Distributed Topology-preserving Collaboration Algorithm against Inference Attack

Zitong Wang, Yushan Li, Chongrong Fang and Jianping He

***Abstract*—Interaction topology through which agents achieve collaboration in multi-agent systems is of fundamental importance. Recently, many efforts have been devoted to the problem of topology inference, e.g., the trajectory information of mobile agents is utilized to regress the topology. In this paper, we develop a distributed topology-preserving collaboration algorithm for multi-agent systems against topology inference attacks. The novelties lie in that: i) By adding well-designed noises to the system states, the irregularity of the state evolution is largely enhanced, making the underlying topology hard to be inferred accurately from the observations over the system; ii) By dividing the added noises into the random and the disturbing terms with mutual compensation properties, the proposed algorithm guarantees the convergence of the system state, which applies to both undirected and directed topology structures. Specifically, the mean square convergence rate and the non-asymptotic error bound are derived. Extensive simulations are conducted to illustrate the effectiveness of our algorithm.**

I. Introduction

Multi-agent systems (MASs) have been widely used in applications such as distributed computing [1], sensor networks [2], multi-robot systems [3], and data aggregation [4]. The interaction topology of MASs, which characterizes the ability of agents to interact with others, is essential for agents to achieve efficient consensus-based collaboration. Besides, the topology will affect the autonomy, adaptation, scalability, and efficiency of the MASs [5]. Due to its significance, the research on the topology of MASs has received great attention from researchers in various areas, including computer science, communication, and control theory [6].

Currently, there are fruitful research results on topology inference. For instance, [7] computes the states of the agents by an iterative root-searching method driven by a maximum likelihood function, and [8] focuses on inferring the directed network topology under unmeasurable latent inputs. In addition, optimization methods are also widely used in topology inference, aiming to infer the topology quickly and accurately. In this context, the outside observers can obtain the topology by collecting a set of observation data and then solving a well-formulated regression problem. Unfortunately, all these methods can also be employed by malicious adversaries to infer the topology within the system (we call it a topology inference attack). Once the topology is accurately inferred, the adversary can launch further premeditated attacks on a certain critical agent in the MASs to drive the system into a state of paralysis. Taking multiple mobile agents as an example [9], an outside attacker can observe the moving trajectory of the agents and infer their internal topology structure by the aforementioned methods. Then, the attacker obtains critical guidance to incapacitate the target agent and largely deteriorate the collaboration performance of the MASs. Therefore, it is necessary for MASs to design the topology-preserving collaboration algorithms against topology inference attacks.

To counter such attacks, researchers have dug deeply and proposed a series of defense mechanisms, where the consensus algorithm is a fundamental tool to ensure MASs' collaboration performance. These methods can be divided into two main categories: dynamic topologies and noise-adding algorithms. In the former kind of studies, the interaction topology of the agents changes from time to time, which notably increases the difficulty for the external attackers to infer an accurate and stable topology. In the pursuit of the consensus in MASs under these topologies, the key is to design appropriate protocols that can guarantee the group of agents reach the consensus point on the shared information with limited and unreliable information exchange and switching topologies [10]–[12]. The drawbacks of these methods origin from their potentially high resource consumption due to the frequent changes of topologies. Additionally, their strong dependence on the systems imposes critical restrictions on the connectivity of the interaction topology. The noise-adding based methods are also commonly used to secure the internal information of MASs [13]–[16]. The main idea is to impose additional noisy signals on agents' states during the collaboration period, thereby hiding the true information from the attackers. Nevertheless, most noise-adding algorithms deal with data privacy that pays more attention to the privacy of every single node rather than the preservation of the system's topology that focuses on hiding the overall network structure from external observers.

Motivated by the above observations, in this paper, we propose a hybrid distributed topology-preserving collaboration algorithm (hybrid DTCA). It is effective in concealing the actual topology structure from the outside inference attack while guaranteeing collaboration convergence. The challenges lie in how to design the extra noises in a distributed way while causing the maximum inference attack degradation. The main contributions of our work are listed as follows:

- To protect the system topology from being accurately

The authors are with the Department of Automation, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai, China. E-mail address: {wangzitong, yushan_li, crfang and jphe}@sjtu.edu.cn. This work was supported by the NSF of China under Grants 62103266 and 61973218, and by the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China (No. ICT2022B07).

inferred by the observers, we investigate a related collaboration algorithm and propose a novel hybrid noise-adding mechanism that leverages a random term and a disturbing term to enhance the irregularity of the state evolution.

- By exploiting the convergence conditions of the system collaboration, we design a distributed disturbing noise that is added in a pair-wise fashion to compensate with former noises. With the noise decaying adaptively with iterations, the mean square convergence rate and the proposed mechanism are derived.
- We further prove that the proposed algorithm largely degrades the topology inference accuracy of the attackers, while maintaining the convergence of the system states. Representative simulation examples demonstrate the effectiveness of the hybrid DTCA.

The rest parts of the paper are organized as follows: Sec. II provides some preliminary knowledge and formulates the problem. The proposed algorithm and its performance analysis are in Sec. III and Sec. IV. Sec. V shows the simulation and comparison results. Finally, Sec. VI concludes the work.

II. PRELIMINARIES

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed and connected graph that models the topology information within the multi-agent system, where $\mathcal{V} = \{1, \dots, N\}$ is the set of nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ denotes the set of edges. Each node represents an agent, and each weighted edge represents an information exchange channel. The adjacency matrix $A = [a_{ij}]_{N\times N}$ of a graph $\mathcal{G}$ with N agents specifies the interconnection topology of the system, where $a_{ij} > 0$ if and only if $(i, j) \in \mathcal{E}$, else $a_{ij} = 0$. Let $\mathcal{N}_i = \{j \in \mathcal{V} : a_{ij} \neq 0\}$ be the neighbor set of agent i, and $d_i = |\mathcal{N}_i|$ as its in-degree. Let $\mathbf{1}$ be an all-one column vector and $\mathbf{0}$ be an all-zero column vector with compatible dimensions. Let $\mathbb{N}^+$ be the set of positive integers, $\|\cdot\|$ and $\|\cdot\|_F$ represent the spectral norm and Frobenius norm of a matrix, respectively. For two real-valued functions f_1 and f_2, $f_1(x) = \mathcal{O}(f_2(x))$ as $x \to x_0$ means $\lim_{x\to x_0} |f_1(x)/f_2(x)| < \infty$.

A. System Model

Consider N agents collaborate to fulfill a common task, where the consensus algorithm is widely adopted to drive all agents to reach a common state, consequently producing collective behaviors. Denote by x_i the state of agent i, and the dynamics of agent i under the consensus-based collaboration algorithm is described by:

$$x_i(k+1) = x_i(k) + \sum_{j\in\mathcal{N}_i} w_{ij}(x_j(k) - x_i(k)), \tag{1}$$

where w_{ij} is the interaction weight between i and j, and is related to a_{ij}. Many popular rules can be adopted to set w_{ij}, using Laplacian rule for example, we have

$$w_{ij} = \begin{cases} \gamma a_{ij}/d_{\max}, & i \neq j, \\ 1 - \sum_{j\in\mathcal{N}_i} w_{ij}, & i = j, \end{cases} \tag{2}$$

where the auxiliary parameter γ satisfies $0 < \gamma < 1$, and the largest in-degree $d_{\max} = \max\{d_i, i \in \mathcal{V}\}$.

Accordingly, the interaction topology matrix is given by $W = [w_{ij}]_{N\times N}$. Then, the global form of the system dynamics is given by

$$x(k+1) = Wx(k),\ y(k) = x(k) + v(k), \tag{3}$$

where $v(k)$ is i.i.d. Gaussian noise, satisfying $v(k) \sim \mathcal{N}(0, \sigma_v^2 I)$. Let λ_i be the i-th eigenvalue of W, ordered as $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_N|$. Since W is a row-stochastic matrix, $\lambda_1 = 1$ and the corresponding right eigenvector of λ_1 is $\mathbf{1}$. By the consensus-based algorithm (3), the states of the agents converge to common value, given by [4]:

$$\lim_{k\to\infty} x_i(k) = x_c = p_1^\mathsf{T} x(0), \forall i \in \mathcal{V}, \tag{4}$$

where p_1 is the normalized left eigenvector associated with λ_1, and x_c is a constant representing the consensus point. Furthermore, if W is a doubly-stochastic matrix, one has $x_c = \frac{1}{N}\sum_{i\in\mathcal{V}} x_i(0)$, i.e., an average consensus is achieved.

Let $z(k) = x(k) - x_c\mathbf{1}$ be the convergence error, and the mean square convergence rate ρ is defined as [16],

$$\rho \triangleq \lim_{k\to\infty} \sup_{z(0)\neq 0} \left(\frac{\mathbb{E}[z(k)^\mathsf{T} z(k)]}{z(0)^\mathsf{T} z(0)} \right)^{\frac{1}{k}}. \tag{5}$$

Based on this definition, the mean square convergence of (3) is characterized by $\rho = \max\{|\lambda_2|^2, |\lambda_n|^2\}$, suggesting that the weighted average consensus is achieved exponentially fast.

B. Topology Inference Mechanism

We consider a scenario where the participating agents are reliable and the external eavesdroppers acquire observation data sets to infer the system's topology. Under the system model (3), the quantity that best reflects the topology would be the interaction matrix W. Estimating W is based on the observations of the system's convergence process. Denote the observation slice matrix as $Y_{a;b} = [y(a), \cdots, y(b)]$ and the noise slice matrix $V_{a;b} = [v(a), \cdots, v(b)]$. It is easy to reach

$$Y_{a+1;b+1} = WY_{a;b} - WV_{a;b} + V_{a+1;b+1}. \tag{6}$$

For simple expressions, let $Y = Y_{k_0;k_n-1}$ and $Z = Y_{k_0+1;k_n}$. Consider that the attackers adopt the classic least squares method [8] [17] to estimate the topology, which aims to solve

$$\min_{\hat{W}} \|\hat{W}Y - Z\|_F^2. \tag{7}$$

When Y^T is specific and column full rank, the optimal solution of (7) is given by

$$\hat{W}^* = ZY^\mathsf{T}(YY^\mathsf{T})^{-1}. \tag{8}$$

Lemma 1. *(Theorem 6 in [18]) The non-asymptotic error bound of the least squares estimator $\hat{W}$ satisfies*

$$\lim_{T\to\infty} \|\hat{W} - W\| = \mathcal{O}(\sigma_v^2), \tag{9}$$

where σ_v^2 is the variance of observation noise, and T is the observation number of the system.

C. Problem Formulation

As is shown in Lemma 1, the topology can be inferred accurately by an attacker, indicating the system's vulnerability. In this paper, we consider preserving the actual topology of the system by adding extra noises to the system states, i.e.,

$$x(k+1) = Wx(k) + \theta(k). \tag{10}$$

Based on (10), we focus on the noise-adding mechanism of $\theta(k)$ to protect the topology from being estimated by the inference model (8), while maintaining the convergence of the system dynamics. Mathematically, it can be formulated as

$$\max_{\theta} \quad \|\hat{W}(\theta) - W\|_F \tag{11a}$$

$$\text{s.t.} \quad \lim_{k\to\infty} x_i(k) = \lim_{k\to\infty} x_j(k), \forall i, j \in \mathcal{V}. \tag{11b}$$

This problem is quite challenging due to the following reasons. First, the adding procedure needs to be implemented in a distributed way, without relying on any global knowledge about the system topology. Second, instead of adding noises with a specific distribution, the added noise should be irregular enough to make the estimated $\hat{W}$ far different from W, while not hindering the convergence of the system. Accordingly, we propose a hybrid distributed topology-preserving collaboration algorithm, applying to both directed and undirected topologies.

III. ALGORITHM DESIGN

In this section, we propose a hybrid distributed topology-preserving algorithm (hybrid DTCA) that runs within finite iterations. The key idea and the detailed algorithm design are contained in this section.

A. Key Idea

To fulfill the requirements of the formulated problem, we consider that the added noise θ is a hybrid noise composed of two independent parts, given by $\theta(k) = \mu(k) + \omega(k)$, where $\mu(k)$ is a random term satisfying a specific distribution, and $\omega(k)$ is a non-random disturbing term used to enhance the irregularity of the system dynamics. This design form is motivated by two aspects. First, it is straightforward and common to introduce $\mu(k)$ to pollute the normal system states, which increases the state variance and thus degrade the inference performance of $\hat{W}$, compared with the situation where no noise is involved. Nevertheless, this kind of noise still has an inherent statistical characteristic (e.g., zero mean and bounded variance), which will bring minor benefits when the observation scale of the attacker grows. Second, considering the noise $\mu(k)$ can be eliminated by outlier handling through the filtering methods towards observation data, we further leverage $\omega(k)$ to intentionally break the theoretical boundaries that the system states should satisfy in normal iteration, while free of the inherent statistical characteristic limitation. Therefore, the combination of the two terms will achieve dual protective effects in terms of the states' statistical randomness and the irregularity of the state evolution. Note that both of the terms will affect the convergence rate of the system states, which will be analyzed in the next section, along with the error analysis of $\hat{W}$ under the proposed hybrid DTCA algorithm.

B. Hybrid DTCA

Design of Random Term. Borrow the idea from [16], the random term $\mu(k)$ can be designed based on a group of i.i.d. Gaussian noises $\vartheta(k)$, where each element is with mean 0 and variance σ_ϑ. Then, $\mu(k)$ is given by

$$\mu(k) = \begin{cases} \vartheta(0), & \text{if} \quad k = 0 \\ \varphi^k \vartheta(k) - \varphi^{k-1}\vartheta(k-1), & \text{otherwise}, \end{cases} \tag{12}$$

where $0 < \varphi < 1$ is an attenuation coefficient that guarantees the convergence of the system.

Remark 1. *Note that the noise sum $\|\sum_{t=0}^{k-1} W^{k-t-1}\mu(t)\|$ is strictly bounded to guarantee the state convergence and $\mu(k)$ is of Gaussian distribution, the convergence rate of $\hat{W}$ is $\mathcal{O}(\sqrt{\frac{\log T}{T}})$, satisfying $\lim\limits_{T\to\infty} \|\hat{W} - W\| = \mathcal{O}(\sigma_v^2)$ [18]. The demand for persevering topology is not fulfilled yet.*

Design of Disturbing Term. The disturbing term is designed with two goals: 1) alternatively add two complementary noises, to increase the regression errors while ensuring the convergence, and 2) set bounds of the noises to improve the flatness of the iteration progress. The two-tuple data (i, k) is selected into a set $\mathbb{B}$, with probability p. Suppose that for the i-th agent, the **additive noise** $\omega_i^+(k)$ is added at time slot k if $(i, k) \in \mathbb{B}$, and an additive noise indicator $b_i(k)$ is given by:

$$b_i(k) = \begin{cases} 1, & (i, k) \in \mathbb{B}, \\ 0, & \text{otherwise}, \end{cases}$$

where "1" represents that there exists an additive noise.

To balance the effect of the additive noise on convergence, a **compensating noise** $\omega_i^-(k + m|k)$ is imposed after m iterations, where $m \in \mathbb{N}^+$ indicates the compensation gap. In general, the expression of the extra noise to the original state $x_i(k)$ will be

$$\omega_i(k) = \omega_i^+(k)b_i(k) + \omega_i^-(k|k-m)b_i(k-m), \tag{13}$$

where k and m can be chosen arbitrarily, and they are not necessarily the same for each agent. As mentioned in Sec. II, an agent's state in the next iteration depends on its state and neighbors' states at present. Denote the state of agent i in time slot $k+1$ based on the regular iteration of $x_i(k)$ as $x_i^r(k+1|k)$. Following the regular iteration process in (3), we have $x_i^r(k+1|k) = W_i x_i(k)$. For simplicity, we denote $x_{\mathcal{N}_i}(k)$ as the set of neighbors' states $x_j(k), \forall j \in \mathcal{N}_i$. Since W is a row stochastic matrix, in the regular iteration, $x_i^r(k+1|k)$ will have an inequality constraint

$$\min\{x_{\mathcal{N}_i}(k), x_i(k)\} \le x_i^r(k+1|k) \le \max\{x_{\mathcal{N}_i}(k), x_i(k)\},$$

which implies that $x_i^r(k+1|k)$ will not exceed the extreme values of $x_{\mathcal{N}_i}(k)$ and $x_i(k)$.

In this way, we could characterize the relative size between the additive noise $\omega_i^+(k)$ and the states by utilizing a scaling parameter α. The upper boundary and the lower boundary of the additive noise become

$$\beta_i^+(k) = \alpha \times (\max\{x_{\mathcal{N}_i}(k), x_i(k)\} - x_i^r(k+1|k)) \tag{14a}$$

$$\beta_i^-(k) = \alpha \times (\min\{x_{\mathcal{N}_i}(k), x_i(k)\} - x_i^r(k+1|k)). \tag{14b}$$

Note that $\alpha = 1$ indicates that when the additive noise takes the boundary value in the relative range, $x_i(k+1)$ can reach extreme value of its neighbors' states, with $\alpha > 1$ beyond and $\alpha < 1$ within. If the designed noises cause the violation of the inequality constraints to $x_i^r(k+1|k)$, the inference error will be especially enlarged. We choose the additive noise between the upper and lower bound to maintain a relative size with the nodes' states and ensure the flatness of the system's convergence. It can be proven that if $\frac{\alpha p}{2} < 1$, $\beta_i^+(k)$ and $\beta_i^-(k)$ are decaying in the sense of expectation. The compensating noise is formulated as

$$\omega_i^-(k+m|k) = -\omega_i^+(k). \tag{15}$$

This part of the algorithm can be done in a distributed manner because the calculation of compensating noise requires only the agent's historical additional noises but not in-degree neighbors' information. The value of m is enormously influential to the overall performance of the algorithm. The smaller m is, the shorter the compensation gap is, resulting in a more frequent and drastic change in the trajectory.

Combining the above two terms, we complete the design of extra noise. In the design, $\mu(k)$ applies to increase the variance of the estimate, which could be eliminated by filtering methods in multi-round observation. On the other hand, $\omega(k)$ offers disturbance, breaking the theoretical boundaries of the states and deceiving the attackers into unveracious links. This kind of noise could be disposed of by the polynomial fitting method [19], while combining $\mu(k)$ helps blur the disturbance, resulting in better concealment of the weight of the links and the links themselves. The following pseudocode presents the details of hybrid DTCA under limited time iteration. Suppose the maximum iteration number is $k_{\max}$.

IV. ALGORITHM PERFORMANCE ANALYSIS

This section presents the performance analysis of the proposed algorithm, including the convergence rate and the non-asymptotic error bound of the inference attack.

A. Convergence Analysis

When the hybrid DTCA is applied to the system, the added noises to the agents will confuse not only the attackers but also the agents in the neighborhood. To ensure the regular operation of the system, convergence to the original consensus point is the first thing to consider.

We can write $x(k)$ in the following form:

$$x(k) = W^k x(0) + \sum_{t=0}^{k-1} W^{k-t-1}\theta(t), \tag{16}$$

which indicates that to ensure the original consensus point, the overall noise $\theta(k)$ must satisfy the following formula: $\lim_{k\to\infty} \sum_{t=0}^{k-1} W^{k-t-1}\theta(t) = \mathbf{0}$.

Algorithm 1: Hybrid DTCA

```
Input: $x(0)$, $w_{ij}$, $\alpha$, $p$, $\mu$, $\varphi$
Output: Observation data set;
Initialization;
for $k = 0 : k_{\max}$ do
    Generate $v(k)$, $\mu(k)$ and $b(k)$;
    $y(k) = x(k) + v(k)$;
    for agent $i$ do
        if $b_i(k) == 1$ then
            Calculate $\beta_i^+(k)$ and $\beta_i^-(k)$ as in (14a) and
              (14b);
            Choose the additive noise $\omega_i^+(k)$ between
              the bounds, i.e., $\omega_i^+(k) \in [\beta_i^-(k), \beta_i^+(k)]$;
        end
        Compute the compensating noise $\omega_i^-(k+m|k)$
          using (15);
    end
    Design $\theta(k)$ as in (12) and (13);
    Update $x(k+1)$ by (10);
end
```

The influence of $\omega(k)$ is $W^m\omega^+(k-m) + \omega^-(k|k-m)$ because the disturbing term shows in pairs. As mentioned in Sec. II, the consensus point $x_c = p_1^\mathsf{T} x(0)$ where p_1 is the first normalized left eigenvector, corresponding to eigenvalue 1. Thus we have $p_1^\mathsf{T} W = p_1^\mathsf{T}$, leading to

$$p_1^\mathsf{T} W^{m+s}\omega^+(k) - p_1^\mathsf{T} W^s \omega^+(k) = 0, \forall s \in \mathbb{N}^+,$$

which shows that adding only the disturbing term $\omega(k)$, the algorithm can be proved to reach the exact consensus point.

Furthermore, adding only the random term $\mu(k)$, an exact consensus is also reached due to the analysis in [16]. Since the two noise terms are linear superposition, it is proven that the hybrid DTCA achieves an exact consensus.

Theorem 1. *Given any $x(0)$, an asymptotic weighted average consensus is achieved exponentially fast using hybrid DTCA, i.e., $\lim_{k\to\infty} x_i(k) = x_c$, where the mean square convergence rate is related with the parameters φ, α and p in the algorithm, and the second largest eigenvalue of $W^\mathsf{T} W$:*

$$\rho = \max\{\varphi^2, (\frac{\alpha p}{2})^2, \lambda_2(W^\mathsf{T} W)\}.$$

Proof. Substitute $x(k)$ for the expression of $z(k)$ and calculate $\mathbb{E}z(k)^\mathsf{T} z(k)$, we would have four terms with convergence speeds $\rho_1 = \lambda_2(W^\mathsf{T} W)$, $\rho_2 = \varphi^2$, $\rho_3 = \max\{\varphi^2, \lambda_2(W^\mathsf{T} W)\}$ and $\rho_4 = (\frac{\alpha p}{2})^2$, respectively. One thus completes the proof of the convergence rate, suggesting the convergence rate is dependent on W and the noise terms. □

B. Inference Error Analysis

Same as the problem formulation, we use the Frobenius norm to describe the precision of the estimation $\hat{W}$. To protect the actual topology, we need to make the approximation error $\|\hat{W} - W\|_F$ as large as possible. Without loss of generality, we suppose the observation data is $Y(0,\cdots,T)$. Obtained by (10), $Y_{1;T} = WY_{0;T-1} + \Phi_{0;T-1}$, where the slice matrix $\Phi_{0;T-1} = [\phi(0),\cdots,\phi(T-1)]$ and $\phi(k) = \theta(k) - Wv(k) + v(k+1)$. As is in Sec. II, the topology's best estimation is:

$$\hat{W} = Y_{1;T}Y_{0;T-1}^{\mathsf{T}}(Y_{0;T-1}Y_{0;T-1}^{\mathsf{T}})^{-1}.$$

Substitute $WY_{0;T-1} + \Phi_{0;T-1}$ for $Y_{1;T}$, then we will have $\hat{W} - W = \Phi_{0;T-1}Y_{0;T-1}^{\mathsf{T}}(Y_{0;T-1}Y_{0;T-1}^{\mathsf{T}})^{-1}$. The problem can be formulated as a constrained optimization problem where the additive noise is selected to maximize the inference error:

$$\begin{aligned} \text{P1:}\quad & \max_{\omega(k)} \left\|\Phi_{0;k}Y_{0;k}^{\mathsf{T}}(Y_{0;k}Y_{0;k}^{\mathsf{T}})^{-1}\right\|_F \\ & \text{s.t.}\quad \beta_i^-(k) \le \omega_i^+(k) \le \beta_i^+(k). \end{aligned} \tag{17}$$

Theorem 2. *(Policy of $\omega(k)$ design) The optimal solution of Problem $\boldsymbol{P1}$ in (17) is equal to the constrained boundary with a larger absolute value, either $\beta^-(k)$ or $\beta^+(k)$.*

Proof. Define $\Upsilon(0;k) = Y_{0;k}^{\mathsf{T}}(Y_{0;k}Y_{0;k}^{\mathsf{T}})^{-1}$ and split it into $\Upsilon_A(0;k)$ whose size is $k \times N$ and $\Upsilon_B(0;k)$ whose size is $1 \times N$. $\Phi_{0;k}$ is split into $\Phi(0;k-1)$ and $\phi(k)$. Thus the optimized objective function of (17) can be broken down as

$$\begin{aligned} & \left\|[\Phi(0;k-1) \mid \phi(k)]\left[\frac{\Upsilon_A(0;k)}{\Upsilon_B(0;k)}\right]\right\|_F \\ =& \sqrt{\sum_{i=1}^{N}\|(\Phi_i(0;k-1)\Upsilon_A(0;k) + \phi_i(k)\Upsilon_B(0;k))\|_F^2}. \end{aligned}$$

Hence, this optimization problem is decomposed into N independent sub-optimization problems, i.e., each row's Frobenius norm optimization problems, which is given by

$$\begin{aligned} \text{P2:}\quad & \max_{\omega_i(k)} \|\Phi_i(0;k-1)\Upsilon_A(0;k) + \phi_i(k)\Upsilon_B(0;k)\|_F \\ & \text{s.t.}\quad \beta_i^-(k) \le \omega_i(k) \le \beta_i^+(k). \end{aligned} \tag{18}$$

Since the objective function of (18) is a convex quadratic function of $\phi_i(k)$, it is maximized when $\phi_i(k)$ reaches its extreme, i.e., $\omega_i(k)$ equals one of the restrictions. Furthermore, as the Frobenius norm is always positive and restricted, the overall target of (17) is accomplished if and only if each independent optimization problem in (18) is maximized. □

Corollary 1. *(Gaussian matrices, deviation; Corollary 5.35 in [20]). Let A be an $N \times n$ Gaussian matrix, and $s_{\min}(A)$, $s_{\max}(A)$ be the smallest and the largest singular values of A, respectively. Then for every $t \ge 0$, with probability at least $1 - 2\exp(-t^2/2)$ one has*

$$\sqrt{N} - \sqrt{n} - t \le s_{\min}(A) \le s_{\max}(A) \le \sqrt{N} + \sqrt{n} + t. \tag{19}$$

Proof. The conclusion follows from Theorem 5.32 and Proposition 5.34 in [20]. □

We define $\Psi_{0;k} = Y_{0;k}Y_{0;k}^{\mathsf{T}}$, and there exists Ψ_{dn} and Ψ_{up} satisfying $0 \prec \Psi_{dn} \preceq \Psi_{0;k} \preceq \Psi_{up}$ to help with the error bound processing. Corollary 1 is needed to deduce the asymptotic bound of this algorithm setting with different observation horizons. The following Theorem infers the relationship between the asymptotic bound and the extra noises, while the hybrid design of $\theta(k) = \mu(k) + \omega(k)$ has been shown to be more effective than either noise term alone at enlarging the mean and variance of the regression error.

TABLE I
COMPARISON OF THE NON-ASYMPTOTIC ERROR BOUNDS

Input Design	$\|\hat{W} - W\|$
$\theta(k) = 0$	$\mathcal{O}(\sigma_v^2)$
$\theta(k) = \mu(k)$	$\mathcal{O}(\sigma_v^2) + \mathcal{O}(\sqrt{\frac{\log T}{T}})$ (Theorem 6 in [18])
$\theta(k) = \mu(k) + \omega(k)$	$\mathcal{O}(\sigma_v^2) + \mathcal{O}\left(\left(\mathbb{E}[\theta] + \sqrt{\mathbb{D}[\theta]/\delta}\right)^2\right)$

Theorem 3. *(Sensitivity; improvement of Lemma 1) Let extra noise $\theta = \{\theta_i(k)\}_{i\in\mathcal{V},k=0,1,\cdots}$ have finite expected value $\mathbb{E}[\theta]$ and variance $\mathbb{D}[\theta]$. Applying hybrid DTCA, the non-asymptotic error bound of the least squares estimator is up to*

$$\|\hat{W} - W\| \sim \mathcal{O}(\sigma_v^2) + \mathcal{O}\left(\left(\mathbb{E}[\theta] + \sqrt{\mathbb{D}[\theta]/\delta}\right)^2\right). \tag{20}$$

Proof. The detailed deduction of (9) can be found in Theorem 6 in [18], and (20) could be seen as an improved version of (9), adding the designed noises $\mu(k)$ and $\omega(k)$. Here we would omit the redundant proof and focus on the effect of $\theta(k)$. Let the error matrix $E_W = \Phi_{0;T-1}Y_{0;T-1}^{\mathsf{T}}(Y_{0;T-1}Y_{0;T-1}^{\mathsf{T}})^{-1}$ and matrix $M = \Phi_{0;T-1}Y_{0;T-1}^{\mathsf{T}} = \sum_{k=1}^{T}\phi(k)y^{\mathsf{T}}(k)$. Using Chebyshev Inequity, for any real number $\delta > 0$, we have $\Pr\left\{|\theta - \mathbb{E}[\theta]| \ge \delta\sqrt{\mathbb{D}[\theta]}\right\} \le 1/\delta^2$. Since $\|M\| \le \|M\|_F$, we have $\|M\| \le N^2(\mathbb{E}[\theta] + \sqrt{\mathbb{D}[\theta]/\delta})^2$ with probability at least $1 - \delta$. Together with $\left\|(Y_{0;T-1}Y_{0;T-1}^{\mathsf{T}})^{-1}\right\| \le \frac{1}{s_{\min}(\Psi_{dn})}$,

$$\begin{aligned} \|E_W\| &\le \|M\|\left\|(Y_{0;T-1}Y_{0;T-1}^{\mathsf{T}})^{-1}\right\| \\ &\le \frac{N^2}{s_{\min}(\Psi_{dn})}(\mathbb{E}[\theta] + \sqrt{\mathbb{D}[\theta]/\delta})^2, \end{aligned}$$

completes the proof of the extra noises. □

Table I shows the comparison of the non-asymptotic error bounds under different noise designs.

V. SIMULATION

A. Simulation Setting

A directed network with ten agents is randomly constructed, reflecting the system's interaction topology. Assign all the agents with random initial states, and start the iteration under hybrid DTCA with different parameter configurations. During the iteration process, we experiment on different parameter configurations in hybrid DTCA, namely α and σ_ϑ, to evaluate the algorithm's performance under inference attacks.

B. Results and Analysis

Without loss of generality, we first verify the proposed algorithm's performance when the attackers collect the information all the time. In the real scenario, the observation time can be limited as the attackers may not be able to keep tracking the system from the beginning to the end. We then verify the algorithm's performance in this situation. Fig. 1 and Fig. 2 present the hybrid DTCA's performance for

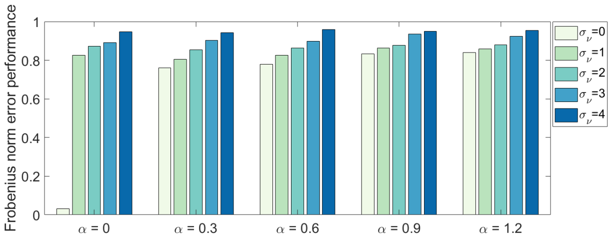

Fig. 1. Error performance with all time slots observation

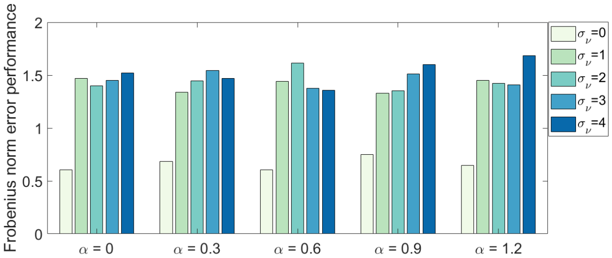

Fig. 2. Error performance with limited time slots observation

all time slots observation and limited time slots observation, respectively. Each figure presents the hybrid algorithm's performance, where the vertical axis shows the Frobenius norm inferred topology's approximation error and the horizontal axis indicates the parameter configurations. In the experiment, we set 5×5 groups of parameter configurations of the random term and the disturbing term for the hybrid DTCA, and the choice of the parameter configuration is based on the initial states of the agents and the maximum iteration number.

All Time Slots Observation. As we can see in Fig. 1, the overall trend of the estimation errors is in line with intuition, which is, the larger the variance of both noise terms is, the more inaccurate the regression will be. In particular, when there is no disturbance, the topology inference using least squares estimation can be highly accurate as the error bars start with zero. The parameters play significant roles in the algorithms as they affect the trajectories of agents' states and the estimation error of topology inference. Also, all the different parameter configurations have decent performance, illustrating the flexibility of the algorithms.

Limited Time Slots Observation. In this part, we select only the data from the $(N+1)$ most recent iterations (minimum data requirements for the least squares method) as the information source of the attackers. As we can see in Fig. 2, the performance of the inference attack when $\alpha = 0$ and $\mu = 0$ is not zero because the limited time slots cause the loss of information. Our algorithm is still effective in limited time slots observation, and the combination of two noise terms outperforms a single noise term alone.

Overall, the results can prove that our topology-preserving algorithm works well and performs better than with only the random term or disturbing term.

VI. Conclusion

In this paper, we propose a distributed topology-preserving collaboration algorithm: hybrid DTCA, to secure the system against topology inference attacks. In our algorithm, designed noises are added into the system to enlarge the topology inference error while guaranteeing the accuracy of the system's convergence. Extensive simulations verify the effectiveness of the proposed algorithm. Future directions include extending the algorithm to a more general dynamical network with switching topology and exploring relations between the algorithm's cost and performance.

References

[1] A. D. Kshemkalyani and M. Singhal, *Distributed computing: Principles, algorithms, and systems*. Cambridge University Press, 2011.

[2] J. He, L. Duan, F. Hou, P. Cheng, and J. Chen, "Multiperiod scheduling for wireless sensor networks: A distributed consensus approach," *IEEE Transactions on Signal Processing*, vol. 63, no. 7, pp. 1651–1663, 2015.

[3] J. Alonso-Mora, S. Baker, and D. Rus, "Multi-robot formation control and object transport in dynamic environments via constrained optimization," *The International Journal of Robotics Research*, vol. 36, pp. 1000–1021, 2017.

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

[5] Q. Zhu, "Topologies of agents interactions in knowledge intensive multi-agent systems for networked information services," *Advanced Engineering Informatics*, vol. 20, no. 1, pp. 31–45, 2006.

[6] Y. Emam, S. Mayya, G. Notomista, A. Bohannon, and M. Egerstedt, "Adaptive task allocation for heterogeneous multi-robot teams with evolving and unknown robot capabilities," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 7719–7725.

[7] D. Spinello and D. J. Stilwell, "Nonlinear estimation with state-dependent gaussian observation noise," *IEEE Transactions on Automatic Control*, vol. 55, no. 6, pp. 1358–1366, 2010.

[8] Q. Jiao, Y. Li, J. He, and L. Shi, "Topology inference for multi-agent cooperation under unmeasurable latent input," *arXiv preprint arXiv:2011.03964*, 2020.

[9] J. Li, J. He, Y. Li, and X. Guan, "Unpredictable trajectory design for mobile agents," in *American Control Conference (ACC)*, 2020, pp. 1471–1476.

[10] W. Ren and R. W. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *IEEE Transactions on automatic control*, vol. 50, no. 5, pp. 655–661, 2005.

[11] W. Ni and D. Cheng, "Leader-following consensus of multi-agent systems under fixed and switching topologies," *Systems & control letters*, vol. 59, no. 3-4, pp. 209–217, 2010.

[12] G. Wen, G. Hu, W. Yu, J. Cao, and G. Chen, "Consensus tracking for higher-order multi-agent systems with switching directed topologies and occasionally missing control inputs," *Systems & Control Letters*, vol. 62, no. 12, pp. 1151–1158, 2013.

[13] T. Dong, X. Bu, and W. Hu, "Distributed differentially private average consensus for multi-agent networks by additive functional laplace noise," *Journal of the Franklin Institute*, vol. 357, no. 6, pp. 3565–3584, 2020.

[14] H. Sun, Z. Wang, J. Xu, and H. Zhang, "Exact consensus error for multi-agent systems with additive noises," *Journal of Systems Science and Complexity*, vol. 33, no. 3, pp. 640–651, 2020.

[15] V. Katewa, A. Chakrabortty, and V. Gupta, "Protecting privacy of topology in consensus networks," in *American Control Conference (ACC)*, 2015, pp. 2476–2481.

[16] Y. Mo and R. M. Murray, "Privacy preserving average consensus," *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 753–765, 2016.

[17] C. Liu, J. He, S. Zhu, and C. Chen, "Dynamic topology inference via external observation for multi-robot formation control," in *2019 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, 2019, pp. 1–6.

[18] Y. Li, J. He, C. Chen, and X. Guan, "On topology inference for networked dynamical systems: Principles and performances," *arXiv preprint arXiv:2106.01031*, 2021.

[19] H. Huang, Y. Cai, H. Xu, and H. Yu, "A multiagent minority-game-based demand-response management of smart buildings toward peak load reduction," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 36, no. 4, pp. 573–585, 2016.

[20] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *arXiv preprint arXiv:1011.3027*, 2010.