IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 May 2023, accepted 13 July 2023, date of publication 19 July 2023, date of current version 31 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3296741

RESEARCH ARTICLE

Reinforcement Learning-Based Control Strategy for Multi-Agent Systems Subjected to Actuator Cyberattacks During Affine Formation Maneuvers

SAMI EL-FERIK[1,2], MUHAMMAD MAARUF[1,2], FOUAD M. AL-SUNNI[1], ABDULWAHID ABDULAZIZ SAIF[1,2], AND MUJAHED MOHAMMAD AL DHAIFALLAH[1,3], (Member, IEEE)

[1]Control and Instrumentation Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
[2]Research Center for Smart Mobility and Logistics, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
[3]Research Center for Renewable Energy and Power Systems, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

Corresponding author: Sami El-Ferik (selferik@kfupm.edu.sa)

This work was supported by the Research Center for Smart Mobility and Logistics, King Fahd University of Petroleum and Minerals, under Project INML2300.

ABSTRACT In this research, we investigate the reinforcement learning-based control strategy for second-order continuous-time multi-agent systems (MASs) subjected to actuator cyberattacks during affine formation maneuvers. In this case, a long-term performance index is created to track the MASs tracking faults using a leader-follower structure. In order to approximate the ideal solution, which is challenging to find for systems vulnerable to cyberattacks during time-varying maneuvers, a critical neural network is used. The distributed control protocol is obtained, and the long-term performance index is minimized, using an actor neural network strengthened with critic signals. The actor-critic neural networks calculate unknown dynamics and the severity of attacks on the MAS actuators. The Nussbaum functions are applied to address this issue since attacks can result in a loss of control direction. The stability of the closed-loop system has been emphasized with the use of a Lyapunov candidate function. The performance of the suggested strategy is then supported by a numerical simulation.

INDEX TERMS Reinforcement learning, actor-critic neural networks, multi-agent systems, affine formation maneuver, actuator attacks, Nussbaum functions.

I. INTRODUCTION

Over the years, the formation control of MASs continued to garner remarkable interest from researchers because of their potential applications in several areas [1]. The application areas include unmanned aerial vehicles, sensor network localization, robotic transportation in smart factories, satellite clusters, traffic flow control, monitoring of sensors, and so on [2], [3]. Formation control of MASs is broadly categorized into two classes: formation geometric shape control and formation maneuver (FM) control [4]. The objective of the formation shape control is to steer a group of agents to achieve a prescribed geometric shape given any initial geometry. On the other hand, the FM control drives the agents to establish the desired formation, and then control the entire formation such that it can translate, rotate, and scale continuously [5]. One important merit of FM control is that it can avoid obstacles.

To drive a group of mobile agents to converge to the desired formation with maneuvers, various formation control strategies have been developed. These approaches are categorized into three groups, namely position-based [6], distance-based [7], [8], and bearing-based [9], [10]. The required formation is formed by specifying several constant constraints on the inter-agent position, distance, and orientation. Due to the invariance of the inter-agent constraints of the desired formation, the position-based scheme can achieve translation maneuvers. However, applying the scaling maneuver to the

The associate editor coordinating the review of this manuscript and approving it for publication was Xiwang Dong.

target formation is difficult because the displacement has to be altered. For distance-based schemes, the control laws can accomplish both orientation and translation maneuvers, but it is difficult to realize time-varying scales. In the case of the bearing-based schemes, the control laws are able to realize both scaling and translation maneuvers but have difficulties in following time-varying orientations. It is clear that these FM schemes cannot accomplish rotation, scaling, and translation concurrently.

To carry out the rotation, scaling, and translation maneuvers simultaneously, several consensus-based control techniques using complex Laplacian matrices have been developed for single-integrator systems [11], [12], two-integrator systems [13], and nonlinear systems [14]. However, this approach is restricted to two-dimensional systems. This limitation motivated new research that led to the establishment of a new technique based on stress matrix [15], [16]. The entries of the stress matrix which correspond to the weights of the edges can be positive, negative, or zero in contrast to the complex Laplacian matrix [16]. The stress matrix-based technique can realize FMs in all dimensions, unlike the complex Laplacian matrix. Moreover, the stress matrix is invariant to any affine transformation of the formation determined by the leaders. The sufficient and necessary conditions for realizing time-varying maneuvers using the stress matrix have been explained in detail in [16].

Based on the properties of the stress matrix, various leader-follower AFM control approaches have been studied for different classes of MASs. For instance, Shiyu [16] developed a generalized approach for AFM of single-integrator and double-integrator MASs. In this scheme, the agents are divided into two groups; the leaders and the followers. The affine transformation of the followers is uniquely obtained from the coordinates of the leaders. A distributed control law makes the followers track the maneuvers of the leaders and ensure the global stability of the whole formation. In [17], an AFM control framework is formulated for single-integrator MASs considering the optimal geometric pattern. To improve the convergence of the single-integrator followers to the affine positions localized by the leaders, a prescribed convergence time function is integrated into the distributed controller [18], [19], [20]. In [21], an AFM control of double-integrator MASs with external disturbances. An adaptive component is added to the controller to estimate and suppress external disturbances. In [22] and [23], the authors designed an affine formation algorithm for linear systems governed by triple-integrator dynamic equations in both continuous-time and sampled-data modes. In [24], a proportional-integral (PI) controller is used to achieve the AFM of high-order-integrator MASs with time-varying communication delays. In [25], The authors investigated the distributed AFM control problem of general linear MASs with dynamic coupling gains and dynamic uncertainties. Then, robust adaptive control protocols have been presented to take care of the dynamic coupling gains and uncertainties.

Even though the above AFM control techniques are effective, they are only applicable to linear MAS. Considering the fact that many physical systems are inherently nonlinear, Hui et al. [26] investigated the AFM control of general second-order uncertain nonlinear MASs. The nonlinearities of the agents were parametrized and adaptive laws were used to estimate the uncertain parameters. Moreover, proportional-integral controllers are employed to guarantee that the followers attained the required affine formation localized by the leaders. Thus, the AFM control of general nonlinear MASs is not well studied and requires enormous attention.

MASs are increasingly connected to wireless networks where information transmission is facilitated for remote supervision and control. As such, they are prone to malicious cyber attacks such as deception attacks, denial of service attacks, false data injection, integrity attacks, malware attacks, replay attacks, and so on [27]. These attacks mostly inject misleading signals into the systems' sensors, actuators, or communication channels, which may lead to the deterioration of the performance of the systems [28]. There exist several control strategies for detecting and mitigating various cyber attacks to secure the consensus of MASs [29], [30], [31], [32], [33] or to secure the geometrical formation acquired by the agents [34], [35], [36], [37], [38], [39], [40], [41]. However, a secure control against cyber attacks is yet to be investigated with respect to AFM of MAS. A false data injection attack on the actuators of the leaders or the followers will lead to loss of collective maneuver and the agents may collide. Therefore, how to secure the leaders as well as the followers so that collective formation maneuvering is maintained in adversarial environments is an interesting area to explore.

As an intelligent technique, a reinforcement learning (RL) scheme has been widely employed for solving optimal adaptive control problems by minimizing a prescribed cost function to maximize a reward received from the environment [42]. For continuous-time systems, integral RL is utilized to compute the solution of the optimal adaptive control problem using the actor-critic framework. Considering the fact that it is difficult to obtain the solution of the optimal adaptive control problem of continuous-time systems using RL, neural networks are commonly used to compute the approximate solution. Approximate RL-based control has been applied to several classes of single agent systems [43], [44], [45], [46], [47], MAS [48], [49], [50], [51]. Recently, approximate RL is employed for formation acquisition control of MASs [52], [53], [54], [55], [56]. The problem of AFM control of MASs using approximate RL deserves to be investigated and related work is not reported yet.

In the existing RL control articles, the AFM control problem of MASs has not been studied yet, let alone simultaneous consideration of malicious cyber attack cases. Motivated by this observation, this article proposes an RL scheme using the actor-critic framework to achieve the AFM control of general nonlinear MASs in an adversarial

environment. Due to the difficulties in obtaining the optimal solution of the long-term performance function for the AFM of MASs with actuator cyber attacks, critic neural network is used to obtain the approximate solution. The critic evaluates the control performance and sends the reinforcement signals to the actor. Then, the actor applies the control signals to the actuators of the agents based on the critic evaluation. Here, two actor networks are used: one for estimating the uncertain nonlinear dynamics of the agents, and the other for estimating the actuator attacks. The attacks on the actuator may alter the control direction of the distributed control protocols. This problem is addressed by using the Nussbaum functions to estimate the control direction. To the authors' best knowledge, the proposed control problem has not been investigated by anyone yet. This work differs from existing AFM control methods as follows:

1) Unlike the leader-follower AFM control approaches in [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], and [26], an RL control based on actor-critic neural networks is proposed to achieve the AFM of the leader-follower MAS in adversarial environments.
2) The control protocols in [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], and [26] will fail to maintain the leader-follower AFM when subjected to cyber attacks on the actuators. Here, this problem is tackled by using the actor neural network to learn and counter the actuator attacks.
3) In [26], the authors considered a class of second-order uncertain nonlinear MAS that can be expressed in the linear-in-the-parameter form, and adaptive laws were used to estimate the uncertain parameters. However, in this study, the uncertain nonlinear MAS doesn't need to be linear-in-the-parameter form since an action neural network is used to approximate it. Moreover, the action neural network which receives reinforced signals from the critic neural network enhances the learning and optimization of the system.
4) The loss of control direction as a result of the actuator attack is tackled using the Nussbaum function.

II. PRELIMINARIES AND PROBLEM FORMULATION

A. GRAPH THEORY

Define $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ as an undirected graph with N nodes. $\mathcal{V} = \{v_1, \ldots, v_N\}$ stands for the set of nodes. $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ represents the set of edges. The set of neighbors of node i is represented by $\mathcal{N}_i = \{j \in \mathcal{V} : (i,j) \in \mathcal{E}\}$. The configuration $q = [q_1^T \; q_2^T \; \ldots \; q_N^T]^T \in \mathbb{R}^{dN}$ consists of the positions of the agents with d being the dimension of the system. The formation $(\mathcal{G}, q)$ is a communication graph together with the corresponding configuration of the agents.

B. STRESS MATRIX

The stress matrix of the formation $(\mathcal{G}, q)$ is a set of scalar weights $\{\omega_{ij}\}_{(i,j)\in\mathcal{E}}$ assigned to each edge such that $\omega_{ij} = \omega_{ji} \in \mathbb{R}$ [2]. The stress matrix $\Omega \in \mathbb{R}^{N\times N}$ can be defined as:

$$[\Omega]_{ij} = \begin{cases} 0, & i \neq j, (i,j) \notin \mathcal{E} \\ -\omega_{ij}, & i \neq j, (i,j) \in \mathcal{E} \\ \sum_{k\in\mathcal{N}_i} \omega_{ik}, & i = j. \end{cases}$$

A stress matrix is said to be equilibrium stress when:

$$\sum_{j\in\mathcal{N}_i} \omega_{ij}(q_j - q_i) = 0, \qquad i \in \mathcal{V} \tag{1}$$

Equation (1) can be transformed to a matrix as:

$$(\Omega \otimes I_d)q = 0 \tag{2}$$

The stress matrix Ω can be partitioned as

$$\bar{\Omega} = \Omega \otimes I_d = \begin{bmatrix} \Omega_{ll} & \Omega_{lf} \\ \Omega_{fl} & \Omega_{ff} \end{bmatrix} \otimes I_d = \begin{bmatrix} \bar{\Omega}_{ll} & \bar{\Omega}_{lf} \\ \bar{\Omega}_{fl} & \bar{\Omega}_{ff} \end{bmatrix}. \tag{3}$$

where $\Omega_{ll} \in \mathbb{R}^{dN_l \times dN_l}$, $\Omega_{ff} \in \mathbb{R}^{dN_f \times dN_f}$ and $\Omega_{fl} \in \mathbb{R}^{dN_l \times dN_f}$

The geometric pattern that the agents are required to form and keep is represented by the constant configuration $r = [r_1^T \; r_2^T \ldots r_N^T]^T = [r_l^T \; r_f^T]^T =\in \mathbb{R}^{dN}$ under a nominal formation $(\mathcal{G}, r)$. The target formation with the desired maneuver is given by:

$$q(t) = [I_N \otimes A(t)]r + 1_N \otimes b(t) \tag{4}$$

where $b(t) \in \mathbb{R}^d$ stands for the translation maneuver of the formation, $A(t) \in \mathbb{R}^{d\times d}$ can be manipulated to achieve the rotation, scaling, and shearing maneuvers of the whole formation with respect to r.

Definition 1: [16] **Affine span** : For a given set of points $\{q_i\}_{i=1}^N \in \mathbb{R}^d$, the affine span (S) of these points is defines as:

$$\mathcal{S} = \left\{ \sum_{i=1}^{N} a_i q_i : a_i \in \mathbb{R} \text{ for all } i \text{ and } \sum_{i=1}^{n} a_i = 1 \right\}. \tag{5}$$

The affine transformations of r are contained in the image of r defined as follows:

$$\mathcal{A}(r) = \{q \in \mathbb{R}^{dN} : q = (I_N \otimes A)r + 1_N \otimes b\}. \tag{6}$$

Definition 2: [23] **Affine localizability** : If the target position of the followers q_f can be uniquely calculated from the position of the leaders q_l for any $q = [q_f^T \; q_l^T]^T \in \mathcal{A}(r)$, then the nominal formation $(\mathcal{G}, r)$ is affinely localizable by q_l.

Assumption 1: [26] For the design of the formation maneuver control, the following assumptions are essential.

1) The nominal configuration r is affinely spanned in $\mathbb{R}^d$
2) The equilibrium stress matrix Ω is semi-definite such that $rank(\Omega) = N - d - 1$
3) The position q_f is affinely localizable by q_l.

Considering Definition 2 and Assumption 1, for any $q = [q_f^T \; q_l^T]^T \in \mathcal{A}(r)$, q_f can be uniquely calculated from q_l as follows:

$$q_f = -(\Omega_{ff}^{-1}\Omega_{fl} \otimes I_d)q_l. \tag{7}$$

where

$$q_l = [q_1^T \; q_2^T \ldots q_{N_l}^T]^T$$
$$q_f = [q_{N_l+1}^T \; q_{N_l+2}^T \cdots q_{N_l+N_f}^T]^T$$

Definition 3: [57] Any continuous function $N(\beta) : \mathbb{R} \longrightarrow \mathbb{R}$ is a Nussbaum function if it satisfies the following properties:

$$\lim_{\theta\to\infty} sup \frac{1}{\theta}\int_0^\theta N(\beta)d\beta = +\infty \tag{8}$$
$$\lim_{\theta\to\infty} inf \frac{1}{\theta}\int_0^\theta N(\beta)d\beta = -\infty \tag{9}$$

For example, the functions $\beta^2 cos(\beta)$ and $\beta^2 sin(\beta)$ are Nussbaum functions. In this work, we choose $\beta^2 cos(\beta)$.

Lemma 1: Let $\beta(t)$ be a smooth function on $[0, \; t_f)$, $L(t)$ be non-negative smooth function, and $N(\beta)$ be a smooth Nussbaum function. If the following inequality is valid.

$$L(t) \leq a_0 + e^{-a_1\tau}\int_0^t [g(.)N(\beta)+1]\dot{\beta}e^{a_1\tau}d\tau \tag{10}$$

where $a_0 > 0$, $a_1 > 0$ are constants, $g(.)$ is a time-varying function, then, $L(t)$, $\beta(t)$, and $\int_0^t g(.)N(\beta)\dot{\beta}d\tau$ must be bounded on $[0, \; t_f]$

Consider second-order nonlinear multi-agent systems consisting of N_l leaders and $N_f = N - N_l$ followers. The nonlinear dynamics of the N_f followers can be described as:

$$\begin{cases} \dot{q}_i = v_i \\ \dot{v}_i = f_i(q_i, v_i) + u_i^c(t) + \varphi_i(q_i, v_i), \qquad i \in \mathcal{V}_f \end{cases} \tag{11}$$

where $q_i \in \mathbb{R}^d$ and $v_i \in \mathbb{R}^d$ denote the position and velocity of the *ith* agent, $u_i^c(t) \in \mathbb{R}^d$ is the compromised actuator output of agent i, $f_i(q_i, v_i) \in \mathbb{R}^d$ is a smooth nonlinear function, $\varphi_i(q_i, v_i) \in \mathbb{R}^d$, $\mathcal{V}_f = \{N_l+1, N_l+2, \ldots, N_l+N_f\}$ is the group of followers.

The trajectories of the N_l leaders are given by:

$$q_i = A(t)r_i + b(t), \qquad i \in \mathcal{V}_l \tag{12}$$

where $\mathcal{V}_l = \{1, 2, \ldots, N_l\}$ is the group of the leaders. The position of the leaders $q_l = [q_1^T \; q_2^T \ldots q_{N_l}^T]^T$ is known in advance and various maneuvers can be obtained by manipulating $A(t)$ and $b(t)$.

C. NEURAL NETWORK APPROXIMATIONS

Owing to the function approximation property of the radial basis function neural network (RBFNN), it has been widely deployed to estimate unknown nonlinear functions. The estimation of continuous nonlinear function with RBFNN is expressed as:

$$\zeta_{nn} = W^T\Psi(X) + \epsilon(X) \tag{13}$$

where $W = [w_1 \; w_2 \ldots w_m]^T \in \mathbb{R}^m$ is the ideal with vector with m number of neurons, $\epsilon(X)$ is the function estimation error satisfying $\|\epsilon(X)\| \leq \bar{\epsilon}$, $\Psi(X) = [\Psi_1(X) \; \Psi_1(X) \ldots \Psi_m(X)]^T$ is the Gaussian basis function vector and $\Psi_i(X)$ is

$$\Psi_i(X) = exp\left(-\frac{(X-c_i)(X-c_i)}{\eta_i^2}\right), \quad i = 1, 2, \ldots, m \tag{14}$$

where c_i is the center of the receptive field, η_i is the width of the Gaussian function.

D. ACTUATOR ATTACKS

The malignant cyber attacks considered here are the false data injection attacks into the actuator. These attacks could corrupt the control command of the *ith* follower. The actuator attacks can be modelled as a state-dependent function as follows:

$$u_i^c(t) = g_i(q_i, v_i)u_i(t) + \delta_i(t, q_i, v_i) \tag{15}$$

where $u_i^c(t)$ is the corrupted actuator output, $u_i(t)$ is the control signal, $\delta_i(t, q_i, v_i)$ is the time-varying state-dependent injected attack, $g_i(q_i, v_i)$ is a state-dependent unknown gain causing loss of effectiveness of the actuator as a result of the attack.

The system (11) under the actuator attacks (15) is rewritten in the following form:

$$\begin{cases} \dot{q}_i = v_i \\ \dot{v}_i = f_i(q_i, v_i) + g_i(q_i, v_i)u_i(t) + \delta_i(t, q_i, v_i) + \varphi_i(q_i, v_i) \end{cases} \tag{16}$$

Assumption 2: [1] The unknown diagonal matrix $g_i(q_i, v_i) \in \mathbb{R}^{d\times d}$ resulting from the false data injection satisfies $\underline{g}_i < \lambda_1(g_i(q_i, v_i)) < \lambda_2(g_i(q_i, v_i) < \infty$, with $\underline{g}_i$ being a constant lower bound of $g_i(q_i, v_i)$, and λ_1 and λ_2 are the Eigen values of $g_i(q_i, v_i)$.

Assumption 3: [58] The disturbance vector φ_i is unknown and satisfies $\|\varphi_i\| \leq \rho_i$, with ρ_i being an unknown constant.

Remark 1: The false data injection model in [59], [60], and [61] is of the form $u_i^c = u_i + \delta_i$, i.e $g_i = 1$. This article extends the results to the situation where the unknown gain g_i is bounded, which can accurately highlight the severity of the attacks on the actuator.

Remark 2: It is worth noting that assumption 1 ensures that $\bar{\Omega}_{ff}$ is positive definite and invertible. Then, the position of the followers q_f can be obtained from the position of the leaders q_l for any $q = [q_f^T \; q_l^T]^T$ [21], [26]. Assumption 2 is widely used in the literature to simplify the complexity of the control law [1], [3]. Instead of formulating a complex algorithm to estimate the unknown $g_i(q_i, v_i)$, using its known upper bound in the control law simplifies the control design process. Assumption 3 is a standard and common assumption in the robust control literature [58], [62]. This assumption is necessary because unbounded disturbances will grow and make the state trajectories of a system explode, i.e. approach infinity.

III. CONTROL DESIGN

The target positions of the followers relative to the leaders can be expressed as:

$$q_f^*(t) = -(\Omega_{ff}^{-1}\Omega_{fl} \otimes I_d)q_l. \tag{17}$$

The control objective of the followers under the actuator attacks is thus:

$$\lim_{t\to\infty}(q_f(t) - q_f^*(t)) = \tag{18}$$

$$\lim_{t\to\infty}(q_f(t) + (\Omega_{ff}^{-1}\Omega_{fl} \otimes I_d)q_l) = 0 \tag{19}$$

Define the following error variables for the followers:

$$\xi_i = \sum_{j=1}^{N}\omega_{ij}(q_i - q_j) + \lambda\sum_{j=1}^{N}\omega_{ij}(v_i - v_j), \qquad i \in \mathcal{V}_f \tag{20}$$

Let $e_i = \sum_{j=1}^{N}\omega_{ij}(q_i - q_j)$. Therefore, one has $\xi_i = e_i + \lambda\dot{e}_i, \ \ i \in \mathcal{V}_f$. Taking the time-derivatives of $\xi_i, \ i \in \mathcal{V}_f$, one can gets:

$$\begin{aligned}\dot{\xi}_f &= \dot{e}_f + \lambda\ddot{e}_f \\ &= \dot{e}_f + \lambda\Big\{\bar{\Omega}_{ff}\big(f(q,v) + g(q,v)u(t) + \delta(t,q,v) \\ &\quad + \varphi(q,v)\big) + (\Omega_{fl} \otimes I_d)\ddot{v}_l\Big\}\end{aligned} \tag{21}$$

where $\xi_f = [\xi_{N_l+1}^T \ \xi_{N_l+1}^T \cdots \xi_{N_l+N_f}^T]^T$, $e_f = [e_{N_l+1}^T \ e_{N_l+1}^T \cdots e_{N_l+N_f}^T]^T$, $f = [f_{N_l+1}^T f_{N_l+2}^T \cdots f_{N_l+N_f}^T]^T$, $u = [u_{N_l+1}^T \ u_{N_l+2}^T \cdots u_{N_l+N_f}^T]^T$, $\delta = [\delta_{N_l+1}^T \ \delta_{N_l+2}^T \cdots \delta_{N_l+N_f}^T]^T$, $\varphi = [\varphi_{N_l+1}^T \ \varphi_{N_l+2}^T \cdots \varphi_{N_l+N_f}^T]^T$.

A. CRITIC DESIGN

A binary utility function $p(\xi_i(t))$, which is the current system performance index is set as [47]:

$$p(\xi_i(t)) = \begin{cases} 0 & \text{if } \xi_i(t)^2 \le th_r \\ 1 & \text{if } \xi_i(t)^2 > th_r \end{cases} \qquad \forall\xi \in [t-T, t) \tag{22}$$

where $th_r > 0$ is a designed threshold. $p(\xi_i(t)) = 0$ indicates acceptable tracking accuracy whereas $p(\xi_i(t)) = 1$ indicates unacceptable tracking performance.

The long-term performance index is described by:

$$\mathrm{I}(t) = \int_t^{\infty}\alpha^{\frac{-\pi+t}{\tau}}p(\xi(\pi))\mathrm{d}\pi \tag{23}$$

where $\alpha \in (0,1)$ is a constant that discounts the future cost, and τ is a small integral reinforcement interval, $p(\xi) = [p(\xi_{N_l+1})^T \ p(\xi_{N_l+2})^T \ldots p(\xi_{N_l+N_f})^T]^T$. If ξ_i is within the allowable threshold, then the control objective is realized and $I(t)$ will not increase. However, if $\xi_i(t)$ is outside the allowable threshold, then the controller should be fine-tuned to make e_f, $\dot{e}_f$, and $I(t)$ smaller.

From (23), the continuous time Bellman error equation can be constructed as:

$$\begin{aligned}I(t-\tau) &= \int_{t-\tau}^{\infty}\alpha^{\frac{-\pi+t-\tau}{\tau}}p(\xi(\pi))\mathrm{d}\pi \\ &= \alpha^{-1}I(t) + \int_{t-\tau}^{t}\alpha^{\frac{-\pi+t-\tau}{\tau}}p(\xi(\pi))\mathrm{d}\pi \\ &\triangleq \alpha^{-1}\left(I(t) + p_c\right)\end{aligned} \tag{24}$$

where $p_c = \int_{t-\tau}^{t}\alpha^{\frac{-\pi+t-\tau}{\tau}}p(\xi(\pi))d\pi$ is the cost value for $[t-\tau, t)$, $p_c = [p_{c_1}^T \ p_{c_2}^T \ldots p_{c_{N_l}}^T]^T$ with

$$p_{c_i} = \int_{t-\tau}^{t}\alpha^{\frac{-\xi+t}{\tau}}p(\xi_i(\pi))\mathrm{d}\pi \tag{25}$$

$$= \begin{cases} 0 & \text{if } \xi_i(t)^2 \le c_{p_i} \\ \dfrac{\tau}{\ln(\alpha)}(\alpha - 1) & \text{if } \xi_i(t)^2 > c_{p_i} \end{cases} \tag{26}$$

Therefore, $0 < p_{c_i} < \tau(\alpha-1)/ln(\alpha)$.

Given that $I(t)$ contains the future information of the system with actuator attacks, it is difficult to obtain its solution. A critic RBFNN is introduced to approximate it.

$$I(t) = {W_c}^T\Psi_c(X_c(t)) + \epsilon_c(X_c(t)) \tag{27}$$

The real-time approximation of $I(t)$ is thus:

$$\hat{I}(t) = \hat{W}_c^T\Psi_c(X_c(t)) \tag{28}$$

Similar to (28), real-time approximation of $I(t-\tau)$ is:

$$\hat{I}(t-\tau) = \hat{W}_c^T\Psi_c(X_c(t-\tau)) \tag{29}$$

From (28) and (29), the temporal difference error is defined as:

$$\begin{aligned}e_c &= \hat{I}(t) - \alpha\hat{I}(t-\tau) + p_c \\ &= \tilde{W}_c^T\Delta\Psi_c(t) + p_c \\ &= \tilde{W}_c^T\Delta\Psi_c(t) + p_c + W_c^T\Delta\Psi_c(t)\end{aligned} \tag{30}$$

where $\tilde{W}_c = W_c - \hat{W}_c$ is the weight error for critic RBFNN, $\Delta\Psi_c(t) = [\Psi_c(X_c(t)) - \alpha\Psi_c(X_c(t-\tau))]$. Define the objective function O_c as:

$$O_c = \frac{1}{2}e_c^Te_c \tag{31}$$

The aim is to find $\hat{W}_c$ that will minimize (31). Using the gradient-decent algorithm, the weight update law is designed as follows:

$$\dot{\hat{W}}_c = \gamma_c\Delta\Psi_c(t)\left[\widehat{W}_c^T\Delta\Psi_c(t) + p_c\right]^T - \theta_c\gamma_c\widehat{W}_c \tag{32}$$

where $\gamma_c = \mathrm{diag}\{\gamma_{c_{N_l+1}} \ \gamma_{c_{N_l+2}} \cdots \gamma_{c_{N_l+N_f}}\}$ is a positive definite diagonal matrix, $\theta_c > 0$ is a small constant.

B. ACTOR DESIGN

Due to the fact that the continuous-time nonlinear model (16) is uncertain, accurate information about $f(q,v)$ is unavailable. Furthermore, the false data $\delta(t,q,v)$ injected into the actuators of the followers is unknown. Then, $f(q,v)$ is approximate with RBFNN as follows:

$$f(q,v) = {W_{a1}}^T\Psi_{a1}(X_{a1}) + \epsilon_{a1}(X_{a1}) \tag{33}$$

The time-varying state-dependent attack $\delta(t, q, v)$ is approximated by the action RBFNN as follows:

$$\delta(t, q, v) = {W_{a2}}^T \Psi_{a2}(X_{a2}) + \epsilon_{a2}(X_{a2}) \tag{34}$$

The real-time approximation of $f(q, v)$ and $\delta(q, v)$ are given as:

$$\hat{f}(q, v) = \hat{W}_{a1}^T \Psi_{a1}(X_{a1}) \tag{35}$$

$$\hat{\delta}(t, q, v) = \hat{W}_{a2}^T \Psi_{a2}(X_{a2}) \tag{36}$$

In the actor network, it is desired that q_f and $\dot{q}_f$ track q_f^* and $\dot{q}_f^*$ with reasonable accuracy, and minimize $\hat{I}(t)$ to the required cost value $I_d = 0$. The actor error is defined as:

$$e_{a1} = \xi_f + \hat{I}(t) - I_d = \xi_f + \hat{W}_c^T \Psi_c) \tag{37}$$

The actor objective function is set as:

$$O_{a1} = \frac{1}{2} e_{a1}^T e_{a1} \tag{38}$$

According to the gradient-decent algorithm, (38) can be minimized by

$$\dot{\widehat{W}}_{a1} = \gamma_{a1} \Psi_{a1}(X_{a1}) \left[\xi + \widehat{W}_c^T \Psi_c(X_c)\right]^T - \theta_{a1} \gamma_{a1} \widehat{W}_{a1} \tag{39}$$

where $\gamma_{a1} = \text{diag}\{\gamma_{a1_{N_l+1}} \ \gamma_{a1_{N_l+2}} \cdots \gamma_{a1_{N_l+N_f}}\}$ is a positive definite diagonal matrix, $\theta_{a1} > 0$ is a small constant. Similarly, the weight update law for $\widehat{W}_{a2}$ is

$$\dot{\widehat{W}}_{a2} = \gamma_{a2} \Psi_{a2}(X_{a2}) \left[\xi + \widehat{W}_c^T \Psi_c(X_c)\right]^T - \theta_{a2} \gamma_{a2} \widehat{W}_{a2} \tag{40}$$

where $\gamma_{a2} = \text{diag}\{\gamma_{a2_{N_l+1}} \ \gamma_{a2_{N_l+2}} \cdots \gamma_{a2_{N_l+N_f}}\}$ is a positive definite diagonal matrix, $\theta_{a2} > 0$ is a small constant.

The RL-based control laws for the nonlinear MAS (16) subjected to actuator attacks are proposed as follows:

$$\begin{aligned}
\bar{u} &= K\xi_f + \underline{g}^{-1} \hat{W}_{a1}^T \Psi_{a1}(X_{a1}) + \underline{g}^{-1} \hat{W}_{a2}^T \Psi_{a2}(X_{a2})] \\
&\quad - \underline{g}^{-1} (\bar{\Omega}_{ff}^{-1} \Omega_{fl} \otimes I_d) \ddot{v}_l \\
u &= N(\beta) \bar{u} \\
\dot{\beta} &= \lambda \xi_f^T [K\xi_f + \hat{W}_{a1}^T \Psi_{a1}(X_{a1}) + \hat{W}_{a2}^T \Psi_{a2}(X_{a2})]
\end{aligned} \tag{41}$$

Remark 3: The essence of the Nussbaum gain $N(\beta)$ in (41) is to determine the control direction that is lost as a result of the actuator attacks.

Theorem 1: Consider the second-order nonlinear MAS (11) with actuator attacks (15). If assumptions (1-4) are satisfied, given the critic neural network (27) and actor neural networks (33) and (34), then the control law (41) with neural network update laws (39), (34) and (32) guarantees that all the closed-loop error signals are semiglobally uniformly ultimately bounded and the leader-follower AFM can be achieved. Moreover, the closed-loop error signals are trapped within a compact set defined as follows:

$$\mathbb{S}_1 = \left\{ e_f \;\middle|\; \|e_f\| \le \sqrt{\frac{2\left(\frac{\mathcal{D}_2}{\mathcal{D}_1} + \mathcal{D}_3 + L(0)\right)}{\lambda_{min}(\bar{\Omega}_{ff})^{-1}}} \right\} \tag{42}$$

$$\mathbb{S}_2 = \left\{ \xi_f \;\middle|\; \|\xi_f\| \le \sqrt{2\left(\frac{\mathcal{D}_2}{\mathcal{D}_1} + \mathcal{D}_3 + L(0)\right)} \right\} \tag{43}$$

$$\mathbb{S}_3 = \left\{ \widetilde{W}_{a1} \;\middle|\; \|\widetilde{W}_{a1}\| \le \sqrt{\frac{2\left(\frac{\mathcal{D}_2}{\mathcal{D}_1} + \mathcal{D}_3 + L(0)\right)}{\lambda_{min}(\gamma_{a1}^{-1})}} \right\} \tag{44}$$

$$\mathbb{S}_4 = \left\{ \widetilde{W}_{a2} \;\middle|\; \|\widetilde{W}_{a2}\| \le \sqrt{\frac{2\left(\frac{\mathcal{D}_2}{\mathcal{D}_1} + \mathcal{D}_3 + L(0)\right)}{\lambda_{min}(\gamma_{a2}^{-1})}} \right\} \tag{45}$$

$$\mathbb{S}_5 = \left\{ \widetilde{W}_c \;\middle|\; \|\widetilde{W}_c\| \le \sqrt{\frac{2\left(\frac{\mathcal{D}_2}{\mathcal{D}_1} + \mathcal{D}_3 + L(0)\right)}{\lambda_{min}(\gamma_c^{-1})}} \right\} \tag{46}$$

Proof: A candidate Lyapunov function is constructed as

$$\begin{aligned}
L &= \frac{1}{2} e_f^T \bar{\Omega}_{ff}^{-1} e_f + \frac{1}{2} \xi_f^T \bar{\Omega}_{ff}^{-1} \xi_f + \frac{\lambda}{2} \left(\widetilde{W}_{a1}^T \gamma_{a1}^{-1} \widetilde{W}_{a1}\right) \\
&\quad + \frac{\lambda}{2} \left(\widetilde{W}_{a2}^T \gamma_{a2}^{-1} \widetilde{W}_{a2}\right) + \frac{\lambda_c}{2} \left(\widetilde{W}_c^T \gamma_c^{-1} \widetilde{W}_c\right) \\
&\triangleq L_1 + L_2 + L_3
\end{aligned} \tag{47}$$

where

$$\begin{aligned}
L_1 &= \frac{1}{2} e_f^T \bar{\Omega}_{ff}^{-1} e_f + \frac{1}{2} \xi_f^T \bar{\Omega}_{ff}^{-1} \xi_f \\
L_2 &= \frac{\lambda}{2} \left(\widetilde{W}_{a1}^T \gamma_{a1}^{-1} \widetilde{W}_{a1}\right) + \frac{\lambda}{2} \left(\widetilde{W}_{a2}^T \gamma_{a2}^{-1} \widetilde{W}_{a2}\right) \\
L_3 &= \frac{\lambda_c}{2} \left(\widetilde{W}_c^T \gamma_c^{-1} \widetilde{W}_c\right)
\end{aligned}$$

It follows that:

$$\begin{aligned}
\dot{L}_1 &= e_f^T \bar{\Omega}_{ff}^{-1} \dot{e}_f + \xi_f^T \bar{\Omega}_{ff}^{-1} \dot{\xi}_f = e_f^T \bar{\Omega}_{ff}^{-1} \dot{e}_f + \xi_f^T \bar{\Omega}_{ff}^{-1} \dot{e}_f \\
&\quad + \lambda \xi_f^T \Big\{ \big(f(q, v) + g(q, v) u(t) + \delta(t, q, v) \\
&\quad + \varphi(q, v)\big) + (\bar{\Omega}_{ff}^{-1} \Omega_{fl} \otimes I_d) \dot{v}_l \Big\} + \dot{\beta} - \dot{\beta} \\
&= e_f^T \bar{\Omega}_{ff}^{-1} \dot{e}_f + (e_f + \lambda \dot{e}_f)^T \bar{\Omega}_{ff}^{-1} \dot{e}_f + \lambda \xi_f^T g(q, v) u(t) \\
&\quad + \lambda \xi_f^T \Big\{ \big(f(q, v) + \delta(t, q, v) + \varphi(q, v)\big) + \ddot{q}^* \Big\} \\
&\quad + \dot{\beta} - \lambda \xi_f^T \left[K\xi_f + \hat{W}_{a1}^T \Psi_{a1}(X_{a1}) + \hat{W}_{a2}^T \Psi_{a2}(X_{a2}) \right] \\
&= e_f^T \bar{\Omega}_{ff}^{-1} \dot{e}_f + (e_f + \lambda \dot{e}_f)^T \bar{\Omega}_{ff}^{-1} \left(\dot{e}_f + \frac{e_f}{\lambda} - \frac{e_f}{\lambda} \right) \\
&\quad + \lambda g(q, v) N(\beta) \dot{\beta} + \dot{\beta} - \lambda \xi_f^T \widetilde{W}_{a1}^T \Psi_{a1}(X_{a1}) \\
&\quad - \lambda \xi_f^T \widetilde{W}_{a2}^T \Psi_{a2}(X_{a2}) - \lambda \xi_f^T K \xi_f + \lambda \xi_f^T \ddot{q}^* \\
&\quad + \lambda \xi_f^T \varphi(q, v) + \lambda \xi_f^T (\epsilon_{a1} + \epsilon_{a2}) \\
&= \lambda \xi_f^T \ddot{q}^* - \lambda \xi_f^T K \xi_f - \frac{1}{\lambda} e_f^T \bar{\Omega}_{ff}^{-1} e_f \\
&\quad + \frac{1}{\lambda^2} (e_f + \lambda \dot{e}_f)^T \bar{\Omega}_{ff}^{-1} (e_f + \lambda \dot{e}_f) + [N(\beta) + 1] \dot{\beta} \\
&\quad - \lambda \xi_f^T \widetilde{W}_{a1}^T \Delta \Psi_{a1}(X_{a1}) - \lambda \xi_f^T \widetilde{W}_{a2}^T \Delta \Psi_{a2}(X_{a2}) \\
&\quad + \lambda \xi_f^T \varphi(q, v) + \lambda \xi_f^T \epsilon_{a1} + \lambda \xi_f^T \epsilon_{a2}
\end{aligned} \tag{48}$$

$$\dot{L}_2 = \lambda \left(\widetilde{W}_{a1}^T \gamma_{a1}^{-1} \dot{\widetilde{W}}_{a1}\right) + \lambda \left(\widetilde{W}_{a2}^T \gamma_{a2}^{-1} \dot{\widetilde{W}}_{a2}\right)$$

$$
\begin{aligned}
&= \lambda\left(\widetilde{W}_{a1}^{T}\gamma_{a1}^{-1}\dot{\widehat{W}}_{a1}\right) + \lambda\left(\widetilde{W}_{a2}^{T}\gamma_{a2}^{-1}\dot{\widehat{W}}_{a2}\right)\\
&= -\theta_{a1}\lambda\widetilde{W}_{a1}^{T}\widehat{W}_{a1} - \theta_{a2}\lambda\widetilde{W}_{a2}^{T}\widehat{W}_{a2}\\
&\quad + \lambda\xi_f^{T}\widetilde{W}_{a1}^{T}\Delta\Psi_{a1}(X_{a1}) + \lambda\xi_f^{T}\widetilde{W}_{a2}^{T}\Delta\Psi_{a2}(X_{a2})\\
&\quad + \lambda\widetilde{W}_{a1}^{T}\Delta\Psi_{a1}(X_{a1})\left[\widehat{W}_c^{T}\Delta\Psi_c(X_c)\right]^{T}\\
&\quad + \lambda\widetilde{W}_{a2}^{T}\Delta\Psi_{a2}(X_{a2})\left[\widehat{W}_c^{T}\Delta\Psi_c(X_c)\right]^{T} \qquad (49)\\
\dot{L}_1+\dot{L}_2 &= \lambda\xi_f^{T}\ddot{q}^{*} - \lambda\xi_f^{T}K\xi_f - \frac{1}{\lambda}e_f^{T}\bar{\Omega}_{ff}^{-1}e_f\\
&\quad + \frac{1}{\lambda^2}\xi_f^{T}\bar{\Omega}_{ff}^{-1}\xi_f + [N(\beta)+1]\dot{\beta} + \lambda\xi_f^{T}\varphi(q,v)\\
&\quad + \lambda\xi_f^{T}\epsilon_{a1} + \lambda\xi_f^{T}\epsilon_{a2}\\
&\quad - \theta_{a1}\lambda\widetilde{W}_{a1}^{T}[\widetilde{W}_{a1}+W_{a1}] - \theta_{a2}\lambda\widetilde{W}_{a2}^{T}[\widetilde{W}_{a2}+W_{a2}]\\
&\quad + \lambda\widetilde{W}_{a1}^{T}\Delta\Psi_{a1}(X_{a1})\left[[\widetilde{W}_c+W_c]^{T}\Delta\Psi_c(X_c)\right]^{T}\\
&\quad + \lambda\widetilde{W}_{a2}^{T}\Delta\Psi_{a2}(X_{a2})\left[[\widetilde{W}_c+W_c]^{T}\Delta\Psi_c(X_c)\right]^{T}\\
&\le -\Big(\lambda\lambda_{min}(K) - 2\lambda - \frac{\lambda_{min}(\bar{\Omega}_{ff})^{-1}}{\lambda^2}\\
&\quad - \frac{\mu_1+\mu_2}{2}\lambda\Big)\|\xi_f\|^2 - \frac{1}{\lambda}\lambda_{min}(\bar{\Omega}_{ff})^{-1}\|e_f\|^2\\
&\quad - \lambda\left(\frac{\theta_{a1}}{2} - \frac{\mu_1}{2} - \mu_1\mu_2\right)\|\widetilde{W}_{a1}\|^2\\
&\quad - \lambda\left(\frac{\theta_{a2}}{2} - \frac{\mu_2}{2} - \mu_1\mu_2\right)\|\widetilde{W}_{a2}\|^2\\
&\quad + \frac{\mu_1\mu_c+\mu_2\mu_c}{2}\lambda\|\widetilde{W}_c\|^2 + [N(\beta)+1]\dot{\beta} + \mathcal{D}_a
\end{aligned}
\qquad (50)
$$

where $\|.\|$ is the Frobenius norm, $\|\Delta\Psi_{a1}\| \le \mu_1$, $\|\Delta\Psi_{a2}\| \le \mu_2$, $\|\Delta\Psi_c\| \le \mu_c$, $\|W_{a_1}\| \le w_{a1}$, $\|W_{a2}\| \le w_{a2}$, $\|W_c\| \le w_c$, $\epsilon_{a1} \le \bar{\epsilon}_{a1}$, $\epsilon_{a2} \le \bar{\epsilon}_{a2}$, $\mathcal{D}_a = \frac{\lambda\bar{\epsilon}_{a1}^2}{2} + \frac{\lambda\bar{\epsilon}_{a2}^2}{2} + \frac{\lambda\|\rho\|^2}{2} + \frac{\lambda\ddot{q}^{*2}}{2} + \frac{\mu_1\mu_c+\mu_2\mu_c}{2}\lambda\|w_c\|^2 + \frac{\lambda\theta_{a1}}{2}\|w_{a1}\|^2 + \frac{\lambda\theta_{a2}}{2}\|w_{a2}\|^2$, w_{a1}, w_{a2}, w_c, μ_1, μ_2, μ_c, $\bar{\epsilon}_{a1}$, and $\bar{\epsilon}_{a2}$ are bound constants. Then, the derivative of L_3 with respect to time gives:

$$
\begin{aligned}
\dot{L}_3 &= \lambda_c\left(\widetilde{W}_c^{T}\gamma_c^{-1}\dot{\widetilde{W}}_c\right) = \lambda_c\left(\widetilde{W}_c^{T}\gamma_c^{-1}\dot{\widehat{W}}_c\right)\\
&= -\lambda_c\widetilde{W}_c^{T}\Delta\Psi_c(t)\left[\widehat{W}_c^{T}\Delta\Psi_c(t) + p_c\right]^{T}\\
&\quad - \theta_c\lambda_c\widetilde{W}_c^{T}\widehat{W}_c\\
&= -\lambda_c\widetilde{W}_c^{T}\Delta\Psi_c(t)\left[\widetilde{W}_c^{T}\Delta\Psi_c(t) + W_c{}^{T}\Delta\Psi_c(t) + p_c\right]^{T}\\
&\quad - \theta_c\lambda_c\widetilde{W}_c^{T}\widetilde{W}_c - \theta_c\lambda_c\widetilde{W}_c^{T}W_c\\
&\le -\lambda_c\Big\{\theta_c + \|\Delta\Psi_c(t)\|^2\Big\}\|\widetilde{W}_c\|^2 - \Big\{\|W_c\|\|\Delta\Psi_c(t)\|^2\\
&\quad + \|\Delta\Psi_c(t)\|\|p_c\| + \theta_c\|W_c\|\Big\}\lambda_c\|\widetilde{W}_c\|\\
&\le -\theta_c\lambda_c\|\widetilde{W}_c\|^2 + \mathcal{D}_c\|\widetilde{W}_c\|\\
&\le -\Big\{\theta_c\lambda_c + \lambda_c\|\Delta\Psi_c(t)\|^2 - \frac{1}{2}\Big\}\|\widetilde{W}_c\|^2 + \frac{\mathcal{D}_c^2}{2}
\end{aligned}
\qquad (51)
$$

where $\mathcal{D}_c = -\lambda_c(w_c\mu_c^2 + \mu_c\mu_{p_c} + \theta_c w_c)$, with $\|p_c\| \le \mu_{p_c}$. The derivative of L with respect to time is thus:

$$
\begin{aligned}
\dot{L} &\le -\Big(\lambda\lambda_{min}(K) - 2\lambda - \frac{\lambda_{min}(\bar{\Omega}_{ff})^{-1}}{\lambda^2}\\
&\quad - \frac{\mu_1+\mu_2}{2}\lambda\Big)\|\xi_f\|^2 - \frac{1}{\lambda}\lambda_{min}(\bar{\Omega}_{ff})^{-1}\|e_f\|^2\\
&\quad - \lambda\left(\frac{\theta_{a1}}{2} - \frac{\mu_1}{2} - \mu_1\mu_2\right)\|\widetilde{W}_{a1}\|^2\\
&\quad - \lambda\left(\frac{\theta_{a2}}{2} - \frac{\mu_2}{2} - \mu_1\mu_2\right)\|\widetilde{W}_{a2}\|^2\\
&\quad - \left(\lambda_c\theta_c + \lambda_c\mu_c^2 - \frac{1}{2} - \frac{\mu_1\mu_c+\mu_2\mu_c}{2}\lambda\right)\|\widetilde{W}_c\|^2\\
&\quad + [N(\beta)+1]\dot{\beta} + \mathcal{D}_a + \frac{\mathcal{D}_c^2}{2}\\
&\le -\mathcal{D}_1 L + [N(\beta)+1]\dot{\beta} + \mathcal{D}_2
\end{aligned}
\qquad (52)
$$

where $\mathcal{D}_2 = \mathcal{D}_a + \frac{\mathcal{D}_c^2}{2}$,

$$
\begin{aligned}
\mathcal{D}_1 = \min\Big\{&\Big(\lambda\lambda_{min}(K) - 2\lambda - \frac{\lambda_{min}(\bar{\Omega}_{ff})^{-1}}{\lambda^2}\\
&- \frac{\mu_1+\mu_2}{2}\lambda\Big), \frac{1}{\lambda}\lambda_{min}(\bar{\Omega}_{ff})^{-1},\\
&\lambda\left(\frac{\theta_{a1}}{2} - \frac{\mu_1}{2} - \mu_1\mu_2\right), \lambda\left(\frac{\theta_{a2}}{2} - \frac{\mu_2}{2} - \mu_1\mu_2\right),\\
&\left(\lambda_c\theta_c + \lambda_c\mu_c^2 - \frac{1}{2} - \frac{\mu_1\mu_c+\mu_2\mu_c}{2}\lambda\right)\Big\}
\end{aligned}
\qquad (53)
$$

To guarantee that $\mathcal{D}_1 > 0$, the parameters λ_c, λ, θ_{a1}, θ_{a2}, θ_c, and the matrix K are selected such that the following conditions are satisfied:

$$
\begin{aligned}
&\lambda_c > 0;\ \lambda > 0;\ \theta_{a1} > 0;\ \theta_{a2} > 0;\ \theta_c > 0\\
&\quad \times \lambda_{min}(K) > 0;\ \ \lambda_{min}(\bar{\Omega}_{ff})^{-1} > 0;\ \ \lambda_{min}(K) > \lambda\\
&\quad \times \left(\lambda\lambda_{min}(K) - 2\lambda - \frac{\lambda_{min}(\bar{\Omega}_{ff})^{-1}}{\lambda^2} - \frac{\mu_1+\mu_2}{2}\lambda\right) > 0\\
&\quad \times \left(\frac{\theta_{a1}}{2} - \frac{\mu_1}{2} - \mu_1\mu_2\right) > 0\\
&\quad \times \left(\frac{\theta_{a2}}{2} - \frac{\mu_2}{2} - \mu_1\mu_2\right) > 0\\
&\quad \times \left(\lambda_c\theta_c + \lambda_c\mu_c^2 - \frac{1}{2} - \frac{\mu_1\mu_c+\mu_2\mu_c}{2}\lambda\right) > 0
\end{aligned}
$$

Having $\mathcal{D}_1 > 0$ will guarantee the convergence of the closed-loop tracking errors into the compact sets (42)-(46).

By multiplying both sides of (53) by $e^{-\mathcal{D}_1 t}$, it can be rewritten as follows:

$$
e^{\mathcal{D}_1 t}\dot{L} \le -\mathcal{D}_1 L e^{\mathcal{D}_1 t} + [N(\beta)+1]\dot{\beta}e^{\mathcal{D}_1 t} + \mathcal{D}_2 e^{\mathcal{D}_1 t} \qquad (54)
$$

The inequality (54) further yields:

$$
\frac{d}{dt}(e^{\mathcal{D}_1 t}L(t)) \le [N(\beta)+1]\dot{\beta}e^{\mathcal{D}_1 t} + \mathcal{D}_2 e^{\mathcal{D}_1 t} \qquad (55)
$$

Computing the integral of (55) over $[0, t)$, we get:

$$
\begin{aligned}
L(t) &\le \frac{\mathcal{D}_2}{\mathcal{D}_1} + L(0)e^{-\mathcal{D}_1 t}\\
&\quad + e^{-\mathcal{D}_1 t}\int_0^t [N(\beta)+1]\dot{\beta}e^{\mathcal{D}_1 \iota}d\iota
\end{aligned}
\qquad (56)
$$

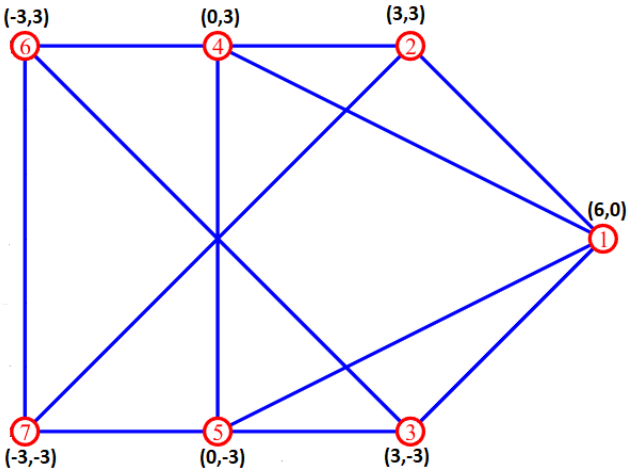

FIGURE 1. The nominal formation showing the inter-agents communications.

Based on Lemma 1, $\beta(t)$, $\int_0^t [N(\beta)+1]\dot{\beta}e^{\mathcal{D}_1 \iota} d\iota$ and and $L(t)$ are bounded. It follows that all the closed-loop error signals, ξ_f, $\widetilde{W}_{a1}$, $\widetilde{W}_{a2}$ and $\widetilde{W}_c$ are bounded. Let $\mathcal{D}_3$ be the upper-bound of $e^{-\mathcal{D}_1 t}\int_0^t [N(\beta)+1]\dot{\beta}e^{\mathcal{D}_1 \iota} d\iota$, then (56) can be simplified as:

$$\begin{aligned} L(t) &\leq \frac{\mathcal{D}_2}{\mathcal{D}_1} + L(0)e^{-\mathcal{D}_1 t} + \mathcal{D}_3 \\ &\leq \frac{\mathcal{D}_2}{\mathcal{D}_1} + \mathcal{D}_3 + L(0) \end{aligned} \tag{57}$$

From the Lyapunov function (47) and the inequality (57), it can be obtain that:

$$\|e_f\| \leq \sqrt{\frac{2}{\lambda_{min}(\bar{\Omega}_{ff})^{-1}}\left(\frac{\mathcal{D}_2}{\mathcal{D}_1} + \mathcal{D}_3 + L(0)\right)} \tag{58}$$

$$\|\xi_f\| \leq \sqrt{2\left(\frac{\mathcal{D}_2}{\mathcal{D}_1} + \mathcal{D}_3 + L(0)\right)} \tag{59}$$

$$\|\widetilde{W}_{a1}\| \leq \sqrt{\frac{2}{\lambda_{min}(\gamma_{a1}^{-1})}\left(\frac{\mathcal{D}_2}{\mathcal{D}_1} + \mathcal{D}_3 + L(0)\right)} \tag{60}$$

$$\|\widetilde{W}_{a2}\| \leq \sqrt{\frac{2}{\lambda_{min}(\gamma_{a2}^{-1})}\left(\frac{\mathcal{D}_2}{\mathcal{D}_1} + \mathcal{D}_3 + L(0)\right)} \tag{61}$$

$$\|\widetilde{W}_c\| \leq \sqrt{\frac{2}{\lambda_{min}(\gamma_c^{-1})}\left(\frac{\mathcal{D}_2}{\mathcal{D}_1} + \mathcal{D}_3 + L(0)\right)} \tag{62}$$

From (53), it can be seen that $\mathcal{D}_1$ can be manipulated by varying the parameters λ_c, λ, θ_{a1}, θ_{a2}, θ_c, and the matrix K. Therefore, the tracking errors ξ_f, $\widetilde{W}_{a1}$, $\widetilde{W}_{a2}$ and $\widetilde{W}_c$ can be made arbitrarily small by carefully adjusting the parameters λ_c, λ, θ_{a1}, θ_{a2}, θ_c, and the matrix K to increase $\mathcal{D}_1$. Large enough $\mathcal{D}_1$ makes the ratio $\mathcal{D}_2/\mathcal{D}_1$ smaller and subsequently reduces the error signals (58)-(62). ■

Remark 4: The problem of AFM control of MAS has been studied in [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], and [26]. However, the authors did not consider the possibility of cyber-attacks on the MAS. These attacks

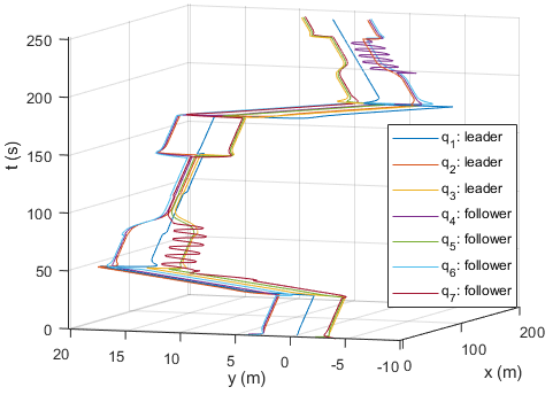

FIGURE 2. Effects of attacks on the trajectories of the leader-follower MAS under the control input proposed in [26].

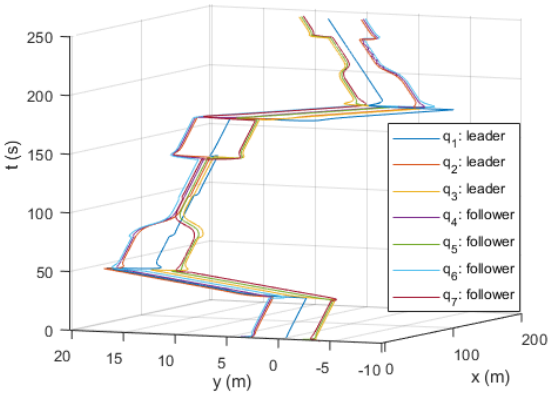

FIGURE 3. The trajectories of the leader-follower MAS under the attack mitigation control protocol (41).

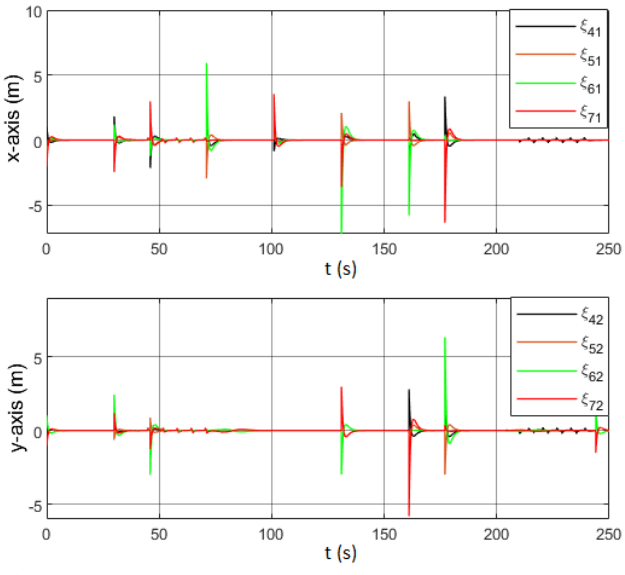

FIGURE 4. The AFM tracking errors between the followers and the leaders under the proposed scheme.

can disrupt the exchange of information among the agents and subsequently leads to the collapse of the MAS. In this

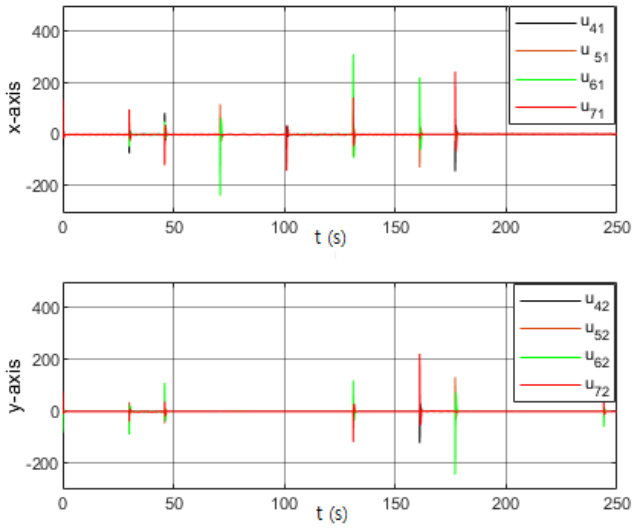

FIGURE 5. The RL control signals of the followers that mitigate actuator attacks and maintain the AFM.

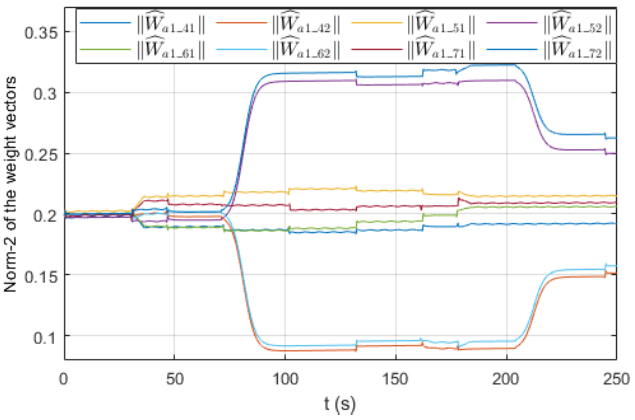

FIGURE 6. The update of the actor weights for the RBFNN approximation of $f(q, v)$.

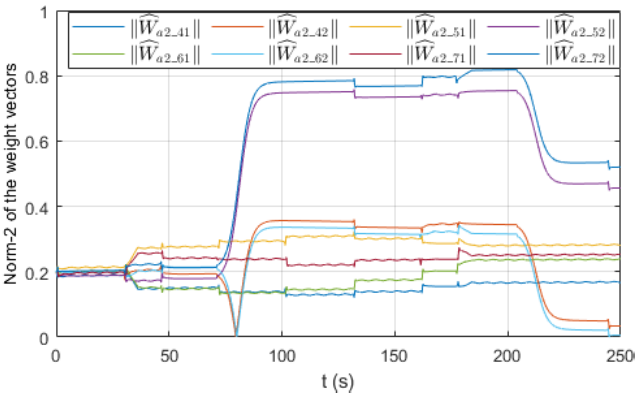

FIGURE 7. The update of the actor weights for the RBFNN approximation of $\delta(q, v)$.

article, for the first time, the secured AFM control of MAS is investigated. The distributed control laws (41) for the MAS (16) have attack estimation and suppression mechanisms via

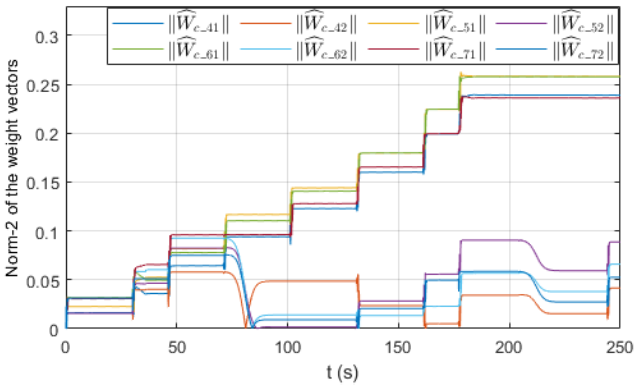

FIGURE 8. The update of the critic weights for the RBFNN approximation of $I(t)$.

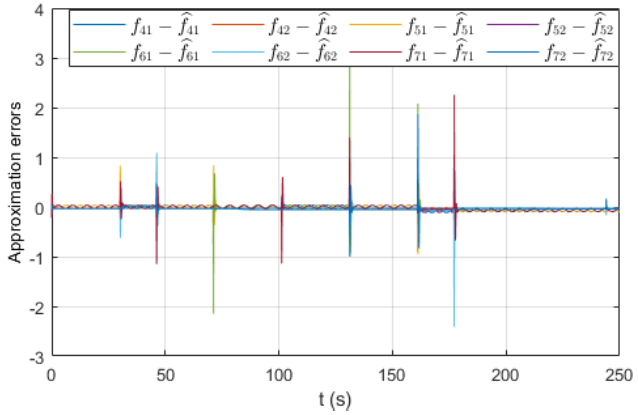

FIGURE 9. The approximation errors between $f(q.v)$ and $\hat{f}(q, v)$.

the actor RBFNN contrary to [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], and [26].

Remark 5: In contrast to [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], and [26], the distributed control laws (41) are formulated to minimize the long-term performance index (23) using the actor-critic RL architecture.

Remark 6: The authors in [26] proposed an AFM control of uncertain second-order nonlinear MAS similar to this article. Nonetheless, the MAS in [26] belongs to the class of nonlinear systems that can be expressed in linear-in-the-parameter form. The uncertain parameters were estimated using adaptive laws. In contrast, the nonlinear MAS in this paper is not required to be in the linear-in-the-parameter form since actor RBFNN are employed to estimate them as shown in (41).

IV. SIMULATION RESULTS

This section presents a numerical simulation to demonstrate the validity of the actor-critic RL scheme for achieving the AFM control of the leader-follower MAS subjected to actuator attacks. The MAS under this study consists of three leaders and four followers. The nominal formation of the agents is depicted in Fig. 1. The nominal configuration of the

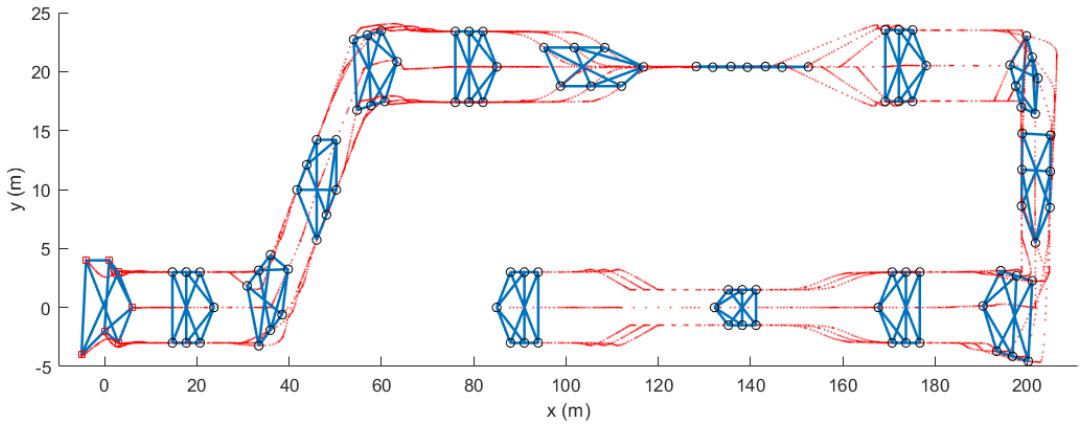

FIGURE 10. Trajectories showing various affine formation maneuvers of the leaders and followers.

agents is given in (63), as shown at the bottom of the page, where $r_1 = [6\ 0]$, $r_2 = [3\ 3]$, $r_3 = [3\ -3]$, $r_4 = [0\ 3]$, $r_5 = [0\ -3]$, $r_6 = [-3\ 3]$ and $r_7 = [-3\ -3]$. Using the method presented in [24], the matrix of the stresses among the agents is computed as (64), shown at the bottom of the page.

The nonlinear dynamic equations of the four followers are given as:

$$\begin{aligned} \dot{q}_i &= v_i \\ \dot{v}_i &= \begin{bmatrix} q_{i1}^2 + 2v_{i1}cos(q_{i1})v_{i1} \\ q_{i2}^2 + 2v_{i2}cos(q_{i2})v_{i2} \end{bmatrix} + \begin{bmatrix} u_{i1}^c \\ u_{i2}^c \end{bmatrix} + \begin{bmatrix} \varphi_{i1}(q_{i1}, v_{i1}) \\ \varphi_{i2}(q_{i2}, v_{i2}) \end{bmatrix} \end{aligned} \tag{65}$$

where $\varphi_{i1}(q_{i1}, v_{i1}) = 0.22v_{i1}\cos(0.7q_{i1}t)$, $\varphi_{i2}(q_{i2}, v_{i2}) = 0.43v_{i2}\cos(2q_{i2}t)$.

The MAS encountered two false data injection attacks on the actuators of the followers. The first attack targeted the actuator of the fourth follower for 40s $\leq t <$ 80s, whereas the second attack targeted the actuator of the first follower for 210s $\leq t <$ 240s. The comprised actuators of the fourth and the first followers as a result of the attacks are expressed as $u_{41}^c = (1.2 + 0.45\cos(q_{41})u_{41} - 62\cos(20\pi t)$, $u_{42}^c = (1.2 + 0.45\cos(q_{42})u_{42} - 62\sin(20\pi t)$, $u_{11}^c = (1.2 + 0.45\sin(q_{11})u_{11} + 62\cos(20\pi t)$, and $u_{12}^c = (1.2 + 0.45\sin(q_{12})u_{12} + 62\cos(20\pi t)$,

The initial positions of the agents are set as $r_1(0) = [6\ 0]^T$, $r_2(0) = [3\ 3]^T$, $r_3(0) = [3\ -3]^T$, $r_4(0) = [1\ 4]^T$, $r_5(0) = [0\ -2]^T$, $r_6(0) = [-4\ 4]^T$, $r_7(0) = [-5\ -4]^T$. The integral reinforcement interval is set as $\tau = 0.2$, The gains of the controllers are carefully chosen as, $\lambda = 2$, $K = 25I$. The parameters of the critic neural network are set as $\gamma_c = 0.2I$, $\theta_c = 0.1$, the width of the Gaussian function $\eta_c = 3$, the center of the receptive field c_c is chosen between $-2, -1.5, -1, \ldots, 2$, number of hidden nodes is chosen as 10, the initial critic weight vector is selected as $W_{c_0} = [0, 0, \ldots, 0]^T$. The parameters of the actor neural network are set as $\gamma_{a1} = 6I$, $\gamma_{a2} = 6I$, $\theta_{a1} = 0.1$, $\theta_{a2} = 0.1$, the width of the Gaussian function $\eta_{a1} = 3$, $\eta_{a1} = 3$, the center of the receptive field c_{a1} and c_{a2} are chosen between $-1, -0.5, \ldots, 2$, number of hidden nodes for the first and second actor neural networks are chosen as 30 and 15, respectively, the initial actor weight vectors are selected as $W_{a1_0} = [0.2, 0.2, \ldots, 0.2]^T$ and $W_{a2_0} = [0.2, 0.2, \ldots, 0.2]^T$.

It is worth noting that the AFM control method presented in [26] did not take into account the impact of cyber-attacks on the MAS. In order to demonstrate the failure of the AFM control approach proposed in [26] in the event of false-data injection attacks, it is applied to the MAS (11). Fig. 2 depicts the distorted trajectories of the followers q_4 and q_7 for 40s $\leq t <$ 80s and 210s $\leq t <$ 240s, respectively as a result of the attacks on their actuators. In practical applications, the

$$r = \begin{bmatrix} r_1^T & r_2^T & r_3^T & r_4^T & r_5^T & r_6^T & r_7^T \end{bmatrix}^T \tag{63}$$

$$\Omega = \begin{bmatrix} -0.2741 & 0.2741 & 0.2741 & -0.1370 & -0.1370 & 0 & 0 \\ 0.2741 & -0.6852 & 0 & 0.5482 & 0 & 0 & -0.1370 \\ 0.2741 & 0 & -0.6852 & 0 & 0.5482 & -0.1370 & 0 \\ -0.1370 & 0.5482 & 0 & -0.7537 & 0.0685 & 0.2741 & 0 \\ -0.1370 & 0 & 0.5482 & 0.0685 & -0.7537 & 0 & 0.2741 \\ 0 & 0 & -0.1370 & 0.2741 & 0 & -0.2741 & 0.1370 \\ 0 & -0.1370 & 0 & 0 & 0.2741 & 0.1370 & -0.2741 \end{bmatrix} \tag{64}$$

attacks will lead to the collision of the followers. Therefore, the AFM control approach proposed in [26] is not suitable for application in adversarial environments. As shown in Fig. 3, the trajectories of the agents are smooth because the proposed control technique is able to detect and mitigate the attacks on the actuators of the followers unlike [26]. The AFM tracking errors of the followers under the proposed scheme are depicted in Fig. 4. The responses of the RL-based distributed control protocols that mitigate the actuator attacks and maintain the AFM of the leader-follower system MAS is depicted in Fig. 5. The evolution of the norm-2 of the RBFNN weight vectors of the two actors and the critic networks is presented in Fig. 6, Fig. 7, and Fig. 8, respectively. It can be observed that the weights are automatically adjusted to adapt to the new collective maneuvers of the agents. In order, to show the accuracy of the RBFNN in approximating the unknown function $f(q, v)$, the approximation errors $\left(f(q, v) - \hat{f}(q, v)\right)$ are presented in Fig. 9. It can be observed that the approximation errors vary between -0.04 and 0.04 during the steady states. Fig. 10 shows the collective maneuvering of the formation of the agents such as translation, collinearity, scaling, shearing, and rotation.

V. CONCLUSION

This article investigated the AFM control problem of MASs under actuator attacks. A reinforcement learning control technique based on the actor-critic architecture is proposed to achieve the affine formation transformation and collective maneuvering of the leader-follower system. Critic neural networks have been utilized to approximate the prescribed long-term performance index which evaluates the control performance of the leader-follower network. Two actor neural networks reinforced with the critic neural network's signals have been used to generate control protocols that minimize the performance index and counter cyber attacks. The closed-loop system has been proven to be semiglobally uniformly and ultimately bounded. Simulation results have shown that the proposed control approach is able to neutralize the effects of the actuator attacks on the followers and maintain the leader-following affine collective maneuvering. Future research will incorporate event-triggered mechanisms into the control laws to lessen the computational burden.

REFERENCES

[1] K. Aryankia and R. R. Selmic, "Neural network-based formation control with target tracking for second-order nonlinear multiagent systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 1, pp. 328–341, Feb. 2022.

[2] R. Wang, X. Dong, Q. Li, and Z. Ren, "Distributed time-varying formation control for multiagent systems with directed topology using an adaptive output-feedback approach," *IEEE Trans. Ind. Informat.*, vol. 15, no. 8, pp. 4676–4685, Aug. 2019.

[3] L. Cao, H. Li, G. Dong, and R. Lu, "Event-triggered control for multiagent systems with sensor faults and input saturation," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 51, no. 6, pp. 3855–3866, Jun. 2021.

[4] X. Liu, X. Wu, Y. Xie, B. Guo, and J. Yan, "Fully distributed adaptive time-varying formation control of singular multiagent systems," *Neurocomputing*, vol. 516, pp. 146–154, Jan. 2023.

[5] L. Tian, Y. Hua, X. Dong, J. Lü, and Z. Ren, "Distributed time-varying group formation tracking for multiagent systems with switching interaction topologies via adaptive control protocols," *IEEE Trans. Ind. Informat.*, vol. 18, no. 12, pp. 8422–8433, Dec. 2022.

[6] X. Wang, B. Zerr, H. Thomas, B. Clement, and Z. Xie, "Pattern formation of multi-AUV systems with the optical sensor based on displacement-based formation control," *Int. J. Syst. Sci.*, vol. 51, no. 2, pp. 348–367, Jan. 2020.

[7] R. Babazadeh and R. R. Selmic, "Distance-based formation control of nonlinear agents over planar directed graphs," in *Proc. Amer. Control Conf. (ACC)*, Jun. 2022, pp. 2321–2326.

[8] F. Mehdifar, C. P. Bechlioulis, F. Hashemzadeh, and M. Baradarannia, "Prescribed performance distance-based formation control of multi-agent systems," *Automatica*, vol. 119, Sep. 2020, Art. no. 109086.

[9] E. Zhang, J. Li, C. Zhu, and B. Huang, "Bearing-based prescribed-time formation control of underactuated autonomous surface vehicles with the interception of attackers," *Ocean Eng.*, vol. 262, Oct. 2022, Art. no. 112187.

[10] S. Zhao and D. Zelazo, "Translational and scaling formation maneuver control via a bearing-based approach," *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 3, pp. 429–438, Sep. 2017.

[11] M. Ranjbar, M. T. H. Beheshti, and S. Bolouki, "Event-based formation control of networked multi-agent systems using complex Laplacian under directed topology," *IEEE Control Syst. Lett.*, vol. 5, no. 3, pp. 1085–1090, Jul. 2021.

[12] H. G. de Marina, "Distributed formation maneuver control by manipulating the complex Laplacian," *Automatica*, vol. 132, Oct. 2021, Art. no. 109813.

[13] X. Li and L. Xie, "Dynamic formation control over directed networks using graphical Laplacian approach," *IEEE Trans. Autom. Control*, vol. 63, no. 11, pp. 3761–3774, Nov. 2018.

[14] S.-J. Chung, S. Bandyopadhyay, I. Chang, and F. Y. Hadaegh, "Phase synchronization control of complex networks of Lagrangian systems on adaptive digraphs," *Automatica*, vol. 49, no. 5, pp. 1148–1161, May 2013.

[15] Z. Lin, L. Wang, Z. Chen, M. Fu, and Z. Han, "Necessary and sufficient graphical conditions for affine formation control," *IEEE Trans. Autom. Control*, vol. 61, no. 10, pp. 2877–2891, Oct. 2016.

[16] S. Zhao, "Affine formation maneuver control of multiagent systems," *IEEE Trans. Autom. Control*, vol. 63, no. 12, pp. 4140–4155, Dec. 2018.

[17] F. Xiao, Q. Yang, X. Zhao, and H. Fang, "A framework for optimized topology design and leader selection in affine formation control," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 8627–8634, Oct. 2022.

[18] J. Wang, X. Ding, C. Wang, L. Liang, and H. Hu, "Affine formation control for multi-agent systems with prescribed convergence time," *J. Franklin Inst.*, vol. 358, no. 14, pp. 7055–7072, Sep. 2021.

[19] K. Gao, Y. Zhou, and Y. Zhao, "Affine formation control for multi-agent systems within an appointed settling time," in *Proc. 40th Chin. Control Conf. (CCC)*, Jul. 2021, pp. 5356–5361.

[20] K. Gao, Y. Liu, Y. Zhou, Y. Zhao, and P. Huang, "Practical fixed-time affine formation for multi-agent systems with time-based generators," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 11, pp. 4433–4437, Nov. 2022.

[21] Z. Luo, P. Zhang, X. Ding, Z. Tang, C. Wang, and J. Wang, "Adaptive affine formation maneuver control of second-order multi-agent systems with disturbances," in *Proc. 16th Int. Conf. Control, Autom., Robot. Vis. (ICARCV)*, Dec. 2020, pp. 1071–1076.

[22] O. Onuoha, H. Tnunay, Z. Li, and Z. Ding, "Affine formation algorithms and implementation based on triple-integrator dynamics," *Unmanned Syst.*, vol. 7, no. 1, pp. 33–45, Jan. 2019.

[23] O. Onuoha, H. Tnunay, and Z. Ding, "Affine formation maneuver control of multi-agent systems with triple-integrator dynamics," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2019, pp. 5334–5339.

[24] Y. Xu, S. Zhao, D. Luo, and Y. You, "Affine formation maneuver control of high-order multi-agent systems over directed networks," *Automatica*, vol. 118, Aug. 2020, Art. no. 109004.

[25] O. Onuoha, H. Tnunay, C. Wang, and Z. Ding, "Fully distributed affine formation control of general linear systems with uncertainty," *J. Franklin Inst.*, vol. 357, no. 17, pp. 12143–12162, Nov. 2020.

[26] H. Zhi, L. Chen, C. Li, and Y. Guo, "Leader–follower affine formation control of second-order nonlinear uncertain multi-agent systems," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 68, no. 12, pp. 3547–3551, Dec. 2021.

[27] W. He, W. Xu, X. Ge, Q.-L. Han, W. Du, and F. Qian, "Secure control of multiagent systems against malicious attacks: A brief survey," *IEEE Trans. Ind. Informat.*, vol. 18, no. 6, pp. 3595–3608, Jun. 2022.

[28] S. Xiao, X. Ge, Q.-L. Han, and Y. Zhang, "Secure and collision-free multi-platoon control of automated vehicles under data falsification attacks," *Automatica*, vol. 145, Nov. 2022, Art. no. 110531.

[29] M. Xing, J. Lu, Y. Liu, and X. Chen, "Event-based bipartite consensus of multi-agent systems subject to DoS attacks," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 1, pp. 68–80, Jan. 2023.

[30] L. Zhou, X. Chen, and H. Wen, "Event-triggered H_∞ consensus for nonlinear multi-agent systems with semi-Markov switching topologies under DoS attacks," *Franklin Open*, vol. 2, Mar. 2023, Art. no. 100006.

[31] D. Zhang, C. Deng, and G. Feng, "Resilient cooperative output regulation for nonlinear multiagent systems under DoS attacks," *IEEE Trans. Autom. Control*, vol. 68, no. 4, pp. 2521–2528, Apr. 2023.

[32] G. Narayanan, M. S. Ali, Q. Zhu, B. Priya, and G. K. Thakur, "Fuzzy observer-based consensus tracking control for fractional-order multi-agent systems under cyber-attacks and its application to electronic circuits," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 2, pp. 698–708, Mar. 2023.

[33] A. Mousavi, K. Aryankia, and R. R. Selmic, "A distributed FDI cyber-attack detection in discrete-time nonlinear multi-agent systems using neural networks," *Eur. J. Control*, vol. 66, Jul. 2022, Art. no. 100646.

[34] S. Wang, S. Zheng, C. Zhao, H. Jian, and H. Li, "Formation control of nonlinear multi-agent systems with actuator and communication attacks," in *Proc. 40th Chin. Control Conf. (CCC)*, Jul. 2021, pp. 2286–2291.

[35] L. Xia, Q. Li, R. Song, and Z. Zhang, "Leader-follower time-varying output formation control of heterogeneous systems under cyber attack with active leader," *Inf. Sci.*, vol. 585, pp. 24–40, Mar. 2022.

[36] Y. Wang, L. Han, X. Li, X. Dong, Q. Li, and Z. Ren, "Resilient time-varying formation control of second-order discrete-time multi-agent systems with actuator faults and attacks on communication link," in *Proc. 5th Chin. Conf. Swarm Intell. Cooperat. Control*. Singapore: Springer, 2023, pp. 470–482.

[37] Y. Yang, Y. Xiao, and T. Li, "Attacks on formation control for multiagent systems," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 12805–12817, Dec. 2022.

[38] Y. Tang, D. Zhang, P. Shi, W. Zhang, and F. Qian, "Event-based formation control for nonlinear multiagent systems under DoS attacks," *IEEE Trans. Autom. Control*, vol. 66, no. 1, pp. 452–459, Jan. 2021.

[39] S. M. Elkhider, S. El-Ferik, and A. A. Saif, "Containment control of multiagent systems subject to denial of service attacks," *IEEE Access*, vol. 10, pp. 48102–48111, 2022.

[40] K. Pan, Y. Lyu, and Q. Pan, "Adaptive formation for multiagent systems subject to denial-of-service attacks," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 69, no. 8, pp. 3391–3401, Aug. 2022.

[41] J. Gong, B. Jiang, Y. Ma, and Z. Mao, "Distributed adaptive fault-tolerant formation control for heterogeneous multiagent systems with communication link faults," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 2, pp. 784–795, Apr. 2023.

[42] Y. Ouyang, C. Sun, and L. Dong, "Actor–critic learning based coordinated control for a dual-arm robot with prescribed performance and unknown backlash-like hysteresis," *ISA Trans.*, vol. 126, pp. 1–13, Jul. 2022.

[43] L. Xi, J. Wu, Y. Xu, and H. Sun, "Automatic generation control based on multiple neural networks with actor-critic strategy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 6, pp. 2483–2493, Jun. 2021.

[44] X. Wang, Q. Wang, and C. Sun, "Prescribed performance fault-tolerant control for uncertain nonlinear MIMO system using actor–critic learning structure," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4479–4490, Sep. 2022.

[45] Y. Sun, J. Xu, H. Wu, G. Lin, and S. Mumtaz, "Deep learning based semi-supervised control for vertical security of maglev vehicle with guaranteed bounded airgap," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4431–4442, Jul. 2021.

[46] D. Hu, Z. Ye, Y. Gao, Z. Ye, Y. Peng, and N. Yu, "Multi-agent deep reinforcement learning for voltage control with coordinated active and reactive power optimization," *IEEE Trans. Smart Grid*, vol. 13, no. 6, pp. 4873–4886, Nov. 2022.

[47] Y. Deng, T. Liu, and D. Zhao, "Event-triggered output-feedback adaptive tracking control of autonomous underwater vehicles using reinforcement learning," *Appl. Ocean Res.*, vol. 113, Aug. 2021, Art. no. 102676.

[48] Z. Peng, Y. Zhao, J. Hu, R. Luo, B. K. Ghosh, and S. K. Nguang, "Input–output data-based output antisynchronization control of multiagent systems using reinforcement learning approach," *IEEE Trans. Ind. Informat.*, vol. 17, no. 11, pp. 7359–7367, Nov. 2021.

[49] Z. Peng, R. Luo, J. Hu, K. Shi, S. K. Nguang, and B. K. Ghosh, "Optimal tracking control of nonlinear multiagent systems using internal reinforce Q-learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 4043–4055, Aug. 2022.

[50] G. Wen and B. Li, "Optimized leader-follower consensus control using reinforcement learning for a class of second-order nonlinear multiagent systems," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 52, no. 9, pp. 5546–5555, Sep. 2022.

[51] L. Zhang, J. Li, Y. Zhu, H. Shi, and K.-S. Hwang, "Multi-agent reinforcement learning by the actor-critic model with an attention interface," *Neurocomputing*, vol. 471, pp. 275–284, Jan. 2022.

[52] H. Liu, F. Peng, H. Modares, and B. Kiumarsi, "Heterogeneous formation control of multiple rotorcrafts with unknown dynamics by reinforcement learning," *Inf. Sci.*, vol. 558, pp. 194–207, May 2021.

[53] Y. Fei, P. Shi, and C.-C. Lim, "Robust formation control for multi-agent systems: A reference correction based approach," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no. 6, pp. 2616–2625, Jun. 2021.

[54] W. Zhao, H. Liu, K. P. Valavanis, and F. L. Lewis, "Fault-tolerant formation control for heterogeneous vehicles via reinforcement learning," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 4, pp. 2796–2806, Aug. 2022.

[55] H. Liu, W. Zhao, and J. Xi, "Optimal formation control for a quadrotor team under switching topologies via reinforcement learning," in *Proc. 4th IEEE Int. Conf. Ind. Cyber-Phys. Syst. (ICPS)*, May 2021, pp. 731–736.

[56] W. Zhao, H. Liu, and F. L. Lewis, "Robust formation control for cooperative underactuated quadrotors via reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 10, pp. 4577–4587, Oct. 2021.

[57] K. Zhao, Y. Song, C. L. P. Chen, and L. Chen, "Adaptive asymptotic tracking with global performance for nonlinear systems with unknown control directions," *IEEE Trans. Autom. Control*, vol. 67, no. 3, pp. 1566–1573, Mar. 2022.

[58] Y. Yin, F. Wang, Z. Liu, and Z. Chen, "Finite-time leader-following consensus of multiagent systems with actuator faults and input saturation," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 52, no. 5, pp. 3314–3325, May 2022.

[59] Y. Zhao, X. Du, C. Zhou, Y.-C. Tian, X. Hu, and D. E. Quevedo, "Adaptive resilient control of cyber-physical systems under actuator and sensor attacks," *IEEE Trans. Ind. Informat.*, vol. 18, no. 5, pp. 3203–3212, May 2022.

[60] B. Chen, Y. Tan, Z. Sun, and L. Yu, "Attack-resilient control against FDI attacks in cyber-physical systems," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 6, pp. 1099–1102, Jun. 2022.

[61] Z. Cao, Y. Niu, and Y. Zou, "Adaptive neural sliding mode control for singular semi-Markovian jump systems against actuator attacks," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 51, no. 3, pp. 1523–1533, Mar. 2021.

[62] Z. Jing, Y. Ma, X. Wu, X. He, and Y. Sun, "Backstepping control for vibration suppression of 2-D Euler–Bernoulli beam based on nonlinear saturation compensator," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 53, no. 5, pp. 2562–2571, May 2023.

SAMI EL-FERIK received the B.Sc. degree in electrical engineering from Laval University, Quebec City, QC, Canada, and the M.Sc. and Ph.D. degrees in electrical and computer engineering from Polytechnique Montreal, Canada. After the completion of the Ph.D. and postdoctoral positions, he was with the Research and Development Center of Systems, Controls, and Accessories, Pratt and Whitney, Canada. He is currently a Professor of control and instrumentation engineering with the Department of Systems Engineering, King Fahd University of Petroleum and Minerals. He is also the Director of the Interdisciplinary Research Center for Smart Mobility and Logistics. His research interests include sensing, monitoring, multiagent systems, and nonlinear control with strong multidisciplinary research and applications.

MUHAMMAD MAARUF received the B.S. degree in electrical engineering from the Kano University of Science and Technology, Kano, Nigeria, in 2016, and the M.Sc. degree in system and control engineering from the King Fahd University of Petroleum and Minerals, Dammam, Saudi Arabia, where he is currently pursuing the Ph.D. degree in control and instrumentation engineering. His current research interests include control theory, multi-agent systems, and artificial intelligence with applications to unmanned aerial vehicles, renewable energy systems, and energy storage systems.

FOUAD M. AL-SUNNI received the B.S. and M.S. degrees in systems engineering from the King Fahd University of Petroleum and Minerals (KFUPM), in 1984 and 1987, respectively, and the Ph.D. degree in electrical engineering from The University of Texas, in 1992. He was the Chairperson of the Systems Engineering Department, KFUPM, from 2006 to 2015, and the General Director of the Industrial Research Cooperation, KFUPM, from 2015 to 2016. He is currently a Full Professor and a part-time Faculty Member with KFUPM. He is also the General Manager of SIBCA automation and fire protection systems company, Saudi Arabia. He is the coauthor of the book *Mechatronic Systems: Analysis, Design and Implementation* and the coauthor of the book *Control and Optimization of Distributed Generation Systems*. He published more than 100 journal articles and conference research papers on optimal control, robust control, fuzzy logic, ANN, and AI in control.

ABDULWAHID ABDULAZIZ SAIF received the B.Sc. degree from the Physics Department, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, the M.Sc. degree from the Systems Engineering Department, King Fahd University of Petroleum and Minerals, and the Ph.D. degree from the Control and Instrumentation Group, Department of Engineering, Leicester University, Leicester, U.K. He was a Research Assistant with the SE Department and a Lecturer with the Electrical Engineering Department and a Lecturer with the Physics Department, King Fahd University of Petroleum and Minerals. After finishing the Ph.D. degree, he joined the Systems Engineering Department. He is currently a Professor of control and instrumentation with the Control and Instrumentation Engineering Department (CIE) and a member of the Interdisciplinary Research Center for Smart Mobility and Logistics (IRC-SML), King Fahd University of Petroleum and Minerals. He taught several courses in modeling and simulation, digital control, digital systems, microprocessor and microcontrollers in automation, optimization, numerical methods, PLC's, process control, and control system design. He has published more than 100 papers in reputable journals and conferences. His research interests include simultaneous and strong stabilization, robust control and H_∞-optimization, and instrumentation and computer control.

MUJAHED MOHAMMAD AL DHAIFALLAH (Member, IEEE) received the B.Sc. and M.Sc. degrees in systems engineering from the King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, and the Ph.D. degree in electrical and computer engineering from the University of Calgary, Calgary, AB, Canada. Since 2020, he has been an Associate Professor of control and instrumentation engineering with the King Fahd University of Petroleum and Minerals. His current research interests include nonlinear systems identification, control systems, optimization, artificial intelligence, and renewable energy.