2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Resilient Projection-based Distributed Leader-Follower Consensus Against Integrity Cyberattacks

Mahdieh S. Sadabadi

Abstract— **This paper focuses on the problem of distributed leader-follower consensus in multi-agent systems in which some agents are subject to adversarial attacks. We develop a resilient distributed leader-follower control strategy subject to integrity attacks, where agents' updates of their states can be compromised by injecting false signals to control inputs. Under such a threat model, we design a resilient distributed leader-follower framework for agents with continuous-time dynamics to resiliently track a reference state propagated by a leader. In the design of the resilient framework, projection-based operators are used as dynamic controllers to estimate the dynamics of uncertainties on the control inputs of each agent. By use of the properties of projection operators and Lyapunov stability theory, the uniform ultimate boundedness of the closed-loop multi-agent system in the presence of integrity attacks is guaranteed. The proposed resilient distributed scheme does not impose any limitations on the maximum tolerable number of cyberattacks and does not require high network connectivity. The effectiveness of the proposed resilient distributed consensus algorithm is verified by a numerical example.**

I. Introduction

Distributed leader-follower consensus algorithms have received significant attention over the last few decades, driven by their numerous wide applications. Their recent applications include frequency synchronization in distributed generation units [1], voltage regulation in DC networks [2], mobile sensor networks [3], and vehicle platoon control [4]. A distributed leader-follower consensus scheme is composed of multiple autonomous agents called followers, which are spatially distributed and communicate over a network to achieve a common goal, provided by a set of agents, called leaders.

Although a distributed algorithm offers several benefits compared to centralized algorithms, in terms of scalability and robustness to a single point of failure, their reliable operations depend on the reliable functionality of cyber resources, particularly sensing, actuating, and communication networks. However, cyber resources in such algorithms are exposed to the risk of remote interference such as cyberattacks.

Ensuring the resilient operation of multi-agent systems to cyberattacks and adversarial misbehavior is essential in several critical applications such as electrical power systems [5]. As part of this need, resilient features must be included in the distributed algorithms for multi-agent systems. Hence, the development of a resilient consensus algorithm is required. In a resilient consensus problem in multi-agent systems, agents seek to reach an agreement on the reference value of a set of leaders in the presence of cyberattacks whose properties are unknown [6].

In recent years, several resilient leader-follower consensus approaches were developed in the literature such as the works in [6]–[11] and references therein. In [7] and [8], resilient leader-follower consensus algorithms were proposed without requiring any assumptions on attackers' behavior. However, these methods restrict the maximum number of compromised agents and/or the connectivity of communication networks. The work in [6] focuses on the problem of resilient leader-follower consensus of multi-agent systems with discrete-time dynamics and develops a sliding window mean-subsequence-reduced-based algorithm. Nevertheless, this work does not consider the case of continuous-time dynamics and also restricts the number of arbitrarily attacked agents. Resilient cooperative control approaches for the leader-follower consensus problem, based on the introduction of a virtual layer and virtual states, are proposed in [9]–[11]. However, the resilience of the proposed algorithms in these papers depends on a parameter that requires to be extremely large to ensure resilient consensus in the presence of cyberattacks. The large value of this parameter might adversely impact the transient response of followers (agents) in multi-agent systems.

This paper focuses on the problem of resilient distributed leader-follower consensus in continuous-time multi-agent systems in the presence of false data injection (FDI) integrity cyberattacks. The paper develops a novel resilient distributed algorithm, which ensures that agents' states converge closely to a reference state provided by a leader, even under unknown but bounded attacks on both the agent's update of its local state and its control inputs. The proposed resilient algorithm is based on projection-based control, which is a robustness-based augmentation technique that bounds the outputs of a controller in sector-bounded sets while conforming to the Lyapunov stability rules [12].

Specifically, the paper provides the following contributions:

1) We propose a novel resilient projection-based distributed algorithm, which ensures that all agents' states converge closely to reference states propagated by a leader.
2) The proposed resilient distributed consensus scheme guarantees the uniform ultimate boundedness of the closed-loop multi-agent system in the presence of bounded FDI integrity attacks on control input channels and asymptotic stability for the case of time-invariant FDI attacks.
3) The proposed algorithm does not impose any limitations

M. S. Sadabadi is with the Department of Electrical and Electronic Engineering, University of Manchester, Manchester, United Kingdom, mahdieh.sadabadi@manchester.ac.uk.

979-8-3503-1632-2/24/$31.00 ©2024 IEEE

on the maximum tolerable number of cyberattacks and does not require high network connectivity.

The performance and effectiveness of the proposed resilient distributed leader-follower consensus algorithm are verified by a numerical example.

The paper is organized as follows. The problem under study is stated in Section II. The proposed cyber-resilience leader-follower consensus scheme is presented in Section III. Section IV is devoted to the stability analysis of the closed-loop system augmented with the proposed resilient distributed consensus approach. The simulation results and numerical analysis are provided in Section V. Conclusions are given in Section VI.

Notation: The notation used in this paper is standard. Specifically, $\mathbf{1}_n$ is an n-dimensional vector of all ones, A^T denotes the transpose of matrix A, and $A = \mathrm{diag}(a_1, \ldots, a_n) \in \mathbb{R}^n$ is a diagonal matrix whose diagonal elements are $a_i \in \mathbb{R}$. For symmetric matrices, $X \succ 0$ ($X \prec 0$) and $X \succeq 0$ ($X \preceq 0$) respectively indicate positive-definiteness (negative-definiteness) and positive semi-definiteness (negative semi-definiteness).

Preliminaries on Projection Operators: The following definition and lemma (Lemma 11.3. in [13]) are used in the proposed control approach in this paper.

Definition 1 (Projection Operation). Suppose that $f(\theta) : \mathbb{R}^n \to \mathbb{R}$ is a scalar-valued continuously differentiable convex function. Let $\theta(t) \in \mathbb{R}^n$ and $y(t) \in \mathbb{R}^n$ be time-varying piecewise continuous vectors. The Projection Operator $\mathrm{Proj} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ for two vectors $\theta(t)$ and $y(t)$ are defined as follows:

$$\mathrm{Proj}(\theta, y) = \begin{cases} y, & \text{if } f(\theta) < 0, \\ & \text{or if } f(\theta) = 0 \text{ and } y^T \nabla f(\theta) \leq 0, \\ y - \dfrac{\nabla f(\theta) \nabla f^T(\theta)}{\nabla f^T(\theta) \nabla f(\theta)} y f(\theta), & \text{otherwise.} \end{cases} \tag{1}$$

where $\nabla f(\theta) = [\frac{\partial f(\theta)}{\partial \theta_1}, \ldots, \frac{\partial f(\theta)}{\partial \theta_n}]^T \in \mathbb{R}^n$.

From the definition of projection operators, starting from any initial condition $\theta(0) = \theta_0$ within a convex set $\mathcal{S}_0 = \{\theta \in \mathbb{R}^n | \ f(\theta) \leq 0\}$, the system trajectory $\theta(t)$ will remain in the set $\mathcal{S}_\theta = \{\theta \in \mathbb{R}^n | \ f(\theta) \leq 1\}$, $\forall t \geq 0$.

Lemma 1. *For $\theta^* \in \mathcal{S}_\theta$, the projection operator* $\mathrm{Proj}(\theta, y)$ *satisfies the following condition:*

$$(\theta - \theta^*)^T (\mathrm{Proj}(\theta, y) - y) \leq 0. \tag{2}$$

II. Leader-Follower Consensus

Consider a cooperative multi-agent system consisting of $n \geq 2$ agents. The dynamics of the agent i, $i \in \{1, \ldots, n\}$, are represented as follows:

$$\dot{x}_i(t) = u_i(t), \tag{3}$$

where $x_i(t) \in \mathbb{R}$ is the state and $u_i(t) \in \mathbb{R}$ is the control input of the agent i. We consider $x_0 \in \mathbb{R}$ as the state of a leader, which is assumed to be constant. Note that in several real-world applications, such as frequency synchronization in electrical power systems [14], x_0 is constant.

Given the dynamics of the agents in (3), the main objective is to design $u_i(t)$ so that the agents reach a consensus, i.e.,

$$\lim_{t \to \infty} x_i(t) = x_0 \tag{4}$$

for $i = 1, \ldots, n$.

It is assumed that x_0 is not accessible to all agents and only a subset of agents have access to it. Hence, to ensure (4), each agent is required to design $u_i(t)$ and update its local state $x_i(t)$ in a distributed manner according to the following update rule:

$$\dot{x}_i(t) = u_i \left(x_i(t), (x_j(t))_{j \in \mathcal{N}_i} \right), \tag{5}$$

where $\mathcal{N}_i$ is the neighbor set of agent i and $x_j(t)$ is the state of the neighboring agent.

Remark 1. *In a multi-agent system whose agents' dynamics are described by a heterogeneous nonlinear or linear system, if the individual dynamic system of agents is input passivity-short (or can become input passivity-short by a local feedback controller), then it is shown that the dynamic behaviors at the network level and network control design can equivalently be transformed to the first-order integral dynamics in* (3) *[15]. Therefore, the focus of the remainder of this paper is on the first-order integral dynamics of agents given by* (3)*.*

A. Conventional Leader-Follower Consensus

It is assumed that the information flow among the agents is modeled by an undirected graph $\mathcal{G} = (\mathcal{V}(\mathcal{G}), \mathcal{E}(\mathcal{G}))$ with a node set $\mathcal{V}(\mathcal{G})$ and an edge set $\mathcal{E}(\mathcal{G})$. Furthermore, we make the following assumption on the communication graph.

Assumption 1. *The undirected graph is connected.*

Under Assumption 1, the conventional distributed control approach for achieving the leader-follower consensus in (4) relies on the following control law [16]:

$$u_i(t) = - \sum_{j \in \mathcal{N}_i} a_{ij} \left(x_i(t) - x_j(t) \right) - \alpha_i (x_i(t) - x_0), \tag{6}$$

for $i = 1, \ldots, n$, where $a_{ij} \geq 0$ and $a_{ij} \neq 0$ if the agent i receives information from the agent $j \in \mathcal{N}_i$; otherwise, $a_{ij} = 0$, $\forall j \notin \mathcal{N}_i$. In (6), $\alpha_i \geq 0$ is a pinning gain and $\alpha_i \neq 0$ if the agent i receives information from the leader; otherwise, $\alpha_i = 0$.

The cooperative multi-agent system with the agent dynamics in (3) and the conventional leader-follower consensus control protocol in (6) can be written in a compact form as follows:

$$\dot{\mathbf{x}}(t) = -(\mathcal{L} + A)(\mathbf{x}(t) - \mathbf{1}_n x_0) \tag{7}$$

where $\mathbf{x}(t) = [x_1(t), \ldots, x_n(t)]^T \in \mathbb{R}^n$ is the closed-loop state vector, $\mathcal{L} \in \mathbb{R}^{n \times n}$ is the Laplacian matrix associated with the communication graph $\mathcal{G}$, and $A = \mathrm{diag}(\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^{n \times n}$ is a pining gain matrix. Assuming that $\alpha_i \neq 0$ for at least one agent $i \in \mathcal{V}(\mathcal{G})$ and Assumption 1 holds, $-(\mathcal{L}+A)$ is a Hurwitz matrix [16]. Hence, it can be shown that the conventional distributed controller in (6) guarantees the leader-follower consensus objective given in (4).

B. Integrity Attack Modeling

The conventional leader-follower consensus control scheme in (6) assumes ideal cyber conditions where sensing, network communication, computation, and actuating are under normal conditions. However, they might be subject to uncertainties due to cyberattacks. Among the different types of attacks, false data injection (FDI) is one of the major types of cyberattacks, where the attacker aims to compromise the integrity of cyber information.

In this paper, we consider a malicious adversary capable of injecting false data to agents' update rules of its local estimate (computational node) and control input channels (actuating node). FDI attackers insert erroneous data into these nodes leading the multi-agent system to generate false control inputs applied to agents.

To provide a mathematical model of such FDI cyberattacks, let us define $\beta_i(t-T_i)$, which is a step function that characterizes the cyberattack time profile with $T_i > 0$ being the unknown cyberattack occurrence time. The function $\beta_i(t)$ operates as $\beta_i(t-T_i)=0$, $\forall t < T_i$ and $\beta_i(t-T_i)=1$, $\forall t \geq T_i$. The FDI cyberattack on agent i's update of its state and/or control inputs can be modeled as

$$\tilde{u}_i(t) = u_i(t) + \beta_i(t-T_i)\delta_i(t) \tag{8}$$

where $u_i(t)$ is the control output, $\tilde{u}_i(t)$ is the compromised signal applied to the actuator of the agent i, and $\delta_i(t)$ is the unknown false data injected to the actuator of the agent i. Note that in the normal situation (no attack case), $\tilde{u}_i(t) = u_i(t)$.

Assumption 2. *It is assumed that the false data injection $\delta_i(t)$ in* (8) *and its rate of change are unknown but bounded.*

It is worth mentioning that Assumption 2 is not restrictive, as a worst-case actuator uncertainty is actuator amplitude saturation in practice. This implies that even if the FDI attacks on control inputs are unbounded, due to actuator saturation, their impact becomes bounded.

Assumption 3. *In addition to Assumption 2, it is assumed that the false signal $\delta_i(t)$ in* (8) *is state-independent and does not depend on the state of the agent i, i.e., $x_i(t)$ or other agents' state.*

To characterize the resilient consensus in the presence of aforementioned integrity attacks, we use a formal definition of resilient consensus as follows:

Definition 2 (Resilient Consensus). If for any possible sets and behaviors of misbehaving agents under attacks, the following condition is satisfied, then the multi-agent system is said to reach a resilient consensus:

$$\left| \lim_{t\to\infty} x_i(t) - x_0 \right| \leq \epsilon, \ \forall i \in \{1,\dots,n\} \tag{9}$$

where ϵ is a small non-negative scalar.

The main objective of this paper is to modify the conventional distributed leader-follower consensus in (7) and to enhance its resilience against FDI cyberattacks modeled by (8).

C. Problem Statement

In the following, the research problem that we seek to address in this paper is stated.

Problem 1. Given the dynamics of the multi-agent system in (3), the integrity attack model in (8), and under Assumption 1, design $u_i(t)$ so that the resilience consensus objective in (9) is met for all FDI attacks on control inputs satisfying Assumption 2 and Assumption 3.

In the next section, a solution to Problem 1 is provided that ensures the resilience consensus objective in (9).

III. Proposed Resilient Leader-Follower Consensus

This section provides a resilient distributed consensus algorithm to address Problem 1.

A. Adaptive Leader-Follower Consensus Control

In the presence of FDI attacks on control inputs, as modeled by (8), the leader-follower consensus controller in (7) is modeled as :

$$\dot{\mathbf{x}}(t) = -(\mathcal{L}+A)(\mathbf{x}(t) - \mathbf{1}_n x_0) + \mathbf{d}(t), \tag{10}$$

where $\mathbf{d}(t) = [d_1(t),\dots,d_n(t)]^T \in \mathbb{R}^n$ indicates the impact of attacks on control inputs and agents' update rule, i.e., $d_i(t) = \delta_i(t)$ in the presence of attacks on $u_i(t)$ and $d_i(t) = 0$ in the absence of attacks on $u_i(t)$ for $i = 1,\dots,n$. Based on Assumption 2, it is assumed that $\|\mathbf{d}(t)\|_2 \leq \bar{d}$ and $\|\dot{\mathbf{d}}(t)\|_2 \leq \bar{\dot{d}}$; $t \geq 0$, where the upper bounds $\bar{d}$ and $\bar{\dot{d}}$ are unknown.

To achieve the resilient consensus in (9) in the face of integrity cyberattacks modeled by (8), the control law $u_i(t)$ in (5) is designed as follows:

$$\begin{aligned} u_i(t) = &-\sum_{j\in\mathcal{N}_i} a_{ij}\left(x_i(t) - x_j(t)\right) - \alpha_i(x_i(t) - x_0) \\ &+ d_i(t) + v_i(t), \end{aligned} \tag{11}$$

where $v_i(t)$ is a correction signal aimed to estimate the uncertainty on the agent i's update rule. The correction signal $v_i(t)$ is designed by

$$v_i(t) = -\hat{d}_i(t) \tag{12}$$

where $\hat{d}_i(t)$ is an estimate of the uncertainty (false data) $d_i(t)$ on the agent i's update rule and its dynamics are represented by

$$\dot{\hat{d}}_i(t) = \gamma_i \text{Proj}(\hat{d}_i(t), \sum_{j\in\mathcal{N}_i} a_{ij}\left(x_i(t) - x_j(t)\right) - \alpha_i(x_i(t) - x_0)) \tag{13}$$

with $\hat{d}_i(0) = \hat{d}_{i,0}$, where γ_i is a positive design gain, referred to as the learning rate [17], and Proj is a projection operation defined in (1). In the projection-based control law in (13), the following continuously differentiable and convex function $f(\hat{d}_i)$ is used:

$$f(\hat{d}_i) = \frac{\hat{d}_i^2 - d_{i,max}^2}{2\epsilon_{d_i} d_{i,max}^2 + \epsilon_{d_i}^2} \tag{14}$$

where $d_{i,max} \in \mathbb{R}$ is a projection norm bound imposed on $\hat{d}_i$ and $\epsilon_{d_i} > 0$ is a projection tolerance bound. Note that $f(\hat{d}_i) = 0$ when $\hat{d}_i^2 = d_{i,max}^2$ and $f(\hat{d}_i) = 1$ when $\hat{d}_i^2 = (d_{i,max} + \epsilon_{d_i})^2$. The projection function f in (14) is commonly used in adaptive systems [17], [18], and the references therein. As shown in [17], the choice of f in (14) results in a bounded estimate $\hat{d}_i(t)$.

Remark 2. *Note that the design of $u_i(t)$ in* (11) *and $\hat{d}_i(t)$ in* (13) *is distributed, since both* (11) *and the dynamic update rule for $\hat{d}_i(t)$ only require the states of neighboring agents $x_j(t)$ for $j \in \mathcal{N}_i$.*

B. Closed-loop Dynamical Systems

Let us define $\hat{\mathbf{d}}(t) = [\hat{d}_1(t), \dots, \hat{d}_n(t)]^T \in \mathbb{R}^n$ as a vector of the estimate of uncertainties on agents' control inputs. The closed-loop multi-agent system with the proposed update rule in (11)-(13) is presented in a compact vector form as follows:

$$\begin{aligned}\dot{\mathbf{x}}(t) &= -(\mathcal{L} + A)(\mathbf{x}(t) - \mathbf{1}_n x_0) + \mathbf{d}(t) - \hat{\mathbf{d}}(t),\\ \dot{\hat{\mathbf{d}}}(t) &= \gamma \mathrm{Proj}_n\left(\hat{\mathbf{d}}(t), (\mathcal{L} + A)(\mathbf{x}(t) - \mathbf{1}_n x_0)\right)\end{aligned} \tag{15}$$

where $\gamma = \mathrm{diag}(\gamma_1, \dots, \gamma_n) \in \mathbb{R}^{n \times n}$ is a positive-definite matrix and $\mathrm{Proj}_n\left(\hat{\mathbf{d}}(t), (\mathcal{L} + A)(\mathbf{x}(t) - \mathbf{1}_n x_0)\right) \in \mathbb{R}^n$ is a vector whose i-th element is $\mathrm{Proj}(\hat{d}_i(t), \sum_{j \in \mathcal{N}_i}^n a_{ij}\,(x_i(t) - x_j(t)) - \alpha_i(x_i(t) - x_0))$ for $i = 1, \dots, n$.

In the next section, the stability and the uniform ultimate boundedness of the closed-loop dynamical system in (15) is analyzed.

IV. Stability Analysis

This section analyzes the stability and the uniform ultimate boundedness of the closed-loop system in (15). To this end, let us define $\mathbf{e}(t) = \mathbf{x}(t) - \mathbf{1}_n x_0$ and $\tilde{\mathbf{d}}(t) = \mathbf{d}(t) - \hat{\mathbf{d}}(t)$. The closed-loop dynamics in (15) are then presented in these new coordinates by the following equations:

$$\begin{aligned}\dot{\mathbf{e}}(t) &= -(\mathcal{L} + A)\mathbf{e}(t) + \tilde{\mathbf{d}}(t),\\ \dot{\tilde{\mathbf{d}}}(t) &= -\gamma \mathrm{Proj}_n\left(\mathbf{d}(t) - \tilde{\mathbf{d}}(t), (\mathcal{L} + A)\mathbf{e}(t)\right) + \dot{\mathbf{d}}(t).\end{aligned} \tag{16}$$

The results about the stability of (16) are given in the following theorem.

Theorem 1. *Let Assumption 1-Assumption 3 hold. The proposed resilient leader-follower consensus scheme in* (11)-(13) *ensures the resilient consensus in* (9) *in the presence of FDI attacks on control inputs modeled by* (8) *with the following ultimate bounds:*

$$\begin{aligned}\|\mathbf{e}\|_2 &\leq \sqrt{\frac{\lambda_{max}(\mathcal{L} + A)}{\lambda_{min}(\mathcal{L} + A)}\eta_1^2 + \frac{1}{\lambda_{min}(\gamma)\lambda_{min}(\mathcal{L} + A)}\eta_2^2}\\ \|\tilde{\mathbf{d}}\|_2^2 &\leq \sqrt{\lambda_{max}(\gamma)\lambda_{max}(\mathcal{L} + A)\eta_1^2 + \frac{\lambda_{max}(\gamma)}{\lambda_{min}(\gamma)}\eta_2^2},\end{aligned} \tag{17}$$

where $\eta_1 = \sqrt{\frac{\xi_2}{\xi_1}}$, $\eta_2 = d_{max} + \bar{d}$, $\xi_1 = (\lambda_{min}(\mathcal{L} + A))^2$, $\xi_2 = \frac{1}{\lambda_{min}(\gamma)}(\bar{d} + d_{max})\bar{\dot{d}}$, and $d_{max} \in \mathbb{R}$ is a projection norm bound, i.e., $d_{max} = \max(d_{1,max}, \dots, d_{n,max})$.

Proof: Consider the following quadratic Lyapunov function:

$$V(\mathbf{e}(t), \tilde{\mathbf{d}}(t)) = \frac{1}{2}\mathbf{e}^T(t)(\mathcal{L} + A)\mathbf{e}(t) + \frac{1}{2}\tilde{\mathbf{d}}^T(t)\gamma^{-1}\tilde{\mathbf{d}}(t) \tag{18}$$

Note that as $\mathcal{L} \succeq 0$ (due to Assumption 1) and $A \succeq 0$, Weyl's inequality [19] implies that $(\mathcal{L} + A) \succ 0$. Hence, $V(\mathbf{e}(t), \tilde{\mathbf{d}}(t)) > 0$, $\forall (\mathbf{e}(t), \tilde{\mathbf{d}}(t)) \neq (0, 0)$, $V(0, 0) = 0$, and $V(\mathbf{e}(t), \tilde{\mathbf{d}}(t))$ is radially unbounded.

The time derivative of the Lyapunov function in (18) along the closed-loop error dynamics in (16) is given by

$$\begin{aligned}\dot{V}(\mathbf{e}(t), \tilde{\mathbf{d}}(t)) = &- \mathbf{e}^T(t)(\mathcal{L} + A)^2\mathbf{e}(t) + \mathbf{e}^T(t)(\mathcal{L} + A)\tilde{\mathbf{d}}(t)\\ &- \tilde{\mathbf{d}}^T(t)\mathrm{Proj}_n\left(\mathbf{d}(t) - \tilde{\mathbf{d}}(t), (\mathcal{L} + A)\mathbf{e}(t)\right)\\ &+ \tilde{\mathbf{d}}^T(t)\gamma^{-1}\dot{\mathbf{d}}(t).\end{aligned} \tag{19}$$

Taking into account the property of the projection operator in (2), one can show that

$$-\tilde{\mathbf{d}}^T(t)\left(\mathrm{Proj}_n\left(\hat{\mathbf{d}}(t), \mathcal{Y}(t)\right) - \mathcal{Y}(t)\right) \leq 0, \quad t \geq 0, \tag{20}$$

where $\mathcal{Y}(t) = (\mathcal{L} + A)\mathbf{e}(t)$. Thus, one can obtain that

$$\begin{aligned}\dot{V}(\mathbf{e}(t), \tilde{\mathbf{d}}(t)) \leq &- \mathbf{e}^T(t)(\mathcal{L} + A)^2\mathbf{e}(t) + \tilde{\mathbf{d}}^T(t)\gamma^{-1}\dot{\mathbf{d}}(t)\\ \leq &- \xi_1\|\mathbf{e}\|_2^2 + \xi_2, \quad t \geq 0,\end{aligned} \tag{21}$$

where $\xi_1 = (\lambda_{min}(\mathcal{L} + A))^2$, $\xi_2 = \frac{1}{\lambda_{min}(\gamma)}(\bar{d} + d_{max})\bar{\dot{d}}$ with a projection norm bound $d_{max} > 0$. Consequently, $\dot{V}(\mathbf{e}(t), \tilde{\mathbf{d}}(t)) < 0$ is outside of the following compact set:

$$\Gamma = \{(\mathbf{e}(t), \tilde{\mathbf{d}}(t)) \in \mathbb{R}^n \times \mathbb{R}^n : \quad \|\mathbf{e}\|_2 \leq \eta_1,\ \|\tilde{\mathbf{d}}\|_2 \leq \eta_2\}. \tag{22}$$

where $\eta_1 = \sqrt{\frac{\xi_2}{\xi_1}}$ and $\eta_2 = d_{max} + \bar{d}$. This proves the uniform boundedness of the solution $(\mathbf{e}(t), \tilde{\mathbf{d}}(t))$ of the closed loop system given by (16) for all $(\mathbf{e}(0), \tilde{\mathbf{d}}(0)) \in \mathbb{R}^n \times \mathbb{R}^n$. To show the ultimate bound for $\mathbf{e}(t)$ and $\tilde{\mathbf{d}}(t)$, note that according to (18) and (22), one can obtain that

$$\begin{aligned}&\lambda_{min}(\mathcal{L} + A)\|\mathbf{e}\|_2^2 + \frac{\|\tilde{\mathbf{d}}\|_2^2}{\lambda_{max}(\gamma)} \leq \lambda_{max}(\mathcal{L} + A)\eta_1^2 + \frac{\eta_2^2}{\lambda_{min}(\gamma)},\\ &\lambda_{min}(\mathcal{L} + A)\|\mathbf{e}\|_2^2 \leq \lambda_{max}(\mathcal{L} + A)\eta_1^2 + \frac{\eta_2^2}{\lambda_{min}(\gamma)},\\ &\frac{\|\tilde{\mathbf{d}}\|_2^2}{\lambda_{max}(\gamma)} \leq \lambda_{max}(\mathcal{L} + A)\eta_1^2 + \frac{\eta_2^2}{\lambda_{min}(\gamma)}.\end{aligned} \tag{23}$$

As a result, one can obtain that

$$\begin{aligned}\|\mathbf{e}\|_2 &\leq \sqrt{\frac{\lambda_{max}(\mathcal{L} + A)}{\lambda_{min}(\mathcal{L} + A)}\eta_1^2 + \frac{1}{\lambda_{min}(\gamma)\lambda_{min}(\mathcal{L} + A)}\eta_2^2}\\ \|\tilde{\mathbf{d}}\|_2^2 &\leq \sqrt{\lambda_{max}(\gamma)\lambda_{max}(\mathcal{L} + A)\eta_1^2 + \frac{\lambda_{max}(\gamma)}{\lambda_{min}(\gamma)}\eta_2^2},\end{aligned} \tag{24}$$

This proves the ultimate bound of the error signals $\mathbf{e}(t)$ and $\tilde{\mathbf{d}}(t)$ whose dynamics are given in (16).

The ultimate bounds of $\mathbf{e}(t)$ and $\tilde{\mathbf{d}}(t)$ in (17) characterize the impact of the learning rate matrix γ on these bounds. Hence, γ should be appropriately designed so that those bounds are as small as possible.

Remark 3. *The boundedness of $\tilde{\mathbf{d}}(t)$ in* (17) *and $\mathbf{d}(t)$ (due to Assumption 2) imply that $\hat{\mathbf{d}}(t)$ is bounded.*

The next theorem analyzes the Lyapunov stability of the closed-loop dynamical system in (16) for the case of time-invariant FDI integrity attacks $\mathbf{d}(t)$ (constant false data injection $\mathbf{d}$).

Theorem 2. *Let Assumption 1-Assumption 3 hold. For the case of time-invariant FDI integrity attacks $\mathbf{d}$, the zero solution $(\mathbf{e}(t), \tilde{\mathbf{d}}(t)) = (0, 0)$ of the closed-loop system given by* (16) *is Lyapunov stable for all $(\mathbf{e}(0), \tilde{\mathbf{d}}(0)) \in \mathbb{R}^n \times \mathbb{R}^n$. Furthermore, $\lim_{t\to\infty} \mathbf{e}(t) = 0$ and $\lim_{t\to\infty} \tilde{\mathbf{d}}(t) = 0$.*

Proof: As the integrity attack $\mathbf{d}$ is time-invariant, $\dot{\mathbf{d}} = 0$ and hence $\dot{\bar{d}} = 0$. Using the same procedure as the proof of Theorem 1, it can be shown that

$$\dot{V}(\mathbf{e}(t), \tilde{\mathbf{d}}(t)) = -\mathbf{e}^T(t)(\mathcal{L} + A)^2\mathbf{e}(t)$$
$$- \tilde{\mathbf{d}}^T(t) \Big(\text{Proj}_n \Big(\mathbf{d}(t) - \tilde{\mathbf{d}}(t), (\mathcal{L} + A)\mathbf{e}(t)\Big) - (\mathcal{L} + A)\mathbf{e}(t) \tag{25}$$

As $-(\mathcal{L} + A)^2 \prec 0$ and due to (20), $\dot{V}(\mathbf{e}(t), \tilde{\mathbf{d}}(t)) \leq 0$, $t \geq 0$. We then apply LaSalle's invariance principle [20] to show that the state trajectories $(\mathbf{e}(t), \tilde{\mathbf{d}}(t))$ converge to the largest invariant set in the following set:

$$\mathcal{D} = \Big\{(\mathbf{e}(t), \tilde{\mathbf{d}}(t)) \in \mathbb{R}^n \times \mathbb{R}^n : \ \dot{V}(\mathbf{e}(t), \tilde{\mathbf{d}}(t)) = 0\Big\} \tag{26}$$

From (25), $\dot{V}(\mathbf{e}(t), \tilde{\mathbf{d}}(t)) = 0$ is equivalent to $\mathbf{e}(t) = 0$ and from (16) this leads to $\tilde{\mathbf{d}}(t) = 0$. As a result, the origin is the largest invariant set in $\mathcal{D}$. As a result, $\lim_{t\to\infty} \mathbf{e}(t) = 0$ and $\lim_{t\to\infty} \tilde{\mathbf{d}}(t) = 0$.

Remark 4. *From the results of Theorem 2, it follows that $\lim_{t\to\infty} \mathbf{x}(t) = \mathbf{1}_n x_0$ despite the existence of a time-invariant FDI integrity attack $\mathbf{d}$. This shows that the projection-based update law in* (13) *accurately estimates the injection of time-invariant false data d_i into the control inputs of the agent i; $\forall\ i \in \{1, \ldots, n\}$.*

V. Simulation Results and Analysis

Consider a multi-agent system represented by an undirected graph depicted in Fig. 1 with $n = 34$ followers. It is assumed that only follower 1 has access to the leader's state $x_0 = 1$ and all agents have random initial conditions $x_i(0)$. In the estimation algorithm in (13), $\hat{d}_i(0) = 0$ is selected for $i = 1, \ldots, n$.

The performance of the proposed resilient distributed leader-follower consensus is evaluated in two case studies.

In the first case study, it is assumed that the control inputs of agents, i.e. $u_i(t)$ for $i = 1, \ldots, 34$, are subject to constant random FDI attacks occurring in $t = 5\ s$.

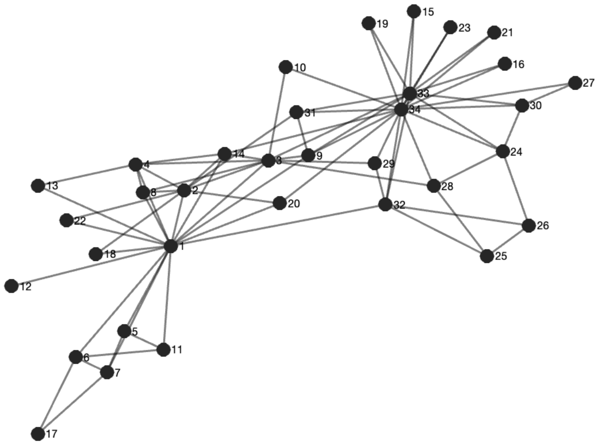

Fig. 1. Undirected graph used in Section V- Gray lines denote communication links and black circles denote follower agents.

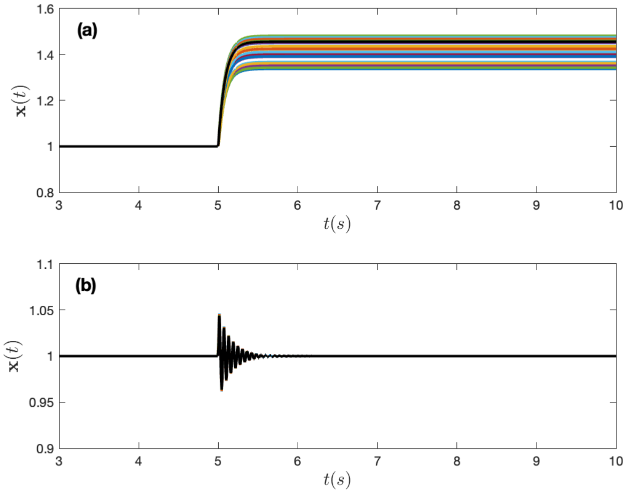

Fig. 2. State trajectories of the multi-agent system in the presence of constant random FDI attacks on control inputs with (a) conventional distributed algorithm in (6) and (b) proposed resilient distributed framework in (15) with $\gamma = 1000\mathbf{I}_n$.

The second case study considers the case of time-varying FDI attacks on the followers' control inputs. For this purpose, the control inputs are subjected to a combination of constant and sinusoidal signals with a random range of frequencies occurring at $t = 5\ s$.

The state trajectories of the follower agents for both case studies are respectively depicted in Fig. 2 and Fig. 3.

As one can observe from both figures, the proposed resilient distributed control scheme can achieve a resilient consensus even when agents are subject to FDI attacks while the conventional distributed leader-follower algorithm in (6) cannot reach a consensus in the presence of FDI attacks.

VI. Conclusions

This paper deals with the problem of leader-follower consensus in multi-agent systems in the presence of cy-

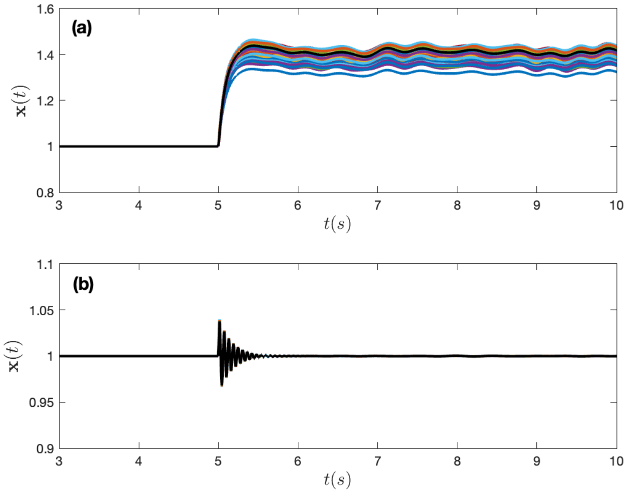

Fig. 3. State trajectories of the multi-agent system in the presence of time-varying attacks on control inputs with (a) conventional distributed algorithm in (6) and (b) proposed resilient distributed framework in (15) with $\gamma = 1000\mathbf{I}_n$.

berattacks on control input channels. The paper develops a novel resilient distributed algorithm for agents to resiliently track a reference state propagated by a set of leaders, despite the existence of unknown but bounded cyberattacks on the agents' local computation and control inputs. A projection-based dynamic scheme is used to estimate the uncertainties on the control input channels of each agent and to ensure the boundedness of the estimation. Our proposed approach guarantees the uniform ultimate boundedness of the consensus and the estimation error in the presence of FDI adversarial integrity attacks. Numerical simulation results are provided to complement the theoretical analysis and demonstrate the effectiveness of the proposed resilient distributed leader-follower consensus scheme. The future scope of this work includes (i) the extension of results to the case of FDI attacks on both sensors and control inputs and (ii) considering the case of time-varying learning rate.

References

[1] M. Jamali, H. R. Baghaee, M. S. Sadabadi, G. B. Gharehpetian, and A. Anvari-Moghaddam, "Distributed cooperative event-triggered control of cyber-physical AC microgrids subject to denial-of-service attacks," *IEEE Transactions on Smart Grid*, vol. 14, no. 6, pp. 4467–4478, 2023.

[2] M. S. Sadabadi, "A resilient-by-design distributed control framework for cyber-physical DC microgrids," *IEEE Transactions on Control Systems Technology*, vol. 32, no. 2, pp. 625–636, 2024.

[3] S. Safavi and U. A. Khan, "Leader-follower consensus in mobile sensor networks," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2249–2253, 2015.

[4] Y. Li, C. Tang, S. Peeta, and Y. Wang, "Nonlinear consensus-based connected vehicle platoon control incorporating car-following interactions and heterogeneous time delays," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2209–2219, 2019.

[5] M. F. Ahern, "Cybersecurity in power systems," *IEEE Potentials*, vol. 36, no. 5, pp. 8–12, 2017.

[6] J. Usevitch and D. Panagou, "Resilient leader-follower consensus to arbitrary reference values in time-varying graphs," *IEEE Transactions on Automatic Control*, vol. 65, no. 4, pp. 1755–1762, 2020.

[7] H. Rezaee, T. Parisini, and M. M. Polycarpou, "Resiliency in dynamic leader-follower multiagent systems," *Automatica*, vol. 125, p. 109384, 2021.

[8] J. Usevitch and D. Panagou, "Resilient leader-follower consensus to arbitrary reference values," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 1292–1298.

[9] A. Gusrialdi, Z. Qu, and M. A. Simaan, "Competitive interaction design of cooperative systems against attacks," *IEEE Transactions on Automatic Control*, vol. 63, no. 9, pp. 3159–3166, Sept. 2018.

[10] H. Dong, C. Li, and Y. Zhang, "Resilient consensus of multi-agent systems against malicious data injections," *Journal of the Franklin Institute*, vol. 357, no. 4, pp. 2217–2231, 2020.

[11] M. S. Sadabadi, M. W. S. Atman, A. Aynala, and A. Gusrialdi, "Resilient design of leader–follower consensus against cyber-attacks," *IEEE Transactions on Control of Network Systems*, vol. 11, no. 2, pp. 1080–1092, 2024.

[12] G. Larchev, S. Campbell, and J. Kaneshige, "Projection operator: A step toward certification of adaptive controllers," in *AIAA Infotech@ Aerospace 2010*, 2010, p. 3366.

[13] E. Lavretsky and K. A. Wise, *Robust and Adaptive Control*. London, UK: Springer, 2012.

[14] M. S. Sadabadi, S. Sahoo, and F. Blaabjerg, "A fully resilient cyber-secure synchronization strategy for AC microgrids," *IEEE Transactions on Power Electronics*, vol. 36, no. 12, pp. 13 372–13 378, 2021.

[15] Z. Qu and M. A. Simaan, "Modularized design for cooperative control and plug-and-play operation of networked heterogeneous systems," *Automatica*, vol. 50, no. 9, p. 2405–2414, 2014.

[16] H. Zhang, F. L. Lewis, and A. Das, "Optimal design for synchronization of cooperative systems: State feedback, observer and output feedback," *IEEE Transactions on Automatic Control*, vol. 56, no. 8, pp. 1948–1952, Aug. 2011.

[17] J. E. Gaudio, A. M. Annaswamy, E. Lavretsky, and M. A. Bolender, "Parameter estimation in adaptive control of time-varying systems under a range of excitation conditions," *IEEE Transactions on Automatic Control*, vol. 67, no. 10, pp. 5440–5447, 2022.

[18] E. Lavretsky, T. E. Gibson, and A. M. Annaswamy, "Projection operator in adaptive systems," Tech. Rep., 2012. [Online]. Available: https://arxiv.org/pdf/1112.4232.pdf

[19] R. A. Horn and C. R. Johnson, *Matrix Analysis*. United States of America: Cambridge University Press, 1990.

[20] H. K. Khalil, *Nonlinear Systems*. New Jersey: Prentice Hall, 2006.