominent solution in the class of additive obfuscation noise methods that guarantees exact convergence while not allowing a malicious agent to obtain the exact reference value of the agents under the connectivity condition stated earlier. However, in the framework of [11] there are disclosure of information about the reference value in two levels. First, as shown in [11], the malicious agent can employ a stochastic observer to obtain an estimate on the reference value of other agents. The normalized variance of this estimator can

be very small compared to the size of the reference value of the agents. Moreover, in [11] agents need to coordinate to choose a common parameter  $\phi$  in their noise generator and also use a common variance. The knowledge about the noise parameters also is another level of disclosure of an estimate on the reference value of the agents, as the transmit message of each agent at the initial step of the algorithm is the reference value plus a value generated by their local noise whose variance is the same for all the agents. Moreover, in choosing the variance value for the noise, agents need to consider the size of their local reference values to yield meaningful obfuscation. We note that using a noise with large variance results in a more perturbed transient response and a slower convergence. The advantage of our proposed solution is to meet the privacy preservation as given in Definition 1 while rendering a similar transient and convergence behavior to that of algorithm (1), having no need for coordination to choose the parameters of the algorithm, and no extra computations. Our notation is standard, though certain pieces of notation are defined as need arises.

## II. PROBLEM SETTING

Our graph theoretic definitions follow [16]. A weighted digraph is a triplet  $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{A})$  where  $\mathcal{V}$  is the node set,  $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$  is the edge set and  $\mathbf{A} = [\mathbf{a}_{ij}]$  is the weighted adjacency matrix defined such that  $a_{ij} > 0$  if  $(i, j) \in \mathcal{E}$ , otherwise  $a_{ij} = 0$ . Given an edge  $(i, j)$ ,  $i$  is called an in-neighbor of  $j$ , and  $j$  is called an out-neighbor of  $i$ . The (weighted) in-degree and out-degree of each node  $i \in \mathcal{V}$  are respectively are  $d_{\text{out}}^i = \sum_{j=1}^N \mathbf{a}_{ij}$  and  $d_{\text{in}}^i = \sum_{j=1}^N \mathbf{a}_{ji}$ . A weighted digraph is undirected if  $\mathbf{a}_{ij} = \mathbf{a}_{ji}$  for all  $i, j \in \mathcal{V}$ . A digraph is weight-balanced if at every node  $i \in \mathcal{V}$ ,  $d_{\text{out}}^i = d_{\text{in}}^i$ . A digraph is strongly connected if there is a directed path from every node to every other node. The Laplacian matrix of a digraph  $\mathcal{G}$  is  $\mathbf{L} = \text{Diag}(d_{\text{out}}^1, \dots, d_{\text{out}}^N) - \mathbf{A}$ .

Let the interaction topology of the agents be a strongly connected and weight-balanced digraph  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$ . We consider the average consensus algorithm

$$v^i(k+1) = v^i(k) + \Delta \sum_{j=1}^N \mathbf{a}_{ij}(x^j(k) - x^i(k)), \quad (2a)$$

$$x^i(k+1) = x^i(k) + \Delta \left( - (x^i(k) - r^i) - \sum_{j=1}^N \mathbf{a}_{ij}(x^j(k) - x^i(k)) - v^i(k) \right), \quad (2b)$$

$$x^i(0), v^i(0) \in \mathbb{R}, \quad i \in \mathcal{V}, \quad \text{s.t.} \quad \sum_{j=1}^N v^j(0) = 0, \quad (2c)$$

proposed in [17], as a solution for privacy preserving average consensus algorithm over  $\mathcal{G}$ . As discussed in [18], (2) yields a similar transient response to that of algorithm (1). In our analysis, the privacy preservation in the sense of Definition 2 is against a malicious agent defined as in Definition 1. Our study is different than the privacy preservation study in [17] where the continuous-time representation of algorithm (2) is considered and the reference values are assumed

to be dynamic time-varying signals. The guarantees provided crucially depend on the time varyingness of the reference values.

Let  $\mathbf{x} = (x^1, \dots, x^N)$ ,  $\mathbf{v} = (v^1, \dots, v^N)$ ,  $\mathbf{\Pi} = \mathbf{I}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\top$ , and  $\bar{\mathbf{r}} = \mathbf{r}^{\text{avg}} \mathbf{1}_N$ . Moreover, let  $\mathbf{r} = \frac{1}{\sqrt{N}} \mathbf{1}_N$ , and  $\mathbf{R}$  such that  $[\mathbf{r} \ \mathbf{R}] \begin{bmatrix} \mathbf{r}^\top \\ \mathbf{R}^\top \end{bmatrix} = \begin{bmatrix} \mathbf{r}^\top \\ \mathbf{R}^\top \end{bmatrix} [\mathbf{r} \ \mathbf{R}] = \mathbf{I}_N$ . Then, using the

change of variable  $\begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_{2:N} \\ \mathbf{p}_1 \\ \mathbf{p}_{2:N} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \mathbf{r}^\top \\ \mathbf{R}^\top \end{bmatrix} & \begin{bmatrix} \mathbf{0} \\ \mathbf{R}^\top \end{bmatrix} \\ \mathbf{0} & \begin{bmatrix} \mathbf{r}^\top \\ \mathbf{R}^\top \end{bmatrix} \end{bmatrix} \begin{bmatrix} \mathbf{v} - \mathbf{\Pi} \mathbf{r} \\ \mathbf{x} - \bar{\mathbf{r}} \end{bmatrix}$ , algorithm (1) reads as  $\begin{bmatrix} \mathbf{p}_1(k+1) \\ \mathbf{p}_{2:N}(k+1) \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} \\ 0 & \mathbf{I} - \Delta \mathbf{L}^+ \end{bmatrix} \begin{bmatrix} \mathbf{p}_1(k) \\ \mathbf{p}_{2:N}(k) \end{bmatrix}$  and algorithm (2) as

$$\begin{bmatrix} \mathbf{q}_1(k+1) \\ \mathbf{q}_{2:N}(k+1) \\ \mathbf{p}_1(k+1) \\ \mathbf{p}_{2:N}(k+1) \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} & 0 & \mathbf{0} \\ 0 & (1-\Delta)\mathbf{I} & 0 & \mathbf{0} \\ -\Delta & \mathbf{0} & (1-\Delta) & \mathbf{0} \\ 0 & -\Delta \mathbf{I} & 0 & \mathbf{I} - \Delta \mathbf{L}^+ \end{bmatrix} \begin{bmatrix} \mathbf{q}_1(k) \\ \mathbf{q}_{2:N}(k) \\ \mathbf{p}_1(k) \\ \mathbf{p}_{2:N}(k) \end{bmatrix},$$

where  $\mathbf{L}^+ = \mathbf{R}^\top \mathbf{L} \mathbf{R}$ . These equivalent representations show the connection between the internal dynamics of algorithms (1) and (2). For a strongly connected and weight-balanced digraph,  $\mathbf{L}$  has a simple zero eigenvalue and the rest of the eigenvalues  $\{\lambda_i\}_{i=2}^N$  have non-zero positive real parts. Then, the eigenvalues of  $\mathbf{L}^+$  are  $\{\lambda_i\}_{i=2}^N$ . Using [18, Lemma S1], the exponential convergence of (1) and (2) is guaranteed, respectively, for any  $\Delta \in (0, \bar{\Delta})$  and  $\Delta \in (0, \min\{2, \bar{\Delta}\})$ , where  $\bar{\Delta} = \min\{2 \frac{\text{Re}(\lambda_i)}{|\lambda_i|^2}\}_{i=2}^N$ . Given a similar performance, an immediate appeal of algorithm (2) over algorithm (1) is that the reference value of agents is not trivially transmitted to the in-neighbors because the transmitted state  $x^i$  is initialized arbitrarily, see (2c). The question we address in the subsequent section is whether the malicious agent can use its knowledge set to compute the reference values of other agents when agents implement algorithm (2).

### III. PRIVACY PRESERVATION ANALYSIS

In this section we show that algorithm (2) intrinsically yields the same privacy preservation guarantees as in [10]–[12] in terms of which agents' privacy can be preserved. However, the guarantees of algorithm (2) are valid also for strongly connected and weight-balanced digraphs. Moreover, unlike the privacy preservation of [10]–[12] which is obtained by using additive noises and comes with disclosure of a stochastic estimate on the private value of the agents, algorithm (2) offers a deterministic and stronger sense of privacy preservation, i.e., it meets the privacy preservation objective defined in Definition 2. In what follows without loss of generality we assume that agent 1 is the malicious agent. The initialization condition  $\sum_{i=1}^N v^i(0) = 0$  is trivially satisfied when every agent  $i \in \mathcal{V}$  uses  $v^i(0) = 0$ . Other choices need coordination among agents with no strong guarantee that the choices are private. Thus, we assume that  $v^i(0) = 0$ ,  $i \in \mathcal{V}$  and is known to the malicious agent. Moreover,  $\mathcal{N}_{\text{in}}^i$  and  $\mathcal{N}_{\text{out}}^i$  are, respectively, the set of in-neighbors and out-neighbors of agent  $i$ . We also define  $\mathcal{N}_{\text{in}}^{i+} = \mathcal{N}_{\text{in}}^i \cup \{i\}$ .

*Lemma 1 (A sufficient condition for privacy preservation when algorithm (2) is implemented):* Let the interaction topology  $\mathcal{G}$  of the agents implementing algorithm (2) with  $\Delta \in (0, \min\{2, \bar{\Delta}\})$ , initialized at  $x^i(0) \in \mathbb{R}$  and  $v^i(0) = 0$ ,  $i \in \mathcal{V}$ , be strongly connected and weight-balanced digraph. Let the

knowledge set of malicious agent 1 be as in Definition 1. The privacy of any agent  $i \in \mathcal{V} \setminus \{1\}$  is preserved from agent 1 if agent  $i$  has at least one out-neighbor that is not an out-neighbor of agent 1.

*Proof:* To prove the statement, we consider the worst case where agent  $i$  whose privacy is being evaluated is an out-neighbor of agent 1, i.e., agent 1 receives the transmitted message of agent  $i$ . Without loss of generality let this agent be labeled agent 2. Let agent 3 be the out-neighbor of agent 2 that is not an out-neighbor of agent 1, i.e.,  $2 \in \mathcal{N}_{\text{in}}^3$  but  $1 \notin \mathcal{N}_{\text{in}}^3$ . To show privacy preservation for agent 2, we show that there are infinitely many executions of algorithm (2) with different values of  $r^2 \in \mathbb{R}$  and  $r^3 \in \mathbb{R}$  that yield exactly the same received signal by agent 1. To this end, consider two executions of algorithm (2). Let  $e_x^i = x_1^i - x_2^i$  and  $e_v^i = v_1^i - v_2^i$ ,  $i \in \mathcal{V}$ . Moreover, let the reference value of the agents in the first execution be  $\{r_1^i\}_{i=1}^N$  and in the second execution be  $\{r_2^i\}_{i=1}^N$  such that  $e_r^i = r_1^i - r_2^i \neq 0$  if  $i \in \mathcal{N}_{\text{in}}^{3+}$ , otherwise  $e_r^i = 0$ . Moreover,  $\sum_{i=1}^N e_r^i = 0$ , i.e.,  $r_1^{\text{avg}} = r_2^{\text{avg}}$ . Lastly,  $e_x^i(0) = 0$ , for  $i \in \mathcal{V} \setminus \{3\}$ . Next, we show that for any  $e_x^3(0) \in \mathbb{R}$  there exists  $e_r^i \neq 0$ ,  $i \in \mathcal{N}_{\text{in}}^{3+}$ , such that  $e_x^i(k) \equiv 0$ ,  $j \in \mathcal{V} \setminus \{3\}$  for  $k \in \mathbb{Z}_{\geq 0}$ . Therefore, the signals received by agent 1 for these two distinct executions is exactly the same and agent 1 cannot distinguish between them.

The error dynamics at each agent  $i \in \mathcal{V}$  reads as

$$e_v^i(k+1) = e_v^i(k) + \Delta \sum_{j=1}^N \mathbf{a}_{ij} (e_x^j(k) - e_x^i(k)), \quad (3a)$$

$$e_x^i(k+1) = e_x^i(k) + \Delta \left( -e_x^i(k) - e_r^i \right. \\ \left. - \sum_{j=1}^N \mathbf{a}_{ij} (e_x^j(k) - e_x^i(k)) - e_v^i(k) \right), \quad (3b)$$

$$e_v^i(0) = 0, \quad \begin{cases} e_x^i(0) = 0, & \text{if } i \neq 3, \\ e_x^i(0) \neq 0 & \text{if } i = 3 \end{cases} \quad (3c)$$

If  $e_x^i(k) \equiv 0$  for any  $i \in \mathcal{V} \setminus \{3\}$ , then  $e_x^i(k) \equiv 0$ , and  $e_v^i(k) \equiv 0$  for  $k \in \mathbb{Z}_{\geq 0}$  satisfy the error dynamics (3) for any  $i \in \mathcal{V} \setminus \mathcal{N}_{\text{in}}^3$ . On the other hand, for  $i \in \mathcal{N}_{\text{in}}^3$

$$e_v^i(k+1) = e_v^i(k) - \Delta \mathbf{a}_{i3} e_x^3(k), \quad (4a)$$

$$0 = e_r^i + \mathbf{a}_{i3} e_x^3(k) - e_v^i(k), \quad (4b)$$

and since  $\sum_{i=1}^N \mathbf{a}_{i3} = \mathbf{d}_{\text{out}}^3$ , (3) for agent 3 reads as

$$e_v^3(k+1) = e_v^3(k) + \Delta \mathbf{d}_{\text{out}}^3 e_x^3(k), \quad (5a)$$

$$e_x^3(k+1) = e_x^3(k) + \Delta (e_r^3 - (1 + \mathbf{d}_{\text{out}}^3) e_x^3(k) - e_v^3(k)). \quad (5b)$$

Note that with adding (4a) and (5a) we have  $\sum_{i \in \mathcal{N}_{\text{in}}^3} e_v^i(k) = 0$ . Considering  $\sum_{i=1}^N \mathbf{a}_{i3} = \mathbf{d}_{\text{in}}^3 = \mathbf{d}_{\text{out}}^3$  and  $\sum_{i \in \mathcal{N}_{\text{in}}^{3+}} e_r^i = 0$  also, it follows from adding up (5b) and (4b) for  $i \in \mathcal{N}_{\text{in}}^3$  that

$$e_x^3(k+1) = (1 - \Delta) e_x^3(k). \quad (6)$$

Then, since  $e_v^3(0) = 0$ , it follows from (5b), (6), and (4b) that

$$e_r^3 = \mathbf{d}_{\text{out}}^3 e_x^3(0), \quad e_r^i = -\mathbf{a}_{i3} e_x^3(0), \quad i \in \mathcal{N}_{\text{in}}^3. \quad (7)$$

Since  $\sum_{i=1}^N \mathbf{a}_{i3} = \mathbf{d}_{\text{in}}^3 = \mathbf{d}_{\text{out}}^3$ , it follows from (7) that indeed  $\sum_{i \in \mathcal{N}_{\text{in}}^{3+}} e_r^i = 0$ . Therefore, we can conclude that  $r_1^{\text{avg}} = r_2^{\text{avg}}$ . It is then proven that with an unbounded error in the initial state of  $e_x^3$ ,  $e_r^2$  varies unboundedly as well yet the malicious agent receives the same signals from its out-neighbors, i.e.,

$e_x^i(k) \equiv 0$ ,  $k \in \mathbb{Z}_{\geq 0}$ , for  $i \in \mathcal{N}_{\text{out}}^1$ . Note here that since  $|1 - \Delta| < 1$ , it follows from (6) that  $e_x^3(k)$  also converges to zero as  $k \rightarrow \infty$  showing that all the agents converge to the same final average point in both execution 1 and 2. ■

Our next result shows that when the malicious agent has access to the messages transmitted to and from an agent  $i$ , similar to the methods proposed in [10]–[12], the privacy of agent  $i$  is not preserved. To establish this result we show that the malicious agent can implement an observer to obtain  $r^i$ .

*Lemma 2 (A Sufficient Condition for Loss of Privacy Preservation When Algorithm (2) is Implemented):* Let the interaction topology  $\mathcal{G}$  of the agents implementing algorithm (2), initialized at  $x^i(0) \in \mathbb{R}$  and  $v^i(0) = 0$ ,  $i \in \mathcal{V}$ , be a strongly connected and weight-balanced digraph. Let the knowledge set of malicious agent 1 be as in Definition 1. If agent 1 is the in-neighbor of agent  $o \in \mathcal{V} \setminus \{1\}$  and all its out-neighbors, then at any  $k \in \mathbb{Z}_{\geq 1}$  agent 1 can obtain  $r^o$  using  $x^o(k)$ ,  $x^o(k-1)$  and  $\{x^j(k-1)\}_{j \in \mathcal{N}_{\text{out}}^o}$ .

*Proof:* Proof of Lemma 2 follows from a simple algebraic manipulation over (2) to obtain  $r^o$  in terms of  $x^o(k)$ ,  $x^o(k-1)$  and  $\{x^j(k-1)\}_{j \in \mathcal{N}_{\text{out}}^o}$ . ■

Lemma 1 and 2 gives us the following main statement on the privacy preservation guarantees of Algorithm (2).

*Theorem 1 (Privacy Preservation Guarantee of Algorithm (2)):* Let the interaction topology  $\mathcal{G}$  of the agents implementing algorithm (2), initialized at  $x^i(0) \in \mathbb{R}$  and  $v^i(0) = 0$ ,  $i \in \mathcal{V}$ , be strongly connected and weight-balanced digraph. Let the knowledge set of malicious agent 1 be as in Definition 1. Privacy of agent  $i \in \mathcal{V} \setminus \{1\}$  is preserved from agent 1 if and only if agent  $i$  has an out-neighbor that is not an out-neighbor of agent 1.

So far, we have shown the intrinsic privacy property of Algorithm (2). Now the natural question is whether we can loosen the topology restriction of Theorem 1 using additive perturbation signals and preserve privacy for agents whose all communication signals (in-coming and out-going) are available to the malicious agent. Herein, we show that this mechanism has no contribution and the malicious agent can still derive local information asymptotically. A comprehensive protection via perturbation signals suggests adding a signal  $g^i(k) : \mathbb{Z}_{\geq 0} \rightarrow \mathbb{R}$  to the transmitted message of every agent  $i \in \mathcal{V}$ , and additive signal  $f_1^i(k) : \mathbb{Z}_{\geq 0} \rightarrow \mathbb{R}$  and  $f_2^i(k) : \mathbb{Z}_{\geq 0} \rightarrow \mathbb{R}$  to the right hand side of (2a) and (2b), respectively. However, without loss of generality, the effect of all these signals can be captured via adding only a dynamic perturbation signal  $f^i(k) : \mathbb{Z}_{\geq 0} \rightarrow \mathbb{R}$ ,  $i \in \mathcal{V}$  to the right hand side of (2b) as follows,

$$v^i(k+1) = v^i(k) + \Delta \sum_{j=1}^N \mathbf{a}_{ij}(x^j(k) - x^i(k)), \quad (8a)$$

$$x^i(k+1) = x^i(k) + \Delta(-(x^i(k) - r^i) - \sum_{j=1}^N \mathbf{a}_{ij}(x^j(k) - x^i(k)) - v^i(k) + f^i(k)). \quad (8b)$$

Instead of using a particular perturbation signal as in [10]–[12], we follow the approach in [19] and first investigate for what set of perturbation signals, which we call

*admissible perturbation signals*, the final convergence point of the algorithm is not perturbed. It is natural that any necessary condition defining the admissible perturbation signal is known to all the agents. Then, we show that by knowing a necessary condition on the perturbation, the malicious agent can employ an observer to obtain the reference input regardless of what the exact additive admissible perturbation signal the agent uses.

*Theorem 2 (A Necessary Condition on Admissible Perturbation Signals):* Let the interaction topology  $\mathcal{G}$  of the agents implementing algorithm (8) with  $\Delta \in (0, \min\{2, \bar{\Delta}\})$ , initialized at  $x^i(0) \in \mathbb{R}$  and  $v^i(0) = 0$ ,  $i \in \mathcal{V}$ , be strongly connected and weight-balanced digraph. Let the knowledge set of malicious agent 1 be as in Definition 1. A necessary condition for preserving the average consensus, i.e.,  $\lim_{k \rightarrow \infty} x^i(k) = r^{\text{avg}}$  for  $i \in \mathcal{V}$  is

$$\lim_{k \rightarrow \infty} \sum_{i=1}^N \sum_{m=0}^k (1 - \Delta)^{k-m} f^i(m) = 0, \quad i \in \mathcal{V}. \quad (9)$$

*Proof:* Using the change of variable introduced in Section II, (8) can be written as

$$\begin{aligned} \mathbf{q}_1(k+1) &= \mathbf{q}_1(k) = 0, \\ \mathbf{q}_{2:N}(k+1) &= (1 - \Delta)\mathbf{q}_{2:N}(k) + \Delta \mathbf{R}^T \mathbf{f}(k), \\ \mathbf{p}_1(k+1) &= (1 - \Delta)\mathbf{p}_1(k) + \Delta \mathbf{r}^T \mathbf{f}(k), \\ \mathbf{p}_{2:N}(k+1) &= (\mathbf{I} - \Delta \mathbf{L}^+) \mathbf{p}_{2:N}(k) - \Delta \mathbf{q}_{2:N}(k) + \Delta \mathbf{R}^T \mathbf{f}(k), \end{aligned}$$

where we used  $\mathbf{q}_1(0) = \sum_{i=0}^N v^i(0) = 0$ . Then,

$$\mathbf{p}_1(k) = (1 - \Delta)^k \mathbf{p}_1(0) - \sum_{m=0}^k (1 - \Delta)^{k-m} \Delta (-\mathbf{r}^T \mathbf{f}(m)). \quad (10)$$

Since  $[\begin{smallmatrix} \mathbf{p}_1 \\ \mathbf{p}_{2:N} \end{smallmatrix}] = [\begin{smallmatrix} \mathbf{r}^T \\ \mathbf{R}^T \end{smallmatrix}] (\mathbf{x} - \mathbf{r}^{\text{avg}} \mathbf{1}_N)$ , the necessary condition for reaching average consensus then is  $\lim_{k \rightarrow \infty} \mathbf{p}(k) = \mathbf{0}$ . Given that  $\lim_{k \rightarrow \infty} (1 - \Delta)^k = 0$ , then it follows from (10) that (9) is a necessary condition for  $\lim_{k \rightarrow \infty} \mathbf{p}_1(k) = \mathbf{0}$ . ■

*Remark 1:* Condition (9) that defines the admissible perturbation signals couples the choices of all the agents. In a decentralized setting, since there is no third-party to assign local perturbation signals  $f^i$ , for agents to satisfy (9) without any coordination among themselves, agents choose their admissible perturbations according to

$$\lim_{k \rightarrow \infty} \sum_{m=0}^k (1 - \Delta)^{k-m} f^i(m) = 0, \quad i \in \mathcal{V}. \quad (11)$$

In our next statement, we investigate whether a malicious agent can derive the local reference value of an agent if it knows condition (11) and all the transmitted messages to and from the agent.

*Theorem 3 (Use of locally chosen perturbation signals in (8) does not increase privacy protection level):* Let the interaction topology  $\mathcal{G}$  of the agents implementing algorithm (8) with  $\Delta \in (0, \min\{2, \bar{\Delta}\})$ , initialized at  $x^i(0) \in \mathbb{R}$  and  $v^i(0) = 0$ ,  $i \in \mathcal{V}$ , be strongly connected and weight-balanced digraph. Suppose agents are implementing locally chosen admissible perturbation signals that satisfy (11), and let the knowledge set of malicious agent 1 be as in Definition 1.

If agent 1 is the in-neighbor of agent  $i \in \mathcal{V} \setminus \{1\}$  and all its out-neighbors, then agent 1 can employ the observer

$$\hat{v}^i(k+1) = \hat{v}^i(k) - \Delta \sum_{j=1}^N \mathbf{a}_{ij}(x^i(k) - x^j(k)), \quad (12a)$$

$$\begin{aligned} \hat{x}^i(k+1) &= \hat{x}^i(k) + \Delta \left( \sum_{j=1}^N \mathbf{a}_{ij}(x^i(k) - x^j(k)) \right. \\ &\quad \left. + \hat{v}^i(k) - \hat{x}^i(k) \right), \end{aligned} \quad (12b)$$

$$\hat{z}^i(k) = \hat{x}^i(k) + x^i(k), \quad (12c)$$

initialized at  $\hat{v}^i(0) = \hat{x}^i(0) = 0$  to have  $\hat{z}^i(k) \rightarrow r^i$  as  $k \rightarrow \infty$ .

*Proof:* By substitution, from (12c) we obtain  $\hat{z}^i(k+1) = \hat{z}^i(k) + \Delta(-\hat{z}^i(k) - (v^i(k) - \hat{v}^i(k)) + r^i + f^i(k))$ , which gives us  $\hat{z}^i(k) = (1 - \Delta)^k \hat{z}^i(0) + \Delta \sum_{m=0}^k (1 - \Delta)^{k-m} (r^i + f^i(m))$ . We note that using the sum of geometric series we can write  $\sum_{m=0}^k (1 - \Delta)^{k-m} = \frac{1 - (1 - \Delta)^{k+1}}{\Delta}$ . Then, since  $|1 - \Delta| < 1$  and given the necessary condition (11) the observer converges to the local reference value of agent  $i$ , i.e.,  $\hat{z}^i(k) \rightarrow r^i$  as  $k \rightarrow \infty$ . ■

We proved that an additive perturbation signal has no contribution in the level of privacy provided by algorithm (2).

#### IV. DEMONSTRATION EXAMPLES

To provide a context to appreciate the implications of our privacy preserving solution for average consensus problem versus the method of [11], we consider an optimal power dispatch (OPD) problem over the undirected connected graph in Fig. 1(b). Reference [11] is a representative of the common method of privacy preservation via additive noises for algorithm (1). We conduct our study over an undirected connected graph since the results in [11] are established only for such graphs. In our OPD problem a group of generators interacting over the graph of Fig. 1(b) are expected to meet the demand of  $\mathcal{P}_D = 1500$  MW ( $p^1 + \dots + p^5 = 1500$ ) while minimizing their collective cost  $\sum_{i=1}^5 f^i(p^i)$ . The parameters of the local cost  $f^i(p^i) = \frac{1}{2\beta^i} (p^i + \alpha^i)^2 + \gamma^i$ ,  $i \in \{1, \dots, 5\}$ , are chosen from the IEEE bus 118 generator list according to corresponding components of  $\{\alpha^i\}_{i=1}^4 = \{188.3, 592.5, 2567.2, 1793.3, 2567.2\}$ ,  $\{\beta^i\}_{i=1}^5 = \{7.17, 45.9, 208.2, 166.6, 208.2\}$ . Here  $\alpha^i$  and  $\beta^i$  are supposed to be the private value of each agent  $i \in \mathcal{V}$ . The optimal solution of this problem for each agent  $i \in \{1, \dots, 5\}$  is given by  $p^{i*} = \beta^i \frac{\mathcal{P}_D + \sum_{j=1}^N \alpha^j}{\sum_j^N \beta^j} - \alpha^i$  [20]. To generate this solution in a distributed manner, let us assume that these  $N = 5$  agents employ two static average consensus algorithms to obtain  $\bar{\alpha} = \frac{1}{N} \sum_{i=1}^N \alpha^i$  and  $\bar{\beta} = \frac{1}{N} \sum_i^N \beta^i$ .

Then, by knowing  $N$  and  $\mathcal{P}_D$ , agents have all the information to obtain  $p^{i*}$ . Fig. 1(b) is used in the numerical example of [11] where agent 5 is the malicious agent, so as we assume here. The privacy of agents  $\{1, 2, 3\}$  is preserved if agents use algorithm (1) augmented by the additive noise method of [11, Corollary 1] (hereafter referred to as method M1) or algorithm (2) (see Theorem 1). Let us assume that when

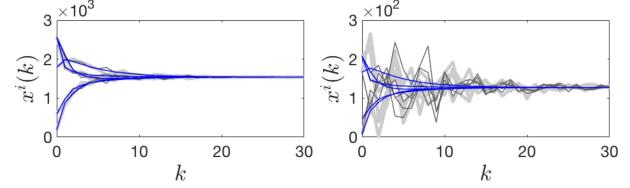


Fig. 2. Trajectories of the average consensus algorithms to obtain  $\bar{\alpha}$  (left plot) and  $\bar{\beta}$  (right plot). The thick gray lines and the thin gray lines show trajectories of two executions of method M1 [11, Corollary 1], each employing a different noise realization. The blue lines show the trajectories of the agreement state of algorithm (2).

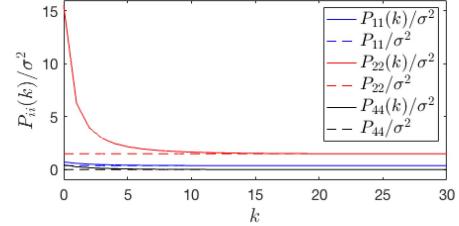


Fig. 3. Normalized covariances of the maximum likelihood estimator that the malicious agent uses to obtain an estimate of reference input of the agents implementing method M1. The dashed line is the steady-state value and the solid lines show the time history. This plot is the same as [11, Fig. 4] where  $\sigma = 1$  is used.

using the method M1, the agents use the same  $\phi = 0.9$  as [11, Sec. 6] and given the value of their parameters, they agree to use an additive Gaussian noise with mean 0 and standard deviation  $\sigma = 100$ . Note that these values should be common for all agents, and thus agents need to coordinate to choose them. Fig. 2 shows the trajectories generated by method M1 for two different executions (each uses a different noise realization) overlaid over the trajectories generated by Algorithm (2). As seen, the trajectories of method M1 are quite noisy and convergence is slower. Using method M1, malicious agent 5 cannot obtain the exact value of  $\{\alpha^i\}_{i=1}^3$  and  $\{\beta^i\}_{i=1}^3$  because as shown in [11, Fig. 4] the covariance of a maximum likelihood estimator that agent 5 uses to obtain the private reference value of  $\{1, 2, 3\}$  has a steady state non-vanishing value, see Fig. 3. However, agent 5 knows that in 99.7% of the times ( $3\sigma$  rule) the error rate to obtain each  $\{\alpha^i\}_{i=1}^3$  and  $\{\beta^i\}_{i=1}^3$  is respectively (0.972%, 0.618%, 0.071%) and (25.512%, 7.972%, 0.879%) according to the computed normalized covariances  $P_{ii}$  as in Fig. 3. Given the numerical values of these parameters, this level of protection gives a good approximate value of the optimal operating point of the supposedly private agents  $\{1, 2, 3\}$  to the malicious agent. On the other hand, the privacy preservation guarantees that algorithm (2) provides for agents  $\{1, 2, 3\}$  is stronger, as the malicious agent cannot obtain any estimate on the private values of the agent. See Fig. 4 for an example scenario, where two alternative cases of reference input signals for  $\alpha^i$  denoted by  $\{\alpha_1^i\}_{i=1}^5 = \{-1311.6, 3592.5, 1067.2, 1793.3, 2567.2\}$ , and  $\{\alpha_3^i\}_{i=1}^5 = \{1688.3, -2407.4, 4067.2, 1793.3, 2567.2\}$ , where  $\frac{1}{N} \sum_{i=1}^5 \alpha_j^i = \frac{1}{N} \sum_{i=1}^5 \alpha_1^i$ ,  $j \in \{2, 3\}$  result in exactly the same transmit message to the malicious agent 5. Therefore, agent 5 cannot distinguish between these different references for agents  $\{1, 2, 3\}$ . Here,  $\{\alpha_1^i\}_{i=1}^5$  are

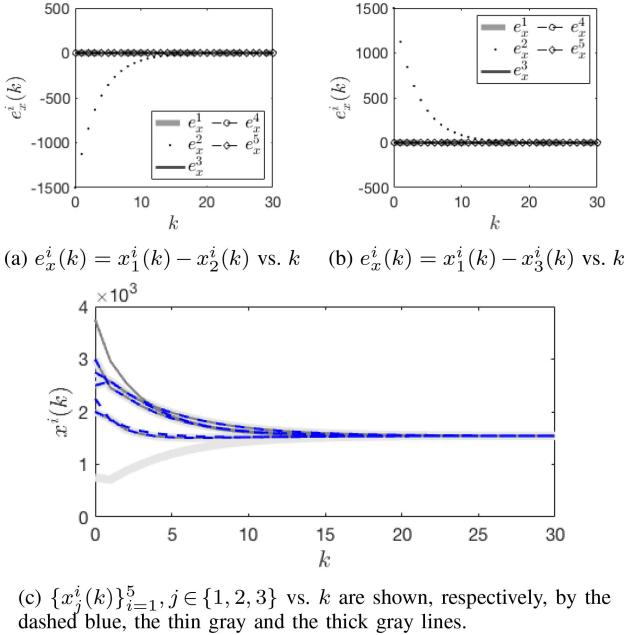


Fig. 4. Three executions of (2) yield exactly the same received signals by the malicious agent 5, while converging to the same average value of  $\frac{1}{N} \sum_{i=1}^5 \alpha^i$ .

the actual inputs given in the OPD problem definition earlier.

## V. CONCLUSION

We considered the problem of privacy preservation in the static average consensus problem. This problem normally is solved by proposing privacy preservation mechanism that are added to the popular first order Laplacian-based algorithm. These mechanisms come with computational overheads or pre-coordinating among the agents to choose the parameters of the algorithm. They also alter the transient response of the algorithm. In this letter we showed that an alternative algorithm that is proposed in the literature in the context of dynamic average consensus problem can be a simple solution for privacy preservation for average consensus problem. The advantage of our proposed solution over existing results in the literature was to provide a stronger notion of privacy while rendering a similar transient and convergence behavior to that of the well-known Laplacian average consensus algorithm, having no need for coordination to choose the parameters of the algorithm, and no extra computations.

## REFERENCES

- [1] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520–1533, Sep. 2004.
- [2] R. E. Kalman, "On the general theory of control systems," in *Proc. 1st Int. Conf. Autom. Control*, pp. 491–502, 1960.
- [3] S. Pequito, S. Kar, S. Sundaram, and A. P. Aguiar, "Design of communication networks for distributed computation with privacy guarantees," in *Proc. 53rd IEEE Conf. Decis. Control*, 2014, pp. 1370–1376.
- [4] C. N. Hadjicostis, "Privacy preserving distributed average consensus via homomorphic encryption," in *Proc. IEEE Conf. Decis. Control (CDC)*, 2018, pp. 1258–1263.
- [5] 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.
- [6] T. Yin, Y. Lv, and W. Yu, "Accurate privacy preserving average consensus," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 4, pp. 690–694, Apr. 2020.
- [7] 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.
- [8] Z. Huang, S. Mitra, and G. Dullerud, "Differentially private iterative synchronous consensus," in *Proc. ACM Workshop Privacy Electron. Soc.*, 2012, pp. 81–90.
- [9] E. Nozari, P. Tallapragada, and J. Cortés, "Differentially private average consensus with optimal noise selection," *IFAC PapersOnLine*, vol. 48, no. 22, pp. 203–208, 2015.
- [10] N. E. Manitara and C. N. Hadjicostis, "Privacy-preserving asymptotic average consensus," in *Proc. IEEE Eur. Control Conf. (ECC)*, 2013, pp. 760–765.
- [11] Y. Mo and R. M. Murray, "Privacy preserving average consensus," *IEEE Trans. Autom. Control*, vol. 62, no. 2, pp. 753–765, 2016.
- [12] J. He, L. Cai, C. Zhao, P. Cheng, and X. Guan, "Privacy-preserving average consensus: Privacy analysis and algorithm design," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 5, no. 1, pp. 127–138, Mar. 2019.
- [13] C. Altafini, "A system-theoretic framework for privacy preservation in continuous-time multiagent dynamics," *Automatica*, vol. 122, Dec. 2020, Art. no. 109253.
- [14] 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.
- [15] 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.
- [16] F. Bullo, J. Cortes, and S. Martinez, *Distributed Control of Robotic Networks: A Mathematical Approach to Motion Coordination Algorithms*, vol. 27. Princeton, NJ, USA: Princeton Univ. Press, 2009.
- [17] S. S. Kia, J. Cortés, and S. Martínez, "Dynamic average consensus under limited control authority and privacy requirements," *Int. J. Robust Nonlinear Control*, vol. 25, no. 13, pp. 1941–1966, 2015.
- [18] S. S. Kia, B. Van Scy, J. Cortes, R. A. Freeman, K. M. Lynch, and S. Martinez, "Tutorial on dynamic average consensus: The problem, its applications, and the algorithms," *IEEE Control Syst. Mag.*, vol. 39, no. 3, pp. 40–72, 2019.
- [19] N. Rezazadeh and S. S. Kia. (2020). *Privacy Preservation in Continuous-Time Average Consensus Algorithm via Deterministic Additive Perturbation Signals*. [Online]. Available: <https://arxiv.org/pdf/1904.05286.pdf>
- [20] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, *Power Generation, Operation, and Control*. Hoboken, NJ, USA: Wiley, 2013.