Accurate Resilient Average Consensus via Detection and Compensation

Wenzhe Zheng, Zhiyu He, Jianping He, and Chengcheng Zhao

Abstract— **We study the problem of resilient average consensus for multi-agent systems with misbehaving nodes. Different from the widely investigated Mean-Subsequence-Reduced-based and detection-isolation-based approaches which guarantee consensus, in this paper, we address this problem by detecting misbehaviors, mitigating corresponding impact and achieving accurate average consensus. General types of misbehaviors are considered, including deception attacks and accidental faults. We characterize the disturbances of misbehaving nodes in a distributed manner via two-hop communication information and develop a deterministic detection-compensation-based consensus (D-DCC) algorithm with a decaying fault-tolerant error bound. Considering scenarios where such information is intermittently available, a stochastic extension named S-DCC is proposed. We prove that D-DCC and S-DCC allow nodes to asymptotically achieve average consensus exactly and in expectation, respectively. Then, the Wasserstein distance is introduced to analyze the accuracy of S-DCC. Finally, extensive simulations verify the effectiveness of the proposed algorithms.**

I. INTRODUCTION

Consensus problems have attracted extensive interests due to their wide applications, e.g., distributed control, estimation [1] and optimization in robotic networks [2], smart grids [3] and wireless sensor networks [4]. The goal of consensus is to enable agents to reach global agreements via local exchanges of information. Under this framework, many interesting topics have been studied, including average consensus [5], consensus with noises [6] or packet drops [7], and the convergence rates [8].

Most of the aforementioned work assumes that all agents faithfully execute predefined protocols. Since distributed systems are usually deployed in open environments, there may exist vulnerabilities that are prone to failures or attacks, thus affecting the overall performance. Hence, it is essential to constrain negative impact to ensure desired agreements. Motivated by this issue, numerous efforts have been devoted to resilient consensus [9], [10], which are mainly divided into the following two categories. The first category is based on Mean-Subsequence Reduced (MSR) algorithms [11]. The main idea is to discard the extreme states of neighbors and update states only by remaining ones. An extended version named Weighted-Mean-Subsequence Reduced (W-MSR) algorithm is developed in [12], where nodes only discard the extreme states strictly larger or smaller than their own states. Based on W-MSR, LeBlanc *et al.* [9] develop a novel graph-theoretic property termed network robustness to characterize the resilience of W-MSR and analyze the sufficient and necessary conditions of resilient asymptotic consensus. Later, the quantized version of the W-MSR handling asynchrony and time-varying time delays is presented in [13]. In applications, Saulnier *et al.* [14] present a hybrid algorithm that enables resilient formation control for mobile robot teams in the presence of noncooperative robots. Similarly, Yan *et al.* [15] solve the resilient multi-dimensional consensus problem by a middle-point-based algorithm where these points are convex combinations at one dimension each time. On the one hand, the above algorithms ensure consensus among normal nodes without detecting malicious nodes. On the other hand, it is difficult to guarantee average consensus due to the straightforward rule of data processing. Also, most of them require graph connectivity strictly, e.g., $(F+1, F+1)$-robust for F-total malicious model and $(2F+1)$-robust for F-local malicious model [9], [13].

Another category is to detect and isolate malicious nodes. The main idea is designing detection algorithms to find malicious nodes once they occur and isolate them immediately. Observer-based techniques are effective ways to achieve detections. Pasqualetti *et al.* [10] solve the fault detection and isolation problem based on observations in almost any linear consensus network, where high connectivity and global knowledge of the network are required. To solve problems with more general attack models, mobile agents are exploited as observers [16]. Fault detection is also achieved by observer-based techniques for interconnected second-order systems [17]. The detection and mitigation in randomized gossiping algorithms based on observation of temporal or spatial difference are investigated in a system of data injection attacks [18]. Multiple communication protocols also contribute to fault detection. In [19], the authors combine two communication protocols, i.e., communication-based and sensing-based model, for fault detection. Based on majority voting, a detection scheme is designed by two-hop communication information for which the constraint on graph structures is less stringent than that for MSR algorithms [20]. He *et al.* [21] propose an enhanced secure consensus-based data aggregation (E-SCDA) algorithm by two-hop information from the aggregator which can preserve bounded average consensus with misbehaving nodes.

The above researches present effective methods to achieve consensus among normal nodes. Due to the considerations of being highly robust and lightweight, most of them cannot guarantee exact convergence to the average of initial values of agents. Specifically, MSR-based algorithms only ensure

W. Zheng, Z. He and J. He 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 200240, China. E-mail address: {wzzheng, hzy970920, jphe}@sjtu.edu.cn.

C. Zhao is with the State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China. Email address: chengchengzhao@zju.edu.cn.

Manuscript 1369 submitted to 2021 60th IEEE Conference on Decision and Control (CDC). Received March 25, 2021.

consensus within the range or convex hull of normal nodes' initial states, and detection-isolation-based algorithms do not eliminate disturbing effects caused by malicious nodes before isolations. Hence, the information of initial states may be polluted. Also, the latter category of algorithms might mistake faulty nodes for malicious ones when link failures or miscalculations occur accidentally, resulting in loss of information and system capacity. Hence, it is essential to design average consensus algorithms equipped with mechanisms to detect and compensate impact of misbehaving nodes and to provide tolerance for faulty nodes, at reasonable costs of storage, communication and computations. The main contributions of this paper are summarized as follows.

- The problem of resilient average consensus with misbehaving nodes is investigated. Not only detection and isolation, but also estimation and compensation are adopted to achieve accurate average consensus.
- By utilizing two-hop communication information, we design schemes to detect and compensate the errors accurately by neighboring normal nodes. Based on the schemes and fault-tolerant error bounds, two detection-compensation-based consensus (i.e., D-DCC and S-DCC) algorithms are proposed to cope with scenarios where information sets are constantly and intermittently available, respectively.
- We prove that average consensus is achieved by D-DCC exactly, and by S-DCC in expectation with less computation costs. We analyze the accuracy of S-DCC by Wasserstein distance, and present evaluation results of the proposed algorithms.

The rest of this paper is organized as follows. Section II introduces the models of networks and attacks. Section III presents the proposed algorithms consisting of the mechanisms of detection and compensation. Then, the effectiveness of the algorithms is shown in Section IV. In Section V, simulation results are provided. Finally, Section VI concludes the paper and discusses the future directions.

II. PRELIMINARIES AND PROBLEM FORMULATION

A. Network Model

Consider a network modeled as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertex set $\mathcal{V} = \{1, 2, \cdots, N\}$ and edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. Note that $(i, j) \in \mathcal{E}$ indicates node i and node j can communicate with each other. The neighbor set of node i is denoted by $\mathcal{N}_i = \{j | (i, j) \in \mathcal{E}\}$. The adjacency matrix is $A_\mathcal{G} = [a_{ij}]_{N\times N}$, and Laplacian matrix is $L = D_\mathcal{G} - A_\mathcal{G}$, where $D_\mathcal{G} = \text{diag}(d_1, \ldots, d_N)$ with $d_i = \sum_{j=1}^{N} a_{ij}$. Let $d_m = \max\{d_1, \ldots, d_N\}$. Let subset $\mathcal{V}_s = \{1, 2, \cdots, n\}$ and $\mathcal{V}_m = \{n+1, n+2, \cdots, N\}$ represent normal nodes set and misbehaving nodes set, respectively. It follows that $\mathcal{V}_s \cup \mathcal{V}_m = \mathcal{V}$ and $\mathcal{V}_s \cap \mathcal{V}_m = \phi$.

B. Consensus Algorithms

Let $x_i(k) \in \mathbb{R}$ be the state of node i at iteration k and $x(k) = [x_1(k), x_2(k), \cdots, x_N(k)]^\top$ be the vector of values. Consensus algorithms are distributed control protocols that drive all states to the same value, $x_i = c$, $\forall i \in \mathcal{V}$ [5]. The final value c is called consensus value. Particularly, if $c = 1/N \sum_{i=1}^{N} x_i(0)$, an average consensus is achieved. The basic discrete-time linear consensus is represented as

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

where W is a doubly stochastic matrix, i.e., both row stochastic and column stochastic. By (1), all nodes' states will exponentially converge to average consensus if $\mathcal{G}$ is a connected graph. The commonly used weight matrix guaranteeing asymptotic convergence includes Metropolis weights [22] and Perron matrix, i.e., $W = I - \gamma L$, where $\gamma < 1/d_m$.

C. Information Set and Attack Model

The update rule (1) implies that each node updates its state by using the states of its own and its neighbors. Such two-hop information can be properly utilized to facilitate the efficient detection of misbehaviors of targeted agents [4], [16]. Specifically, for node i at time k, the corresponding information set $\Psi_i(k)$ is denoted by

$$\Psi_i(k) = \{i, x_i(k), \pi_i(k), \varepsilon_i(k-1), \{j, x_j^{(i)}(k-1), j \in \mathcal{N}_i\}\},$$

where $x_j^{(i)}(k-1)$ is the state value (at time $k-1$) of node j sent by node i to neighbors at time k. $\varepsilon_i(k)$ is the compensation added by normal nodes and will be discussed in detail later. Let $\pi_i(k)$ be the attack detection flag taking values of either "1" or "0", representing "attack" or "no attack", respectively. Note that $\varepsilon_i(k)$ is allowed to be non-zero for normal nodes if $\pi_i(k) = 1$. All nodes transmit the information set to all neighbors at each time. Under Perron weights, the updating coefficients of node i, i.e., ω_{ij}, $\forall j \in \mathcal{N}_i$, are known by neighbors if the number of neighbors $|\mathcal{N}_i|$ and N are known by neighbors, which provide bases for the error detection.

Misbehaving nodes considered in this paper can be either malicious nodes or faulty nodes. Faulty nodes may cause disturbances to the system because of accident faults, e.g., miscalculations. A node is said to be malicious if it sends the same information to all of its neighbors at each time by manipulating the information set. If the network is realized by broadcast communication, it is natural to assume that any agent sends the same value to all of its neighbors. We introduce three assumptions regarding the specific attacks that malicious nodes can generate.

Assumption 1. *Any two misbehaving nodes do not neighbor with each other.*

Assumption 2. *A malicious node can manipulate its information set by changing the state values of its own and its neighbors, and delete the IDs and states of its neighbors, but cannot add any entries.*

Assumption 3. *A normal node will cut off all future communication with a certain node that is to be isolated.*

Assumption 1 is reasonable when the attack capability is limited, e.g., when the number of misbehaving nodes is much

Manuscript 1369 submitted to 2021 60th IEEE Conference on Decision and Control (CDC). Received March 25, 2021.

less than that of normal nodes, or misbehaving nodes are sparsely distributed in the networks [4]. The types of attacks considered are specified in Assumption 2. Malicious nodes prefer not to add any entries because it will be easily detected by normal nodes if two-hop neighbors are known. With Assumption 3, the misbehaving nodes will be effectively isolated by all normal nodes, which are verified in Sec. IV.

Since misbehaving nodes will cause disturbances and normal nodes will add compensation input, the discrete-time linear updating rule is written as follows

$$x(k+1) = Wx(k) + \varepsilon(k), \tag{2}$$

where $\varepsilon(k)$ is the input vector. It holds that $\exists k < \infty, i \in \mathcal{V}_m$, $\varepsilon_i(k) \neq 0$, because of malicious attacks or random faults. The input of normal nodes, i.e., $\varepsilon_i(k)$, $i \in \mathcal{V}_s$, may be non-zero because of compensation.

D. Problem of Interest

We consider a multi-agent system described by $\mathcal{G}$, where each node owns an initial value $x_i(0)$ and updates its state by (2). This paper aims to develop a distributed detection and compensation mechanism where normal nodes detect the errors of neighboring misbehaving nodes by examining the information set from neighbors and mitigate the impact by adding compensating input to own states.

Under deterministic scenarios where all information sets from neighbors are available, we design a fault-tolerant algorithm to achieve average consensus for the system with malicious nodes and faulty nodes, i.e.,

$$\lim_{k\to\infty} x_j(k) = \frac{1}{|\mathcal{V}_r|}\sum_{u\in\mathcal{V}_r} x_u(0), \quad \forall j \in \mathcal{V}_r, \tag{3}$$

where $\mathcal{V}_r$ is a subset of $\mathcal{V}$ including normal nodes and faulty nodes within the regulation (error bound). It is assumed that the subgraph $\mathcal{G}_r = (\mathcal{V}_r, \mathcal{E}_r)$ is connected, where $\mathcal{E}_r$ denotes the set of edges connecting nodes in $\mathcal{V}_r$.

Under scenarios where information is intermittently available, we extend our algorithm to achieve average consensus in expectation, i.e.,

$$E\big\{\lim_{k\to\infty} x_j(k)\big\} = \frac{1}{|\mathcal{V}_r|}\sum_{u\in\mathcal{V}_r} x_u(0), \quad \forall j \in \mathcal{V}_r. \tag{4}$$

III. ALGORITHM DESIGN

In this section, we propose algorithms to enable nodes to detect misbehaviors, compensate negative impact and achieve average consensus. The basic idea is to design communication protocols to achieve detection and compensation. By two-hop information set, the attacks can be characterized in a distributed manner, which leads to corresponding four types of compensation. Considering information sets are required constantly by D-DCC, a stochastic scheme is introduced. Specifically, a deterministic algorithm (D-DCC) is first presented, followed by a stochastic algorithm (S-DCC) more efficient in complex scenarios.

A. Detection Strategies

According to Assumption 2, a malicious node i can manipulate states of neighbors in the information set, i.e., $x_j^{(i)}(k-1)$, or updates its own state with arbitrary disturbance. The detection strategies are characterized by the corresponding two types:

- *Detection Strategy I*: node j detects whether misbehaving nodes change the state values of j in information set, i.e., $x_j^{(i)}(k) \neq x_j(k), j \in \mathcal{N}_i$. If the malicious node deletes the ID and state of j in the information set, it can be regarded as changing the corresponding state value to zero.
- *Detection Strategy II*: node j detects that misbehaving nodes do not follow the update rule, i.e., $x_i(k+1) \neq \sum_{j\in\mathcal{N}_i} w_{ij}x_j^{(i)}(k)$.

Detection Strategy I and Detection Strategy II can detect basic deception attacks such as spoofing attack and false-data injection attack [23]. Nodes could be detected by Detection Strategy I, II or both. The update rule is as

$$\begin{aligned} x_i(k+1) &= \sum_{j\in\mathcal{N}_i} w_{ij}x_j(k) + \varepsilon_i(k) \\ &= \sum_{j\in\mathcal{N}_i} w_{ij}x_j(k) + \sum_{j\in\mathcal{N}_i} \varepsilon_i^{j(1)}(k) + \varepsilon_i^{(2)}(k), \end{aligned} \tag{5}$$

where $\varepsilon_i^{j(1)}(k)$ and $\varepsilon_i^{(2)}(k)$ are the disturbances detected by Detection Strategy I and II, respectively, i.e.,

$$\varepsilon_i^{j(1)}(k) = w_{ij}(x_j^{(i)}(k) - x_j(k)), \tag{6a}$$

$$\varepsilon_i^{(2)}(k) = x_i(k+1) - \sum_{j\in\mathcal{N}_i} w_{ij}x_j^{(i)}(k). \tag{6b}$$

We divide the disturbances into two parts in a distributed manner, which correspond to the two types of detections.

B. D-DCC Algorithm

In this part, a deterministic detection-compensation-based consensus (D-DCC) algorithm is proposed. We provide the following lemma to show the sufficient condition of resilient average consensus on dynamic system (2).

Lemma 1 ([21])**.** *For the system (2), if the added input vectors are bounded, i.e., $||\varepsilon(k)||_\infty \leq \alpha\rho^k$ for certain $\alpha > 0$ and $\rho \in [0, 1)$, and the sum of inputs satisfies*

$$\sum_{k=0}^{\infty}\sum_{i=1}^{n} \varepsilon_i(k) = 0, \tag{7}$$

then average consensus is achieved exponentially.

It is obvious that the existence of misbehaving nodes can lead to the violation of (7), which is the necessary condition of average consensus (if (7) does not hold, $\lim_{k\to\infty}\sum_{i=1}^{N} x_i(k) = \sum_{i=1}^{N} x_i(0)$ will not hold). To achieve exact average consensus, we need to compensate the impact of misbehaviors by introducing an error compensator η_j for each normal node j. The compensation values to be added is stored in error compensator η_j. We define the following three types of compensation.

Manuscript 1369 submitted to 2021 60th IEEE Conference on Decision and Control (CDC). Received March 25, 2021.

- *Compensation Scheme I*: to compensate the impact detected by Detection Strategy I, i.e.,

$$\eta_j^{i(1)}(k+1) = -w_{ij}(x_j^{(i)}(k) - x_j(k)). \tag{8}$$

- *Compensation Scheme II*: to compensate the impact detected by Detection Strategy II, i.e.,

$$\eta_j^{i(2)}(k+1) = -\varepsilon_i^{(2)}(k)/|\mathcal{N}_i|. \tag{9}$$

- *Compensation Scheme III*: to compensate the impact of isolation, i.e.,

$$\eta_j^{i(3)}(k+1) = (x_i(k+1) - x_i(0))/|\mathcal{N}_i|. \tag{10}$$

Note that all the $|\mathcal{N}_i|$ neighbors of node i detect the misbehaviors by Detection Strategy II of i. Hence, each node averagely compensates the error. Considering that there may be misbehaving nodes with low data utility, e.g., malicious nodes who constantly cause errors and faulty nodes with severe malfunction, isolation are adopted to thoroughly eliminate negative effects. Hence, Compensation Scheme III is adopted when node i is isolated by neighbors. Each neighbor will equally compensate the impact on average consensus.

Inspired by Lemma 1, we adopt a distributed exponential decaying bound of errors, i.e., $\alpha_j\rho_j^k$, $j \in \mathcal{V}_s$, to guarantee the convergence. Let $\alpha = N\max_{i\in\mathcal{V}}\alpha_i$, $\rho = \max_{i\in\mathcal{V}}\rho_i$. Then, the condition $||\varepsilon(k)||_\infty \leq \alpha\rho^k$ will hold. Node j detects the error of its neighbor i. If the error is in the bound $\alpha_j\rho_j^k$, node j will compensate the error by Compensation Scheme I and II, which is a resilient mechanism for finite errors and accidental errors such as computation error and actuator error. Otherwise, node j will cut off the future communication with i and Compensation Scheme III will be used by the neighbors.

By means of the above three types of compensation, we propose D-DCC as Algorithm 1. Node j detects by Strategy I at steps 4-7 and by Strategy II at steps 8-11. At steps 12-14, node j checks whether the error of i is out of local bound $\alpha_i\rho_i^k$. At steps 15-17, node j estimates whether node i is isolated at last step or by other nodes at this time, then calculate $\eta_j^{i(3)}(k+1)$ and send the new $|\mathcal{N}_j|$ to neighbors if it is. At step 18, node j updates its error compensator. Then, at last of time $k+1$, node j updates its state with designed $\varepsilon_i(k)$. The compensation input $\varepsilon_j(k+1)$ should guarantee the secure state and non-increasing property of $|\eta_j|$.

Algorithm 1: D-DCC Algorithm

Input: $x_j(k), \Psi_i(k+1), i \in \mathcal{N}_j$.
Output: The IDs of misbehaving nodes.

1: **Initialize**: set compensator $\eta_j = 0$, parameters α_j, ρ_j; node j exchanges its true initial value $x_j(0)$ and the number of neighbors $|\mathcal{N}_j|$ with neighbors, and store them;
2: **for** $k = 0$: Max_time **do**
3: **for** $i \in \mathcal{N}_j$ **do**
4: $\varepsilon_i^{j(1)}(k+1) = w_{ij}(x_j^{(i)}(k) - x_j(k))$.
5: **if** $\varepsilon_j^{i(1)}(k) \neq 0$ **then**
6: $\eta_j^{i(1)}(k+1) = -\varepsilon_i^{j(1)}(k)$.
7: **end if**
8: $\varepsilon_i^{(2)}(k) = x_i(k+1) - \sum_{j\in N_i} w_{ij}x_j^{(i)}(k)$.
9: **if** $\varepsilon_i^{(2)}(k) \neq 0$ **then**
10: $\eta_j^{i(2)}(k+1) = -\varepsilon_i^{(2)}(k)/|\mathcal{N}_i|$.
11: **end if**
12: **if** $|\varepsilon_i^{j(1)}(k) + \varepsilon_i^{(2)}(k)| > \alpha_j\rho_j^k$ **then**
13: node j cut off the communication with node i.
14: **end if**
15: **if** i is isolated at this step **then**
16: $\eta_j^{i(3)}(k+1) = (x_i(k+1) - x_i(0))/|\mathcal{N}_i|$.
17: **end if**
18: $\eta_j = \eta_j + \eta_j^{i(1)}(k+1) + \eta_j^{i(2)}(k+1) + \eta_j^{i(3)}(k+1)$.
19: **end for**
20: $\eta_j = \eta_j - \varepsilon_j(k+1)$.
21: update its state $x_j(k+2)$ with $\varepsilon_j(k+1)$ by (5).
22: **end for**

C. S-DCC Algorithm

The proposed D-DCC consumes considerable computing resources and requires the information sets at each time because it calculates all neighbors' states at each time according to the information sets. Hence, considering that information sets are intermittently available and malicious nodes may attack randomly with a certain probability, a stochastic detection-compensation-based consensus (S-DCC) algorithm is proposed as Algorithm 2. The stochasticity of S-DCC consists in two aspects: (1) the detection is stochastic at each time of probability p; (2) The compensation is stochastic because of random misbehaviors.

The main difference of S-DCC from D-DCC is that normal nodes detect the neighbor nodes at random times. As a result, some attacks may not be detected. To handle this issue, we propose the following rule of compensation based on the mean value of estimation.

- *Compensation Scheme IV*: to compensate the impact of undetected misbehaviors, i.e.,

$$\eta_j^{i(4)}(k) = -(k - m_j)\bar{\varepsilon}_i^j,$$

where $\bar{\varepsilon}_i^j = \sum_{\varepsilon_i^j(k)\in\Phi_j^{(i)}} \varepsilon_i^j(k)/m_j$ and m_j is the number of times that j detects. The intuition behind Compensation Scheme IV is that the mean of errors which are detected may represent the mean effect by misbehaving nodes in a time window, because malicious nodes must follow a certain scheme to attack and misbehaviors of faulty nodes may follow some certain rules. If a faulty node is not isolated, the error may be accidental. Therefore, no additional compensation is adopted.

Specifically, the detection is enabled at each time independently of probability p. When the detection is enabled, node j will detect and compensate the possible errors. In order to analyze the effect of misbehaving nodes, node j will store the detected error $\varepsilon_i^j(k)$ in the error set $\Phi_j^{(i)}$ corresponding to node i. When node i is isolated, both Compensation Scheme III and IV will be adopted.

IV. Performance Analysis

In this section, we prove that for D-DCC, all misbehaving nodes will be detected and average consensus will be achieved. Additionally, for S-DCC, we demonstrate that all

Algorithm 2: S-DCC Algorithm

Input: $x_j(k), \Psi_i(k+1), i \in \mathcal{N}_j$.
Output: The IDs of misbehaving nodes.
Initialize: The same as Algorithm 1;
set detection probability p, detection times $m_j = 0$.
for $k = 0$: Max_time **do**
  if Detection is enabled **then**
    set $m_j = m_j + 1$.
    for $i \in \mathcal{N}_j$ **do**
      Execute steps 4-14 in Algorithm 1.
      if node j is malicious **then**
        store $\varepsilon_i^j(k) = \varepsilon_i^{j(1)}(k) + \varepsilon_i^{(2)}(k)/|\mathcal{N}_i|$ in $\Phi_j^{(i)}$.
      end if
      if i is isolated at this step **then**
        $\eta_j^{i(3)}(k+1) = (x_j(k+1) - x_j(0))/|\mathcal{N}_i|$.
        $\eta_j^{i(4)}(k+1) = -(k-m)\overline{\varepsilon}_i^j$.
      end if
      $\eta_j = \eta_j + \eta_j^{i(1)}(k) + \eta_j^{i(2)}(k) + \eta_j^{i(3)}(k) + \eta_j^{i(4)}(k)$.
    end for
  end if
  $\eta_j = \eta_j - \varepsilon_j(k+1)$.
  update its state $x_j(k+2)$ with $\varepsilon_j(k+1)$ by (5).
end for

malicious nodes will be detected with probability one and average consensus in expectation will be achieved, and we analyze the accuracy of S-DCC.

A. Analysis of D-DCC

First, the detection performance of D-DCC is evaluated. The following lemma shows the effectiveness of D-DCC.

Lemma 2. *If Assumptions 1-3 hold, then all misbehaving nodes will be detected and some of them will be isolated by Algorithm 1.*

Proof. Suppose that the misbehaving node i changes the value of its neighbors' (j) in the information set $\Psi_i(k+1)$. According to Assumption 1, node j is normal. Hence, when node j receives the information value from i, it will find out that $x_j^{(i)}(k) \neq x_j(k)$ and node i will be detected by Detection Stratery I. Similarly, if node i deletes the ID and state of j, it will be detected by node j.

If the misbehaving node j does not follow the update rule based on the information set, i.e., $x_i(k+1) \neq \sum_{j\in\mathcal{N}_i} w_{ij} x_j^{(i)}(k)$, it will be detected by all neighbors by Detection Strategy II, because w_{ij} is known by all neighbors.

According to Algorithm 1, if the error is in the local bound, the misbehaving node will not be isolated. However, once the error is out of the bound, the misbehaving node will be isolated. □

Next, we evaluate the performance of D-DCC in compensation and consensus by the following theorem.

Theorem 1. *If Assumptions 1-3 hold, then D-DCC achieves average consensus, i.e., (3) holds.*

Proof. First, we illustrate that consensus will be achieved by Algorithm 1. Since W is doubly stochastic and $\varepsilon(k)$ satisfies the condition $||\varepsilon(k)||_\infty \leq \alpha\rho^k$ due to the error bound. the consensus will be achieved according to Lemma 1. If node i is isolated, let $W_{\{i\}}$ be the matrix obtained by deleting i-th row and column of W. It follows that $W_{\{i\}}$ is still a matrix with Perron weight or Metropolis weights. Hence, consensus will be achieved among remaining nodes.

Next, we prove the limit value is average consensus among remaining nodes. Without loss of generality, we consider a subsystem composed by misbehaving node i and its neighbors. According to (2), and W is doubly stochastic, we have

$$\sum_{j\in\mathcal{V}} x_j(k+1) = \sum_{j\in\mathcal{V}} x_j(0) + \sum_{l=1}^{k}\sum_{j\in\mathcal{V}} \varepsilon_j(l), \tag{11}$$

(Case 1) If node i is not isolated, according to Compensation Scheme I and II, we have

$$\sum_{l=0}^{\infty}\{\varepsilon_i(l) + \sum_{j\in\mathcal{N}_i}(\eta_j^{i(1)}(l+1) + \eta_j^{i(2)}(l+1))\} = 0. \tag{12}$$

Hence, (7) holds.

(Case 2) If node i is isolated at time $k+1$, we can regard it as staying at the value $x_i(k+1)$. Since W is a doubly stochastic matrix at each-step, we have that (11) holds. Node i is regarded to stay at the value $x_i(k+1)$, i.e. $\varepsilon_i(l) = 0, \forall l > k$. Because compensation (1) and (2) will compensate the impact before isolation, we have that (12) holds. Compensation (3) will compensate the impact of node isolation. Thus, we have

$$\sum_{l=1}^{\infty}\sum_{j\in\mathcal{V}} \varepsilon_j(l) = x_i(k+1) - x_i(0). \tag{13}$$

Combining (11) with (13) and noting that node i is regarded as staying at the value $x_i(k)$ after isolation, it follows that

$$\lim_{l\to\infty} \sum_{j\in\mathcal{V}/\{i\}} x_j(l) = \sum_{j\in\mathcal{V}/\{i\}} x_j(0). \tag{14}$$

Generally, we have

$$\lim_{l\to\infty} \sum_{j\in\mathcal{V}_r} x_j(l) = \sum_{j\in\mathcal{V}_r} x_j(0). \tag{15}$$

Hence, (3) holds and an average consensus among remaining nodes is achieved. □

Theorem 1 guarantees the accurate average consensus even if there are misbehaving nodes in the system. Due to the use of local error bound, D-DCC provides tolerance for accidental miscalculations and transmission errors. If isolation is not adopted in these scenarios, it may contribute to the overall efficiency of the system, e.g., an agent malfunctions transitorily. The tolerance depends on the parameters α_i and ρ_i. Note that a larger $\alpha_i\rho_i$ will improve the fault tolerance but reduce the convergence speed.

B. Analysis of S-DCC

Before the analysis, some notations and assumptions are provided. To evaluate the performance of S-DCC, it is supposed that malicious node i attacks the system with probability $\theta_i \in [0, 1]$. Malicious node i decides to attack or

Manuscript 1369 submitted to 2021 60th IEEE Conference on Decision and Control (CDC). Received March 25, 2021.

not at each time with a certain probability θ_i. Let $X_i(k)$ be the random variable of attack or not, i.e., $X_i(k) \sim \mathcal{B}(1, \theta_i)$, where $\mathcal{B}$ represents the Bernoulli distribution. Further, the malicious nodes attack the system with a certain mechanism. Hence, if node i attacks the system at time k, the random variable of the error, i.e., $Y_i(k)$, obeys a certain distribution with expectation μ_i and variance σ_i^2. Hence, $\varepsilon_i(k) = X_i(k)Y_i(k)$. Then, we have

$$\begin{aligned} E\{\varepsilon_i(k)\} &= \theta_i\mu_i, \\ D\{\varepsilon_i(k)\} &= \theta_i\sigma_i^2 + (1-\theta_i)\theta_i\mu_i^2 \triangleq \sigma_{\varepsilon_i}^2 \end{aligned}$$

The average compensation of $\varepsilon_i(k)$ is $\sum_{j\in N_i} \overline{\varepsilon}_i^j \triangleq \overline{\varepsilon}_i$, where $\overline{\varepsilon}_i$ is the average error of detecting times. Note that there may be some faulty nodes that will not be isolated. These faulty nodes may cause errors due to accidental miscalculations or transmission failures with multiple uncertainties. Hence, it is reasonable to assume that the errors of faulty nodes are zero in expectation. We provide the following theorem to analyze the performance of S-DCC.

Theorem 2. *If the accidental error of the faulty nodes is zero in expectation, then S-DCC achieves average consensus in expectation, i.e.,* $\forall j \in \mathcal{V}_r$,

$$E\big\{\lim_{l\to\infty} x_j(l)\big\} = \frac{1}{|\mathcal{V}_r|}\sum_{u\in\mathcal{V}_r} x_u(0), \tag{16}$$

$$D\big\{\lim_{l\to\infty} x_j(l)\big\} \le \sum_{i\in\mathcal{V}_m} \frac{(k_i^{iso} - M_i)(1+1/M_i)\sigma_{\varepsilon_i}^2}{|\mathcal{V}_r|^2}, \tag{17}$$

where k_i^{iso} is the time of isolation of node i and $M_i = \min_{j\in\mathcal{N}_i} m_j$. Also, the consensus value is bounded, i.e.,

$$\big|\lim_{l\to\infty} x_j(l) - \frac{1}{|\mathcal{V}_r|}\sum_{u\in\mathcal{V}_r} x_u(0)\big| \le \frac{\alpha\rho|\mathcal{V}_m|}{(1-\rho)|\mathcal{V}_r|}. \tag{18}$$

Proof. First, we illustrate that all malicious nodes will be detected with probability one. For each malicious node i, it attacks the system with probability $\theta_i, 0 < \theta_i \le 1$. For each normal node j in the neighborhood of i, it detects i with probability $p, 0 < p \le 1$. The probability of the event that node i is detected by j in no later than time k is

$$P(k) = 1 - (1 - p\theta_i)^k. \tag{19}$$

By taking limit on both sides of (19), we have

$$\lim_{k\to\infty} P(k) = \lim_{k\to\infty}(1 - (1-p\theta_i)^k) = 1.$$

Similarly, all malicious nodes will be isolated with probability one.

Next, we will show

$$E\big\{\lim_{l\to\infty}\sum_{j\in\mathcal{V}_r} x_j(l)\big\} = \sum_{j\in\mathcal{V}_r} x_j(0). \tag{20}$$

Considering the subsystem composed by the malicious node i which is isolated at time k_i^{iso} and its neighbors. Note that $X_i(l)$ and $Y_i(l)$ are independent. For malicious node i, the expectation of the sum of its error within time k is

$$E\big\{\sum_{l=1}^{k}\varepsilon_i(l)\big\} = E\big\{\sum_{l=1}^{k} X_i(l)Y_i(l)\big\} = k\theta_i\mu_i. \tag{21}$$

Without loss of generality, we consider all the neighbors of i detect at the same time. The expectation of $\sum_{j\in\mathcal{N}_i}\overline{\varepsilon}_i^j$ satisfies

$$\begin{aligned} E\big\{\sum_{j\in\mathcal{N}_i}\overline{\varepsilon}_i^j\big\} &= E\big\{\frac{1}{m_j}\sum_{j\in\mathcal{N}_i}\sum_{\varepsilon_i^j(k)\in\Phi_j^{(i)}}\varepsilon_i^j(k)\big\} \\ &= E\big\{\varepsilon_i(l)\big\} = \theta_i\mu_i. \end{aligned}$$

The expectation of the Compensation Scheme I, II, IV is

$$\begin{aligned} &E\big\{\sum_{j\in\mathcal{N}_i}\sum_{l=1}^{k}(\eta_{ji}^{(1)}(l) + \eta_{ji}^{(2)}(l) + \eta_{ji}^{(4)}(l))\big\} \\ &= E\big\{\sum_{j\in\mathcal{N}_i}\sum_{\varepsilon_i^j(k)\in\Phi_j^{(i)}}\varepsilon_i^j(k) - (k-m_j)\sum_{j\in\mathcal{N}_i}\overline{\varepsilon}_i^j\big\} \\ &= E\big\{-k\sum_{j\in\mathcal{N}_i}\overline{\varepsilon}_i^j\big\} = -k\theta_i\mu_i. \end{aligned} \tag{22}$$

Let $k+1 = k_i^{\text{iso}}$ for (10). Therefore, combining (10), (21) and (22), we have

$$E\big\{\sum_{l=1}^{k}(\varepsilon_i(l) + \sum_{j\in\mathcal{N}_i}\eta_j(l))\big\} = x_i(k+1) - x_i(0).$$

The sum of the errors of the faulty nodes which is not isolated is zero in expectation. Similar with the proof in Theorem 3, we have (20). Hence, (4) holds and an average consensus in expectation is achieved.

Since $\overline{\varepsilon}_i$ is the average value of sampling, we have $D\{\overline{\varepsilon}_i\} \le \sigma_{\varepsilon_i}^2/M_i$. Because the detected errors (m_j times) will be compensated accurately by j, we have the variance of the consensus value,

$$\begin{aligned} D\big\{\sum_{u\in\mathcal{V}_r} x_u(k)\big\} &= \sum_{i\in\mathcal{V}_m} D\big\{\sum_{l=1}^{k_i^{\text{iso}}}\varepsilon_i(l) - k_i^{\text{iso}}\sum_{j\in\mathcal{N}_i}\overline{\varepsilon}_i^j\big\} \\ &\le \sum_{i\in\mathcal{V}_m}(k_i^{\text{iso}} - M_i)(1+1/M_i)\sigma_{\varepsilon_i}^2. \end{aligned}$$

Each malicious node causes disturbances. Hence, (17) holds.

According to the proof of Theorem 1, $\varepsilon(k)$ satisfies the condition $||\varepsilon(k)||_\infty \le \alpha\rho^k$. For misbehaving nodes, we have

$$\sum_{i\in\mathcal{V}_m}\sum_{l=1}^{\infty}\varepsilon_i(l) \le \frac{|\mathcal{V}_m|\alpha\rho}{(1-\rho)}.$$

Hence, (18) is proved. □

Theorem 2 ensures average consensus in expectation. Note that if $M_i = k_i^{\text{iso}}$, $D\big\{\lim_{l\to\infty} x_j(l)\big\} = 0$, which corresponds to deterministic conditions by D-DCC.

Next, we analyze the accuracy of mean-based Compensation Scheme III, i.e., the distance between the mean-based compensation and actual errors. Let $F_{Y_i(k)}(x)$ be the cumulative distribution function (CDF) of $Y_i(k)$. Because $\varepsilon_i(k) = X_i(k)Y_i(k)$, the CDF of $\varepsilon_i(k)$ is

$$F_{\varepsilon_i(k)}(x) = \begin{cases} \theta_i F_{Y_i(k)}(x) & x < 0 \\ 1 - \theta_i + \theta_i F_{Y_i(k)}(x) & x \ge 0. \end{cases} \tag{23}$$

Without loss of generality, consider all the detection numbers $m_j, j \in \mathcal{N}_i$ are the same. According to the Central Limit Theorem [24], we have

$$\bar{\varepsilon}_i \sim N(\theta_i \mu_i, \sigma^2_{\bar{\varepsilon}_i}/M_i).$$

The CDF of $\bar{\varepsilon}_i$ is

$$F_{\bar{\varepsilon}_i}(x) = \int_{-\infty}^{x} \frac{\sqrt{M_i}}{\sqrt{2\pi}\sigma_{\varepsilon_i}} \exp\big(\frac{-M_i(x-\theta_i\mu_i)^2}{2\sigma^2_{\varepsilon_i}}\big)\mathrm{d}x.$$

The characteristic of proximity of two probability distributions can be described by the Wasserstein distance [25]. In the case of one-dimensional space with the Euclidean metric, the Wasserstein distance is calculated by

$$R(\mathcal{P}, \mathcal{Q}) = \int_{-\infty}^{\infty} |F(x) - G(x)|\mathrm{d}x,$$

where F and G are the CDF of distributions $\mathcal{P}$ and $\mathcal{Q}$, respectively [25]. Hence, we have the Wasserstein distance between $\varepsilon_i(k)$ and $\bar{\varepsilon}_i$:

$$R(\varepsilon_i(k), \bar{\varepsilon}_i) = \int_{-\infty}^{\infty} |F_{\varepsilon_i(k)}(x) - F_{\bar{\varepsilon}_i}(x)|\mathrm{d}x. \tag{24}$$

We can use the Wasserstein distance to show the expectation of absolute error between the mean-based compensation and actual errors. Let $\Phi(\cdot)$ be the CDF of normal distribution. We provide the following theorem to illustrate the bound of $R(\varepsilon_i(k), \bar{\varepsilon}_i)$ when $Y_i(k)$ obeys the normal distribution.

Theorem 3. *If $Y_i(k) \sim \mathcal{N}(\mu_i, \sigma_i)$, then we have*

$$R(\varepsilon_i(k), \bar{\varepsilon}_i) \leq (1-\theta_i)\mu_i + |\sigma_i - \sigma_{\varepsilon_i}| + (1-\theta_i)E\{|Y_i|\}$$

where $E\{|Y_i|\} = \{\sqrt{\frac{2}{\pi}}\sigma_i \exp\big(\frac{-\mu_i^2}{2\sigma_i^2}\big) + \mu_i[1 - 2\Phi(\frac{-\mu_i}{\sigma_i})]\}$.

Proof. Referring to the absolute value inequality, we have

$$\begin{aligned} R(\varepsilon_i(k), \bar{\varepsilon}_i) &= \int_{-\infty}^{\infty} |F_{\varepsilon_i(k)}(x) - F_{\bar{\varepsilon}_i}(x)|\mathrm{d}x \\ &\leq \int_{-\infty}^{\infty} |F_{\varepsilon_i(k)}(x) - F_{Y_i}(x)|\mathrm{d}x + \int_{-\infty}^{\infty} |F_{\bar{\varepsilon}_i}(x) - F_{Y_i}(x)|\mathrm{d}x. \end{aligned}$$

According to (23), we have

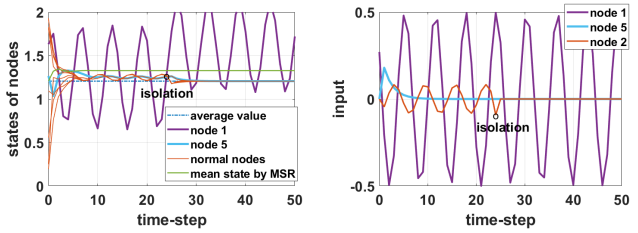
$$\begin{aligned} &\int_{-\infty}^{\infty} |F_{\varepsilon_i(k)}(x) - F_{Y_i}(x)|\mathrm{d}x \\ &= \int_{-\infty}^{0} |(1-\theta_i)F_{Y_i}(x)|\mathrm{d}x + \int_{0}^{+\infty} |(1-\theta_i)(1-F_{Y_i}(x))|\mathrm{d}x \\ &= (1-\theta_i)\big[\int_{-\infty}^{0} F_{Y_i}(x)\mathrm{d}x + \int_{0}^{+\infty} (1-F_{Y_i}(x))\mathrm{d}x\big] \\ &= (1-\theta_i)E\{|Y_i|\}. \end{aligned}$$

In addition, according to the Wasserstein distance between two normal distributions [26], we have

$$\int_{-\infty}^{\infty} |F_{\bar{\varepsilon}_i}(x) - F_{Y_i}(x)|\mathrm{d}x \leq (1-\theta_i)\mu_i + |\sigma_i - \sigma_{\varepsilon_i}|.$$

Combining the above aspects, we complete the proof. □

Remark 1. *Theorem 3 shows the Wasserstein distance under the normal distribution. Hence, the expectation of absolute error between the mean-based compensation and actual errors is bounded. Generally, the explicit bound is difficult to formulate under other distributions, but the Wasserstein distance is bounded as long as $E\{|Y_i(k)|\}$ and $E\{|\bar{\varepsilon}_i|\}$ exist. The quantitative evaluations can be found in Sec. V-B.*



(a) (b)

Fig. 1. Performance of D-DCC. (a) States of nodes. (b) Errors of misbehaving nodes (1 and 5) and compensation input of node 2.

V. Numerical Evaluations

In this section, we conduct numerical evaluations to illustrate the performance of D-DCC and S-DCC. Consider a Erdös-Rényi Random graph (probability for edge creation is 0.7) with $N = 10$ nodes which update values by (2), where W is designed by Perron weights. In the network, there are two misbehaving nodes who are not neighbors, i.e., malicious node 1 that intends to break average consensus and faulty node 5. We set $\alpha_i = 5$, $\rho_i = 0.9$, $\forall i \in \mathcal{V}$.

A. Performance of D-DCC

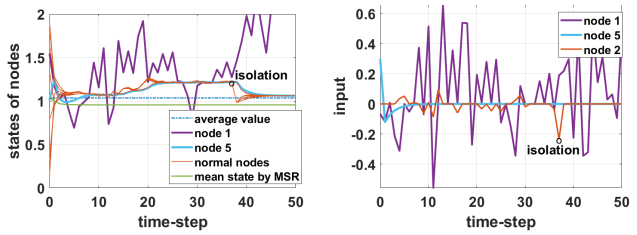
At this part, we set the disturbances of node 1 and 5 to satisfy $\varepsilon_1(k) = 0.5\cos(k)$ and $\varepsilon_5(k) = 0.5 \times 0.6^k$. Fig. 1(a) shows that all nodes except node 1 achieves consensus. The consensus value is the average value of initial states of remaining nodes showing as the blue dotted line. As a contract, we plot the average state of normal nodes by MSR algorithm [11] as the green line. The exact average consensus is achieved by D-DCC, but MSR algorithm does not guarantee average consensus. Fig. 1(b) shows the errors of node 1 and 5. Node 5 has not been isolated because its error is exponentially decaying and in the local bound. Node 5 is isolated at time 24 because the error is out of the bound.

B. Performance of S-DCC

The error of node 1 is set to obey the normal distribution if $X_1(k) = 1$, i.e., $Y_1(k) \sim \mathcal{N}(0.1, 1)$. The error of node 1 is exponentially decaying and zero-sum. We set the detection probability $p = 0.5$ and attack probability $\theta_1 = 0.8$.

Fig. 2(a) shows that all nodes except node 1 achieve consensus and node 1 is isolated by other nodes but node 5 is not. The consensus value is close to the average value of initial states of remaining nodes. Though the limit value is the exact average consensus in expectation, in practice it may vary from it. Compared with MSR [11], S-DCC achieves more accurate average consensus. Fig. 2(b) shows the compensation of node 2 and the errors of node 1 and 5.

According to the Wasserstein distance, we have $R(\varepsilon_1(k), \bar{\varepsilon}_1) = 0.1109$. The CDF of $\varepsilon_1(k)$ and $\bar{\varepsilon}_1$ are shown as Fig. 3. The expectation of absolute error between the disturbance and the mean-based compensation at



(a) (b)

Fig. 2. Performance of S-DCC. (a) States of nodes. (b) Errors of misbehaving nodes (1 and 5) and compensation input of node 2.

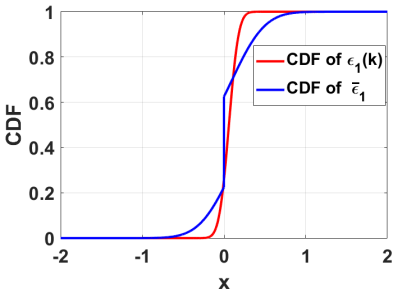


Fig. 3. CDF of $\varepsilon_1(k)$ and $\bar{\varepsilon}_1$

each time is within a small range. The two distributions have excellent similarity, which shows the accuracy of mean-based compensation.

VI. CONCLUSION

In this paper, we investigated the resilient average consensus problem against misbehaving nodes in multi-agent systems. We first presented the D-DCC to compensate the disturbances caused by misbehaving nodes. The exponential decaying bound provides fault tolerance for faulty nodes and guarantee the convergence. We proved that the detection scheme is effective and an average consensus is achieved by D-DCC. Furthermore, we proposed S-DCC with mean-based compensation to adapt for scenarios where information are intermittently available. It was proved that an average consensus in expectation is achieved by S-DCC, and the absolute error between mean-based compensation and actual disturbance was analyzed by Wasserstein distance. Finally, simulations were conducted to illustrate the effectiveness of the proposed algorithms.

There are still many issues worthy of further investigations. First, achieving exact resilient consensus over time-varying and directed networks will be considered in future. Second, the extension of resilient average consensus for high-dimension systems with general linear dynamics is left for future work. Third, applications of resilient average consensus including formation control and flocking of multi-robot systems can be possible directions.

REFERENCES

[1] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3575–3605, 2012.

[2] W. Wang, J. Huang, C. Wen, and H. Fan, "Distributed adaptive control for consensus tracking with application to formation control of nonholonomic mobile robots," *Automatica*, vol. 50, no. 4, pp. 1254–1263, 2014.

[3] J. Li, F. Liu, Z. Wang, S. H. Low, and S. Mei, "Optimal power flow in stand-alone dc microgrids," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5496–5506, 2018.

[4] J. He, P. Cheng, L. Shi, and J. Chen, "Sats: Secure average-consensus-based time synchronization in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 61, no. 24, pp. 6387–6400, 2013.

[5] 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, 2004.

[6] 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, 2019.

[7] C. N. Hadjicostis, N. Vaidya, and D. D. Alejandro, "Robust distributed average consensus via exchange of running sums," *IEEE Trans. Autom. Control*, vol. 61, no. 6, pp. 1492–1507, 2015.

[8] A. Olshevsky and J. N. Tsitsiklis., "Convergence speed in distributed consensus and averaging," *SIAM J. Control Optim.*, vol. 48, no. 1, pp. 33–55, 2009.

[9] 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, 2013.

[10] F. Pasqualetti, A. Bicchi, and F. Bullo, "Consensus computation in unreliable networks: A system theoretic approach," *IEEE Trans. Autom. Control*, vol. 57, no. 1, pp. 90–104, 2012.

[11] R. M. Kieckhafer and M. H. Azadmanesh, "Reaching approximate agreement with mixed-mode faults," *IEEE Trans. Parallel Distrib. Syst.*, vol. 5, no. 1, pp. 53–63, 1994.

[12] H. Zhang and S. Sundaram, "Robustness of information diffusion algorithms to locally bounded adversaries," in *Proc. Amer. Control. Conf.*, pp. 5855-5861, 2012.

[13] S. M. Dibaji, H. Ishii, and R. Tempo, "Resilient randomized quantized consensus," *IEEE Trans. Autom. Control*, vol. 63, no. 8, pp. 2508–2522, 2018.

[14] K. Saulnier, D. Saldaña, A. Prorok, G. J. Pappas, and V. Kumar, "Resilient flocking for mobile robot teams," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 1039–1046, 2017.

[15] J. Yan, X. Li, Y. Mo, and C. Wen, "Resilient multi-dimensional consensus in adversarial environment," *arXiv preprint arXiv:2001.00937*, 2020.

[16] C. Zhao, J. He, and J. Chen, "Resilient consensus with mobile detectors against malicious attacks," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 1, pp. 60–69, 2018.

[17] I. Shames, A. M. H. Teixeira, H. Sandberg, and K. H. Johansson, "Distributed fault detection for interconnected second-order systems," *Automatica*, vol. 47, no. 12, pp. 2757–2764, 2011.

[18] R. Gentz, S. X. Wu, H. Wai, A. Scaglione, and A. Leshem, "Data injection attacks in randomized gossiping," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 4, pp. 523–538, 2016.

[19] M. Guo, D. Dimarogonas, and K. Johansson, "Distributed real-time fault detection and isolation for cooperative multi-agent systems," in *Proc. Amer. Control. Conf.*, pp. 5270–5275, 2012.

[20] L. Yuan and H. Ishii, "Secure consensus with distributed detection via two-hop communication," *arXiv preprint arXiv:2101.05087*, 2021.

[21] J. He, L. Cai, P. Cheng, J. Pan, and L. Shi, "Distributed privacy-preserving data aggregation against dishonest nodes in network systems," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1462–1470, 2019.

[22] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. Int. Conf. Inf. Process. Sensor Netw.*, IEEE, pp. 63-70, 2005.

[23] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, "A systems and control perspective of cps security," *Annu. Rev. Control*, vol. 47, pp. 394–411, 2019.

[24] G. R. Grimmett and D. R. Stirzaker, *Probability and random processes*. Oxford Univ. Press, 2003.

[25] S. Vallender, "Calculation of the wasserstein distance between probability distributions on the line," *Theory Probab. Appl.*, vol. 18, no. 4, pp. 784–786, 1974.

[26] D. Chafai and F. Malrieu, "On fine properties of mixtures with respect to concentration of measure and sobolev type inequalities," in *Annales de l'IHP Probabilités et statistiques*, pp. 72-96, 2010.

Manuscript 1369 submitted to 2021 60th IEEE Conference on Decision and Control (CDC). Received March 25, 2021.