IEEE Control Systems Letters paper presented at
2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

An LMI-based Risk Assessment of Leader-Follower Multi-Agent System Under Stealthy Cyberattacks

Sounghwan Hwang, Minhyun Cho, Sungsoo Kim, and Inseok Hwang

Abstract— **This paper proposes a methodology for assessing the potential risk from cyberattacks in multi-agent systems (MASs). MASs inherently rely on communication between agents, rendering them more vulnerable to cyberattacks than single-agent systems. The impact of cyberattacks could lead to performance disruption and safety violations. To address these concerns, we propose a risk assessment method for MASs using reachability analysis which computes the reachable set of the MASs via a Lyapunov function and its corresponding linear matrix inequalities (LMIs). The proposed method can quantify the potential risk against cyberattacks at agent and entire system levels by deriving ellipsoidal over-approximated reachable sets. An illustrative example is provided to validate the potency of the method, which shows the risk associated with the formation control of a leader-following MAS in an adversarial environment with scattered obstacles.**

I. INTRODUCTION

Multi-agent systems (MASs) have received significant attention due to their ability to perform intricate missions that single-agent systems (SASs) cannot achieve, thanks to the intercommunication between agents [1]. Each agent of a MAS has a *distributed* control input relying on the local information obtained from its neighbors, which allows the MAS to be scalable and efficient. These advantages have led to the emergence of MAS engineering applications, including multi-robots and unmanned aerial vehicles (UAVs) [1], [2].

However, the heavy reliance of MASs on the communication protocol between agents introduces *vulnerabilities* against cyberattacks [3]–[9]. Specifically, MASs are more susceptible to cyberattacks than SASs, as they have more attack vectors exposed to adversaries, including the failures in feedback control loops [3]–[5] and communication links [6]. Due to the detrimental impacts of cyberattacks, ensuring their safety against cyberattacks has emerged as an active area of research.

Research on the cybersecurity of MASs focuses on two areas: (1) *system resiliency* [3]–[7], and (2) *attack detection* [8], [9]. In terms of system resiliency, Pirani et al. [3] proposed a game-theoretic method to enhance the system resiliency under false-data-injection (FDI) attacks. To accommodate stochasticity, a distributed output-feedback controller was developed to address random attacks [4], [5]. Besides, a data-driven strategy for MAS formation control was provided in [6] to handle sophisticated threats with denial-of-service (DoS) attacks and FDI attacks. A distributed resilient controller was developed in [7] to mitigate network failures by a network centroid reconfiguration. Regarding attack detection, a distributed model-based strategy to detect stealthy attacks was studied in [8] while a neural network-based strategy was addressed in [9].

Despite the above studies, quantifying the *potential risk* of cyberattacks on MASs still remains a challenge and has yet to be properly addressed. Previous research primarily focused on developing reactive mitigation and detection algorithms which are activated *after* cyber threats occur. However, MASs are generally exposed to high-risk environments with sophisticated cyber threats, where attackers can vary their attack strategies. Thus, reactive defense strategies may not be sufficient, making it critical to evaluate the potential risk of cyber threats beforehand and mitigate their impact proactively [10]–[13]. While a tool for evaluating the potential risk of cyber threats was developed only for SASs [13], it is limited in its applicability in that the characteristics of MASs, like the informational dependency among agents, are not taken into account. For instance, a cyberattack targeting a single agent in a MAS may prevent the system from reaching a consensus. Given the above discussions, expanding this risk assessment framework to MASs is necessary, albeit more complex than that in SASs.

In this work, we develop a proactive *risk assessment* method for MASs under stealthy cyberattacks. We employ reachability analysis based on a Lyapunov function and the corresponding linear matrix inequalities (LMIs). The computed reachable set allows an ellipsoidal over-approximation of the actual reachable set containing all possible compromised states within a predefined time window. The over-approximated reachable set can be regarded as a *security metric* to show how much potential cyber threats affect a system. We apply geometric operations, such as the intersection and union of ellipsoids, to assess the potential risk between individual agents and entire systems, allowing the quantification of the security of MASs.

The rest of this letter is organized as follows: Section II gives preliminaries, Section III presents the problem formulation, Section IV provides main results, and an illustrative example is given in Section V. Finally, Section VI concludes this study.

II. PRELIMINARIES

We denote the set of real numbers and integers by $\mathbb{R}$ and $\mathbb{Z}$, respectively. The superscript $+$ on $\mathbb{R}$ and $\mathbb{Z}$ stands for non-negativeness. X^T denotes the transpose of matrix X, the symbol $He\{A\}$ stands for $A + A^T$, and $\mathbb{R}^{n \times m}$ denotes all real matrices with dimension $n \times m$. The set containing

The authors are with the School of Aeronautics and Astronautics, Purdue University, West Lafayette, IN 47907, USA (email:hwang214@purdue.edu; cho515@purdue.edu; kim4021@purdue.edu; ihwang@purdue.edu)

Copyright ©2023 IEEE

natural numbers $\{1, 2, \cdots, N\}$ is defined as $\mathcal{I}_N$. For any vector $x(k) \in \mathbb{R}^n$, we use $|x(k)|_2 = \sqrt{x^T(k)x(k)}$ to represent the Euclidean norm of a vector. Then, $\|x(k)\|_\infty = \sup_{k\geq 0}|x(k)|_2$ represents the infinity norm of a signal $\{x(k)\}_{k\in\mathbb{Z}^+}$. The symbol $*$ stands for a symmetric term for notation simplicity, I_n denotes a n-dimensional identity matrix, $\mathbf{0}_{n\times m} \in \mathbb{R}^{n\times m}$ is a zero matrix, and $\otimes$ represents the Kronecker product.

Consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ to represent communication links, where $\mathcal{V} = \{v_1, \cdots, v_N\}$ is the set of agents, and $N \in \mathbb{R}$ is the number of agents. The set $\mathcal{E} \subseteq \mathcal{V}\times\mathcal{V}$ is the set of links. $\mathcal{N}_i = \{j \in \mathcal{V}, |(v_j, v_i) \in \mathcal{E}\}$ is the set of neighbors of agent i. The adjacency matrix $\mathcal{A} = [a_{ij}] \in \mathbb{R}^{N\times N}$ is defined such that $a_{ij} > 0$ if $(v_j, v_i) \in \mathcal{E}$, and $a_{ij} = 0$ otherwise. The Laplacian matrix $\mathcal{L} = [\mathcal{L}_{ij}] \in \mathbb{R}^{N\times N}$ is defined as $\mathcal{L}_{ii} = \sum_{j\neq i} a_{ij}$ and $\mathcal{L}_{ij} = -a_{ij}$ for $i \neq j$.

III. Problem Formulation

Consider the dynamics of a discrete linear time-invariant (DLTI) multi-agent system (MAS) described as follows:

$$\begin{aligned} x_i(k+1) &= Ax_i(k) + Bu_i(k) + Ew_i(k), \\ y_i(k) &= Cx_i(k) + Fv_i(k) + \Gamma\delta_i(k), \end{aligned} \tag{1}$$

where i is the index of agent i for $i \in \mathcal{I}_N$, i.e., N denotes the number of agents, $x_i(k) \in \mathbb{R}^n$ denotes the system state, $u_i(k) \in \mathbb{R}^m$ is the control input, and $y_i(k) \in \mathbb{R}^p$ is the measurement output. The matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $C \in \mathbb{R}^{p\times n}$ denote the system matrix, the input matrix, and the output matrix, respectively. The system pairs (A, B) and (A, C) are assumed to be stabilizable and detectable. $w_i(k) \in \mathbb{R}^{\bar{w}}$ and $v_i(k) \in \mathbb{R}^{\bar{v}}$ are norm-bounded process and sensor noises with perturbation matrices $E \in \mathbb{R}^{n\times\bar{w}}$ and $F \in \mathbb{R}^{p\times\bar{v}}$. $\delta_i(k) \in \mathbb{R}^{\bar{\delta}}$ with the matrix $\Gamma \in \mathbb{R}^{p\times\bar{\delta}}$ denotes the cyberattack that can stealthily impact the system. An undirected graph $\mathcal{G}$ represents the communication among the agents. The following assumptions on the noises and graph structure of the MAS (1) are held throughout the paper.

Assumption 1: The noises $w_i(k)$ and $v_i(k)$ for $i \in \mathcal{I}_N$ satisfy the following norm-bounded conditions: $|w_i|_2^2 = w_i^T(k)w_i(k) \leq W_i$ and $|v_i|_2^2 = v_i^T(k)v_i(k) \leq V_i$, where $W_i \in \mathbb{R}^+$, $V_i \in \mathbb{R}^+$, and $k \in \mathbb{Z}^+$.

Assumption 2: The graph $\mathcal{G}$ is strongly connected. The element a_{ij} of the adjacency matrix $\mathcal{A}$ is set as 1, i.e., the weight for the information flow among all agents is the same.

Remark 1: If the DLTI system (1) is assumed to be noiseless, the noises $w_i(k)$ and $v_i(k)$ can be considered as random attacks on the actuators and sensors, respectively.

Afterward, an observer-based leader-follower control protocol is presented, which enables the agents, i.e., followers of the MAS (1), to track the reference trajectory generated by a leader. The dynamics of the leader is given by:

$$x_l(k+1) = Ax_l(k) + Bu_l(k), \tag{2}$$

where $x_l(k) \in \mathbb{R}^n$ is the state of the leader, $u_l(k) \in \mathbb{R}^m$ is its control input, and the system matrices are the same as the other agent's matrices. According to a leader-follower MAS [15], the dynamics (2) can be regarded as a virtual system that generates the desired trajectory for the agents to follow.

The synchronization error of agent i is defined as $q_i(k) \triangleq x_i(k) - x_l(k)$, where $i \in \mathcal{I}_N$. By combining (1) and (2), the synchronization error dynamics can be obtained as:

$$q_i(k+1) = Aq_i(k) + B\left(u_i(k) - u_l(k)\right) + Ew_i(k). \tag{3}$$

From the distributed control protocol [15], an observer-based leader-follower control input of agent i can be formulated as:

$$u_i(k) = K_c \sum_{j\in\mathcal{N}_i} \hat{x}_{ij}(k) + \alpha_i K_c \hat{x}_{il}(k) + u_l(k), \tag{4}$$

where $\hat{x}_{ij}(k) = \hat{x}_i(k) - \hat{x}_j(k)$ and $\hat{x}_{il}(k) = \hat{x}_i(k) - x_l(k)$ denote the differences between the estimated state of agent i and its neighbor agent j, and between the estimated state of agent i and the leader state $x_l(k)$, respectively. The control gain $K_c \in \mathbb{R}^{m\times n}$ is to be designed. The indicator $\alpha_i \in \mathbb{R}$ is defined to be 1 if agent i is connected to the leader, and 0 otherwise. Substituting (4) to (3), the stacked synchronization error dynamics for the MAS (1) is obtained as follows:

$$\begin{aligned} q(k+1) = &\left(I_N \otimes A\right) q(k) + \left(\bar{L} \otimes BK_c\right) \hat{q}(k) \\ &+ \left(I_N \otimes E\right) w(k), \end{aligned} \tag{5}$$

where $\bar{L} = \mathcal{L} + \Lambda$, $\Lambda = \operatorname{diag}[\alpha_1, \cdots, \alpha_N] \in \mathbb{R}^{N\times N}$, $q(k) = [q_1^T(k), \cdots, q_N^T(k)]^T$, $\hat{q}(k) = [\hat{q}_1^T(k), \cdots, \hat{q}_N^T(k)]^T$, $\hat{q}_i(k) = \hat{x}_i(k) - x_l(k)$, and $w(k) = [w_1^T(k), \cdots, w_N^T(k)]^T$.

Lemma 1 ([13]): The matrix $\bar{L}$ is positive definite, i.e., $0 < \lambda_1(\bar{L}) < \cdots < \lambda_N(\bar{L})$. For the rest of the paper, the i-th eigenvalue of $\bar{L}$ is denoted as $\bar{\lambda}_i$ for notational simplicity.

To consider realistic scenarios, we assume that $x_i(k)$ for $i \in \mathcal{I}_N$ is not directly available. Thus, the control input (4) utilizes the estimated state from the following state observer:

$$\begin{aligned} \hat{x}_i(k+1) &= A\hat{x}_i(k) + Bu_i(k) + L_o\left(y_i(k) - \hat{y}_i(k)\right), \\ \hat{y}_i(k) &= C\hat{x}_i(k), \end{aligned} \tag{6}$$

where $L_o \in \mathbb{R}^{n\times p}$ is the observer gain to be designed, and $\hat{y}_i(k) \in \mathbb{R}^p$ is the estimated output of agent i. Let us define the state estimation error of agent i as $e_i(k) \triangleq x_i(k) - \hat{x}_i(k)$. By combining (1) and (6), the stacked state estimation error dynamics can be represented in the following manner:

$$\begin{aligned} e(k+1) = &\left(I_N \otimes (A - L_oC)\right) e(k) + \left(I_N \otimes E\right) w(k) \\ &- \left(I_N \otimes L_oF\right) v(k) - \left(I_N \otimes L_o\Gamma\right) \delta(k), \end{aligned} \tag{7}$$

where we have $e(k) = [e_1^T(k), \cdots, e_N^T(k)]^T$, $v(k) = [v_1^T(k), \cdots, v_N^T(k)]^T$, and $\delta(k) = [\delta_1^T(k), \cdots, \delta_N^T(k)]^T$.

Finally, we can obtain an augmented DLTI error dynamics by combining the dynamics (5) and (7) as follows:

$$\zeta(k+1) = \mathcal{A}_1\zeta(k) + \mathcal{A}_2w(k) - \mathcal{A}_3v(k) - \mathcal{A}_4\delta(k), \tag{8}$$

where $\zeta(k) = [q^T(k), e^T(k)]^T$, and we have

$$\mathcal{A}_1 = \begin{bmatrix} I_N \otimes A + \bar{L} \otimes BK_c & -\bar{L} \otimes BK_c \\ \mathbf{0}_{nN\times nN} & I_N \otimes (A - L_oC) \end{bmatrix},$$

$$\mathcal{A}_2 = \begin{bmatrix} I_N \otimes E \\ I_N \otimes E \end{bmatrix}, \mathcal{A}_3 = \begin{bmatrix} \mathbf{0}_{nN\times\bar{v}N} \\ I_N \otimes L_oF \end{bmatrix}, \mathcal{A}_4 = \begin{bmatrix} \mathbf{0}_{nN\times\bar{\delta}N} \\ I_N \otimes L_o\Gamma \end{bmatrix}.$$

Remark 2: In the absence of the stealthy attack $\delta(k)$, the leader-follower consensus of the MAS (1) and (2) will be achieved by designing proper gains K_c and L_o. However, the stealthy attack $\delta(k)$ can disrupt the consensus and cause the agents to deviate from their desired states. The deviations can be considered as a potential risk to both individual agents and the entire system. Our objective is to develop a risk assessment method that can quantify the potential risk against cyberattacks, enabling us to reinforce the safety of the MAS.

IV. Main Result

This section consists of two major parts: (a) providing design preliminaries for the controller, estimator, and residual-based attack detector; and then (b) presenting an LMI-based risk assessment method for the leader-follower MAS.

A. Design Preliminaries

1) Stabilization in Stealthy-Attack-Free: This subsection presents the derivation of the controller gain K_c and observer gain L_o ensuring the stability of the leader-follower MAS (1) and (2) in the stealthy-attack-free (SAF) case, i.e., when $\delta(k) = 0$. The lemma below is proposed to design K_c which can be determined by the LMI-based sufficient condition.

Lemma 2: Consider the MAS dynamics regarding $q(k)$:

$$q(k+1) = \left(I_N \otimes A + \bar{L} \otimes BK_c\right) q(k). \tag{9}$$

The MAS dynamics (9) under a graph determined by $\bar{L}$ can be stabilized if there exist proper matrices $X = X^T \in \mathbb{R}^{n\times n}$ and $Y \in \mathbb{R}^{m\times n}$ such that the following LMIs are satisfied:

$$X > 0, \begin{bmatrix} X & * \\ XA^T + \bar{\lambda}_i Y^T B^T & X \end{bmatrix} > 0, \text{ for } i \in \mathcal{I}_N, \tag{10}$$

where $\bar{\lambda}_i > 0$ is an i-th eigenvalue of the modified Laplacian matrix $\bar{L}$ and we can obtain the controller gain $K_c = YX^{-1}$.

Proof: Consider a Lyapunov function as: $V_1(k) = q^T(k)(I_N \otimes P_1)q(k)$, where $P_1 = P_1^T \in \mathbb{R}^{n\times n} > 0$. From Lemma 1, the similarity transformation $\bar{L} = TJT^{-1}$ is applied, where $T \in \mathbb{R}^{N\times N}$, and the diagonal matrix $J = \text{diag}[\bar{\lambda}_1, \cdots, \bar{\lambda}_N]$. Then, the coordinate transformation yields $\rho_1(k) \triangleq \mathcal{M}q(k)$, where $\mathcal{M} = \left(T^{-1} \otimes I_n\right)$, and $\rho_1(k) = [\rho_{11}^T(k), \cdots, \rho_{1N}^T(k)]^T$. Using the coordinate transform with Kronecker product properties (see Theorem 1 in [13]), $\Delta V_1(k) \triangleq V_1(k+1) - V_1(k) < 0$ can be represented as $\Delta V_1(k) = \sum_{i=1}^N \rho_{1i}^T(k)\Omega_i\rho_{1i}(k) < 0$, where $\Omega_i = \left(A + \bar{\lambda}_i BK_c\right)^T P_1 \left(A + \bar{\lambda}_i BK_c\right) - P_1$. By applying the Schur complement and congruence transformation [15] with P_1^{-1} to Ω_i, the condition $\Omega_i < 0$ is equivalent to the condition described in (10). Finally, the MAS dynamics (9) is asymptotically stable if the LMIs in (10) are satisfied. ■

Based on the results from Lemma 2, our next step is to design the observer gain L_o via the H_∞ approach from [17]. This approach specifies two conditions for L_o stabilizing the augmented DLTI error dynamics (8): (a) ensuring the asymptotic stability of (8) when $w(k) = 0$ and $v(k) = 0$; (b) given for some $\gamma \in \mathbb{R} > 0$ and under the zero initial condition, the dynamics (8) satisfies $\lim_{k\to\infty} \psi(k) \le \gamma$, where $\psi(k) = \sum_{\tau=0}^k |e(\tau)|_2^2 / \sum_{\tau=0}^k \left(|w(\tau)|_2^2 + |v(\tau)|_2^2\right)$.

Theorem 1: Let the LMIs from Lemma 2 be satisfied. For some $\gamma \in \mathbb{R} > 0$, the augmented DLTI error dynamics (8) is stabilizable with the H_∞ criterion if there exist matrices $P_2 = P_2^T \in \mathbb{R}^{n\times n}$ and $Z \in \mathbb{R}^{n\times p}$ such that the following LMI-based optimization is satisfied:

$$\min_{\gamma} \gamma \text{ s.t. } P_2 > 0, \ \bar{\Xi}_i < 0 \text{ for } i \in \mathcal{I}_N, \tag{12}$$

where $\bar{\Xi}_i$ is provided in (11), and $L_o = P_2^{-1}Z$.

Proof: Consider a Lyapunov function: $V_2(k) = q^T(k)(I_N \otimes P_1)q(k) + e^T(k)(I_N \otimes P_2)e(k)$, where $P_1 = P_1^T$ is obtained from Lemma 2, and $P_2 = P_2^T \in \mathbb{R}^{n\times n} > 0$. To derive a sufficient condition for the stabilizability with the H_∞ criterion, the following condition should be held [17]:

$$\Delta V_2(k) + |e(k)|_2^2 - \gamma\left(|w(k)|_2^2 + |v(k)|_2^2\right) < 0, \tag{13}$$

where $\Delta V_2(k) \triangleq V_2(k+1) - V_2(k)$. Likewise Lemma 2, we can apply the coordinate transformation as $\rho_2(k) \triangleq \mathcal{M}e(k)$, $\rho_3(k) \triangleq \mathcal{M}w(k)$, $\rho_4(k) \triangleq \mathcal{M}v(k)$; and $\rho_2(k) = \left[\rho_{21}^T(k), \cdots, \rho_{2N}^T(k)\right]^T$, $\rho_3(k) = \left[\rho_{31}^T(k), \cdots, \rho_{3N}^T(k)\right]^T$, $\rho_4(k) = \left[\rho_{41}^T(k), \cdots, \rho_{4N}^T(k)\right]^T$. Subsequently, we can rewrite the condition (13) as $\sum_{i=1}^N z_i^T(k)\Xi_i z_i(k) < 0$, where $z_i(k) = \left[\rho_{1i}^T(k), \rho_{2i}^T(k), \rho_{3i}^T(k), \rho_{4i}^T(k)\right]^T$, and detailed description for Ξ_i is omitted due to space constraints. To maintain $\sum_{i=1}^N z_i^T(k)\Xi_i z_i(k) < 0$, the condition $\Xi_i < 0$ for $i \in \mathcal{I}_N$ should be held. With the Schur complement and congruence transformation on the matrix P_2, $\Xi_i < 0$ is equivalent to the condition in (12), where $\bar{\Xi}_i$ is provided in (11). If the LMI-based optimization (12) is held, we can obtain the observer gain L_o that guarantees H_∞ criterion. ■

2) Residual-based Attack Detector Design: A residual-based attack detector (RAD) for agent $i \in \mathcal{I}_N$ is designed subsequently. The input of the RAD, $r_i(k) \triangleq y_i(k) - \hat{y}_i(k) \in \mathbb{R}^p$, is a residual between the compromised output $y_i(k)$ and the estimated output $\hat{y}_i(k)$. Then, the dynamics of $r_i(k)$ is:

$$r_i(k) = Ce_i(k) + Ev_i(k) + \Gamma\delta_i(k). \tag{14}$$

Remark 3: Given (8) and (14), the noises $w_i(k)$ and $v_i(k)$ can perturb the residual $r_i(k)$ in the SAF case.

In detail, the RAD employs a distance measure, $d_i(k) = r_i^T(k)\Pi_i r_i(k) \in \mathbb{R}$, which can evaluate the deviation caused by cyberattacks. The structure of the RAD is given as:

$$d_i(k) = r_i^T(k)\Pi_i r_i(k), \tag{15}$$

where $\Pi_i = \Pi_i^T \in \mathbb{R}^{p\times p}$ is a design parameter to be decided in a subsequent section. If $d_i(k) > 1$, an alarm is triggered to notify cyberattacks. Note that the stealthy attacks can hide their impacts and do not trigger an alarm from the RAD by keeping $d_i(k) \le 1$. The positive semi-definite matrix Π_i can be determined as follows: the ellipsoid $r_i^T(k)\Pi_i r_i(k) = 1$ should enclose all residual $r_i(k)$ that the noises $w_i(k)$ and $v_i(k)$ can induce (see (14) in the SAF case, i.e., $\delta_i(k) = 0$).

To derive Π_i, we first need to obtain the minimal upper bound of $e_i(k)$ in the SAF case when the system reaches a steady state. To this end, the following lemma is derived:

Lemma 3: Consider the state estimation error dynamics of agent i in the SAF case: $e_i(k+1) = (A - L_oC)\, e_i(k) +$

$$\bar{\Xi}_i = \begin{bmatrix} \begin{array}{c} A^TP_1A+\bar{\lambda}_i He\left\{A^TP_1BK_c\right\} \\ +\bar{\lambda}_i^2K_c^TB^TP_1BK_c-P_1 \end{array} & * & * & * & * \\ -\bar{\lambda}_iK_c^TB^TP_1A-\bar{\lambda}_i^2K_c^TB^TP_1BK_c & \bar{\lambda}_i^2K_c^TB^TP_1BK_c-P_2+I_n & * & * & * \\ E^TP_1A+\bar{\lambda}_iE^TP_1BK_c & -\bar{\lambda}_iE^TP_1BK_c & E^TP_1E-\gamma I_{\bar{w}} & * & * \\ \mathbf{0}_{\bar{w}\times n} & \mathbf{0}_{\bar{w}\times n} & \mathbf{0}_{\bar{w}\times\bar{w}} & -\gamma I_{\bar{w}} & * \\ \mathbf{0}_{n\times n} & P_2A-ZC & P_2E & ZF & -P_2 \end{bmatrix}, \tag{11}$$

$Ew_i(k) - L_oFv_i(k)$, where $A - L_oC$ is Hurwitz. For given constant $\alpha \in (0,1)$, positive definite matrix $P_2 = P_2^T \in \mathbb{R}^{n\times n}$ and constant matrix $Z \in \mathbb{R}^{n\times p}$ satisfying $L_o = P_2^{-1}Z$, if there exist positive constants $\mu_1, \mu_2 \in \mathbb{R}^+$ such that the following LMI-based optimization is satisfied:

$$\begin{aligned} &\min_{\mu_1,\mu_2} \ \mu_1+\mu_2 \\ &\text{s.t.} \begin{bmatrix} (\alpha-1)P_2 & * & * & * \\ \mathbf{0}_{n\times n} & -\alpha\mu_1 I_n & * & * \\ \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & -\alpha\mu_1 I_p & * \\ P_2A-ZC & P_2E & ZF & -P_2 \end{bmatrix} \leq 0, \\ &\quad P_2 - \mu_2^{-1}I_n \geq 0, \end{aligned} \tag{16}$$

then the upper bound of $e_i(k)$ holds the following: $|e_i(k)|_2 \leq c\lambda^k \left\|e_i(0)\right\|_2 + \sqrt{\mu_1\mu_2}\left(\left\|w_i(k)\right\|_\infty + \left\|v_i(k)\right\|_\infty\right)$, with some positive constant $c > 0$ and design parameter $\lambda \in (0,1)$.

Proof: The reader may refer to Chapter 7 in [16]. ■

With Lemma 3 and S-procedure, we can obtain optimal Π_i (in terms of the minimal boundary) by solving the following:

Theorem 2: Assume that Lemma 3 is solved. Consider the upper bound of $e_i(k)$ from Lemma 3 and the residual $r_i(k)$ from (14). If there exist positive constants $\kappa_1, \kappa_1 \in \mathbb{R}^+$ and a positive semi-definite matrix $\Pi_i = \Pi_i^T \in \mathbb{R}^{p\times p}$ such that the following LMI-based optimization is satisfied:

$$\begin{aligned} &\min_{\kappa_1,\kappa_2,\Pi_i} -\log\det\left(\Pi_i\right) \\ &\text{s.t.} \begin{bmatrix} \mathcal{T}_{11} & * & * \\ \mathcal{T}_{21} & \mathcal{T}_{22} & * \\ \mathbf{0}_{1\times\bar{v}} & \mathbf{0}_{1\times\bar{v}} & \mathcal{T}_{33} \end{bmatrix} \geq 0, \quad \begin{array}{l} \Pi_i \geq 0, \\ \kappa_1 > 0, \\ \kappa_2 > 0, \end{array} \end{aligned} \tag{17}$$

where $\mathcal{T}_{11} = \kappa_1 I_{n\times n} - C^T\Pi_iC$, $\mathcal{T}_{21} = -F^T\Pi_iC$, $\mathcal{T}_{22} = \kappa_2 I_{\bar{v}\times\bar{v}} - F^T\Pi_iF$, and $\mathcal{T}_{33} = 1 - \kappa_1\mu_1\mu_2\left(W_i + V_i\right) - \kappa_2V_i$, then, the RAD (15) can enclose all possible residuals $r_i(k)$ in the SAF case where $w_i(k)$ and $v_i(k)$ hold the norm bounded conditions $w_i^T(k)w_i(k) \leq W_i$ and $v_i^T(k)v_i(k) \leq V_i$.

Proof: Consider the LMI condition: $\Psi_i \leq 0$, where $\Psi_i := \left(Ce_i(k) + Fv_i(k)\right)^T \Pi_i \left(Ce_i(k) + Fv_i(k)\right) - 1 - \kappa_1\left(e_i^T(k)e_i(k) - \mu_1\mu_2(W_i+V_i)\right) - \kappa_2\left(v_i^T(k)v_i(k) - V_i\right)$. By using the preliminaries $e_i^T(k)e_i(k) \leq \mu_1\mu_2(W_i+V_i)$, $v_i^T(k)v_i(k) \leq V_i$, and S-procedure, $\Psi_i \leq 0$ is equivalent to the condition in (17); thus, the proof is completed. ■

From (14), we can represent the stealthy attack $\delta_i(k)$ as: $\delta_i(k) = \Gamma^+\left(r_i(k) - Ce_i(k) - Ev_i(k)\right)$, where Γ^+ denotes the Moore–Penrose inverse of Γ. By using it, the augmented DLTI error dynamics (8) can be reformulated as follows:

$$\zeta(k+1) = \mathcal{B}_1\zeta(k) + \mathcal{A}_2w(k) + \mathcal{B}_2v(k) + \mathcal{B}_3r(k), \tag{18}$$

where $r(k) = [r_1^T(k), \cdots, r_N^T(k)]^T$, $U = (I_p - \Gamma\Gamma^+)$,

$$\mathcal{B}_1 = \begin{bmatrix} I_N\otimes A + \bar{L}\otimes BK_c & -\bar{L}\otimes BK_c \\ \mathbf{0}_{nN\times nN} & I_N\otimes(A - L_oUC) \end{bmatrix},$$

$$\mathcal{B}_2 = \begin{bmatrix} \mathbf{0}_{nN\times\bar{v}N} \\ -I_N\otimes L_o\left(F - \Gamma\Gamma^+F\right) \end{bmatrix}, \mathcal{B}_3 = \begin{bmatrix} \mathbf{0}_{nN\times pN} \\ -I_N\otimes L_o\Gamma\Gamma^+ \end{bmatrix},$$

Remark 4: We substitute the residual $r_i(k)$ for the stealthy attack $\delta_i(k)$, resulting in the dynamics (18) with the bounded inputs $w(k)$, $v(k)$ and $r(k)$. The stealthy attack $\delta_i(k)$ in (14) can lead to the significant estimation error $e_i(k)$ while generating the output $y_i(k)$ close to the estimated output $\hat{y}_i(k)$. This implies that the stealthy attack $\delta_i(k)$ can be executed in a way that evades the detection. Note that the method does not require a specific sequence of the residual $r_i(k)$ and the estimated output $\hat{y}_i(k)$.

B. Risk Assessment via Lyapunov-based Reachability

This subsection presents an LMI-based risk assessment method for the leader-follower MAS (1) and (2) under cyberattacks that uses LMIs and Lyapunov function-based reachability. The following definition and lemma are employed for deriving the main theorem in this subsection:

Definition 1 (Reachable set [13]): Consider the DLTI system as: $x(k+1) = \bar{\mathcal{A}}x(k) + \sum_{i=1}^{M}\bar{\mathcal{B}}_id_i(k)$, where $x(k) \in \mathbb{R}^n$ is the system state disturbed by $d_i(k) \in \mathbb{R}^m$ where $d_i^T(k)\bar{D}_id_i(k) \leq 1$ for $i \in \mathcal{I}_M$ (M is the number of perturbations); and $\bar{\mathcal{A}} \in \mathbb{R}^{n\times n}$, and $\bar{\mathcal{B}}_i \in \mathbb{R}^{n\times m}$ are the system and perturbation matrices. The reachable set $\mathcal{R}_k^x$ at time step k, from the initial state $x(1)$, is the set of all states reachable in k time steps, i.e., $\mathcal{R}_k^x := \{x(k) \in \mathbb{R}^n | x(k+1) = \bar{\mathcal{A}}x(k) + \sum_{i=1}^{M}\bar{\mathcal{B}}_id_i(k),\ \text{and}\ d_i^T(k)\bar{D}_id_i(k) \leq 1, \forall i \in \mathcal{I}_M\}$.

Remark 5: Referring to Definition 1, we can consider the reachable set $\mathcal{R}_k^\zeta$ for (18), where $\mathcal{R}_k^\zeta$ is the set of all states $\zeta(k)$ impacted by norm-bounded $w(k)$, $v(k)$, and $r(k)$.

Lemma 4 ([13]): For a given $a \in (0,1)$, if there exist a design parameter $a_i \in (0,1)$ for $i \in \mathcal{I}_M$, and a function $V : \mathbb{R}^n \to \mathbb{R} > 0$ satisfying the following inequality: $V(x(k+1)) - aV(x(k)) - \sum_{i=1}^{M}\left(1-a_i\right)d_i^T(k)\bar{D}_id_i(k) \leq 0$ with $\sum_{i=1}^{M}a_i \geq a$ and $d_i^T(k)\bar{D}_id_i(k) \leq 1$ for $i \in \mathcal{I}_M$, then, $V(x(k)) \leq a^{k-1}V(x(1)) + \left((M-a)(1-a^{k-1})\right)/(1-a)$, and $\lim_{k\to\infty}V(x(k)) = \left((M-a)(1-a^{k-1})\right)/(1-a)$.

The following theorem provides the ellipsoidal over-approximation of $\mathcal{R}_k^\zeta$ computed from the Lyapunov function.

Theorem 3: Consider the compromised augmented DLTI system (18), and its reachable set $\mathcal{R}_k^\zeta$. For a given design parameter $a \in (0,1)$, if there exist constants $a_1, a_2, a_3 \in \mathbb{R}^+$

and a positive definite matrix $\mathcal{P} = \mathcal{P}^T \in \mathbb{R}^{2nN \times 2nN}$ derived from the following LMI-based optimization:

$$\begin{aligned} &\min_{\mathcal{P}, a_1, a_2, a_3} \; -\log\det(\mathcal{P}) \\ &\text{s.t.} \begin{bmatrix} a\mathcal{P} & * & * \\ \mathbf{0}_{(p+2N)\times 2nN} & \mathcal{W} & * \\ \mathcal{P}\mathcal{B}_1 & \mathcal{P}\mathcal{D} & \mathcal{P} \end{bmatrix} \geq 0, \; \mathcal{P} > 0, \\ &\qquad a_1, a_2, a_3 \in (0,1), \; a_1 + a_2 + a_3 \geq a, \end{aligned} \tag{19}$$

where $\mathcal{W} = \text{diag}\left[(1-a_1)\mathcal{C}_1, (1-a_2)\mathcal{C}_2, (1-a_3)\mathcal{C}_3\right] \in \mathbb{R}^{(p+2)N \times (p+2)N}$, $\mathcal{C}_1 = \text{diag}\left[W_1^{-1}/N, \cdots, W_N^{-1}/N\right] \in \mathbb{R}^{N\times N}$, $\mathcal{C}_2 = \text{diag}\left[V_1^{-1}/N, \cdots, V_N^{-1}/N\right] \in \mathbb{R}^{N\times N}$, $\mathcal{C}_3 = \text{diag}\left[\Pi_1/N, \cdots, \Pi_N/N\right] \in \mathbb{R}^{pN\times pN}$, and $\mathcal{D} = [\mathcal{A}_2 \; \mathcal{B}_2 \; \mathcal{B}_3] \in \mathbb{R}^{2nN\times(p+2)N}$; then, we have $\mathcal{R}_k^\zeta \subseteq \mathcal{Y}_k^\zeta := \{\zeta(k) \in \mathbb{R}^{2nN} | \zeta^T(k)\mathcal{P}\zeta(k) \leq \bar{Y}_k^\zeta\}$, where the upper bound $\bar{Y}_k^\zeta := a^{k-1}\zeta^T(1)\mathcal{P}\zeta(k) + \left((3-a)(1-a^{k-1})\right)/(1-a)$.

Proof: Consider a Lyapunov function as: $V(\zeta(k)) = \zeta^T(k)\mathcal{P}\zeta(k)$, where $\mathcal{P} = \mathcal{P}^T \in \mathbb{R}^{2nN\times 2nN}$. By exploiting Lemma 4, suppose that $V(\zeta(k+1)) - aV(\zeta(k)) - (1-a_1)w^T(k)\mathcal{C}_1 w(k) - (1-a_2)v^T(k)\mathcal{C}_2 v(k) - (1-a_3)r^T(k)\mathcal{C}_3 r(k) \leq 0$ is held. Substituting $\zeta(k)$ from (18) into this inequality, we can obtain: $p^T(k)\bar{\mathcal{Q}}p(k) \geq 0$, where $p(k) = \left[\zeta^T(k), w^T(k), v^T(k), r^T(k)\right]^T$, and we have

$$\bar{\mathcal{Q}} = \begin{bmatrix} a\mathcal{P} - \mathcal{B}_1^T\mathcal{P}\mathcal{B}_1 & * \\ -\mathcal{D}^T\mathcal{P}\mathcal{B}_1 & \mathcal{W} - \mathcal{D}^T\mathcal{P}\mathcal{D} \end{bmatrix}. \tag{20}$$

By applying the Schur complement to (20), $\bar{\mathcal{Q}} > 0$ is equivalent to the condition described in (19). Finally, we obtain $\zeta^T(k)\mathcal{P}\zeta(k) \leq \bar{Y}_k^\zeta$ and the compromised state trajectories of $\zeta(k)$ from (18) are encapsulated in the $\mathcal{Y}_k^\zeta$. ■

Remark 6: Theorem 3 can provide the ellipsoidal over-approximated reachable set $\mathcal{Y}_k^\zeta$, which contains the reachable set of $q(k)$ and $e(k)$ for all N agents. To obtain an individual reachable set for agent i, we can apply ellipsoidal projection into the matrix $\mathcal{P}$ (see Corollary 2 from [13]). Given ellipsoidal over-approximated reachable sets for individual agents, we can apply two geometric operations: (a) a *union*; (b) an *intersection*. These operations can measure potential risks associated with an agent and the entire system. Due to page constraints, we refer readers to Chapter 3.7 of [16] for more technical details of the geometric operations.

V. Illustrative Example

In this section, the proposed method is applied to a leader-follower MAS performing the formation control of three agents (e.g., UAVs) whose system matrices are given as:

$$A = \begin{bmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{bmatrix}, \; B = \begin{bmatrix} \frac{\Delta t^2}{2} & 0 \\ \Delta t & 0 \\ 0 & \frac{\Delta t^2}{2} \\ 0 & \Delta t \end{bmatrix}, \; C = I_4,$$

$$E = 0.1I_4, \; F = I_4, \; \Gamma = [0.25, 0.25, -0.53, -0.77]^T.$$

The total simulation time is $30s$, and $\Delta t = 0.5s$. $x_i(k) \triangleq [p_{Xi}^T(k), v_{Xi}^T(k), p_{Yi}^T(k), v_{Yi}^T(k)]^T \in \mathbb{R}^4$ is the state of agent i, where $x_i(k)$ contains the position/velocity in the X-axis/Y-axis direction. In this example. all agents are connected, i.e., $a_{ij} = 1$ for $i \neq j$, and each agent is connected to the virtual leader, where $\Lambda = I_N \in \mathbb{R}^{N\times N}$. Figure. 1 (*left*) depicts the trajectories of the agents tracking the leader's reference trajectory and maintaining the formation in the attack-free case. Now, we implement stealthy attacks in 1000 Monte-Carlo simulations. The attack sequences linearly increase over 30 seconds by following $\delta_i(k) = U([-0.07, 0.07]) \times k$, where $U([a,b])$ gives a random number from the uniform distribution over $[a,b]$ $(a, b \in \mathbb{R})$ and k denotes the discrete time step where $k \in [0, 60]$ $(k \in \mathbb{Z}^+)$. Note that all attack sequences satisfy the stealthiness condition, i.e., $d_i(k) \leq 1$. Hence, the potential risk against stealthy cyberattacks should be quantified to enhance safety proactively using reachable sets and geometric operations on the sets.

While the MAS can accomplish the mission without safety violations in the attack-free case as shown in Fig. 1 (*middle*), stealthy cyberattacks may lead to collisions between agents and obstacles as depicted in Fig. 1 (*right*). In detail, the orange, green, and purple ellipses denote the projected ellipsoidal over-approximated reachable sets of each agent, respectively. These ellipses encapsulate all trajectories from Monte-Carlo simulations at the given time instances. The blue and red ellipses show the union and intersections of the ellipses of individual agents. It is noted that, at $t = 21s$, the size of blue ellipse in the stealthy-attack case is $66.95m^2$, which is 2.5 times larger than that of the attack-free case. For both time instances at $t = 12.5s$ and $t = 21s$, the blue ellipses in the stealthy-attack case overlap with the obstacles, which implies that the collective behaviors of the MAS can violate the safe conditions. This fact can also be validated by Fig. 2 (*right*) in that the minimum distances between obstacles and these blue ellipses become zero, i.e., potential collisions. In Fig. 2, the solid lines denote the distance between the union of reachable sets and the obstacles in the attack-free case, while the dotted lines denote that in the stealthy attack case. The red ellipses in Fig. 1 (*right*) denote the intersections of the projected ellipsoidal over-approximated reachable sets of the agents. Figure. 2 (*left*) shows that, in the stealthy attack case, the sum of the areas of these red ellipses grows over time, which implies a higher probability of collisions between agents. The results shown in Fig. 2 suggest that increasing the inter-agent distance can prevent collisions between the agents. Yet, this may increase the chance of safety violations from other perspectives, i.e., collisions between the union and the obstacles. Our proposed method will be extended to the trade-off study to minimize the risk caused by stealthy cyberattacks.

VI. CONCLUSIONS

In this letter, we developed an LMI-based risk assessment method for the leader-follower MAS under stealthy cyberattacks. Our proposed method employed reachability analysis based on a Lyapunov function and computed ellipsoidal over-approximated reachable sets. We showed how the obtained reachable sets could be utilized to evaluate the potential risks at the agent and system levels. Finally, the efficacy of our method was demonstrated via an illustrative example.

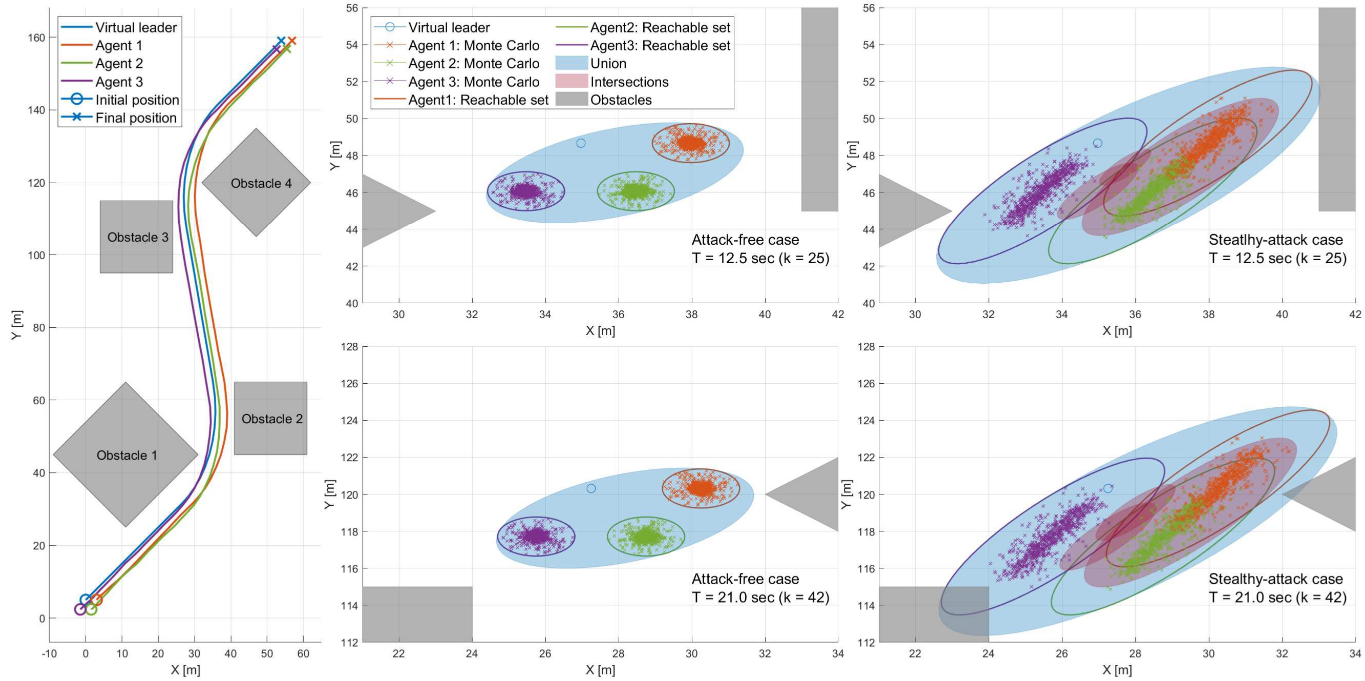

Fig. 1: Formation control of a leader-follower MAS in the attack-free case (*left*); the ellipsoidal over-approximated reachable sets computed at $t = 12.5s$ and $t = 21s$ in the attack-free case (*middle*); the ellipsoidal over-approximated reachable sets computed at $t = 12.5s$ and $t = 21s$ in the stealthy-attack case (*right*).

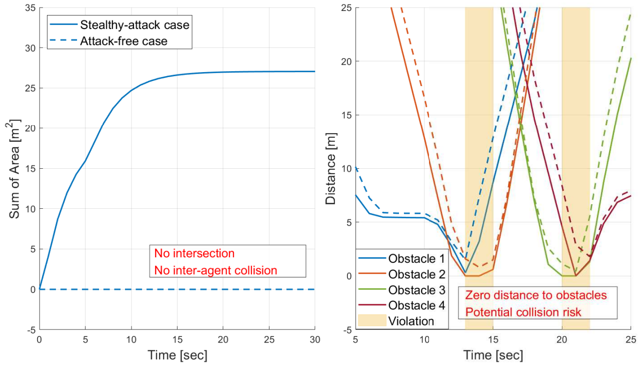

Fig. 2: Impact of stealthy attacks over time: the summation of the ellipsoidal intersections (*left*); the minimum distance between obstacles and the ellipsoidal union (*right*).

ACKNOWLEDGMENT

The authors would like to acknowledge that this work was supported by NASA University Leadership Initiative (ULI) under Grant 80NSSC20M0161.

REFERENCES

[1] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and Cooperation in Networked Multi-Agent Systems", in *Proceeding of the IEEE*, vol. 95, no. 1, pp. 215-233, Jan. 2007.

[2] R. Owoputi and S. Ray, "Security of Multi-Agent Cyber-Physical Systems: A Survey," in *IEEE Access*, vol. 10, 2022.

[3] M. Pirani, et al. "Design of Attack-Resilient Consensus Dynamics: A Game-Theoretic Approach," *2019 18th European Control Conference*, Naples, Italy, 2019.

[4] Y. Yang, et al. "Observer-Based Distributed Secure Consensus Control of a Class of Linear Multi-Agent Systems Subject to Random Attacks", *IEEE Trans. Circuits. Syst. I*, vol. 66, no. 8, pp. 3089-3099, Aug. 2019.

[5] S. Huo, et al. "Observer-Based Resilient Consensus Control for Heterogeneous Multi-Agent Systems Against Cyber-Attacks", *IEEE Trans. Control. Netw. Syst.*, Early access.

[6] O. Thapliyal and I. Hwang "Data-driven Cyberattack Synthesis against Network Control Systems", *ArXiv preprint arXiv:2211.05203*, 2022

[7] S. Clarke, O. Thapliyal, S. Hwang, and I. Hwang, "Attack-Resilient Distributed Optimization-based Control of Multi-Agent Systems with Dual Interaction Networks", *AIAA SciTech Forum*, Jan. 2022, pp. 2342.

[8] F. Boem, et al. "A distributed attack detection method for multi-agent systems governed by consensus-based control", *Proc. IEEE Conf. Decis. Control.*, Dec. 2017.

[9] A. Mousavi, K. Aryankia, and R. R. Selmic "Cyber-Attack Detection in Discrete-Time Nonlinear Multi-Agent Systems Using Neural Networks", *Proc. IEEE Conf. Control. Tec. Appli*, Aug. 2021.

[10] N. Sun, et al. "Cyber Threat Intelligence Mining for Proactive Cybersecurity Defense: A Survey and New Perspectives", *IEEE Commun. Surv. Tutor.*, Early Access, May. 2023.

[11] J. H. Cho, et al. "Toward Proactive, Adaptive Defense: A Survey on Moving Target Defense", *IEEE Commun. Surv. Tutor.*, vol. 22, 2022.

[12] R. Ferrari, and A. Teixeira, "Safety, Security, and Privacy for Cyber-Physical Systems", 1st ed. Basel, Switzerland: Springer, 2021.

[13] C. Murguia, et al. "Security Metrics of Networked Control Systems", *Automatica*, vol. 115, May. 2020.

[14] J. Trejo et. al, "Robust observer-based leader-following consensus for a class of nonlinear multi-agent systems: application to UAV formation control", in *Proc. Int. Conf. Unmanned. Aircraft. Syst.*, June. 2021.

[15] X. Xu, S. Chen, W. Huang, and L. Gao, "Leader-following consensus of discrete-time multi-agent systems with observer-based protocols", *Neurocomputing*, vol. 118, pp. 334-341, 2013.

[16] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, "Linear matrix inequalities in system and control theory", *Society for industrial and applied mathematics (SIAM)*, 1994.

[17] Y. Luo, et al. "Event-Triggered Finite-Time Guaranteed Cost H-Infinity Consensus for Nonlinear Uncertain Multi-Agent Systems," in *IEEE Trans. Netw. Sci.*, vol. 9, no. 3, pp. 1527-1539, June 2022.