Differential Privacy-Based Resilient Bipartite Consensus for Multi-Agent Systems Under Adversarial Nodes

Lijuan Zha, Jianan Wang, and Jinde Cao, *Fellow, IEEE*

***Abstract*—This article investigates the bipartite consensus problem in multi-agent systems (MASs) characterized by both cooperative and competitive interactions. To address this problem, a unified strategy is developed that integrates differential privacy mechanisms with malicious attack mitigation techniques. First, we optimize the noise injection mechanism by separately injecting Laplacian noise with different distributions into cooperative and competitive nodes. This approach not only enables more flexible adjustment of differential privacy levels but also effectively accelerates the convergence process of agents achieving bipartite consensus. Then, by introducing a random label verification mechanism and combining it with *r*-robust graph theory, a new resilient bipartite consensus algorithm is proposed, which can effectively resist Sybil attacks and Byzantine attacks, and broaden the scope of its application. Finally, simulation results confirm the effectiveness of the proposed method.**

***Index Terms*—Differential privacy, multi-agent system, bipartite consensus, Sybil attack, byzantine attack.**

I. Introduction

In RECENT decades, the cooperative control of multi-agent systems (MASs) has attracted widespread attention, owing to their broad applications in areas such as robot collaboration, intelligent transport systems, and satellite formation. Significant achievements have been made in this field [1], [2], [3]. The work by Jiao et al. [4] contributed to the optimal consensus control of continuous-time nonlinear MASs by introducing a parallel control method leveraging adaptive dynamic programming. In [5], Song et al. designed a fully distributed event-triggered secure control scheme for nonlinear multi-agent systems, enabling the achievement of leader-follower asymptotic consensus.

However, the aforementioned studies are predicated on the assumption of exclusively cooperative interactions. In practical scenarios, cooperative and competitive relationships often coexist, driving collaborative agents toward agreement and competitive ones toward opposition. Consequently, the bipartite consensus problem has received significant attention, leading to various methodologies designed for complex conditions, such as actuator faults [6] and switching topologies [7].

In practical deployments, MASs face significant threats from adversarial attacks that aim to compromise consensus. Primary threats include Denial-of-Service (DoS) and deception attacks, which exploit operational vulnerabilities [8], [9]. Defensive strategies against these attacks typically operate under a critical assumption that the number of malicious agents remains below a specific threshold. The very nature of a Sybil attack fundamentally violates this premise. By allowing adversaries to create numerous fake identities, Sybil attacks artificially inflate the number of malicious entities, causing severe performance degradation in conventional control algorithms. While recent research has developed Sybil attack detection mechanisms [10], [11], including the establishment of Doppler shift-based frameworks [10] and the analysis of degree changes in undirected graphs [11], their applicability diminishes significantly in environments with excessively high Sybil node density. Moreover, Sybil attacks are often combined with Byzantine attacks [12], which manipulate nodes to disseminate false information to legitimate ones [13]. This combination makes defense exceptionally challenging and significantly increases potential damage.

In addition, MASs are also prone to eavesdropping during operation [14], [15], [16], which can compromise sensitive system information and create significant security risks. Consequently, privacy preservation in MASs has become an increasingly vital research domain. As an effective method for preventing data leakage, differential privacy is widely adopted, the core mechanism of which involves injecting unrelated noise into datasets or query responses. Due to its proven effectiveness in preventing data leakage, the differential privacy scheme has become a mainstream approach for protecting the privacy of MASs [17], [18]. Gao et al. [17] introduced a unified quantization communication strategy based on the existing differential privacy consensus algorithm and added an additional exponentially decaying dynamic factor to mitigate the impact of quantization error on average consensus. A framework combining bipartite consensus with differential privacy protection was developed by Zuo et al. [18] for MASs operating over digraphs.

Received 23 February 2026; revised 4 April 2026; accepted 25 April 2026. Date of publication 29 April 2026; date of current version 28 August 2026. This work was supported in part by the National Natural Science Foundation of China under Grant 62273174 and in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20230063. *(Corresponding author: Lijuan Zha.)*

Lijuan Zha and Jianan Wang are with the School of Science, Nanjing Forestry University, Nanjing 210037, China (e-mail: zhalijuan@vip.163.com; wangjianan623@163.com).

Jinde Cao is with the School of Mathematics, Southeast University, Nanjing 210096, China (e-mail: jdcao@seu.edu.cn).

Digital Object Identifier 10.1109/TCE.2026.3688314

Inspired by existing research, this paper presents a solution for bipartite consensus in MASs operating in a security-critical and privacy-sensitive environment threatened by Sybil and Byzantine attacks. Our approach integrates differential privacy to safeguard sensitive information throughout the process. The principal contributions are outlined below:

1) The problem of bipartite consensus in MASs with cooperative-competitive interactions is investigated, in which agents operate under a more complex and potent adversarial environment that involves both Sybil and Byzantine attacks. This new form of hybrid attack renders existing resilient bipartite consensus algorithms ineffective. This limitation motivates the development of novel attack resilient scheme and control algorithm specifically designed for such adversarial network environment.

2) To mitigate the impact of the aforementioned hybrid attacks on bipartite consensus, a resilient bipartite consensus algorithm based on random label is designed. Compared with prior research such as [11] that requires Sybil nodes to be a strict minority, the proposed algorithm overcomes this restriction, maintaining resilience even in adversarial majority scenarios with unlimited Sybil child nodes.

3) A new differential privacy noise injection scheme is proposed. Unlike conventional uniform noise injection methods [17], [18], our approach introduces group-specific noise allocation, where agents in different groups receive tailored perturbation levels. By strategically distributing noise, our method minimizes unnecessary interference while maintaining rigorous privacy guarantees, leading to faster convergence in bipartite consensus compared to existing techniques.

II. Preliminary and Problem Description

A. Notations

Let $\mathbb{R}$ and $\mathbb{N}$ denote the set of real numbers and non-negative integers, respectively. The symbols $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ represent the n-dimensional real vector space and the set of $m \times n$ real matrices, respectively. For a matrix X belonging to $\mathbb{R}^{m\times n}$, X^T denotes its transpose. Let $\mathbf{1}_N$ be a column vector in which every element is equal to 1. For two sets Y_1 and Y_2, the notation $Y_1 \setminus Y_2$ denotes their set difference, i.e., elements in Y_1 but not in Y_2. The cardinality of a set Y, representing the number of its elements, is denoted by $|Y|$. The symbol $\varnothing$ represents the empty set. For a random variable, $\mathbb{P}[\cdot]$, $\mathbb{E}[\cdot]$, $\mathbb{V}[\cdot]$ denote its probability, expected value and variance. The p-norm of a vector is represented by $\|\cdot\|_p$. The diag$(\cdot)$ operation is used to create a diagonal matrix based on the given elements, and sign$(\cdot)$ is the sign function which returns the sign of a number.

B. Graph Theory

This paper considers the MASs composed of N agents. Its communication topology is modeled by a signed digraph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{A}\}$, where $\mathcal{V} = \{1, 2, \ldots, N\}$ is the node set, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set, and $\mathcal{A} = [a_{ij}] \in \mathbb{R}^{N\times N}$ is the adjacency matrix. The entries of $\mathcal{A}$ satisfy: $a_{ij} \neq 0$ if and only if $(i, j) \in \mathcal{E}$, and $a_{ij} = 0$ otherwise. A cooperative relationship between agents i and j is indicated by $a_{ij} > 0$, while a competitive one is indicated by $a_{ij} < 0$. The graph has no self-loops, i.e., $a_{ii} = 0$ for all $i \in \{1, 2, \ldots, N\}$.

The in-neighbor set and out-neighbor set of agent i are defined as $\mathcal{N}_i^+ = \{j \in \mathcal{V} \mid (i, j) \in \mathcal{E}\}$ and $\mathcal{N}_i^- = \{j \in \mathcal{V} \mid (j, i) \in \mathcal{E}\}$, respectively. The digraph is said to contain a directed spanning tree if there exists at least one root agent that has a directed path to every other agent. The Laplacian matrix $\mathcal{L} = [l_{ij}] \in \mathbb{R}^{N\times N}$ is defined by $l_{ii} = \sum_{j\neq i} |a_{ij}|$ and $l_{ij} = -a_{ij}$ for $i \neq j$.

Definition 1: [19]: A signed digraph $\mathcal{G}$ is structurally balanced if and only if there exists a bipartition $\mathcal{V}_1, \mathcal{V}_2$ of $\mathcal{V}$ such that $a_{ij} \geq 0$ for $i, j \in \mathcal{V}_p (p = 1, 2)$ and $a_{ij} \leq 0$ otherwise. This is equivalent to the existence of a diagonal matrix $S = \text{diag}\{s_1, \ldots, s_N\}$, $s_i \in \{\pm 1\}$, for which the transformed Laplacian $\mathcal{L}_S = S\mathcal{L}S$ is positive semidefinite.

Definition 2: (r-Reachable Set and r-Robust Graph [20]): In a graph $\mathcal{G}$ with $N \geq 2$ nodes, a nonempty set $\bar{\mathcal{V}} \subset \mathcal{V}$ is r-reachable ($r \in \mathbb{N}$) if $\exists i \in \bar{\mathcal{V}}$ such that $|\mathcal{N}_i^+ \setminus \bar{\mathcal{V}}| \geq r$. The graph $\mathcal{G}$ itself is r-robust if for any pair of nonempty, disjoint sets $\mathcal{V}_1, \mathcal{V}_2 \subset \mathcal{V}$, at least one is r-reachable.

C. Attack Model

In this study, inter-agent data transmission occurs through a shared communication network vulnerable to three primary security threats: eavesdropping, Sybil attacks, and Byzantine attacks. The formal characterization of these attack models is presented below.

An eavesdropper aims to covertly extract sensitive system data without interfering with control algorithms. Typically, the eavesdropper can obtain the complete network topology, monitor all inter-node communication signals, and infer target states through signal analysis.

In the Byzantine attack model [21], a subset of agents, referred to as Byzantine nodes, may be either compromised by an adversary or exhibit inherent malicious behavior. These nodes deviate from predefined protocols and may transmit arbitrary forged or false messages to other agents, with the goal of disrupting collective coordination and preventing the system from reaching correct consensus.

Sybil attacks involve adversaries manipulating compromised agents to fabricate numerous fake identities. Even highly secure agents can be manipulated to propagate forged identities. The primary objective of Sybil attacks is to maximize adversarial influence within the targeted system, ultimately undermining system control. For clarity, we refer to the original compromised node as the Sybil parent node, and the counterfeit identities cloned from it as Sybil child nodes. The formal definition of Sybil nodes can be found in reference [11].

Given the existence of malicious nodes in the network, the nodes in the graph $\mathcal{G}$ can be partitioned into two parts. We denote the set of malicious nodes as $\mathcal{V}_M$ and the set of normal nodes as $\mathcal{V}_N$.

Definition 3 (f-local attack model): The MASs denoted by the triple $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{A}\}$ is subjected to the f-local attack precisely when, at every instant, the number of Byzantine nodes and Sybil parent nodes adjacent to any node within the graph does not exceed f.

Assumption 1: Suppose the graph $\mathcal{G}$ is directed, structurally balanced and has a spanning tree.

Assumption 2: Suppose that all the attacks in this paper follow the f-local attack model.

D. Differential Privacy

In MASs, agents exchange information to achieve bipartite consensus. However, directly transmitting raw data can lead to significant privacy leakage risks. Differential privacy offers a rigorous framework to mitigate such risks by introducing carefully calibrated noise (typically drawn from Laplace or Gaussian distributions) into the information shared among agents, which ensures that individual data remain indistinguishable, thereby preserving privacy without compromising the overall consensus process. The formal definition of the differential privacy mechanism is presented below.

Define a randomized mechanism $\mathcal{H}(\cdot) : \mathcal{D} \to \mathcal{O}$, where $\mathcal{D}$ is the space of the initial dataset and $\mathcal{O}$ is the space of all the possible observations.

Definition 4 (σ-adjacent): Given $\sigma > 0$, the initial state sets $D = \{x_i(0), i \in \mathcal{V}_N\}$ and $D' = \{x'_i(0), i \in \mathcal{V}_N\}$ are σ-adjacent if there exists $i_0 \in \mathcal{V}_N$, such that

$$|x_i(0) - x'_i(0)| \leq \begin{cases} \sigma, & i = i_0 \\ 0, & i \neq i_0 \end{cases} \tag{1}$$

Definition 5 (Differential privacy): Given $\epsilon > 0$, a randomized mechanism $\mathcal{H}(\cdot)$ is ϵ-differential private if for any two σ-adjacent initial states $x_i(0), x'_i(0)$ and $\mathcal{O} \subseteq (\mathbb{R}^N)^N$

$$\mathbb{P}\{\mathcal{H}(D) \in \mathcal{O}\} \leq e^{\epsilon}\mathbb{P}\{\mathcal{H}(D') \in \mathcal{O}\} \tag{2}$$

where ϵ is the privacy level.

Remark 1: The differential privacy mechanism ensures that the presence or absence of any single agent has a negligible impact on the query result, effectively preventing eavesdroppers from inferring sensitive individual or network information. Notably, a smaller value of the privacy level ϵ corresponds to a higher level of privacy protection.

E. Problem Formulation and Algorithm Design

Consider the following linear discrete-time MASs consisting of N agents:

$$x_i(k+1) = x_i(k) + u_i(k) \tag{3}$$

where $x_i(k) \in \mathbb{R}$ is the state of the agent i, $u_i(k) \in \mathbb{R}$ is the control input to be designed.

This paper aims to develop a control scheme $u_i(k)$ that ensures mean-square bipartite consensus among normal nodes, while maintaining ϵ_i-differential privacy across the entire system.

To enhance system resilience against Sybil attacks, we develop a randomized label verification mechanism. At time k, the label $L_i(k)$ generated by agent i will be sent to its neighboring agents along with the transmitted state value. The label generation model operates through a random sampling process based on a truncated normal distribution. The total length of all randomly sampled labels from the N agents sums to 1, with each node having an identical sampling range of $r_{L_i} = \frac{1}{N}, \forall i \in N$ where N denotes the number of normal nodes in the initial network. Meanwhile, the maximum and minimum values of the label of node i are specified as $L_{i,max} = \frac{i}{N}$ and $L_{i,min} = \frac{i-1}{N}$ respectively. That is to say, the label value $L_i(k)$ generated by node i satisfies $L_i(k) \in \left(\frac{i}{N}, \frac{i-1}{N}\right)$. To avoid different nodes generating the same label value, the truncated normal distribution function for node label sampling is given by:

$$L_i(k) \sim \psi\left(\mu_i, \sigma_i^2, L_{i,min}, L_{i,max}\right) \tag{4}$$

where $\mu_i = \frac{2i-1}{2N}$ denotes the mean of the distribution, indicating the expected value or central location of the label $L_i(k)$, $\sigma_i^2 = \left(\frac{1}{2N}\right)^2$ represents the variance of the distribution, quantifying the dispersion of the label values $L_i(k)$ around the mean μ_i.

Remark 2: By employing truncated normal distributions with strictly non-overlapping sampling intervals, we guarantee that the probability of label collisions among legitimate agents is mathematically zero. Since Sybil nodes lack physical identities and merely duplicate parent labels, the detection of identical local labels provides a reliable mechanism to identify and exclude such malicious actors.

Remark 3: Under the f-local attack model, our approach restricts only the number of Sybil parent nodes, imposing no limit on Sybil child nodes. Unlike [11], which strictly limits total Sybil nodes to a minority, our model handles adversarial majority scenarios. Since child nodes lack authentic identities and merely replicate parent labels, our random label verification mechanism can simultaneously filter unlimited clones, offering enhanced resilience against environments dominated by Sybil identities.

Define $\mathcal{V}_N$ as the set of all normal nodes. The noisy state $\hat{x}_j(k)$ for any $j \in \mathcal{V}_N$ at time k is generated according to:

$$\begin{aligned} \hat{x}_j(k) = x_j(k) &+ \frac{1 - \text{sign}(a_{ij}(k))}{2}\eta_j(k) \\ &+ \frac{1 + \text{sign}(a_{ij}(k))}{2}\theta_j(k) \end{aligned} \tag{5}$$

where $\eta_j(k) \sim \text{Lap}(b_j(k))$, $\theta_j(k) \sim \text{Lap}(\bar{b}_j(k))$, $b_j(k) = c_j q_j^k$, $\bar{b}_j(k) = \bar{c}_j q_j^k$, $0 < \bar{c}_j < c_j$, $q_j \in (0,1)$. It follows from the properties of the Laplace distribution that $\mathbb{E}(\eta(k)) = \mathbb{E}(\theta(k)) = 0$, $\mathbb{V}(\eta(k)) = 2b_j(k)^2$, $\mathbb{V}(\theta(k)) = 2\bar{b}_j(k)^2$.

After generating the label value $L_j(k)$, the node j transmits its state $\tilde{x}_j(k) = (\hat{x}_j(k), L_j(k))$ along with the label to its out-neighbors.

To mitigate the impact of outliers in the state values received by agent i, we implement a filtering mechanism to eliminate potentially corrupted state values transmitted by malicious nodes, as formally specified in Algorithm 1.

Based on the type of attacker and the attack mode (single or hybrid attack), we first exclude all state values that may be sent by Sybil nodes according to the label values (if the attack involves only a Sybil attack, all malicious nodes will be eliminated at this step). Then, the maximum number of local Byzantine nodes $\hat{f}$ is calculated based on the f-local attack model, followed by a further outlier removal process.

Remark 4: The proposed algorithm efficiently mitigates hybrid and single-type attacks with minimal computational burden. By relying solely on scalar sampling for label generation and straightforward local sorting for filtering, it ensures

Algorithm 1 Neighbor State Filtering

Input: Neighbor state $\tilde{x}_j(k)$, normal agent state $x_i(k)$, parameter f.

1: **group** nodes that share identical labeled state values together; identify Sybil node set $N_s(k)$
2: **record** the total number of such groups as $\bar{f}$.
3: **set** $\hat{f} = \max\{0, f-\bar{f}\}$. If $\hat{f} = 0$, output $\mathcal{Q}_i(k) = \mathcal{N}_i^+ \setminus N_s(k)$.
4: **Let** $\mathcal{R}_i(k) = \mathcal{N}_i^+ \setminus N_s(k)$. Define upper/lower sets:

$$\overline{\mathcal{R}}_i(k) = \{\hat{x}_j(k) \in \mathcal{R}_i(k) \mid \hat{x}_j(k) > x_i(k)\}$$
$$\underline{\mathcal{R}}_i(k) = \{\hat{x}_j(k) \in \mathcal{R}_i(k) \mid \hat{x}_j(k) < x_i(k)\}$$

5: **remove** $\min\{\hat{f}, |\overline{\mathcal{R}}_i(k)|\}$ nodes with largest values from $\overline{\mathcal{R}}_i(k)$
6: **remove** $\min\{\hat{f}, |\underline{\mathcal{R}}_i(k)|\}$ nodes with smallest values from $\underline{\mathcal{R}}_i(k)$

Output: Filtered neighbor set $\mathcal{Q}_i(k)$, filtered states $\hat{x}_j(k)$

high scalability. Furthermore, label verification operates independently of the f-local bound, completely filtering cloned Sybil nodes even if adversaries exceed f and compromise overall consensus.

The control input $u_i(k)$ can be computed based on the filtered state values:

$$u_i(k) = \mu \sum_{j\in\mathcal{Q}_i(k)} a_{ij}(k)(\hat{x}_j(k) - \text{sign}(a_{ij}(k))x_i(k)) \tag{6}$$

where $\mu > 0$ is a constant.

From (3) and (6), we get

$$x_i(k+1) = x_i(k) - \mu \sum_{j\in\mathcal{Q}_i(k)} a_{ij}(k)\text{sign}(a_{ij}(k))x_i(k) + \mu \sum_{j\in\mathcal{Q}_i(k)} a_{ij}(k)\hat{x}_j(k) \tag{7}$$

Remark 5: Conventionally, differential privacy noise injection entails the injection of noise following identical distributions into all information-transmitting nodes, regardless of whether nodes are cooperative or competitive [18]. Although this approach effectively safeguards the privacy of system information, it imposes deleterious impacts on the bipartite consensus dynamics, notably impairing the convergence rate. In this paper, we propose injecting noise with distinct distributions into competitive and cooperative nodes, respectively. This strategy accelerates convergence because allocating distinct noise levels reduces the overall perturbation variance within the network, allowing agents to track true state values more smoothly and precisely while preserving privacy.

III. Main Results

A. Resilient Bipartite Consensus

As established in the preceding analysis, the process of valid neighbor value selection in Algorithm 1 involves retaining only a subset of neighbor values for state updates. Specifically, at each time instant k, the algorithm prunes certain edges from the initial communication graph $\mathcal{G}$. Let the resulting communication graph after edge pruning be denoted as $\mathcal{G}'$. The following lemma establishes that $\mathcal{G}'$ maintains structural balance.

Lemma 1 [22]: The reduced signed digraph $\mathcal{G}'$ remains structurally balanced provided that the original digraph $\mathcal{G}$ is structurally balanced.

Notably, while the aforementioned algorithm filters out state values of multiple malicious nodes based on the magnitude of received state values, there remains a possibility that malicious nodes could be admitted into the effective neighbor set $\mathcal{Q}_i(k)$. This arises because neighbor values might lie within the interval defined by the maximum and minimum initial state values of normal agents. To further mitigate the impact of malicious nodes, we put forward the following lemma.

Lemma 2 [23]: If a graph $\mathcal{G}$ is r-robust, then the graph $\mathcal{G}'$ obtained by removing at most $r-1$ incoming edges from each node in $\mathcal{G}$ still contains a spanning tree.

In Algorithm 1, outlier filtering is performed by eliminating the extreme maximum and minimum values. Since Byzantine nodes disrupt consensus by transmitting arbitrary state values that violate the control protocol, the filtered set $\mathcal{Q}_i(k)$ may still contain residual malicious nodes.

As known from [23], when the graph satisfies $(2f+1)$-robust, the value of any malicious node in the set $\mathcal{Q}_i(k)$ can be expressed as a convex combination of the state values of normal agents in $\mathcal{N}_i^+ \cap \mathcal{V}_N$. Thus, we present the following lemma:

Lemma 3 [23]: Consider the MASs operating under the f-local attack model. Provided that the communication topology $\mathcal{G}$ is $(2f+1)$-robust, the state update law for any normal agent $i \in \mathcal{V}_N$ can be equivalently expressed as:

$$\begin{aligned} x_i(k+1) &= x_i(k) - \mu \sum_{j\in\mathcal{N}_i^+\cap\mathcal{V}_N} \bar{a}_{ij}(k)(\text{sign}(\bar{a}_{ij}(k))x_i(k) - x_j(k)) \\ &\quad + \frac{\mu}{2} \sum_{j\in\mathcal{N}_i^+\cap\mathcal{V}_N} \bar{a}_{ij}(k)(1 - \text{sign}(\bar{a}_{ij}(k)))\eta_j(k) \\ &\quad + \frac{\mu}{2} \sum_{j\in\mathcal{N}_i^+\cap\mathcal{V}_N} \bar{a}_{ij}(k)(1 + \text{sign}(\bar{a}_{ij}(k)))\theta_j(k) \end{aligned} \tag{8}$$

where the non-negative weight $\bar{a}_{ij}(k)$ corresponds to the adjusted weighting factor assigned to the state values of the remaining normal nodes, $\bar{a}_{ii}(k)$ represents the adjusted weight assigned by agent i to its own state value, satisfying:

$$\bar{a}_{ii}(k) + \mu \sum_{j\in\mathcal{N}_i^+\cap\mathcal{V}_N} \bar{a}_{ij}(k) = 1,\ \forall k \in \mathbb{N}$$

Subsequently, we proceed to prove the resilient bipartite consensus property of the algorithm. To formalize the proof framework, the following definitions and lemmas are herein presented.

Definition 6 (Mean-Square Bipartite Consensus): For MASs (3) affected by Sybil nodes, normal nodes are considered capable of achieving mean-square bipartite consensus if, for any initial values, the following conditions are met.

$$\lim_{k\to\infty} \mathbb{E}[|x_i(k) - s_i x^*|]^2 = 0,\ s_i \in \{\pm 1\},\ \forall i \in \mathcal{V}_N$$

Lemma 4 [24]: The matrix $U \in Z_n$ is a non-singular M-matrix if and only if one can find a positive diagonal matrix $\Phi = \text{diag}\{\phi_1, \cdots, \phi_n\}$ for which the symmetric matrix $\Omega = \Phi U + U^T\Phi > 0$ is positive definite.

Theorem 1: If the topological structure is $(2f+1)$- robust and structurally balanced, the normal agents (3) can achieve bipartite consensus when $0 < \mu < \dfrac{\delta}{r_{max}\|\mathcal{L}'_S\|^2}$, where $r_{max} = max\{r_1, r_2, \ldots, r_{N-N_A}\}$, $\mathcal{L}'_S = S\mathcal{L}'S$, $\delta > 0$.

Proof: Suppose the set of the normal agents is $\{1, 2, \ldots, N-N_A\}$, define $X(k) = [x_1^T(k), x_2^T(k), \ldots, x_{N-N_A}^T(k)]^T$, $\eta(k) = [\eta_1^T(k), \eta_2(k), \ldots, \eta_{N-N_A}^T(k)]^T$, $\theta(k) = [\theta_1^T(k), \theta_2^T(k), \ldots, \theta_{N-N_A}^T(k)]^T$, $\gamma = 1 - \dfrac{\delta}{r_{max}}\mu + \mu^2\|\mathcal{L}'_S\|^2$, $\lambda_1 = \dfrac{\mu^2}{4}r_{max}\|\bar{\mathcal{A}}\|^2$, $\lambda_2 = \dfrac{\mu^2}{4}r_{max}\|\widetilde{\mathcal{A}}\|^2$. Then the system (8) can be written in matrix form as follows

$$X(k+1) = X(k) - \mu\mathcal{L}'X(k) + \frac{\mu}{2}(\mathcal{A}' - |\mathcal{A}'|)\eta(k) + \frac{\mu}{2}(\mathcal{A}' + |\mathcal{A}'|)\theta(k) \tag{9}$$

where $\mathcal{A}'$ and $\mathcal{L}'$ denote the adjacency matrix and Laplacian matrix of the reduced signed digraph $\mathcal{G}'$, and $|\mathcal{A}'|$ represents the matrix obtained by taking the absolute value of all elements in $\mathcal{A}'$.

Let $Z(k) = SX(k)$ and $\mathcal{L}'_S = S\mathcal{L}'S$, note that $SS^{-1} = I_{N-N_A}$, then (9) can be rewritten as follows.

$$\begin{aligned} &Z(k+1) \\ &= (I_{N-N_A} - \mu\mathcal{L}'_S)Z(k) + \frac{\mu}{2}S(\mathcal{A}' - |\mathcal{A}'|)\eta(k) \\ &\quad + \frac{\mu}{2}S(\mathcal{A}' + |\mathcal{A}'|)\theta(k) \end{aligned} \tag{10}$$

According to Lemma 1 in [18], there exists a vector $\boldsymbol{r} = [r_1, \ldots, r_n]^T$ such that $\boldsymbol{r}^T\mathbf{1}_N = 1$, $\boldsymbol{r}^T\mathcal{L}'_S = \mathbf{0}_N^T$ and $(I_N - \mathbf{1}_N\boldsymbol{r}^T)\mathcal{L}'_S = \mathcal{L}'_S(I_N - \mathbf{1}_N\boldsymbol{r}^T) = \mathcal{L}'_S$. Let $J = \left(\frac{1}{N-N_A}\right)\mathbf{1}_{N-N_A}\mathbf{r}_{N-N_A}^T$, $\zeta(k) = (I_{N-N_A} - J)Z(k)$, $V(k) = \zeta(k)^T R\zeta(k)$, where $\mathbf{r}_{N-N_A} = [r_1, \ldots, r_{N-N_A}]^T$, $R = \text{diag}\{r_1, r_2, \ldots, r_{N-N_A}\}$, $\sum_{i=1}^{N-N_A} r_i = 1, i = 1, 2, \ldots, N-N_A, r_i \geq 0$. Then we can get

$$\begin{aligned} &\zeta(k+1) = (I_{N-N_A} - J)Z(k+1) \\ &= (I_{N-N_A} - \mu\mathcal{L}'_S)\zeta(k) + \frac{\mu}{2}(I_{N-N_A} - J)S(\mathcal{A}' - |\mathcal{A}'|)\eta(k) \\ &\quad + \frac{\mu}{2}(I_{N-N_A} - J)S(\mathcal{A}' + |\mathcal{A}'|)\theta(k) \end{aligned} \tag{11}$$

For the convenience of description, denote $(I_{N-N_A} - J)S(\mathcal{A}' - |\mathcal{A}'|)$ and $(I_{N-N_A} - J)S(\mathcal{A}' + |\mathcal{A}'|)$ as $\bar{\mathcal{A}}$ and $\widetilde{\mathcal{A}}$, respectively.

Since $\zeta(k)$, $\eta(k)$ and $\theta(k)$ are mutually independent, from (11), it follows that

$$\begin{aligned} &\mathbb{E}[V(k+1)] = \mathbb{E}[\zeta^T(k+1)R\zeta(k+1)] \\ &= \mathbb{E}[\zeta(k)^T R\zeta(k)] - \mu\mathbb{E}[\zeta^T(k)(R\mathcal{L}'_S + \mathcal{L}'^T_S R)\zeta(k)] \\ &\quad + \mu^2\mathbb{E}[\zeta^T(k)\mathcal{L}'^T_S R\mathcal{L}'_S\zeta(k)] + \frac{\mu^2}{4}\mathbb{E}[\eta^T(k)\bar{\mathcal{A}}^T R\bar{\mathcal{A}}\eta(k)] \\ &\quad + \frac{\mu^2}{4}\mathbb{E}[\theta^T(k)\widetilde{\mathcal{A}}^T R\widetilde{\mathcal{A}}\theta(k)] \end{aligned} \tag{12}$$

Since the topological structure is $(2f+1)$-robust, according to Lemma 4, there exists a constant $\delta > 0$ such that $R\mathcal{L}'_S + \mathcal{L}'^T_S R \geq \delta I_{N-N_A}$. Then we can get

$$\zeta^T(k)(R\mathcal{L}'_S + \mathcal{L}'^T_S R)\zeta(k) \geq \delta\zeta^T(k)\zeta(k) \geq \frac{\delta}{r_{max}}V(k) \tag{13}$$

According to (12) and (13), we can derive

$$\begin{aligned} &\mathbb{E}[V(k+1)] \\ &\leq \gamma\mathbb{E}[V(k)] + \lambda_1\mathbb{E}[\eta(k)^T\eta(k)] + \lambda_2\mathbb{E}[\theta(k)^T\theta(k)] \end{aligned} \tag{14}$$

According to the conditions of Theorem 1, $0 < \mu < \dfrac{\delta}{r_{max}\|\mathcal{L}'_S\|^2}$, then we can derive that $0 < \gamma < 1$. Then we obtain that the first term of equation (14) converges to 0 as $k \to \infty$. What is more, $\mathbb{E}[\eta(k)^T\eta(k)] = \mathbb{E}\left[\sum_{i=1}^{N-N_A}\eta_i^T(k)\eta_i(k)\right] = \mathbb{V}\left[\sum_{i=1}^{N-N_A}\eta_i^T(k)\eta_i(k)\right] = 2\sum_{i=1}^{N-N_A}c_i^2q_i^{2k} \to 0, k \to \infty$. Similarly, we can obtain that $\mathbb{E}[\theta(k)^T\theta(k)] \to 0, k \to \infty$. In conclusion, $\mathbb{E}[V(k)] \to 0$, as $k \to \infty$.

Since $\mathbf{r}_{N-N_A}^T\mathcal{L}'_S = \mathbf{0}_{N-N_A}^T$, and there is

$$\begin{aligned} &\mathbf{r}_{N-N_A}^T Z(k+1) \\ &= \mathbf{r}_{N-N_A}^T Z(k) + \frac{\mu}{2}\mathbf{r}_{N-N_A}^T S(\mathcal{A}' - |\mathcal{A}'|)\eta(k) \\ &\quad + \frac{\mu}{2}\mathbf{r}_{N-N_A}^T S(\mathcal{A}' + |\mathcal{A}'|)\theta(k) \end{aligned} \tag{15}$$

Let $Z(k) = [z_1^T(k), z_2(k)^T, \ldots, z_{N-N_A}^T(k)]^T$. Continuously expanding and iterating the above-mentioned equation leads to

$$\begin{aligned} &\mathbf{r}_{N-N_A}^T Z(k) \\ &= \sum_{i=1}^{N-N_A} r_i z_i(0) + \frac{\mu}{2}\sum_{l=0}^{k-1}\sum_{i,j=1}^{N-N_A} r_j s_j \tilde{a}_{ij}(l)\eta_j(l) \\ &\quad + \frac{\mu}{2}\sum_{l=0}^{k-1}\sum_{i,j=1}^{N-N_A} r_j s_j \tilde{a}_{ij}(l)\theta_j(l) \end{aligned} \tag{16}$$

where $\tilde{a}_{ij}(l) = a_{ij}(l) - |a_{ij}(l)|$.

Furthermore, when $k \to \infty$, we can obtain

$$\begin{aligned} \lim_{k\to\infty}\mathbf{r}^T Z(k) &= \sum_{i=1}^{N-N_A} r_i z_i(0) + \frac{\mu}{2}\sum_{l=0}^{\infty}\sum_{i,j=1}^{N-N_A} r_j s_j \tilde{a}_{ij}(l)\eta_j(l) \\ &\quad + \frac{\mu}{2}\sum_{l=0}^{\infty}\sum_{i,j=1}^{N-N_A} r_j s_j \tilde{a}_{ij}(l)\theta_j(l) \end{aligned} \tag{17}$$

Since $\mathbb{E}[V(k)] \to 0, k \to \infty$, we derive

$$\begin{aligned} &\lim_{k\to\infty}\mathbb{E}[s_i x_i(k) - x^*]^2 \\ &\leq \lim_{k\to\infty}\mathbb{E}\left[s_i x_i(k) - \frac{1}{N-N_A}\mathbf{r}^T Z(k)\right]^2 \\ &\quad + \lim_{k\to\infty}\mathbb{E}\left[\frac{1}{N-N_A}\mathbf{r}^T Z(k) - x^*\right]^2 = 0 \end{aligned} \tag{18}$$

where

$$\begin{aligned} x^* &= \frac{1}{N-N_A}\sum_{i=1}^{N-N_A} r_i z_i(0) \\ &\quad + \frac{\mu}{2(N-N_A)}\sum_{l=0}^{\infty}\sum_{i,j=1}^{N-N_A} r_j s_j \tilde{a}_{ij}(l)\left(\eta_j(l) + \theta_j(l)\right) \end{aligned} \tag{19}$$

Moreover, since $\eta_i(k)$ and $\theta_i(k)$, $i \in \mathcal{V}_N$ are independent and identically distributed (i.i.d.), it follows that

$$\mathbb{E}(x^*) = \frac{1}{N - N_A} \sum_{i=1}^{N-N_A} r_i s_i x_i(0) \tag{20}$$

The proof is completed.

B. Privacy Analysis

In this subsection, we analyze the privacy level of the proposed algorithm. First, the definition of sensitivity will be given as follow.

Definition 7 (Sensitivity): The sensitivity of a randomized mechanism $\mathcal{H}(\cdot)$ at time $k \in \mathbb{N}$ is defined for any adjacent datasets D and D' by the following equation:

$$\Delta(k) = \sum_{D,D' \in \mathcal{D}} \|x_i(k) - x_i'(k)\|_1 \tag{21}$$

Following [25], the following lemma is given.

Lemma 5: For the given adjacency relation (7), the sensitivity of the designed randomized mechanism $\mathcal{H}(\cdot)$ at time instant $k \in \mathbb{N}$ satisfies

$$\Delta(k) = \sigma(1 - \mu\phi_{min})^k \tag{22}$$

where $\phi_i = \sum_{\mathcal{N}_i^+ \cap \mathcal{V}_N} |a_{ij}(k)|$.

Proof: Based on (8), we can derive

$$x_i(k+1) = (1 - \mu\phi_i)x_i(k) + \mu \sum_{\mathcal{N}_i^+ \cap \mathcal{V}_N} \bar{a}_{ij}(k)\hat{x}_j(k) \tag{23}$$

Similarly, for dataset D', we can obtain

$$x_i'(k+1) = (1 - \mu\phi_i)x_i'(k) + \mu \sum_{\mathcal{N}_i^+ \cap \mathcal{V}_N} \bar{a}_{ij}(k)\hat{x}_j(k) \tag{24}$$

Given adjacent datasets $D = \{x_i(0), i \in \mathcal{V}_N\}$ and $D' = \{x_i'(0), i \in \mathcal{V}_N\}$, we respectively define $\mathcal{H}(D) = \{\hat{x}_i(k), i \in \mathcal{V}_N\}$ and $\mathcal{H}(D') = \{\hat{x}_i'(k), i \in \mathcal{V}_N\}$ to represent the mappings from private information to observation sequences. Since the observation sequences of D and D' are the same, according to (23) and (24), we can obtain

$$x_i'(k) - x_i(k) = (1 - \mu\phi_i)^k (x_i'(0) - x_i(0)) \tag{25}$$

Then we have that for $k \geq 1$

$$\begin{aligned} \|x_i(k) - x_i'(k)\|_1 &= \sum_{i \in \mathcal{V}_N} (1 - \mu\phi_i)^k |x_i'(0) - x_i(0)| \\ &\leq \sigma(1 - \mu\phi_{min})^k \end{aligned} \tag{26}$$

The proof is completed.

Theorem 2: For system (3), when the system state satisfies the σ-adjacency relation, the proposed algorithm can achieve ϵ_i-differential privacy, and the privacy protection level is

$$\epsilon_i = \epsilon_i^\eta + \epsilon_i^\theta \tag{27}$$

where $\epsilon_i^\eta = \dfrac{\sigma(1 - \mu\phi_{\min})}{c_{i_0}(q_{i_0} - 1 + \mu\phi_{\min})}$, $\epsilon_i^\theta = \dfrac{\sigma(1 - \mu\phi_{\min})}{\bar{c}_{i_0}(q_{i_0} - 1 + \mu\phi_{\min})}$.

Proof: Given adjacent datasets D and D', define $\mathcal{P} = \{\rho(D, \mathcal{H}(D))\}$ and $\mathcal{P}' = \{\rho(D', \mathcal{H}(D'))\}$ as the sets of possible state evolution sequences, which have corresponding probability density functions $f(D, \rho(D, \mathcal{H}(D))$ and $f(D', \rho(D', \mathcal{H}(D'))$ in the observation set $\mathcal{O}$, respectively. From this, the probability density function of $\mathcal{H}(D)$ can be derived as follows:

$$\begin{aligned} &f(D, \rho(D, \mathcal{H}(D))) \\ &= f(\hat{x}_i(0), \hat{x}_i(1), \ldots) \\ &= \prod_{k \in \mathbb{N}} f(\hat{x}_i(k) | \hat{x}_i(0), \hat{x}_i(1), \ldots, \hat{x}_i(k-1)) \\ &= \prod_{k \in \mathbb{N}} \left[\prod_{i \in \mathcal{V}_\eta} f_\eta(\hat{x}_i(k) - x_i(k)) \prod_{i \in \mathcal{V}_\theta} f_\theta(\hat{x}_i(k) - x_i(k)) \right] \end{aligned} \tag{28}$$

where $\mathcal{V}_\eta = \{i \in \mathcal{V}_N | (i,j) \in \mathcal{E}_\eta\}$ and $\mathcal{V}_\theta = \{i \in \mathcal{V}_N | (i,j) \in \mathcal{E}_\theta\}$ represent the competitive edges and cooperative edges of agent i respectively. And f_η and f_θ respectively correspond to the Laplacian noise probability density functions of competitive edges and cooperative edges.

Since the observations at time T are the same, there exists a bijective function $u(\cdot)$: $\mathcal{P} \to \mathcal{P}'$, such that for any pair $\rho(D, \mathcal{H}(D)) \in \mathcal{P}$ and $\rho(D', \mathcal{H}(D')) \in \mathcal{P}'$, $u(\rho(D, \mathcal{H}(D))) = \rho(D', \mathcal{H}(D'))$. Thus, we can derive

$$\begin{aligned} &\frac{\mathbb{P}[\mathcal{H}(D) \in \mathcal{O}]}{\mathbb{P}[\mathcal{H}(D') \in \mathcal{O}]} \\ &= \lim_{k \to \infty} \frac{\int_{\rho(D, \mathcal{H}(D)) \in \mathcal{P}} f(\rho(D, \mathcal{H}(D)))d\tau}{\int_{\rho(D', \mathcal{H}(D')) \in \mathcal{P}'} f(\rho(D', \mathcal{H}(D')))d\tau} \\ &= \lim_{k \to \infty} \frac{\int_{\rho(D, \mathcal{H}(D)) \in \mathcal{P}} f(\rho(D, \mathcal{H}(D)))d\tau}{\int_{\rho(D, \mathcal{H}(D)) \in \mathcal{P}} f(u(\rho(D, \mathcal{H}(D))))d\tau} \end{aligned} \tag{29}$$

As can be seen from (23) and (24), the observation sequence depends on D, D' and the noise sequence. Then, according to (29) and the properties of Laplacian noise, we can obtain the following equation.

$$\begin{aligned} &\frac{f(D, \rho(D, \mathcal{H}(D)))}{f(D', \rho(D', \mathcal{H}(D')))} \\ &= \frac{\prod_{k \in \mathbb{N}} \left[\prod_{i \in \mathcal{V}_\eta} f_\eta(\hat{x}_i(k) - x_i(k)) \prod_{i \in \mathcal{V}_\theta} f_\theta(\hat{x}_i(k) - x_i(k)) \right]}{\prod_{k \in \mathbb{N}} \left[\prod_{i \in \mathcal{V}_\eta} f_\eta(\hat{x}_i(k) - x_i'(k)) \prod_{i \in \mathcal{V}_\theta} f_\theta(\hat{x}_i(k) - x_i'(k)) \right]} \\ &\leq \prod_{k \in \mathbb{N}} \prod_{i \in \mathcal{V}_\eta} \exp\left(\frac{|x_i(k) - x_i'(k)|}{b_i(k)} \right) \prod_{i \in \mathcal{V}_\theta} \exp\left(\frac{|x_i(k) - x_i'(k)|}{\bar{b}_i(k)} \right) \\ &\leq \exp\left(\sum_{l=0}^{k} \frac{\sigma(1 - \mu\phi_{min})^l}{c_{i_0} q_{i_0}^l} + \sum_{l=0}^{k} \frac{\sigma(1 - \mu\phi_{min})^l}{\bar{c}_{i_0} q_{i_0}^l} \right) \end{aligned} \tag{30}$$

Integrate both sides with respect to the probability density function, and let $k \to \infty$, then we can obtain

$$\mathbb{P}[\mathcal{H}(D) \in \mathcal{O}] \leq e^{\left(\epsilon_{i_0}^\eta + \epsilon_{i_0}^\theta\right)} \mathbb{P}[\mathcal{H}(D') \in \mathcal{O}] \tag{31}$$

where i_0 can be any normal agent in the entire system.

The proof is completed.

C. Accuracy Analysis

In this subsection, the calculation of the (p, r)-accuracy of the algorithm proposed in this paper will be introduced.

Definition 8: Given $p \in [0, 1]$, $r \in \mathbb{R}_{\geq 0}$, and any initial state vector $x \in \mathbb{R}^N$ that satisfies the σ - adjacency relation. When $k \to \infty$, if the system state $x(k)$ converges to a random

variable x^* such that $\mathbb{E}[x^*] < \infty$ and $\mathbb{P}\{|x^* - \mathbb{E}[x^*]| \le r\} \ge 1 - p$ hold, then the random algorithm $\mathcal{M}$ is said to satisfy (p, r)-accuracy.

Theorem 3: The algorithm proposed in this paper can achieve (p, r)-accuracy, where

$$r = \sqrt{\frac{\mu^2\Phi^2 r_{max}^2}{2(N-N_A)^2}\sum_{j=1}^{N-N_A}\frac{c_j^2+\bar{c}_j^2}{1-q_j^2}}\Big/\sqrt{p},$$

$\Phi = \tilde{a}_{\max}\phi_{\max}$ and $\tilde{a}_{\max} = \max_{i,j}\{\tilde{a}_{ij}(k)\}$.

Proof: According to the expression of x^* in (19), we have

$$\begin{aligned}\mathbb{V}\left[x^*\right] = &\mathbb{V}\left[\frac{\mu}{2(N-N_A)}\sum_{l=0}^{\infty}\sum_{i,j=1}^{N-N_A} r_j s_j \tilde{a}_{ij}(l)\eta_j(l)\right]\\ &+\mathbb{V}\left[\frac{\mu}{2(N-N_A)}\sum_{l=0}^{\infty}\sum_{i,j=1}^{N-N_A} r_j s_j \tilde{a}_{ij}(l)\theta_j(l)\right]\end{aligned} \tag{32}$$

Utilize the relationship between variance and expectation, it follows that

$$\begin{aligned}&\mathbb{V}\left[\frac{\mu}{2(N-N_A)}\sum_{l=0}^{\infty}\sum_{i,j=1}^{N-N_A} r_j s_j \tilde{a}_{ij}(l)\eta_j(l)\right]\\ &= \mathbb{E}\left[\left(\frac{\mu}{2(N-N_A)}\sum_{l=0}^{\infty}\sum_{i,j=1}^{N-N_A} r_j s_j \tilde{a}_{ij}(l)\eta_j(l)\right)^2\right]\\ &= \frac{\mu^2}{4(N-N_A)^2}\mathbb{E}\left[\sum_{l=0}^{\infty}\sum_{j=1}^{N-N_A} r_j^2 s_j^2 \tilde{a}_{jj}(l)^2\eta_j(l)^2\right]\\ &\le \frac{\mu^2\Phi^2 r_{max}^2}{2(N-N_A)^2}\sum_{j=1}^{N-N_A}\frac{c_j^2}{1-q_j^2}\end{aligned} \tag{33}$$

Similarly, we have

$$\begin{aligned}&\mathbb{V}\left[\frac{\mu}{2(N-N_A)}\sum_{l=0}^{\infty}\sum_{i,j=1}^{N-N_A} r_j s_j \tilde{a}_{ij}(l)\theta_j(l)\right]\\ &\le \frac{\mu^2\Phi^2 r_{max}^2}{2(N-N_A)^2}\sum_{j=1}^{N-N_A}\frac{\bar{c}_j^2}{1-q_j^2}\end{aligned} \tag{34}$$

According to Chebyshev's inequality, we can obtain

$$\mathbb{P}\{|x^* - \mathbb{E}[x^*]| \le r\} \ge 1 - \frac{\mathbb{V}[x^*]}{r^2} \tag{35}$$

Furthermore, let $r = \sqrt{\frac{\mu^2\Phi^2 r_{max}^2}{2(N-N_A)^2}\sum_{j=1}^{N-N_A}\frac{c_j^2+\bar{c}_j^2}{1-q_j^2}}/\sqrt{p}$, then we have $\mathbb{P}\{|x^*-\mathbb{E}[x^*]| \le r\} \ge 1-p$. That is to say, the algorithm proposed in this paper can satisfy the (p, r)-accuracy.

IV. Simulation

In this section, we verify the effectiveness of the proposed algorithm through numerical simulations.

The communication topology graph is shown in Fig. 1. The blue solid lines represent the cooperative interactions within the group, and the red dashed lines represent the competitive interactions between groups. The graph $\mathcal{G}$ is structurally balanced and thus admits a bipartition into two disjoint node subsets, where $\mathcal{V}_1 = \{1, 3, 5, 7, 9\}$ and $\mathcal{V}_2 = \{2, 4, 6, 8, 10\}$. The graph is 3-robust and under 1-local attack. Let node 11 be the Sybil child node of agent 9, and agent 10 be the Byzantine agent respectively. Neither of them follows the normal state update law, and agent 10 sends the message $\hat{x}_{10}(k) = -2 + 6\cos(0.25k) + \hat{\eta}_{10}(k)$, and the node 11 sends the message $\hat{x}_{11}(k) = 6$ with the same label $L_9(k)$, where $\hat{\eta}_j(k) \sim Lap(0, \hat{c}_j\hat{q}_j^k)$, $\hat{c}_j = 0.9$, $\hat{q}_j = 0.95$. The initial states $x(0) = [-4, -1.5, 2.58, -3.89, 3, -4.67, 2.78, -3.66, 5, -2]^T$. According to the convergence condition of Theorem 1, we set the control parameters $\mu = 0.3$, $c_i = 0.95$, $\bar{c}_i = 0.1$, $q_i = 0.85$.

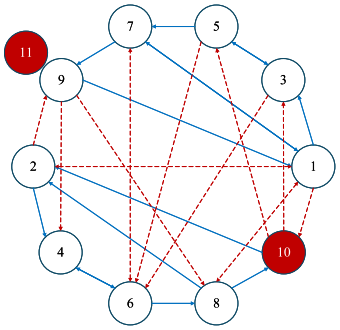


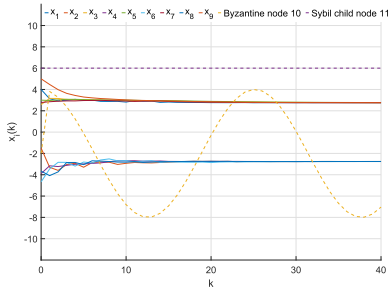
Fig. 1. Structurally balanced signed graph $\mathcal{G}$ containing a spanning tree.

As evident from the outlier filtering step in Algorithm 1, only the state values of neighbors retained in the set $Q_i(k)$ are utilized for updating the state of normal agent i. Consequently, each element $a_{ij}(k)$ in the adjacency matrix A takes a value of 0 or 1, determined solely by whether neighbor j is included in agent i's filtered neighbor set $Q_i(k)$. From Fig. 1, the adjacency matrix A is constructed based on the topology in Fig. 1 with edge weights ± 1.

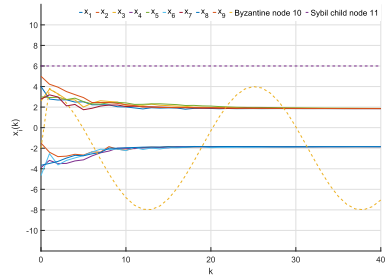
As demonstrated in Fig. 2 and Fig. 3, the proposed method in Theorem 1 mitigates the impact of Sybil attacks and Byzantine attacks, ensuring that the agent states achieve mean-square bipartite consensus. Moreover, Fig. 2 reveals that, in contrast to the uniform noise injection approach in [18], injecting another kind of noise with a lower intensity on the basis of the original noise can effectively accelerate the convergence speed of the state values.

To demonstrate the characteristics of differential privacy, we select the σ-adjacent initial conditions $x(0) = [-4, -1.5, 2.58, -3.89, 3, -4.67, 2.78, -3.66, 5, -2]^T$ and $x'(0) = [-4, -1.5, 2.58, -3.89, 3, -4.67, 2.78, -3.66, 5, -2]^T$. This implies that only the internal state of the first agent differs between $x(0)$ and $x'(0)$. That is to say, $\sigma = 1$. Based on the σ-adjacent initial conditions $x(0)$ and $x'(0)$, the resulting observation sequences among the agents are shown in Fig. 4, which demonstrates that individual privacy is effectively protected, as a significant deviation exists between any original data point and its corresponding protected data point. Injecting tailored Laplace noise prevents adversaries from accurately inferring the true values of the original sensitive state $x_i(k)$ from the transmitted state $\hat{x}_i(k)$, thereby achieving the desired privacy preservation.

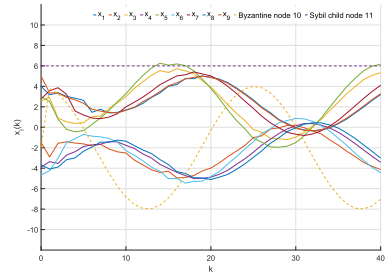
To rigorously assess the resilience of our proposed algorithm against Sybil attacks, we construct a challenging



(a) State trajectories $x_i(k)$ with the different noise injection



(b) State trajectories $x_i(k)$ with the same noise injection



(c) State trajectories $x_i(k)$ without the resilient algorithm

Fig. 2. State trajectories $x_i(k)$.

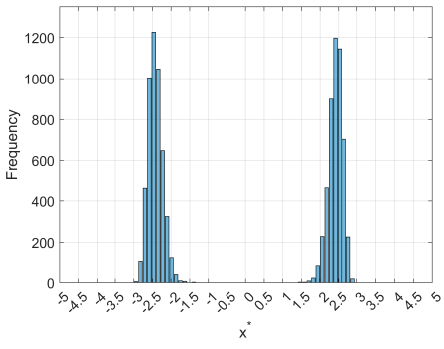


Fig. 3. Histogram of the convergence points.

adversarial scenario within the baseline network topology shown in Fig. 1. Nodes 9 and 10 are assumed to be fully compromised by the adversary. Node 10 is elevated from a standard Byzantine node to a Sybil parent node. Node 9 generates replicas 9a, 9b, 9c. Node 10 generates replicas 10a, 10b, 10c. With all other parameters remaining unchanged, the initial

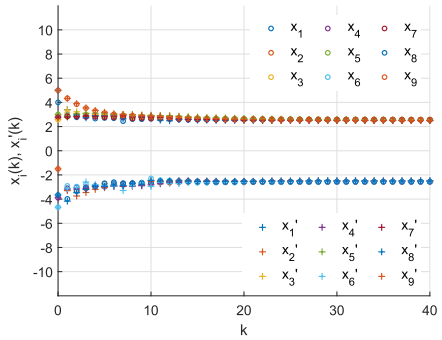


Fig. 4. Observation sequences $x_i(k), x_i'(k)$ corresponding to the σ-adjacent $x_i(0), x_i'(0)$.

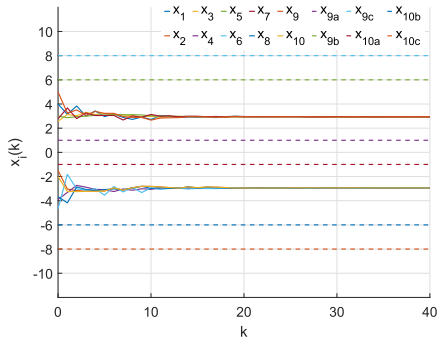


Fig. 5. State trajectories of normal agents $x_i(k)$ with a high proportion of Sybil nodes.

state vector is set to $x(0) = [4,\text{-}1.5, 2.58,\text{-}3.89, 3,\text{-}4.67, 2.78,\text{-}3.66, 5,\text{-}2]^T$. All Sybil child nodes are configured to transmit constant extreme values to their neighbors: $x_{9a}(k) \equiv 1$, $x_{9b}(k) \equiv 6$, $x_{9c}(k) \equiv 8$, $x_{10a}(k) \equiv -1$, $x_{10b}(k) \equiv -6$, $x_{10c}(k) \equiv -8$. As shown in Fig. 5, the MASs successfully achieve bipartite consensus even under intense Sybil attacks, validating the effectiveness of the bipartite consensus algorithm proposed in this paper.

V. Conclusion

This paper addresses the problem of bipartite consensus in MASs under Sybil attacks and Byzantine attacks, introducing an innovative analytical framework. When the directed communication topology satisfies the $(2f+1)$-robust and f-local attack conditions, the proposed resilient bipartite consensus algorithm effectively mitigates the impact of Sybil nodes and Byzantine nodes on system consensus. Notably, injecting distinct differential privacy noise between different agent groups offers greater flexibility in adjusting the privacy level according to the context, while simultaneously preserving system privacy, compared to traditional noise injection methods. By reducing the noise intensity for specific groups, the system achieves accelerated bipartite consensus. Future research will focus on extending the proposed resilient bipartite consensus framework to nonlinear MASs, investigating dynamic and time-varying privacy budget allocation strategies to further optimize the trade-off between security, privacy, and system performance.

REFERENCES

[1] P. Shi and B. Yan, "A survey on intelligent control for multiagent systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 161–175, Jan. 2021.

[2] S. Arunima and B. Subudhi, "A multiagent-based distributed flocking controller for an AC microgrid," *IEEE Trans. Consum. Electron.*, vol. 70, no. 3, pp. 5229–5237, Aug. 2024.

[3] K. Li, Y. Hu, J. Wang, L. Li, S. Zhang, and G. Luo, "Task prioritization in multiagent environments: A novel approach using Nash Q-learning," *IEEE Trans. Consum. Electron.*, vol. 71, no. 2, pp. 4206–4220, May 2025.

[4] S. Jiao, Q. Wei, and F.-Y. Wang, "A novel parallel control method for optimal consensus of nonlinear multiagent systems," *IEEE Trans. Cybern.*, vol. 54, no. 10, pp. 5912–5925, Oct. 2024.

[5] W. Song, D. Zhao, L. Xu, X. Li, and F. Tong, "Distributed consensus of Lipschitz nonlinear multiagent networks under DoS attacks: A strict zeno-free event-triggered approach," *IEEE Trans. Consum. Electron.*, early access, Mar. 2, 2026, doi: 10.1109/TCE.2026.3669225.

[6] Y. Zhou, G. Wen, J. Zhou, and T. Yang, "Data-driven fault-tolerant bipartite consensus tracking for multi-agent systems with a non-autonomous leader," *IEEE/CAA J. Autom. Sinica*, vol. 12, no. 1, pp. 279–281, Jan. 2025.

[7] X. Song, X. Sun, S. Song, and S. Xu, "A novel event-triggered bipartite consensus for PDE-based multiagent systems with switching topologies and antagonistic interactions," *IEEE Trans. Cybern.*, vol. 54, no. 10, pp. 5759–5769, Oct. 2024.

[8] Y. Shang, Z. Liu, J. Kang, M. S. Hossain, and Y. Wu, "Adversarial attacks on vision-language model-empowered chatbots in consumer electronics," *IEEE Trans. Consum. Electron.*, vol. 70, no. 3, pp. 6075–6083, Aug. 2024.

[9] M. Sayad Haghighi, F. Farivar, A. Jolfaei, A. B. Asl, and W. Zhou, "Cyber attacks via consumer electronics: Studying the threat of covert malware in smart and autonomous vehicles," *IEEE Trans. Consum. Electron.*, vol. 69, no. 4, pp. 825–832, Nov. 2023.

[10] S. Dogan-Tusha, S. Althunibat, and M. Qaraqe, "Doppler-shift-based Sybil attack detection for mobile IoT networks," *IEEE Internet Things J.*, vol. 11, no. 1, pp. 1136–1147, Jan. 2024.

[11] Y. Wu, C. Ying, N. Zheng, W.-A. Zhang, and S. Zhu, "Whole-process privacy-preserving and Sybil-resilient consensus for multiagent networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 7, pp. 12027–12039, Jul. 2025.

[12] S. Zhang and J.-H. Lee, "Double-spending with a Sybil attack in the Bitcoin decentralized network," *IEEE Trans. Ind. Informat.*, vol. 15, no. 10, pp. 5715–5722, Oct. 2019.

[13] Y. Shang, "Resilient cluster consensus of multiagent systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 1, pp. 346–356, Jan. 2022.

[14] J. Liu, Z. Liu, Y. Li, E. Tian, and C. Peng, "Privacy-protected formation control for unmanned surface vehicles: A neural predictor-based non-cooperative game framework," *IEEE Trans. Veh. Technol.*, early access, Jan. 12, 2026, doi: 10.1109/TVT.2026.3651837.

[15] J. Liu, S. Mao, L. Zha, E. Tian, and C. Peng, "Privacy-preserving and collision-free formation control for heterogeneous multiagent systems," *IEEE Trans. Consum. Electron.*, early access, Feb. 24, 2026, doi: 10.1109/TCE.2026.3667772.

[16] J. Liu, E. Ma, L. Zha, and E. Tian, "Distributed energy management for microgrids: When privacy preservation meets multiple mechanisms," *IEEE Trans. Consum. Electron.*, vol. 72, no. 1, pp. 1514–1523, Feb. 2026.

[17] L. Gao, S. Deng, W. Ren, and C. Hu, "Differentially private consensus with quantized communication," *IEEE Trans. Cybern.*, vol. 51, no. 8, pp. 4075–4088, Aug. 2021.

[18] Z. Zuo, R. Tian, Q. Han, Y. Wang, and W. Zhang, "Differential privacy for bipartite consensus over signed digraph," *Neurocomputing*, vol. 468, pp. 11–21, Jan. 2021.

[19] C. Altafini, "Consensus problems on networks with antagonistic interactions," *IEEE Trans. Autom. Control*, vol. 58, no. 4, pp. 935–946, Apr. 2013.

[20] H. J. LeBlanc, H. Zhang, X. Koutsoukos, and S. Sundaram, "Resilient asymptotic consensus in robust networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 4, pp. 766–781, Apr. 2013.

[21] B. Turan, C. A. Uribe, H.-T. Wai, and M. Alizadeh, "Resilient primal–dual optimization algorithms for distributed resource allocation," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 1, pp. 282–294, Mar. 2021.

[22] X. Lu and Y. Jia, "Bipartite Byzantine-resilient event-triggered consensus control of heterogeneous multi-agent systems," *Int. J. Robust Nonlinear Control*, vol. 33, no. 1, pp. 282–310, 2022.

[23] S. Sundaram and B. Gharesifard, "Distributed optimization under adversarial nodes," *IEEE Trans. Autom. Control*, vol. 64, no. 3, pp. 1063–1076, Mar. 2019.

[24] W. Ren and Y. Cao, *Distributed Coordination of Multi-agent Networks: Emergent Problems, Models, and Issues*. Cham, Switzerland: Springer, 2011.

[25] Z. Huang, R. Hu, Y. Guo, E. Chan-Tin, and Y. Gong, "DP-ADMM: ADMM-based distributed learning with differential privacy," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1002–1012, 2020.