Towards Resilient Design of Leader-following Consensus with Attack Identification and Privacy Preservation Capabilities

Azwirman Gusrialdi, Muhammad Iqbal, and Zhihua Qu

***Abstract*— This paper considers a leader-following consensus in the presence of unknown but bounded cyber-attacks. Specifically, we consider the following cyber-attack scenarios: (i) an attacker aims to destabilize the consensus dynamics by injecting exogenous signals to both the actuators of the followers and/or the communication network, (ii) an eavesdropper adversary aims to obtain information on the physical state of the agents. To this end, a novel resilient leader-following consensus algorithm based on a competitive interaction method is proposed. In addition, it is demonstrated that by appropriately choosing the information exchanged between the agents, the proposed control framework also enables the cooperative system to either distributively identify the compromised communication links in real-time or to protect the privacy of the physical state of the agents from the eavesdropper. A numerical example is provided to illustrate the proposed resilient control algorithms.**

***Index Terms*— Resilient control, attack identification, leader-following consensus, distributed algorithm, privacy.**

I. Introduction

Leader-following consensus is a class of cooperative control systems which has been widely used in various applications such as intelligent transportation systems, smart grid, and robotic network [1]–[4], mainly due to its scalability and robustness to a single point of failure. While the use of communication network enables the design and implementation of the leader-following consensus algorithm, it comes at a price of making the cooperative system vulnerable to cyber-attacks [5]. Specifically, the adversary may intercept and observe the information being exchanged between the agents/nodes (violate the data confidentiality) or he/she may inject malicious signals into the communication links which deteriorate the system's performance and in the worst case destabilize the overall system by taking advantage of the tight coupling between the communication network and the physical system. Therefore, it is of importance to ensure resilient operation of the leader-following consensus in the presence of unknown cyber-attacks and further protect the data of interest from the adversary.

The main focus of this paper is on cyber-attacks on both the actuator of agents and the communication links used to exchange information among the agents. A solution to this problem is by identifying the compromised nodes or communication links followed by their removal [6], [7]. However, the method requires the network to be highly connected to guarantee the existence of a spanning tree after the links removal. Furthermore, the system's stability may already have been compromised before the attack is detected. A mean subsequence reduced (MSR) algorithm has been proposed for leader-following consensus in presence of misbehaving agents [8], [9]. However, this strategy imposes a restriction on the number of compromised nodes and the network connectivity. Furthermore, this method is not effective in ensuring consensus among the agents when only the communication network is being compromised and all the agents are not adversarial. Various resilient leader-following consensus algorithms have been developed in the literature which remove the requirement of high network connectivity, attack detection, and restriction on the number of attacks, see e.g., [10]–[15]. Note that the strategies proposed in the above work rely on exchanging the physical states of followers and the leader with their neighbors which disclosure the agents' physical state. Hence, the cooperative systems is at the risk of potential privacy threats by the eavesdropper adversaries. While there exists a line of work on privacy-preserving leader-following consensus, e.g., [16], the resiliency of the systems is not guaranteed. The work [17] proposes a differentially private MSR algorithm to ensure both resiliency and differential privacy requirements. However, the result is limited to a leaderless consensus and requires a high network connectivity and knowledge on the upper bound of the number of attacks.

This paper presents a novel resilient leader-following consensus algorithm in presence of cyber-attacks on both the actuator and communication network using the idea of competitive interaction method [12], [18], see Section III. In addition to ensuring resilient operation, the proposed control framework also equips the cooperative system with one of the following capabilities, depending on the choice of the information being exchanged between the agents:

1) real-time and distributed identification of the compromised communication links (Section III-A). Hence, the proposed resilient control unifies the attack-detection based method and resilient control approach described previously. A unique feature about our result is that the stability of the cooperative system is ensured during the attack identification process. Note that a similar idea has been used in our previous work [19] for the leaderless consensus. However, the attack model considered in [19] is more restrictive, that is limited to

A. Gusrialdi and M. Iqbal are with Faculty of Engineering and Natural Sciences, Tampere University, Tampere 33014, Finland. Z. Qu is with Department of Electrical and Computer Engineering, University of Central Florida, Orlando 32816, USA. Emails: *azwirman.gusrialdi@tuni.fi*, *muhammad.iqbal@tuni.fi* and *qu@ucf.edu*. The work of A. Gusrialdi and M. Iqbal was supported by Academy of Finland under academy project decision no. 330073. The work of Z. Qu was supported in part by US Department of Energy's awards DE-EE0007998, DE-EE0009028, DE-EE0009152, and DE-EE0009339.

attack on the communication links where the adversary can only insert the same injection signals to the links connected to a particular node.

2) protection of the physical state of the agents from the eavesdropper adversaries (Section III-B). Note that in contrast to the related work on privacy-preserving leader-following consensus discussed previously, the proposed method in this paper simultaneously guarantees the resiliency of the cooperative system.

II. Problem Formulation

In this section, we first provide a brief overview of graph theory followed by describing the problem formulation.

A. Notation and Preliminaries

Let $\mathbb{R}$ be the set of real numbers; vector $\mathbf{1}_n \in \mathbb{R}^n$ denotes the vector of all ones. Cardinality of a set $\mathcal{N}$ is denoted by $|\mathcal{N}|$. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph with a set of nodes $\mathcal{V} = \{1, 2, \cdots, p\}$ and a set of edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. An edge $(j, i) \in \mathcal{E}$ denotes that node i can receive information from node j. The graph $\mathcal{G}$ is undirected if and only if $(j, i) \in \mathcal{E} \Leftrightarrow (i, j) \in \mathcal{E}$ for all edges in $\mathcal{E}$. The neighbor set of node i is defined as $\mathcal{N}_i = \{j | (j, i) \in \mathcal{E}, j \neq i\}$, that is the set of nodes from which node i can receive information.

B. Problem Statement

Consider a cooperative system consisting of $n + 1$ nodes where a leader node is labeled by 0 and the follower nodes are labeled by $i = 1, \cdots, n$. Let $x_i \in \mathbb{R}$ denote the physical state of follower node i whose dynamics is given by

$$\dot{x}_i = u_i\left(J_i, \{I_j\}\right), \;\; i = \{1, \cdots, n\}, \;\; j \in \mathcal{N}_i, \qquad (1)$$

where u_i is the control input, J_i denotes the local information of node i and I_j denotes the information sent by nodes j (including the leader node) which are the neighbours of node i . The communication network topology among the follower nodes and between the leader node and the follower nodes is given by the following assumption.

Assumption 1: The communication network topology among the follower nodes is given by a connected undirected graph. In addition, there exists communication link(s) from the leader node to at least one of the follower nodes.

Assumption 1 can be found for example in the intelligent transportation system [1].

In practice, the actuators of the follower nodes and the communication channels are vulnerable to cyber-attacks. In this paper we consider the following class of cyber-attacks as illustrated in Fig 1.

A1. False data injection (FDI) attacks on the actuators and communication channels. The attack on the actuator of the follower node can be modeled as

$$\tilde{u}_i(t) = u_i(t) + \delta_{ui}(t), \qquad (2)$$

where $\tilde{u}_i(t)$ is the compromised control input under unknown injection $\delta_{ui}(t)$. In the case of FDI attacks on the communication channel, follower node i may not receive the true information from its neighboring node j, that is the possibly corrupted information that node i is receiving from its neighbour, that is node j, including the leader node, takes the following form

$$I_i^j = I_j + \delta_{ij}(t), \; j \in \mathcal{N}_i, \qquad (3)$$

where $\delta_{ij}(t)$ is the malicious injection into the communication channel.

A2. Eavesdropping attack on the communication channel where the goal of the adversary is to obtain information on the physical state x_i of both the leader and follower nodes.

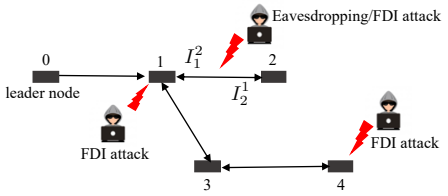

Fig. 1: Multi-agent system with a leader and followers under FDI attack and eavesdropping attack: Nodes and links with red color are attacked

Next, we introduce the following assumption on the injections δ_{ui}, δ_{ij} in (2) and (3).

Assumption 2: The injections δ_{ui}, δ_{ij} and their derivatives are both uniformly bounded.

This assumption is reasonable in practice and has been considered for example in power system applications [20]. Note that no restrictions are made on the number of attacks and in contrast to [9], [10], [21] in this paper we assume that the communication link from the leader to the follower nodes can also be compromised.

The objective of this paper is to design the control input $u_i(t)$ such that the cooperative system in (1) reaches an approximate consensus, that is

$$\left| \lim_{t \to \infty} x_i(t) - x_0 \right| \leq \epsilon, \; i = 1, \cdots, n, \qquad (4)$$

where the constant value of $x_0 \in \mathbb{R}$ is the state of the leader node and ϵ is a small non-negative scalar. Furthermore, the control input $u_i(t)$ should also meet one of the following additional goals:

1) detect and identify in a distributed manner all the compromised communication links
2) protect the privacy of all the nodes' physical states x_i against the eavesdropping adversaries

III. Main Results: Resilient Design

We propose the following control law $u_i(t)$ for the i-th follower node

$$\begin{aligned} u_i &= -|\mathcal{N}_i|[x_i - \beta_1 \alpha(t) z_i] + \sum_{j=0}^{n} a_{ij}[x_j - \beta_1 \alpha(t) z_j], \\ \dot{z}_i &= -|\mathcal{N}_i|[\beta_1 \alpha(t) x_i + \beta_2 z_i] + \sum_{j=0}^{n} a_{ij}[\beta_2 z_j + \beta_1 \alpha(t) x_j], \end{aligned} \qquad (5)$$

where $a_{ij} = 1$ if follower node i can receive information from node j and $a_{ij} = 0$ otherwise. Furthermore, $\beta_1, \beta_2 > 0$

are scalar gains and scalar $\alpha(t) > 0$ is any time-varying function chosen such that $\alpha(t)$ is uniformly bounded away from zero and that $\dot{\alpha}(t)$ exists and is uniformly bounded. Finally, $z_i(t) \in \mathbb{R}$ is an auxiliary state of the i-th follower node. In contrast to the physical state $x_i(t)$, the state $z_i(t)$ does not have any physical meaning and its initial value $z_i(0)$ can be set to any arbitrary values. Hence, we also call $z_i(t)$ as the virtual state of node i. Referring to the notations in (1), the local information of a follower node i are given by

$$J_i = \{x_i - \beta_1\alpha(t)z_i, \;\; \beta_1\alpha(t)x_i + \beta_2 z_i\}$$

while the information being sent by node j can take one of the following possibilities, depending on the design objectives (as will be discussed in more details later).

C1. $I_j \in \{I_{j,1}, I_{j,2}, I_{j,3}\}$ where

$$I_{j,1} = x_j, \;\; I_{j,2} = \beta_1\alpha(t)z_j, \;\; I_{j,3} = \beta_2 z_j + \beta_1\alpha(t)x_j \quad (6)$$

C2. $I_j \in \{I_{j,a}, I_{j,b}\}$ where

$$I_{j,a} = x_j - \beta_1\alpha(t)z_j, \;\; I_{j,b} = \beta_2 z_j + \beta_1\alpha(t)x_j. \quad (7)$$

Defining vectors $x = [x_1, \cdots, x_n]^T$ and $z = [z_1, \cdots, z_n]^T$ we can write in a compact form the closed-loop system (1) for all the follower nodes under control law (5) and in the presence of unknown attacks δ_{ui}, δ_{ij} as

$$\begin{aligned} \dot{x} &= Ax + Bx_0 - \beta_1\alpha(t)Az - \beta_1\alpha(t)Bz_0 + d_x \\ \dot{z} &= \beta_2 Az + \beta_1\alpha(t)Ax + \beta_1\alpha(t)Bx_0 + \beta_2 Bz_0 + d_z, \end{aligned} \quad (8)$$

where the injections δ_{ui}, δ_{ij} are lumped into the attack vectors d_x and d_z. The symmetric matrix $A \in \mathbb{R}^{n\times n}$ is defined as

$$A = \begin{bmatrix} -\sum_{j=0}^{n} a_{1j} & a_{12} & \cdots & a_{1n} \\ a_{21} & -\sum_{j=0}^{n} a_{2j} & \cdots & a_{2n} \\ \vdots & & \ddots & \\ a_{n1} & a_{n2} & \cdots & -\sum_{j=0}^{n} a_{nj} \end{bmatrix}$$

while vector $B \in \mathbb{R}^{n\times 1}$ is defined as $B = [a_{10}, \cdots, a_{n0}]^T$. It can be observed that matrix A is Hurwitz [22]. Furthermore, we have the following relationship $A\mathbf{1}_n = -B$.

Before proceeding, we first analyze the stability of closed-loop system (8) in the absence of cyber-attacks. The following lemma provides condition on the gains β_1, β_2 to ensure the convergence of physical state x_i to the leader value x_0 in the absence of attacks.

Lemma 1: Consider the overall system (8) whose communication network topology is given by Assumption 1 and injections $d_x = d_z = 0$. For any $\beta_1, \beta_2 > 0$, the physical state x_i converges to x_0 for all $i = \{1, \cdots, n\}$, that is $x \to \mathbf{1}_n x_0$ as $t \to \infty$.

Proof: Let us define the following error vectors

$$\overline{x} = x - \mathbf{1}_n x_0, \;\; \overline{z} = z - \mathbf{1}_n z_0. \quad (9)$$

By noting $A\mathbf{1}_n = -B$, dynamics of the error $\overline{x}$ can be calculated as

$$\begin{aligned} \dot{\overline{x}} &= A\overline{x} + (A\mathbf{1}_n + B)x_0 - \beta_1\alpha(t)A\overline{z} - \beta_1\alpha(t)(A\mathbf{1}_n + B)z_0 \\ &= A\overline{x} - \beta_1\alpha(t)A\overline{z}. \end{aligned}$$

Similarly, dynamics of the error $\overline{z}$ can also be calculated as

$$\dot{\overline{z}} = \beta_2 A\overline{z} + \beta_1\alpha(t)A\overline{x}.$$

Error dynamics of $\overline{x}, \overline{z}$ can be written in a compact form as

$$\begin{bmatrix} \dot{\overline{x}} \\ \dot{\overline{z}} \end{bmatrix} = \left(\begin{bmatrix} 1 & -\beta_1\alpha(t) \\ \beta_1\alpha(t) & \beta_2 \end{bmatrix} \otimes A \right) \begin{bmatrix} \overline{x} \\ \overline{z} \end{bmatrix}. \quad (10)$$

Since A is Hurwitz, the system (10) has a unique equilibrium equal to zero if and only if $\beta_1^2\alpha^2(t) + \beta_2 \neq 0$ which is satisfied for the choices of $\beta_1, \beta_2, \alpha(t)$ described previously. Next, let us choose the following Lyapunov function candidate

$$V = \overline{x}^T\overline{x} + \overline{z}^T\overline{z}.$$

Computing the derivative of V along the trajectories of $\overline{x}, \overline{z}$ and for any values $\beta_1, \beta_2 > 0$ yields

$$\begin{aligned} \dot{V} &= 2\overline{x}^T[A\overline{x} - \beta_1\alpha(t)A\overline{z}] + 2\overline{z}^T[\beta_2 A\overline{z} + \beta_1\alpha(t)A\overline{x}] \\ &= 2\overline{x}^T A\overline{x} + 2\overline{z}^T\beta_2 A\overline{z} < 0. \end{aligned}$$

Therefore, it can be concluded that $\overline{x} \to 0$, that is $x \to \mathbf{1}_n x_0$ as $t \to \infty$ which completes the proof. ■

The following theorem shows that the proposed resilient control can ensure the follower nodes to achieve approximate consensus in the presence of bounded but unknown injections.

Theorem 1: Consider the overall system (8) whose communication network topology is given by Assumption 1 and the unknown injections satisfy Assumption 2. For a large value of $\beta_1 > 0$ and a small value of $\beta_2 > 0$, the approximate consensus (4) is ensured for all the follower nodes.

Proof: Similar to the proof of Lemma 1, defining the errors $\overline{x}, \overline{z}$ as in (9) and taking its derivatives results in the following error dynamics

$$\begin{aligned} \dot{\overline{x}} &= A\overline{x} - \beta_1\alpha(t)A\overline{z} + d_x \\ \dot{\overline{z}} &= \beta_2 A\overline{z} + \beta_1\alpha(t)A\overline{x} + d_z. \end{aligned} \quad (11)$$

Let us now take the following Lyapunov function

$$V' = \beta_1\overline{x}^T\overline{x} + \beta_1\overline{z}^T\overline{z} - 2\overline{z}^T d'_x + 2\overline{x}^T d'_z$$

where $d'_x = \alpha(t)^{-1}A^{-1}d_x$ and $d'_z = \alpha(t)^{-1}A^{-1}d_z$. It follows that d'_x, d'_z and $\dot{d}'_x, \dot{d}'_z$ are all uniformly bounded as $d_x, d_z, \alpha(t)$ and $\dot{\alpha}(t)$ are uniformly bounded.

Taking the derivative of V' along trajectories (11) yields

$$\begin{aligned} \dot{V}' &= 2\beta_1\overline{x}^T A\overline{x} - \cancel{2\beta_1^2\alpha(t)\overline{x}^T A\overline{z}} + \cancel{2\beta_1\overline{x}^T d_x} + 2\beta_1\beta_2\overline{z}^T A\overline{z} \\ &\quad + \cancel{2\beta_1^2\alpha(t)\overline{x}^T A\overline{z}} + \cancel{2\beta_1\overline{z}^T d_z} - 2\overline{z}^T\dot{d}'_x - 2\beta_2\overline{z}^T A d'_x \\ &\quad \underbrace{-2\beta_1\alpha(t)\overline{x}^T A d'_x}_{\cancel{-2\beta_1\overline{x}^T d_x}} - 2d_z^T d'_x + 2\overline{x}^T\dot{d}'_z + 2\overline{x}^T A d'_z \\ &\quad \underbrace{-2\beta_1\alpha(t)\overline{z}^T A d'_z}_{\cancel{-2\beta_1\overline{z}^T d_z}} + 2d_x^T d'_z \\ &= 2\beta_1\overline{x}^T A\overline{x} + 2\beta_1\beta_2\overline{z}^T A\overline{z} - 2\overline{z}^T\dot{d}'_x - 2\beta_2\overline{z}^T A d'_x \\ &\quad + 2\overline{x}^T\dot{d}'_z + 2\overline{x}^T A d'_z + 2d_x^T d'_z - 2d_z^T d'_x. \end{aligned}$$

Next, substituting $d'_x = \alpha(t)^{-1}A^{-1}d_x$ and $d'_z = \alpha(t)^{-1}A^{-1}d_z$ into the last two terms of $\dot{V}'$ above yields

$$\begin{aligned} \dot{V}' &= 2\beta_1\overline{x}^T A\overline{x} + 2\beta_1\beta_2\overline{z}^T A\overline{z} - 2\overline{z}^T\dot{d}'_x - 2\beta_2\overline{z}^T A d'_x \\ &\quad + 2\overline{x}^T\dot{d}'_z + 2\overline{x}^T A d'_z. \end{aligned}$$

Since d'_x, d'_z and their derivatives are bounded and noting that the matrix A is Hurwitz, for a large value of β_1 and small value of β_2 we have $\dot{V}' < -Q(x) < 0$ where $Q(x)$ is positive definite which shows that approximate consensus is achieved in presence of unknown injections d_x, d_z. This completes the proof. ■

Remark 3.1: If an estimate of the worst case attack is available, one can fix β_1 and β_2 while keeping the value of β_2 relatively small compared to β_1. Otherwise, one could make β_1 adaptive which is a subject of future work. When the gains β_1, β_2 are constant their values can be fixed before the deployment of the cooperative system. Furthermore, since the time-varying function $\alpha(t)$ is independent of the node's states, it can also be designed in advance and later during the execution of the resilient control each node can individually update the values of $\alpha(t)$ in real-time. Hence, the proposed resilient leader-following consensus algorithm can be implemented distributively.

Remark 3.2: The control law (5) also ensures stability of the cooperative system under noisy communication channel as the resulting system can also be written as in (8).

In addition to ensuring approximate consensus in presence of unknown injections and depending on the information being exchanged between the nodes, the proposed resilient cooperative control (5) also allows one to either identify the compromised communication link in real-time and distributed manner or to preserve the privacy of the physical state x_i from the eavesdropping attacks as discussed in the following subsections.

A. Real-time and Distributed Attack Identification

Let us consider the case where node j sends the information $I_j \in \{I_{j,1}, I_{j,2}, I_{j,3}\}$ as defined in (6) to its neighbors. We will demonstrate how each node can identify the compromised communication link using those information.

Before proceeding, let us introduce the following definition of a *stealthy* attack.

Definition 1: An attack launched at $t = t_a$ on the link (j, i) is *stealthy* if $\hat{I}_i^j(t) = I_i^j(t)$, $\forall t \geq t_a$ where $\hat{I}_i^j(t)$ is the estimation of the information $I_j(t)$ at node i.

First, from the (possibly corrupted) information $I_{i,2}^j, I_{i,3}^j$ received from node j in (6) the i-th node can estimate the information $I_{j,1}, j \in \mathcal{N}_i$, i.e., x_j given by

$$\hat{I}_{i,1}^j = \frac{1}{\beta_1 \alpha(t)} \left[I_{i,3}^j - \frac{\beta_2}{\beta_1} \frac{I_{i,2}^j}{\alpha(t)} \right]. \tag{12}$$

We then propose the following detection test for each node where the idea is to compare the estimated information $\hat{I}_{i,1}^j$ with the possibly corrupted one, i.e., $I_{i,1}^j$ directly obtained via the communication network.

$$\textbf{Detection test:} \quad \hat{I}_{i,1}^j = I_{i,1}^j. \tag{13}$$

If $\hat{I}_{i,1}^j \neq I_{i,1}^j$, follower node i can identify that the communication link (j, i) is being attacked. On the other hand, when $\hat{I}_{i,1}^j = I_{i,1}^j$ then there are two possible conclusions that node i can take. The first conclusion is that there exists no attacks on the communication link (j, i), that is $\delta_{ij} = 0$. The second one is that the attacker launches a harmful but stealthy attack on the communication link (j, i), see Definition 1. However, it is very challenging for the adversary to launch a harmful and stealthy attack as he/she needs to know the structure of the detection test in (12) together with the values of scalar gains β_1, β_2 and time varying function $\alpha(t)$. Note that the values of β_1, β_2 and time varying function $\alpha(t)$ are local information to the individual nodes and not directly communicated via the communication network as can be observed in (6). Hence, it is not possible for the adversary to learn both the structure in (12) and the time-varying function $\alpha(t)$ in real-time using solely the information on I_j.

Remark 3.3: A clock synchronization is required for the nodes (and assumed to be achieved) in order to effectively detect and identify the compromised communication link using detector test (13).

B. Privacy Preservation against Eavesdropping Attacks

Next, let us consider the case where the node sends information $I_i \in \{I_{i,a}, I_{i,b}\}$ as defined in (7) to its neighbors. The virtual state $z_i(t)$ together with the scalar gains β_1, β_2 and time-varying function $\alpha(t)$ act as dynamic masks on the physical state x_i. In addition, it can be observed from (7) that each node, including the leader, does not send directly the physical state x_i to its neighbors but instead it sends the masked physical states given by $I_{i,a}, I_{i,b}$. Note again that the structure of (7), the gains β_1, β_2 and the time-varying function $\alpha(t)$ are local knowledge to individual node and unknown to the adversary. Therefore, the adversary will not be able to learn the physical state $x_i(t)$ using only the information on I_i and without having the knowledge on β_1, β_2 and $\alpha(t)$. In addition, when there is no FDI attack even though the follower nodes converge to the leader state x_0 (see Lemma 1) one can design $\alpha(t)$ so that the information $I_{i,a}, I_{i,b}$ will not converge to any values and thus the consensus value x_0 can also be shielded from the adversary.

Remark 3.4: As shown in Theorem 1, the follower nodes can still track the leader's value x_0 even though the eavesdropper adversary inject malicious cyber-attack signals into the communication links. However, the follower nodes will not be able to identify the compromised communication links using both the information $I_{i,a}^j$ and $I_{i,b}^j$.

The discussions in Sections III-A and III-B demonstrate that by exchanging appropriate information, in addition to ensuring approximate consensus in presence of FDI attacks the cooperative system can also either identify in real-time and distributed manner the compromised communication links or enhance the privacy of the physical states against eavesdropping attack.

IV. A Numerical Example

In this section, we provide simulation results, which demonstrate our theoretical results. In Subsection IV-A, we demonstrate that the consensus protocol (5) solves resilient consensus problem in the presence of adversaries.

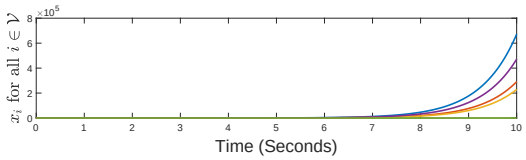

Fig. 2: Attack on multi-agent systems with a standard leader-following consensus protocol.

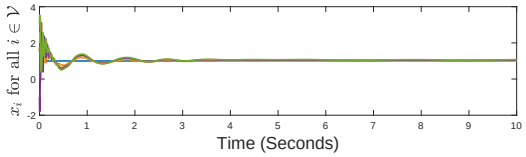

Fig. 3: Followers converge to the state value of the leader (denoted by the blue line) in the presence of cyber-attacks where the agents using the strategy (5) with $\alpha(t)$ given in (15), $\beta_1 = 40$ and $\beta_2 = 10$.

Subsection IV-B discusses how the information exchanging mechanism in (6) can be used to identify cyber-attacks. In Subsection IV-C, we demonstrate that the privacy of the physical state can be preserved using (7) from an eavesdropping attacker.

We will use the following setup in the subsequent discussion. Consider the leader-following multi-agent system given in Fig. 1, where the number of followers are $n = 4$. The value of the leader's state is set to $x_0 = 1$. Moreover, the dynamics of the attack in (8) is given below:

$$\dot{d}_x = F_1 d_x + B_a x, \quad \dot{d}_z = F_2 d_z + B_a x, \tag{14}$$

where $x = [x_1, \cdots, x_4]^{\mathrm{T}}$ is the state vector of the followers agent, $F_1 = -\mathbb{I}_n$, $F_2 = -2\mathbb{I}_n$ $d_x \in \mathbb{R}^n$ and $d_z \in \mathbb{R}^n$ are attack vectors on both the actuators and communication links. Here,

$$B_a = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note that the adversary has an access to the physical state, and therefore can destabilize any multi-agent system with a standard leader-following consensus protocol, that is by setting $\beta_1 = 0$ in (8), as shown in Fig. 2.

A. Resilient Leader-Following Consensus

Next, we apply the proposed resilient leader-following consensus algorithm (8) where the cyber-attacks d_x, d_z are also given in (14). That is, each agent takes decision based on the strategy given in (5), where $\beta_1 = 40$, $\beta_2 = 10$, and time varying function $\alpha(t)$ is set to

$$\alpha(t) = 0.1(\sin(t) + \cos(2t) + \sin(3t)) + 1. \tag{15}$$

Fig. 3 shows four followers converge to the static value of the leader's state. For a fixed $\beta_1 = 40$ and low $\beta_2 = 1$, Fig. 4 shows that the oscillations' tail is longer due to the fact that β_2 is low, which is an integral gain. Note that in order to make the overall system stable, the value of β_1 should be kept high as with low $\beta_1 = 1$, the state trajectories diverge as shown in Fig. 5.

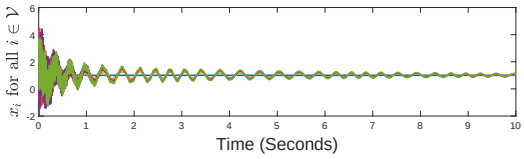

Fig. 4: Followers converge to the state value of the leader (denoted by the blue line) in the presence of cyber-attacks where the agents using the strategy (5) with $\alpha(t)$ given in (15), $\beta_1 = 100$ and $\beta_2 = 1$.

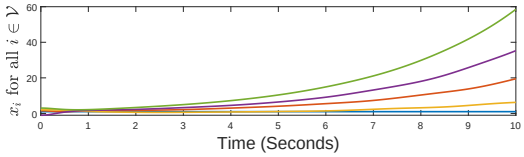

Fig. 5: Followers diverging from the state value of the leader in the presence of cyber-attacks with low β_1.

B. Real-time and Distributed Attack Detection

By exchanging the information I_j as given in (6), each agent is able to detect the polluted information in real-time and distributed manner. In Fig. 6, we demonstrate that Agent 1 detects the cyber-attack on link $(1, 2)$ using detection test (13), that is by comparing the information $\hat{I}^2_{1,1}$ and $I^2_{1,1}$, where $\beta_1 = 40$, $\beta_2 = 10$ and $\alpha(t)$ is given in (15). It is worth to note that during the attack identification the resiliency of the cooperative system is ensured using the proposed method. Furthermore, Agent 1 can also conclude that the communication link $(1, 3)$ is not being compromised as the difference $\hat{I}^3_{1,1} - I^3_{1,1}$ is equal to zero as shown in Fig. 7.

C. Privacy Preservation of the Physical States

With the competitive-based interaction method given in (5), one can mask the physical state's information sent to the neighbors by exchanging the information as defined in (7). This allows us to protect the multi-agent systems from eavesdropping attack, where an attacker is curious to know the physical state information x_i. Note that, without the loss of generality, the attack in (14) is not considered in this subsection. In Fig. 8, we see that the information x_0, $I_{0,a}$ and $I_{0,b}$ are drastically changed from one another. That is, even though the leader's information x_o is constant, the information being sent by the leader is time-varying. As $I_{0,a}$ and $I_{0,b}$ are sent instead of x_0, and that the gains β_1, β_2, time-varying function $\alpha(t)$ including how the information in (7) is defined, are local information to each agent, thus making it extremely difficult for the eavesdropping attack to predict the physical state x_0. Next, in Figs. 9 and 10, we see how the information x_3 is masked using $I_{3,a}$ and $I_{3,b}$ with $\beta_1 = 40$, $\beta_2 = 10$ and $\alpha(t)$ given in (15). It can be observed in Fig. 9 that even though the physical state x_3 settles down quickly to the leader's state x_0, the information being observed by the eavesdropper adversary, i.e., $I_{3,a}$ and $I_{3,b}$, do not converge to any value (Fig. 10) and thus the privacy of the physical state can be protected from the adversary.

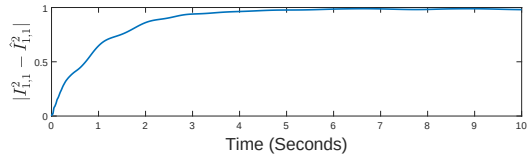

Fig. 6: Cyber-attack detection by observing $|I^2_{1,1} - \hat{I}^2_{1,1}|$ with $\alpha(t)$ given in (15), $\beta_1 = 40$, and $\beta_2 = 10$. Node 1 can identify that the information from node 2 is being compromised since $I^2_{1,1} \neq \hat{I}^2_{1,1}$.

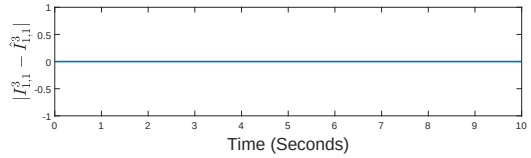

Fig. 7: Cyber-attack detection by observing $I^3_{1,1} - \hat{I}^3_{1,1}$ with $\alpha(t)$ given in (15), $\beta_1 = 40$, and $\beta_2 = 10$. Due to no attack on the link between Agent 1 and Agent 3, the difference $I^3_{1,1} - \hat{I}^3_{1,1}$ is zero.

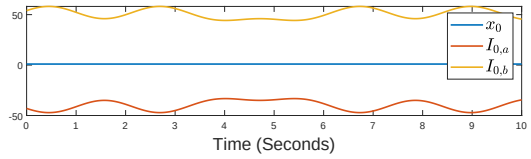

Fig. 8: Masked information $I_{0,a}$ and $I_{0,b}$ sent by the leader using $\beta_1 = 40$, $\beta_2 = 10$ and $\alpha(t)$ as given in (15).

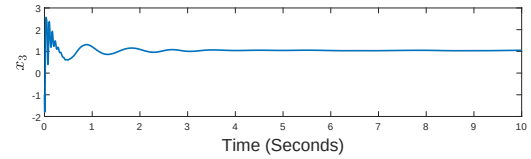

Fig. 9: State information of Agent 3 using $\beta_1 = 40$, $\beta_2 = 10$ and $\alpha(t)$ given (15).

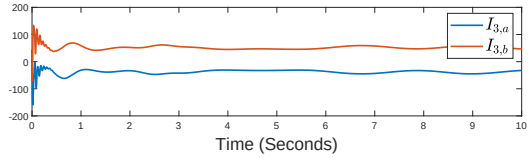

Fig. 10: Masked information $I_{3,a}$ and $I_{3,b}$ sent by Agent 3 using $\beta_1 = 40$, $\beta_2 = 10$ and $\alpha(t)$ given (15).

V. Conclusion & Future Work

In this paper, a resilient leader-following consensus is proposed in the presence of unknown attacks on both the actuators of the followers and the communication network. It is also demonstrated that by appropriately choosing the information being exchanged between the agents, the resilient control also enable the agents to either identify in real-time and a distributed manner the compromised links or to protect the privacy of the physical state from the eavesdropper adversaries. In the future, we aim to extend the approach to allow us in designing a unified control framework for achieving resilient leader-following consensus, real-time attack identification, and privacy-preservation (also from curious agents) in a unified manner.

References

[1] A. Gusrialdi, Z. Qu, and M. A. Simaan, "Distributed scheduling and cooperative control for charging of electric vehicles at highway service stations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 10, pp. 2713–2727, 2017.

[2] J. Hu, P. Bhowmick, F. Arvin, A. Lanzon, and B. Lennox, "Cooperative control of heterogeneous connected vehicle platoons: An adaptive leader-following approach," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 977–984, 2020.

[3] A. Gusrialdi and C. Yu, "Exploiting the use of information to improve coverage performance of robotic sensor networks," *IET Control Theory & Applications*, vol. 8, no. 13, pp. 1270–1283, 2014.

[4] G. Wen, X. Yu, Z.-W. Liu, and W. Yu, "Adaptive consensus-based robust strategy for economic dispatch of smart grids subject to communication uncertainties," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 6, pp. 2484–2496, 2017.

[5] A. Gusrialdi and Z. Qu, "Smart grid security: Attacks and defenses," in *Smart Grid Control: An Overview and Research Opportunities* (J. Stoustrup, A. Annaswamy, A. Chakrabortty, and Z. Q. (Eds.), eds.), pp. 199–223, Springer Verlag, 2018.

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

[7] A. Eslami, F. Abdollahi, and K. Khorasani, "Stochastic fault and cyber-attack detection and consensus control in multi-agent systems," *International Journal of Control*, pp. 1–19, 2021.

[8] 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.

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

[10] M. S. Sadabadi and A. Gusrialdi, "On resilient design of cooperative systems in presence of cyber-attacks," in *Proceedings of European Control Conference*, pp. 946–951, Rotterdam, the Netherlands, 2021.

[11] M. Iqbal, Z. Qu, and A. Gusrialdi, "Distributed resilient consensus on general digraphs under cyber-attacks," *European Journal of Control*, p. 100681, 2022.

[12] 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, 2018.

[13] J. Wang, Y. Li, Z. Duan, and J. Zeng, "A fully distributed robust secure consensus protocol for linear multi-agent systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2022.

[14] L. Wang, C. Fan, C. Xie, and W. Zhou, "Leader-following consensus of multiple euler-lagrange systems under deception attacks," *IEEE Access*, vol. 9, pp. 100548–100557, 2021.

[15] Z. Zuo, X. Cao, Y. Wang, and W. Zhang, "Resilient consensus of multiagent systems against denial-of-service attacks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 4, pp. 2664–2675, 2022.

[16] H. Xu, Y.-H. Ni, Z. Liu, and Z. Chen, "Privacy-preserving leader-following consensus via node-augment mechanism," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 6, pp. 2117–2121, 2021.

[17] D. Fiore and G. Russo, "Resilient consensus for multi-agent systems subject to differential privacy requirements," *Automatica*, vol. 106, pp. 18–26, 2019.

[18] A. Gusrialdi, Z. Qu, and M. Simaan, "Robust design of cooperative systems against attacks," in *Proceedings of American Control Conference*, pp. 1456–1462, 2014.

[19] A. Gusrialdi and Z. Qu, "Cooperative systems in presence of cyber attacks: A unified framework for resilient control and attack identification," in *Proceedings of American Control Conference*, pp. 330–335, 2022.

[20] M. Shi, X. Chen, M. Shahidehpour, Q. Zhou, and J. Wen, "Observer-based resilient integrated distributed control against cyberattacks on sensors and actuators in islanded ac microgrids," *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 1953–1963, 2021.

[21] A. Mustafa and H. Modares, "Attack analysis and resilient control design for discrete-time distributed multi-agent systems," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 369–376, 2020.

[22] Z. Qu, *Cooperative Control of Dynamical Systems*. London: Springer Verlag, 2009.