SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

July 2026, Vol. 69, Iss. 7, 172209:1–172209:14
https://doi.org/10.1007/s11432-025-4941-x

Resilient human-in-the-loop cooperative optimal control for constrained-input multiagent systems under DoS attacks via reinforcement learning

Zongsheng HUANG[1], Tieshan LI[1,2,3*], Yue LONG[1] & Hongjing LIANG[1]

[1]*School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China*
[2]*Yangtze Delta Region Institute, University of Electronic Science and Technology of China, Huzhou 313000, China*
[3]*Laboratory of Electromagnetic Space Cognition and Intelligent Control, Beijing 100089, China*

Received 4 December 2025/Revised 28 February 2026/Accepted 4 April 2026/Published online 4 June 2026

Abstract The resilient human-in-the-loop (HiTL) cooperative optimal control problem for constrained-input multiagent systems (MASs) under denial-of-service (DoS) attacks is studied in this work. First, the HiTL structure allows the supervisor to enhance the safety of MASs by transmitting commands to the leader in emergency scenarios. However, the occurrence of DoS attacks periodically disrupts communication between agents and their neighbors, leading to intermittent topology connectivity. To counteract these disruptions, a switching control strategy is proposed. Specifically, agents execute optimal consensus control during dormant attack intervals; during active attack periods, the attacked agents switch to optimal stabilizing control, while the non-attacked ones maintain the optimal consensus control. Then, to achieve optimal performance under physical limitations, a nonquadratic cost function is designed to accommodate asymmetric input constraints, leading to the formulation of a Hamilton-Jacobi-Bellman (HJB) equation. Subsequently, the reinforcement learning method is developed to derive the optimal control strategy, in which the critic neural network (NN) approximates the solution of the HJB equation, with its weights updated via the gradient descent method. The boundedness of all signals in the closed-loop system is proven. Finally, the simulation results verify the effectiveness of the control scheme.

Keywords reinforcement learning, human-in-the-loop control, optimal control, denial-of-service attacks, constrained-input multiagent systems

Citation Huang Z S, Li T S, Long Y, et al. Resilient human-in-the-loop cooperative optimal control for constrained-input multiagent systems under DoS attacks via reinforcement learning. Sci China Inf Sci, 2026, 69(7): 172209, https://doi.org/10.1007/s11432-025-4941-x

1 Introduction

Inspired by the cooperative behaviors of animals, scholars have been fascinated by the study of multiagent systems (MASs) [1]. One key reason for this interest is the applications of MASs in engineering areas such as spacecraft formation flying [2] and cooperative surveillance [3]. Over the past decades, a wide range of cooperative control issues in MASs have been thoroughly explored. Consensus control, as a key area of cooperative control for MASs, has been favored by scholars. In particular, the majority of recent studies on MASs focus on fully autonomous operation. However, accidents involving Boeing 737 jetliners and Tesla's autonomous driving have served as critical reminders for the control of autonomous systems. Consequently, it is crucial to develop schemes for monitoring the entire system to ensure task completion when MASs are at risk or in unexpected situations. As a result, the human-in-the-loop (HiTL) control approach was first introduced for MASs in [4], where a human operator supervised the system in response to sudden changes by broadcasting a signal to the leader. Later, based on [4], abundant results associated with HiTL control for MASs popped up [5–10]. In [5], an asynchronous edge-based event-triggered HiTL control scheme was proposed to achieve time-varying formation of MASs, subject to the unknown input of the human-manipulated leader. The problem of fuzzy dynamic event-triggered containment control for HiTL MASs with error constraints was explored in [7]. Using its inherent strengths in safety and reliability, the HiTL method has achieved significant success in various practical applications, including multi-UAV systems [9] and marine vehicles [10]. These practical implementations highlight the considerable promise of the HiTL method.

Based on this, optimal control has emerged as a pivotal framework for MASs. To integrate the benefits of optimal control and adaptive control, concepts from the field of machine learning, namely reinforcement learning

* Corresponding author (email: litieshan073@uestc.edu.cn)

(RL), are adopted. The adaptive dynamic programming (ADP) approach is similar to RL. Thus, they are two interchangeable concepts in the field of optimal control [11]. For nonlinear systems, the optimal solution is associated with the Hamilton-Jacobi-Bellman (HJB) equation [12]. However, solving the HJB equation numerically is not straightforward. To address this issue, Werbos in [13] introduced the ADP method. The core concept of ADP is to approximate the solution to the HJB equation through a function approximation structure. Building on pioneering work, several results have been made for MASs. For example, the RL-based event-triggered optimal consensus control problem for cooperative-competitive MASs was solved in [14]. In [15], a novel policy iteration-based RL method was designed, in which distributed optimal control policies can be learned based only on the agent's and its neighbors' states. In the standard ADP architecture [13], the actor network typically improves the control policy, while the critic network learns the value function. For control-affine systems, this structure can be simplified to a critic-only network structure, which alleviates the computational burden and eliminates the approximation errors caused by the action network [16]. In [17], the problem of secure containment control for multiple autonomous aerial vehicles was solved using the fixed-time convergent RL method with the critic-only NN structure. To investigate the optimal consensus control problem for nonlinear MASs with unknown dynamics and disturbances, an RL method with a critic-only structure was designed in [18]. For MASs under false data injection attacks, the RL-based secure formation control scheme using a critic-only structure was developed in [19].

However, in practical systems, there are many physical constraints, such as input or actuator constraints. These have led to extensive discussions regarding the safety of the RL-based optimal controller design under such physical constraints. Consequently, the study of the optimal controller design subject to input constraints has become a prominent research topic. In [20], an RL-based method was first constructed to solve the optimal control problem for nonlinear systems with saturating actuators. Subsequently, in [21], a novel RL-based algorithm was proposed to address the constrained-input robust control problem for uncertain nonlinear systems. Furthermore, an RL algorithm was developed to solve the tracking control problem for partially unknown systems with input constraints in [22]. In [23], the robust tracking control problem for constrained-input MASs was addressed, and a novel function was constructed to tackle asymmetric constraints.

MASs operating in open communication networks are vulnerable to malicious cyber attacks due to inherent security vulnerabilities. Typical types of cyber attacks include deception attacks and denial-of-service (DoS) attacks [24]. DoS attacks can induce switching of communication topologies, potentially causing transient instability during the switching process [25]. Additionally, when an agent is targeted, the information flow between that agent and its neighbors is blocked, thereby disrupting the connectivity of the communication graph. More critically, such a disruption of communication topology may ultimately compromise the stability of the MASs. Therefore, cooperative control for MASs under DoS attacks has been extensively discussed. In [26], the problem of containment control for MASs with output saturation and aperiodic DoS attacks was considered. In [27], the bipartite leader-follower consensus control problem for MASs under DoS attacks and completely unknown system dynamics was studied. In [28], the stabilization problem for switched systems in the presence of DoS attacks was solved, where multiple Lyapunov functions were constructed capturing the joint effects of DoS attacks and controller modes. For microgrids facing time-constrained DoS attacks, a mitigation adaptive secondary control method was proposed in [29]. However, most existing studies on DoS attacks rely on the assumption of continuous topology connectivity, which is often impractical. Thus, developing an RL-based HiTL optimal control scheme for input-constrained MASs under DoS attacks, which cause intermittent topology connectivity, remains a challenging research direction.

To address these issues, this work aims to develop a resilient HiTL optimal control strategy for constrained-input MASs under DoS attacks, relaxing the assumption of continuous topology connectivity. However, achieving this objective presents at least two significant technical challenges.

(1) Intermittent DoS attacks block communication between agents, resulting in discontinuous topology connectivity. This renders the control scheme designed in previous studies, such as [25, 26], which rely on continuous connectivity, inapplicable, and poses a significant challenge to optimal control design.

(2) The presence of asymmetric input constraints in practical systems imposes additional complexity, which makes the design of an optimal control scheme for MASs particularly challenging.

The main contributions are summarized below.

(1) A supervisor is integrated to monitor the MASs and send commands to the leader, thus enhancing the security and emergency response capabilities for accident prevention. Moreover, this work generalizes existing HiTL consensus results from regular environments, as in [5, 8], to scenarios involving malicious attacks, thus broadening the scope of HiTL research.

(2) This work distinguishes itself from existing studies on DoS attacks, such as [25, 26], which rely on the assumption of continuous topology connectivity. To address this gap, this work proposes a more practical control scheme capable of handling communication interruptions induced by intermittent DoS attacks. Specifically, the

scheme stabilizes the attacked agents while maintaining cooperative control for the unaffected agents, thereby reducing the conservatism in the controller design.

(3) To address practical application requirements, this work incorporates asymmetric input constraints. By employing a nonquadratic cost function tailored for such constraints, the proposed approach achieves greater generality than existing methods that only address symmetric constraints, as seen in [21, 22].

The structure is listed below. In Section 2, the considered system and some assumptions are given. In Section 3, the main results regarding controller design and stability analysis are provided. The simulation results are given in Section 4. Finally, the conclusion is presented in Section 5.

Notations: Throughout the article, $\lambda_{\min}(\cdot)$ denotes the minimal eigenvalue of the matrix.

2 Problem formulation and preliminaries

2.1 Communication topology

The communication topology containing N followers is represented by a directed graph $\mathcal{G} = \{\mathcal{V}, \varepsilon, \mathcal{A}\}$, where $\mathcal{V} = \{\mathcal{V}_1, \mathcal{V}_2, \cdots, \mathcal{V}_N\}$ represents the vertex set, $\varepsilon \subseteq \mathcal{V} \times \mathcal{V}$ represents the edge set of N followers. Let $\mathcal{A} = [a_{ij}] \in \mathbb{R}^{N\times N}$ be the weight matrix, where $a_{ij} > 0$ if $(\mathcal{V}_i, \mathcal{V}_j) \in \varepsilon$; otherwise, $a_{ij} = 0$. Define $\mathcal{L} = \mathcal{D} - \mathcal{A} \in \mathbb{R}^{N\times N}$ as the Laplacian matrix of $\mathcal{G}$, where $\mathcal{D} = \mathrm{diag}(d_1, d_2, \cdots, d_N) \in \mathbb{R}^{N\times N}$ denotes the degree matrix with $d_i = \sum_{j=1}^{N} a_{ij}$. The augmented graph consisting of one leader and N followers is denoted by $\tilde{\mathcal{G}} = \{\tilde{\mathcal{V}}, \tilde{\varepsilon}\}$, in which $\tilde{\mathcal{V}} = \{\mathcal{V}_0, \mathcal{V}_1, \mathcal{V}_2, \cdots, \mathcal{V}_N\}$ and $\tilde{\varepsilon} \subseteq \tilde{\mathcal{V}} \times \tilde{\mathcal{V}}$. Let $\mathcal{B} = \mathrm{diag}\{b_1, b_2, \cdots, b_N\} \in \mathbb{R}^{N\times N}$, where $b_i = 1$ indicates that leader's information is available for the ith node; otherwise, $b_i = 0$.

2.2 Problem formulation

Assume that the nonlinear MAS is composed of N ($\geqslant 2$) followers and one leader. The dynamic model of the ith follower is provided as

$$\dot{x}_i = f_i(x_i) + g_i(x_i)u_i, i = 1, 2, \cdots, N, \tag{1}$$

where $x_i(t) \in \mathbb{R}^n$ denotes the state, $f_i(x_i) \in \mathbb{R}^n$ is the internal dynamics, $g_i(x_i) \in \mathbb{R}^{n\times m}$ is the input dynamics, and $u_i(t) \in \mathcal{U} \subset \mathbb{R}^m$ is the control input suffering from asymmetric constraints, where $\mathcal{U} = \{(u_{i,1}, u_{i,2}, \cdots, u_{i,m}) \in \mathbb{R}^m : u_{i,\min} \leqslant u_{i,c} \leqslant u_{i,\max}, c = 1, 2, \cdots, m\}$ with $u_{i,\min}$ and $u_{i,\max}$ being the minimum and maximum limits and $|u_{i,\min}| \neq |u_{i,\max}|$.

To improve the security of the MASs in emergency situations, a supervisor is introduced to monitor the entire system and send commands u_0^s to the leader. The dynamic model of the supervisor-guided leader is given by

$$\dot{x}_0^s = f_0^s(x_0^s) + u_0^s, \tag{2}$$

where $x_0^s(t) \in \mathbb{R}^n$ denotes the state, $f_0^s(x_0^s) \in \mathbb{R}^n$ represents the internal dynamics, and $u_0^s(t) \in \mathbb{R}^m$ is the trajectory adjustment signal that the supervisor sends to the leader.

Remark 1. When MASs encounter emergencies during operation, such as a sudden obstacle or collision, the supervisor sends a bounded trajectory adjustment signal $u_0^s(t) \neq 0$ to the leader. The leader then adjusts its trajectory. Simultaneously, this adjustment signal is indirectly transmitted to followers via the communication topology, allowing MASs to collaboratively avert the emergency and improve their safety and reliability.

Then, some necessary assumptions are provided.

Assumption 1. There exist known constants $\bar{u}_0^s$ and $\underline{u}_0^s$ that satisfy $\underline{u}_0^s \leqslant u_0^s \leqslant \bar{u}_0^s$.

Assumption 2. For any x_i, $g_{i,\min} \leqslant \|g_i(x_i)\| \leqslant g_{i,\max}$ holds, where $g_{i,\min} > 0$ and $g_{i,\max} > 0$ are known constants.

Remark 2. Assumption 1 is reasonable in practical applications, as the control input of the non-autonomous leader is bounded. Similar assumptions can be found in the literature, e.g., [5–8]. For Assumption 2, the boundedness of $\|g_i(x_i)\|$ is standard in nonlinear control design. The positive lower bound $g_{i,\min} > 0$ ensures that the control direction is well-defined and avoids singularity, while the upper bound $g_{i,\max}$ ensures that the control effort remains bounded. Similar assumptions can be found in [30, 31].

2.3 DoS attacks modeling

In this paper, we consider the topology in which agents are subjected to intermittent DoS attacks. When an agent is subjected to the DoS attack, it cannot send or receive information and temporarily ceases communication, relying solely on its own inherent dynamics to evolve. Since the energy of the attacks is limited, the communication can

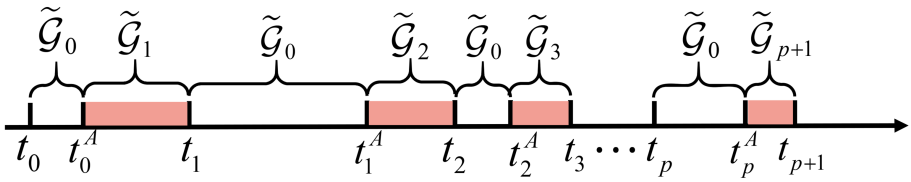

Figure 1 (Color online) The effect of DoS attacks on topology structure.

quickly recover during the intervals when the attack is dormant. For simplicity, let $0 = t_0 < t_1 < \cdots < t_p < \cdots$ denote the time sequence of dormant attacks and $t_0^A < t_1^A < \cdots < t_p^A < \cdots$ denote the time sequence of active attacks. Therefore, the time sequence of topology switching is denoted by $0 = t_0 < t_0^A < t_1 < t_1^A < \cdots < t_p < t_p^A < t_{p+1} \cdots$. The intervals $[t_p, t_p^A)$ and $[t_p^A, t_{p+1}), p = 0, 1, 2, \cdots$ represent periods without and with DoS attacks, respectively, where $t_p < t_p^A < t_{p+1}$, and t_p^A marks the occurrence of the $(p+1)$-th DoS attack. Once the attack is dormant, the topology recovers to $\tilde{\mathcal{G}}_0$ with $\tilde{\mathcal{G}}_0 = \tilde{\mathcal{G}}$ until the next attack occurs. When certain agents are attacked, the topology reorganizes into a new graph described by a finite set $\{\tilde{\mathcal{G}}_1, \tilde{\mathcal{G}}_2, \ldots, \tilde{\mathcal{G}}_q\}$. Thus, the switching signal $\varsigma(t) : [0, \infty) \to \Delta = \{0, 1, 2, \ldots, q\}$ is used to describe the evolution of the topology, where $q \geqslant 1$. The set of graphs $\tilde{\mathcal{G}}_{\varsigma(t)} = \{\tilde{\mathcal{G}}_0, \tilde{\mathcal{G}}_1, \tilde{\mathcal{G}}_2, \ldots, \tilde{\mathcal{G}}_q\}$ captures all potential topologies. Figure 1 describes the effect of DoS attacks on topology.

To develop the controller, the assumption is given below.

Assumption 3. For $\varsigma(t) \in \Delta$ and $T \geqslant t > 0$, the constant γ represents the dwell time of the switching signal $\varsigma(t)$ if the following condition is satisfied: $N_\varsigma(t, T) \leqslant N_0 + \frac{T-t}{\gamma}$, where $N_0 > 0$, $\gamma > 0$, and $N_\varsigma(t, T)$ is the number of switchings of the signal $\varsigma(t)$ within the interval $[t, T]$.

Remark 3. Inspired by [12], a restriction is imposed on the switching topology: the number of the switchings of the signal $\varsigma(t)$ is limited. Therefore, Assumption 3 describes a restriction on the frequency of switching. According to [32], the durations of the attacks and the dormant periods satisfy $t_{p+1} - t_p^A > \gamma$ and $t_p^A - t_p > \gamma, p = 0, 1, 2, \cdots$.

3 Main results

3.1 Optimal controller design under DoS attacks

The optimal controller design under DoS attacks is divided into two parts: (1) the dormant periods of attacks $[t_p, t_p^A)$; (2) the active periods of attacks $[t_p^A, t_{p+1}), p = 0, 1, 2, \cdots$.

Part I: Controller design during the dormant periods of attacks $[t_p, t_p^A)$.

During $t \in [t_p, t_p^A)$, all agents implement cooperative control. The consensus error of agent i during the dormant periods of the attacks is denoted by $\varpi_{i,1} = \sum_{j=1}^N a_{ij}(x_i - x_j) + b_i(x_i - x_0^s)$.

Considering that Eq. (1) is affected by asymmetric input constraints, the cost function is defined as

$$\mathcal{V}_i(\varpi_{i,1}) = \int_t^\infty (\varpi_{i,1}^{\mathrm{T}} Q_{ii} \varpi_{i,1} + \mathcal{R}_i(u_{\bar{N}_i}))d\tau, \tag{3}$$

where Q_{ii} is a positive symmetric definite matrix, $\bar{N}_i = \{i\} \cup \mathcal{N}_i$ and $\mathcal{N}_i$ is the set of neighbors of the ith follower. Following [30], $\mathcal{R}_i(u_{\bar{N}_i})$ is defined as

$$\mathcal{R}_i(u_{\bar{N}_i}) = 2\alpha_i \sum_{c=1}^m \int_{\beta_i}^{u_{i,c}} \iota_i^{-1}\left(\alpha_i^{-1}(k_{i,c} - \beta_i)\right) dk_{i,c} + \sum_{j \in \mathcal{N}_i} 2\alpha_j \sum_{c=1}^m \int_{\beta_j}^{u_{j,c}} \iota_j^{-1}\left(\alpha_j^{-1}(k_{j,c} - \beta_j)\right) dk_{j,c},$$

where ι_i^{-1} denotes a bounded monotonic odd function that satisfies $\iota_i^{-1}(0) = 0$, with $\alpha_i = (u_{i,\max} - u_{i,\min})/2$ and $\beta_i = (u_{i,\max} + u_{i,\min})/2$. Without loss of generality, we set $\iota_i(\cdot) = \tanh(\cdot)$ (or equivalently, $\iota_i^{-1}(\cdot) = \tanh^{-1}(\cdot)$). Although $\tanh(\cdot)$ is symmetric, the mapping $\mathcal{R}_i(u_{\bar{N}_i})$ in (3) incorporates asymmetric input constraints. This comes from the condition $|u_{i,\min}| \neq |u_{i,\max}|$, which implies $\beta_i \neq 0$.

The Bellman function, as well as the Hamiltonian function, is defined as

$$\mathcal{H}_{i,1}(\varpi_{i,1}, u_{\bar{N}_i}, \nabla \mathcal{V}_i^*) = \varpi_{i,1}^{\mathrm{T}} Q_{ii} \varpi_{i,1} + \mathcal{R}_i(u_{\bar{N}_i}) + \nabla \mathcal{V}_i^{*\mathrm{T}} l_{i,1} = 0, \tag{4}$$

where $\nabla\mathcal{V}_i^* = \frac{\partial \mathcal{V}_i^*}{\partial \varpi_{i,1}}$ and $l_{i,1} = d_i(f_i(x_i) - f_j(x_j)) + b_i(f_i(x_i) - f_0^s(x_0^s)) + (d_i + b_i)g_i(x_i)u_i - d_i g_j(x_j)u_j - b_i u_0^s$. The optimal controller u_i^* during dormant periods is derived by differentiating (4) with respect to u_i, which produces

$$u_i^*(\varpi_{i,1}) = -\alpha_i \tanh\left(\frac{d_i + b_i}{2\alpha_i} g_i^{\mathrm{T}}(x_i)\nabla\mathcal{V}_i^*(\varpi_{i,1})\right) + \imath_{\beta i}, \tag{5}$$

where $\imath_{\beta i} = [\beta_i, \ldots, \beta_i]^{\mathrm{T}} \in \mathbb{R}^m$.

Note that $\nabla\mathcal{V}_i^*$ in (5) is a nonlinear term, which makes it difficult to obtain directly by solving (4). The critic NN is applied to approximate the optimal cost function $\mathcal{V}_i^*$ as

$$\mathcal{V}_i^*(\varpi_{i,1}) = \theta_i^{\mathrm{T}}\phi_i(\varpi_{i,1}) + \nu_i(\varpi_{i,1}),$$

where $\theta_i \in \mathbb{R}^{n_{hi}}$ is the ideal weight vector and n_{hi} denotes the number of neurons in the hidden layer, $\phi_i(\varpi_{i,1})$ is the activation function and $\nu_i(\varpi_{i,1})$ is the approximate error.

The gradient form can be expressed as $\nabla\mathcal{V}_i^*(\varpi_{i,1}) = \nabla\phi_i^{\mathrm{T}}(\varpi_{i,1})\theta_i + \nabla\nu_i(\varpi_{i,1})$, where $\nabla\phi_i(\varpi_{i,1}) = \partial\phi_i(\varpi_{i,1})/\partial\varpi_{i,1}$ and $\nabla\nu_i(\varpi_{i,1}) = \partial\nu_i(\varpi_{i,1})/\partial\varpi_{i,1}$.

Assumption 4. The critic weight, the gradient of the activation function, the approximation error, and the term $l_{i,1}$ are bounded as $\|\theta_i\| \leqslant \theta_{i,\max}$, $\|\phi_i\| \leqslant \phi_{i,\max}$, $\|\nu_i\| \leqslant \nu_{i,\max}$, and $\|l_{i,1}\| \leqslant l_{i,1,\max}$.

The optimal controller u_i^* in (5) becomes

$$u_i^*(\varpi_{i,1}) = -\alpha_i \tanh(\mathcal{K}_1(\varpi_{i,1})) + \chi_i(\varpi_{i,1}) + \imath_{\beta i}, \tag{6}$$

where

$$\mathcal{K}_1(\varpi_{i,1}) = \frac{d_i + b_i}{2\alpha_i} g_i^{\mathrm{T}}(x_i)\nabla\phi_i^{\mathrm{T}}(\varpi_{i,1})\theta_i,$$
$$\chi_i(\varpi_{i,1}) = -\frac{d_i + b_i}{2}\left(\mathbb{I}_m - D(v(\varpi_{i,1}))\right) g_i^{\mathrm{T}}(x_i)\nabla\nu_i(\varpi_{i,1}),$$

with $D(v(\varpi_{i,1})) = \mathrm{diag}\left(\tanh^2(v_c(\varpi_{i,1}))\right), c = 1, 2, \ldots, m, v(\varpi_{i,1}) = [v_1(\varpi_{i,1}), v_2(\varpi_{i,1}), \ldots, v_m(\varpi_{i,1})]^{\mathrm{T}} \in \mathbb{R}^m$, and $v(\varpi_{i,1}) \in \mathbb{R}^m$ selected between $\left(\frac{d_i+b_i}{2\alpha_i} g_i^{\mathrm{T}}(x_i)\nabla\mathcal{V}_i^*(\varpi_{i,1})\right)$ and $\mathcal{K}_1(\varpi_{i,1})$.

Remark 4. We make the following simplified procedure to explain (6). Define

$$\mathrm{C}(\mathcal{K}_o(\varpi_{i,1})) = -\alpha_i \tanh(\mathcal{K}_o(\varpi_{i,1})), o = 0, 1,$$

where

$$\mathcal{K}_0(\varpi_{i,1}) = \frac{d_i + b_i}{2\alpha_i} g_i^{\mathrm{T}}(x_i)\nabla\mathcal{V}_i^*(\varpi_{i,1}).$$

Using the mean value theorem, one has

$$\mathrm{C}(\mathcal{K}_0(\varpi_{i,1})) - \mathrm{C}(\mathcal{K}_1(\varpi_{i,1})) = -\alpha_i(\tanh(\mathcal{K}_0) - \tanh(\mathcal{K}_1)) = -\frac{d_i + b_i}{2}\left(\mathbb{I}_m - D(v(\varpi_{i,1}))\right) g_i^{\mathrm{T}}(x_i)\nabla\nu_i(\varpi_{i,1}).$$

Thus, we obtain

$$u_i^*(\varpi_{i,1}) = -\alpha_i \tanh\left(\mathcal{K}_0(\varpi_{i,1})\right) + \imath_{\beta i} = -\alpha_i \tanh\left(\mathcal{K}_1(\varpi_{i,1})\right) + \imath_{\beta i} - \frac{d_i + b_i}{2}\left(\mathbb{I}_m - D(v(\varpi_{i,1}))\right) g_i^{\mathrm{T}}(x_i)\nabla\nu_i(\varpi_{i,1}). \tag{7}$$

This confirms that Eq. (6) is valid.

As mentioned above, θ_i in (6) is typically not available. Therefore, $u^*(\varpi_{i,1})$ in (6) cannot be applied directly. Thus, the estimated form of (6) is expressed as

$$\hat{u}_i^* = -\alpha_i \tanh(\mathcal{K}_2(\varpi_{i,1})) + \imath_{\beta i}, \tag{8}$$

where $\hat{\theta}_i$ is the estimate of θ_i, and $\mathcal{K}_2(\varpi_{i,1}) = \frac{d_i+b_i}{2\alpha_i} g_i^{\mathrm{T}}(x_i)\nabla\phi_i^{\mathrm{T}}(\varpi_{i,1})\hat{\theta}_i$. The weight estimation error is defined as $\tilde{\theta}_i = \theta_i - \hat{\theta}_i$. Then, the estimated Hamiltonian function is given by

$$\begin{aligned}\hat{\mathcal{H}}_{i,1}(\varpi_{i,1}, \hat{\theta}_i) =& \varpi_{i,1}^{\mathrm{T}} Q_{ii}\varpi_{i,1} + \mathcal{R}_i(\hat{u}_{\bar{N}_i}) + \nabla\phi_i^{\mathrm{T}}(\varpi_{i,1})\hat{\theta}_i((d_i + b_i)g_i(x_i)\hat{u}_i^* - d_i g_j(x_j)\hat{u}_j^*) \\ &+ \nabla\phi_i^{\mathrm{T}}(\varpi_{i,1})\hat{\theta}_i(d_i(f_i(x_i) - f_j(x_j)) + b_i(f_i(x_i) - f_0^s(x_0^s) - u_0^s)),\end{aligned}$$

and $\epsilon_{i,1}^B = \hat{\mathcal{H}}_{i,1}$ is defined as the Bellman residual error.

The objective function is defined as $E_{i,1}^B = \frac{1}{2}(\epsilon_{i,1}^B)^2$, which is minimized by updating $\hat{\theta}_i$. The gradient descent method is applied to obtain the NN weight updating law as

$$\dot{\hat{\theta}}_i = -\frac{\mu_{i,1}\check{\hbar}_{i,1}}{\Pi_{i,1}}\epsilon_{i,1}^B, \tag{9}$$

where the learning rate $0 < \mu_{i,1} < 1$,$\Pi_{i,1} = 1 + \hbar_{i,1}^{\mathrm{T}}\hbar_{i,1}$, $\check{\hbar}_{i,1} = \hbar_{i,1}/\Pi_{i,1}$, and $\hbar_{i,1} = \nabla\phi_i^{\mathrm{T}}(\varpi_{i,1})(d_i(f_i(x_i) - f_j(x_j)) + b_i(f_i(x_i) - f_0^s(x_0^s)) + (d_i + b_i)g_i(x_i)\hat{u}_i^* - d_i g_j(x_j)\hat{u}_j^* - b_i u_0^s)$.

Part II: Controller design in active attack periods $[t_p^A, t_{p+1})$.

During $t \in [t_p^A, t_{p+1})$, when a portion of the agents is attacked, the remaining unaffected agents execute the optimal consensus strategy, while the attacked agents focus on stabilizing control. For clarity, let $\mathcal{N}^A$ and $\mathcal{N}^C$ denote the sets of attacked agents and non-attacked agents, respectively.

First, the consensus error of the non-attacked agent i is defined as $\varpi_{i,2} = \sum_{j\in\mathcal{N}_i^C} a_{ij}(x_i - x_j) + b_i(x_i - x_0^s)$, where $\mathcal{N}_i^C$ is the set of neighbors of non-attacked agent i. The cost function for the non-attacked agent is constructed as

$$\mathcal{V}_i(\varpi_{i,2}) = \int_t^\infty (\varpi_{i,2}^{\mathrm{T}} Q_{ii}\varpi_{i,2} + \mathcal{R}_i(u_{\bar{\mathcal{N}}_i}))d\tau, i \in \mathcal{N}^C,$$

where $\bar{\mathcal{N}}_i = \{i\} \cup \mathcal{N}_i^C$. Let $\check{d}_i = \sum_{j\in\mathcal{N}_i^C} a_{ij}$. The Hamiltonian function is defined as

$$\mathcal{H}_{i,2}(\varpi_{i,2}, u_{\bar{\mathcal{N}}_i}, \nabla\mathcal{V}_i^*) = \varpi_{i,2}^{\mathrm{T}} Q_{ii}\varpi_{i,2} + \mathcal{R}_i(u_{\bar{\mathcal{N}}_i}) + \nabla\mathcal{V}_i^{*\mathrm{T}} l_{i,2} = 0, \tag{10}$$

where $\nabla\mathcal{V}_i^* = \frac{\partial\mathcal{V}_i^*}{\partial\varpi_{i,2}}$, $l_{i,2} = \check{d}_i(f_i(x_i) - f_j(x_j)) + b_i(f_i(x_i) - f_0^s(x_0^s)) + (\check{d}_i + b_i)g_i(x_i)u_i - \check{d}_i g_j(x_j)u_j - b_i u_0^s$ and $\|l_{i,2}\| \leqslant l_{i,2,\max}$. The optimal controller u_i^* is given by

$$u_i^*(\varpi_{i,2}) = -\alpha_i \tanh\left(\frac{\check{d}_i + b_i}{2\alpha_i} g_i^{\mathrm{T}}(x_i)\nabla\mathcal{V}_i^*(\varpi_{i,2})\right) + \iota_{\beta i}.$$

The critic NN is designed as $\mathcal{V}_i^*(\varpi_{i,2}) = \theta_i^{\mathrm{T}}\phi_i(\varpi_{i,2}) + \nu_i(\varpi_{i,2})$. The gradient form is $\nabla\mathcal{V}_i^*(\varpi_{i,2}) = \nabla\phi_i^{\mathrm{T}}(\varpi_{i,2})\theta_i + \nabla\nu_i(\varpi_{i,2})$, where $\nabla\phi_i(\varpi_{i,2}) = \partial\phi_i(\varpi_{i,2})/\partial\varpi_{i,2}$ and $\nabla\nu_i(\varpi_{i,2}) = \partial\nu_i(\varpi_{i,2})/\partial\varpi_{i,2}$. Therefore, the estimated control law $u_i^*(\varpi_{i,2})$ is given by

$$\hat{u}_i^*(\varpi_{i,2}) = -\alpha_i \tanh(\mathcal{K}_2(\varpi_{i,2})) + \iota_{\beta i}, \tag{11}$$

where $\mathcal{K}_2(\varpi_{i,2}) = \frac{\check{d}_i + b_i}{2\alpha_i} g_i^{\mathrm{T}}(x_i)\nabla\phi_i^{\mathrm{T}}(\varpi_{i,2})\hat{\theta}_i$.

The estimated form of the Hamiltonian equation is derived as

$$\begin{aligned}\hat{\mathcal{H}}_{i,2}(\varpi_{i,2}, \hat{\theta}_i) =& \varpi_{i,2}^{\mathrm{T}} Q_{ii}\varpi_{i,2} + \mathcal{R}_i(\hat{u}_{\bar{\mathcal{N}}_i}) + \nabla\phi_i^{\mathrm{T}}(\varpi_{i,1})\hat{\theta}_i((\check{d}_i + b_i)g_i(x_i)\hat{u}_i^* - \check{d}_i g_j(x_j)\hat{u}_j^*) \\ &+ \nabla\phi_i^{\mathrm{T}}(\varpi_{i,2})\hat{\theta}_i(\check{d}_i(f_i(x_i) - f_j(x_j)) + b_i(f_i(x_i) - f_0^s(x_0^s) - u_0^s)),\end{aligned}$$

and $\epsilon_{i,2}^B = \hat{\mathcal{H}}_{i,2}$ is defined as the Bellman residual error.

By repeating the process in Part I, the NN weight updating law is obtained as

$$\dot{\hat{\theta}}_i = -\frac{\mu_{i,2}\check{\hbar}_{i,2}}{\Pi_{i,2}}\epsilon_{i,2}^B, \tag{12}$$

where the learning rate $0 < \mu_{i,2} < 1$, $\Pi_{i,2} = 1 + \hbar_{i,2}^{\mathrm{T}}\hbar_{i,2}$, $\check{\hbar}_{i,2} = \hbar_{i,2}/\Pi_{i,2}$, and $\hbar_{i,2} = \nabla\phi_i^{\mathrm{T}}(\varpi_{i,2})(\check{d}_i(f_i(x_i) - f_j(x_j)) + b_i(f_i(x_i) - f_0^s(x_0^s)) + (\check{d}_i + b_i)g_i(x_i)\hat{u}_i^* - \check{d}_i g_j(x_j)\hat{u}_j^* - b_i u_0^s)$.

Second, the optimal cost function for the attacked agent i is defined as

$$\mathcal{V}_i^*(x_i) = \int_t^\infty (x_i^{\mathrm{T}} Q_{ii}x_i + \mathcal{R}_i(u_i))d\tau, i \in \mathcal{N}^A. \tag{13}$$

Similarly, $\mathcal{R}_i(u_i)$ is defined as $\mathcal{R}_i(u_i) = 2\alpha_i\sum_{c=1}^m \int_{\beta_i}^{u_{i,c}} \iota_i^{-1}\left(\alpha_i^{-1}(k_{i,c} - \beta_i)\right)dk_{i,c}$, where the symbols are as previously defined. The corresponding Hamiltonian function is given by

$$\mathcal{H}_{i,3}(x_i, u_i, \nabla\mathcal{V}_i^*) = x_i^{\mathrm{T}} Q_{ii}x_i + \mathcal{R}_i(u_i) + \nabla\mathcal{V}_i^{*\mathrm{T}}\dot{x}_i = 0. \tag{14}$$

The optimal stabilizing controller u_i^* for the attacked agent is given by

$$u^*(x_i) = -\alpha_i \tanh(\mathcal{K}_1(x_i)) + \chi_i(x_i) + \imath_{\beta i},$$

where $\mathcal{K}_1(x_i) = \frac{1}{2\alpha_i} g_i^{\mathrm{T}}(x_i)\nabla\phi_i^{\mathrm{T}}(x_i)\theta_i$, $\chi_i(x_i) = -\frac{1}{2}(\mathbb{I}_m - D(v(x_i)))g_i^{\mathrm{T}}(x_i)\nabla\nu_i(x_i)$, with $D(v(x_i)) = \mathrm{diag}(\tanh^2(v_c(x_i)))$, for $c = 1, 2, \ldots, m$, and $v(x_i) = [v_1(x_i), v_2(x_i), \ldots, v_m(x_i)]^{\mathrm{T}} \in \mathbb{R}^m$. Here, $v(x_i)$ is selected between $(\frac{1}{2\alpha_i} g_i^{\mathrm{T}}(x_i)\nabla\mathcal{V}_i^*(x_i))$ and $\mathcal{K}_1(x_i)$. Consequently, the estimated form of u_i^* is

$$\hat{u}_i^* = -\alpha_i \tanh(\mathcal{K}_2(x_i)) + \imath_{\beta i}, \tag{15}$$

where $\mathcal{K}_2(x_i) = \frac{1}{2\alpha_i} g_i^{\mathrm{T}}(x_i)\nabla\phi_i^{\mathrm{T}}(x_i)\hat{\theta}_i$.

The Bellman error becomes

$$\epsilon_{i,3}^B = x_i^{\mathrm{T}} Q_{ii} x_i + \mathcal{R}_i(\hat{u}_i) + \nabla\phi_i^{\mathrm{T}}(x_i)\hat{\theta}_i(f_i(x_i) + g_i\hat{u}_i^*).$$

The critic NN weight updating law is

$$\dot{\hat{\theta}}_i = -\frac{\mu_{i,3}\check{\hbar}_{i,3}}{\Pi_{i,3}}\epsilon_{i,3}^B, \tag{16}$$

where the learning rate $0 < \mu_{i,3} < 1$, $\Pi_{i,3} = 1 + \hbar_{i,3}^{\mathrm{T}}\hbar_{i,3}$, $\check{\hbar}_{i,3} = \hbar_{i,3}/\Pi_{i,3}$, and $\hbar_{i,3} = \nabla\phi_i^{\mathrm{T}}(x_i)(f_i(x_i) + g_i\hat{u}_i^*)$.

3.2 Stability analysis

The stability analysis of the proposed control scheme will be given in the theorem.

Theorem 1. Consider the nonlinear MASs consisting of the followers (1) and the supervisor-guided leader (2) under Assumptions 1–4. If the optimal strategies under DoS attacks are chosen as (8), (11) and (15), the weight update laws are designed as (9), (12) and (16), and the dwell time satisfies $\gamma > \ln(\sigma)/\xi$ with $\sigma > 1, \xi > 0$, then the following results can be reached.

(1) All signals in the closed-loop system are bounded.

(2) All agents can perform consensus control during the dormant periods of the attacks.

(3) During the active periods of the attacks, the attacked agents maintain stabilizing control, while the unaffected agents achieve consensus.

Proof. **Part I:** For the dormant periods of the attacks $[t_p, t_p^A)$, construct the following Lyapunov function:

$$\mathcal{L}_i^{\varsigma(t)} = \mathcal{V}_i^*(\varpi_{i,1}) + \frac{1}{2\mu_i}\tilde{\theta}_i^{\mathrm{T}}\tilde{\theta}_i. \tag{17}$$

Taking the derivative of $\mathcal{V}_i^*(\varpi_{i,1})$, it yields

$$\begin{aligned}\dot{\mathcal{V}}_i^*(\varpi_{i,1}) =& \nabla\mathcal{V}_i^{*\mathrm{T}}(d_i(f_i(x_i) - f_j(x_j)) + b_i(f_i(x_i) - f_0^s(x_0^s)) + (d_i + b_i)g_i(x_i)u_i^* - d_i g_j(x_j)u_j^* - b_i u_0^s)\\ &+ \nabla\mathcal{V}_i^{*\mathrm{T}}((d_i + b_i)g_i(x_i)(\hat{u}_i - u_i^*) - d_i g_j(x_j)(\hat{u}_j - u_j^*)).\end{aligned} \tag{18}$$

From (4) and (5), one has

$$\nabla\mathcal{V}_i^{*\mathrm{T}} g_i(x_i) = -\frac{2\alpha_i}{d_i + b_i}\left(\tanh^{-1}\left(\frac{u_i^* - \imath_{\beta i}}{\alpha_i}\right)\right)^{\mathrm{T}}, \tag{19}$$

$$\begin{aligned}-\varpi_{i,1}^{\mathrm{T}} Q_{ii}\varpi_{i,1} - \mathcal{R}_i(u_{\bar{N}_i}^*) =& \nabla\mathcal{V}_i^{*\mathrm{T}}(d_i(f_i(x_i) - f_j(x_j)) + b_i(f_i(x_i) - f_0^s(x_0^s)) + (d_i + b_i)g_i(x_i)u_i^*\\ &- d_i g_j(x_j)u_j^* - b_i u_0^s).\end{aligned} \tag{20}$$

Substituting (19) and (20) into (18), the derivative of $\mathcal{V}_i^*(\varpi_{i,1})$ becomes

$$\dot{\mathcal{V}}_i^*(\varpi_{i,1}) = -\varpi_{i,1}^{\mathrm{T}} Q_{ii}\varpi_{i,1} - \mathcal{R}_i(u_{\bar{N}_i}^*) + \mathcal{S}_{i,1} + \mathcal{S}_{j,1}, \tag{21}$$

where

$$\begin{aligned}\mathcal{S}_{i,1} &= -2\alpha_i\left(\tanh^{-1}\left(\frac{u_i^* - \imath_{\beta i}}{\alpha_i}\right)\right)^{\mathrm{T}}(\hat{u}_i - u_i^*),\\ \mathcal{S}_{j,1} &= \frac{2d_i\alpha_i g_j}{(d_i + b_i)g_i}\left(\tanh^{-1}\left(\frac{u_i^* - \imath_{\beta i}}{\alpha_i}\right)\right)^{\mathrm{T}}(\hat{u}_j - u_j^*).\end{aligned}$$

Using Young's inequality, the following results hold by employing (6) and (8):

$$\mathcal{S}_{i,1} \leqslant \alpha_i^2 \sum_{c=1}^{m} \left(\tanh^{-1} \left(\frac{u_{i,c}^* - \beta_i}{\alpha_i} \right) \right)^2 + \underbrace{\left\| \alpha_i \left(\tanh(\mathcal{K}_1(\varpi_{i,1})) - \tanh(\mathcal{K}_2(\varpi_{i,1})) \right) - \chi_i(\varpi_{i,1}) \right\|^2}_{\mathcal{S}_{i,2}},$$

$$\mathcal{S}_{j,1} \leqslant \mathcal{T}_{ij}^2 \sum_{c=1}^{m} \left(\tanh^{-1} \left(\frac{u_{i,c}^* - \beta_i}{\alpha_i} \right) \right)^2 + \underbrace{\left\| \mathcal{T}_{ij} \left(\tanh(\mathcal{K}_1(\varpi_{j,1})) - \tanh(\mathcal{K}_2(\varpi_{j,1})) \right) - \chi_j(\varpi_{j,1}) \right\|^2}_{\mathcal{S}_{j,2}},$$

where $\mathcal{T}_{ij} = \frac{d_i \alpha_i g_{j,\max}}{(d_i + b_i) g_{i,\min}}$.

Thus, following [30] and using the facts that

$$\|\tanh(\mathcal{K}_w(\varpi_{\varphi,1}))\| \leqslant \sqrt{m}, (w = 1, 2, \varphi = i, j), \left\| \mathbb{I}_m - \operatorname{diag}(\tanh^2(v_c(\varpi_{i,1}))) \right\|_{c=1}^{m} \leqslant 2,$$

it follows that

$$\begin{aligned}
\mathcal{S}_{i,2} \leqslant & 2\alpha_i^2 \left\| \tanh(\mathcal{K}_1(\varpi_{i,1})) - \tanh(\mathcal{K}_2(\varpi_{i,1})) \right\|^2 + 2\|\chi_i(\varpi_{i,1})\|^2 \\
\leqslant & 4\alpha_i^2 \left(\|\tanh(\mathcal{K}_1(\varpi_{i,1}))\|^2 + \|\tanh(\mathcal{K}_2(\varpi_{i,1}))\|^2 \right) + 2 \left\| \frac{d_i + b_i}{2} \left(\mathbb{I}_m - D(v(\varpi_{i,1})) \right) g_i^{\mathrm{T}}(x_i) \nabla \nu_i(\varpi_{i,1}) \right\|^2 \\
\leqslant & 8\alpha_i^2 m + 2(d_i + b_i)^2 g_{i,\max}^2 \nu_{i,\max}^2, \\
\mathcal{S}_{j,2} \leqslant & 8\mathcal{T}_{ij}^2 m + 2(d_j + b_j)^2 g_{j,\max}^2 \nu_{j,\max}^2.
\end{aligned} \tag{22}$$

Similarly, referring to [21], $\mathcal{R}_i(u_{\bar{N}_i}^*)$ has the following results:

$$\begin{aligned}
-\mathcal{R}_i(u_{\bar{N}_i}^*) \leqslant & \frac{1}{2} g_{i,\max}^2 \mathcal{V}_{i,\max}^2 + \frac{1}{2} g_{j,\max}^2 \mathcal{V}_{j,\max}^2 - \alpha_i^2 \sum_{c=1}^{m} \left(\tanh^{-1} \left(\frac{u_{i,c}^* - \beta_i}{\alpha_i} \right) \right)^2 \\
& - \sum_{j \in \mathcal{N}_i} \alpha_j^2 \sum_{c=1}^{m} \left(\tanh^{-1} \left(\frac{u_{j,c}^* - \beta_j}{\alpha_j} \right) \right)^2,
\end{aligned} \tag{23}$$

where $\mathcal{V}_{i,\max}$ is the upper bound of $\mathcal{V}_i^*$, i.e., $||\mathcal{V}_i^*|| \leqslant \mathcal{V}_{i,\max}$. Then, from (9) and using $\hat{\theta}_i = \theta_i - \tilde{\theta}_i$, it can be deduced that $\dot{\tilde{\theta}}_i = -\mu_{i,1} \check{h}_{i,1} \check{h}_{i,1}^{\mathrm{T}} \tilde{\theta}_i + \frac{\mu_{i,1} \check{h}_{i,1}}{\Pi_{i,1}} \mathcal{W}_{i,1}$, where $\mathcal{W}_{i,1} = -\nabla \nu_i^{\mathrm{T}}(\varpi_{i,1})(d_i(f_i(x_i) - f_j(x_j)) + b_i(f_i(x_i) - f_0^s(x_0^s)) + (d_i + b_i) g_i(x_i) \hat{u}_i^* - d_i g_j(x_j) \hat{u}_j^* - b_i u_0^s)$. Then, it follows that

$$\frac{d(\frac{1}{2\mu_i} \tilde{\theta}_i^{\mathrm{T}} \tilde{\theta}_i)}{dt} = -\tilde{\theta}_i^{\mathrm{T}} \check{h}_{i,1} \check{h}_{i,1}^{\mathrm{T}} \tilde{\theta}_i + \frac{\check{h}_{i,1}}{\Pi_{i,1}} \tilde{\theta}_i^{\mathrm{T}} \mathcal{W}_{i,1} \tag{24}$$

with $\frac{1}{\Pi_{i,1}} \leqslant 1$ and $\frac{\check{h}_{i,1}}{\Pi_{i,1}} \tilde{\theta}_i^{\mathrm{T}} \mathcal{W}_{i,1} \leqslant \frac{1}{2} \tilde{\theta}_i^{\mathrm{T}} \check{h}_{i,1} \check{h}_{i,1}^{\mathrm{T}} \tilde{\theta}_i + \frac{1}{2} \mathcal{W}_{i,1}^{\mathrm{T}} \mathcal{W}_{i,1}$. As a result, we obtain

$$\frac{d(\frac{1}{2\mu_i} \tilde{\theta}_i^{\mathrm{T}} \tilde{\theta}_i)}{dt} \leqslant -\frac{1}{2} \lambda_{\min}(\check{h}_{i,1} \check{h}_{i,1}^{\mathrm{T}}) \left\| \tilde{\theta}_i \right\|^2 + \frac{1}{2} \left\| \mathcal{W}_{i,1} \right\|^2. \tag{25}$$

Invoking (21)–(23) and (25), it yields

$$\dot{\mathcal{L}}_i^{\varsigma(t)} \leqslant -\lambda_{\min}(Q_{ii}) \left\| \varpi_{i,1} \right\|^2 - \frac{1}{2} \lambda_{\min}(\check{h}_{i,1} \check{h}_{i,1}^{\mathrm{T}}) \left\| \tilde{\theta}_i \right\|^2 + \upsilon_{i,1}, \tag{26}$$

where $\upsilon_{i,1} = 8\alpha_i^2 m + 2(d_i + b_i)^2 g_{i,\max}^2 \nu_{i,\max}^2 + 8\mathcal{T}_{ij}^2 m + 2(d_j + b_j)^2 g_{j,\max}^2 \nu_{j,\max}^2 + (\frac{1}{2} + \frac{\mathcal{T}_{ij}^2}{4\alpha_i^2}) g_{i,\max}^2 \mathcal{V}_{i,\max}^2 + \frac{1}{4} g_{j,\max}^2 \mathcal{V}_{j,\max}^2 + \frac{1}{2} \left\| \mathcal{W}_{i,1} \right\|^2$.

Based on the above analysis, one has

$$\dot{\mathcal{L}}_i^{\varsigma(t)} \leqslant -\xi_{i,1} \mathcal{L}_i^{\varsigma(t)} + \upsilon_{i,1}, \tag{27}$$

where $\xi_{i,1} = \min\{2\lambda_{\min}(Q_{ii}), \lambda_{\min}(\check{h}_{i,1} \check{h}_{i,1}^{\mathrm{T}})\}$.

Part II: For the active periods of the attacks $[t_p^A, t_{p+1})$, we first analyze the stability of the attacked agents. Similarly to the above process, develop the following Lyapunov function:

$$\mathcal{L}_i^{\varsigma(t)} = \mathcal{V}_i^*(x_i) + \frac{1}{2\mu_i}\tilde{\theta}_i^{\mathrm{T}}\tilde{\theta}_i. \tag{28}$$

Taking the derivative of $\mathcal{V}_i^*(x_i)$, one obtains

$$\dot{\mathcal{V}}_i^* = \nabla\mathcal{V}_i^{*\mathrm{T}}(f_i(x_i) + g_i(x_i)u_i^*) + \nabla\mathcal{V}_i^{*\mathrm{T}}g_i(x_i)(\hat{u}_i - u_i^*).$$

Calculating $\mathcal{R}_i(u_i^*) \leqslant \frac{1}{2}g_{i,\max}^2\mathcal{V}_{i,\max}^2$, we have

$$\dot{\mathcal{V}}_i^*(x_i) \leqslant -\lambda_{\min}(Q_{ii})\left\|x_i\right\|^2 + 8\alpha_i^2 m + 2g_{i,\max}^2\nu_{i,\max}^2 + \frac{1}{2}g_{i,\max}^2\mathcal{V}_{i,\max}^2. \tag{29}$$

According to (16), the dynamics of the weight estimation error is $\dot{\tilde{\theta}}_i = -\mu_{i,2}\check{h}_{i,2}\check{h}_{i,2}^{\mathrm{T}}\tilde{\theta}_i + \frac{\mu_{i,2}\check{h}_{i,2}}{\Pi_{i,2}}\mathcal{W}_{i,2}$, where $\mathcal{W}_{i,2} = -\nabla\nu_i^{\mathrm{T}}(x_i)(f_i(x_i) + g_i(x_i)\hat{u}_i^*)$. Then, we have

$$\dot{\mathcal{L}}_i^{\varsigma(t)} \leqslant -\lambda_{\min}(Q_{ii})\left\|x_i\right\|^2 - \frac{1}{2}\lambda_{\min}(\check{h}_{i,2}\check{h}_{i,2}^{\mathrm{T}})\left\|\tilde{\theta}_i\right\|^2 + \upsilon_{i,2} \leqslant -\xi_{i,2}\mathcal{L}_i^{\varsigma(t)} + \upsilon_{i,2}, \tag{30}$$

where $\xi_{i,2} = \min\{2\lambda_{\min}(Q_{ii}), \lambda_{\min}(\check{h}_{i,2}\check{h}_{i,2}^{\mathrm{T}})\}$ and $\upsilon_{i,2} = 8\alpha_i^2 m + 2g_{i,\max}^2\nu_{i,\max}^2 + \frac{1}{2}g_{i,\max}^2\mathcal{V}_{i,\max}^2 + \frac{1}{2}\left\|\mathcal{W}_{i,2}\right\|^2$.

Furthermore, the stability of the unaffected agents is investigated. By following a similar procedure, the Lyapunov function is constructed as

$$\mathcal{L}_i^{\varsigma(t)} = \mathcal{V}_i^*(\varpi_{i,2}) + \frac{1}{2\mu_i}\tilde{\theta}_i^{\mathrm{T}}\tilde{\theta}_i.$$

Repeating the above process, we can obtain the following result:

$$\dot{\mathcal{L}}_i^{\varsigma(t)} \leqslant -\xi_{i,3}\mathcal{L}_i^{\varsigma(t)} + \upsilon_{i,3}, \tag{31}$$

where $\xi_{i,3} = \min\{2\lambda_{\min}(Q_{ii}), \lambda_{\min}(\check{h}_{i,3}\check{h}_{i,3}^{\mathrm{T}})\}$, $\upsilon_{i,3} = 8\alpha_i^2 m + 2(d_i + b_i)^2 g_{i,\max}^2\nu_{i,\max}^2 + 8\mathcal{T}_{ij}^2 m + 2(d_j + b_j)^2 g_{j,\max}^2\nu_{j,\max}^2 + (\frac{1}{2} + \frac{\mathcal{T}_{ij}^2}{4\alpha_i^2})g_{i,\max}^2\mathcal{V}_{i,\max}^2 + \frac{1}{4}g_{j,\max}^2\mathcal{V}_{j,\max}^2 + \frac{1}{2}\left\|\mathcal{W}_{i,3}\right\|^2$ and $\mathcal{W}_{i,3} = -\nabla\nu_i^{\mathrm{T}}(\varpi_{i,2})(\breve{d}_i(f_i(x_i) - f_j(x_j)) + b_i(f_i(x_i) - f_0^s(x_0^s)) + (\breve{d}_i + b_i)g_i(x_i)\hat{u}_i^* - \breve{d}_i g_j(x_j)\hat{u}_j^* - b_i u_0^s)$.

Integrating (27), (30) and (31), the following result can be derived:

$$\dot{\mathcal{L}}_i^{\varsigma(t)} \leqslant -\xi_i\mathcal{L}_i^{\varsigma(t)} + \upsilon_i, \tag{32}$$

where $\xi_i = \min\{\xi_{i,1}, \xi_{i,2}, \xi_{i,3}\}$ and $\upsilon_i = \max\{\upsilon_{i,1}, \upsilon_{i,2}, \upsilon_{i,3}\}$. Then, we have $\dot{\mathcal{L}}^{\varsigma(t)} \leqslant -\xi\mathcal{L}^{\varsigma(t)} + \upsilon$, in which $\mathcal{L}^{\varsigma(t)} = \sum_{i=1}^{N}\mathcal{L}_i^{\varsigma(t)}, \xi = \min\{\xi_1, \xi_2, \cdots, \xi_N\}, \upsilon = \sum_{i=1}^{N}\upsilon_i$.

To facilitate stability analysis, we introduce the following coordinate transformation from [12]:

$$\mathcal{P}(t) = e^{\xi t}\mathcal{L}^{\varsigma(t)}(t). \tag{33}$$

From (32) and (33), for each interval $[t_p, t_p^A)$ and $[t_p^A, t_{p+1})$, we have

$$\dot{\mathcal{P}}(t) = \xi e^{\xi t}\mathcal{L}^{\varsigma(t)}(t) + e^{\xi t}\dot{\mathcal{L}}^{\varsigma(t)}(t) \leqslant \upsilon e^{\xi t}.$$

Subsequently, considering the dormant period $[t_p, t_p^A)$, it follows that

$$\mathcal{P}(t_p^A) = e^{\xi t_p^A}\mathcal{L}^{\varsigma(t_p^A)}(t_p^A) \leqslant \sigma e^{\xi t_p^A}\mathcal{L}^{\varsigma(t_p)}(t_p^A) \leqslant \sigma\left(\mathcal{P}(t_p) + \int_{t_p}^{t_p^A}\upsilon e^{\xi t}dt\right). \tag{34}$$

Applying (34) iteratively from $\kappa = 0$ to $\kappa = N_\varsigma(0,T) - 1$, the result is

$$\mathcal{P}(T^-) \leqslant \sigma^{N_\varsigma(0,T)}\mathcal{P}(0) + \int_{t^{N_\varsigma(0,T)}}^{T}\upsilon e^{\xi t}dt + \sum_{\kappa=0}^{N_\varsigma(0,T)-1}\sigma^{N_\varsigma(0,T)-\kappa} + \int_{t_p}^{t_p^A}\upsilon e^{\xi t}dt. \tag{35}$$

Given $\gamma > \ln(\sigma)/\xi$, there exists a constant η satisfying $0 < \eta < \xi - \ln(\sigma)/\gamma$. One has $N_\varsigma(0,T) \leqslant N_0 + \frac{(\xi-\eta)T}{\ln(\sigma)}$. It holds that $N_\varsigma(0,T) - \kappa \leqslant 1 + N_\varsigma(t_p^A, T), \kappa = 0, \ldots, N_\varsigma(0,T)$ such that

$$\sigma^{N_\varsigma(0,T)-\kappa} \leqslant \sigma^{1+N_0+\frac{(\xi-\eta)(T-t_p^A)}{\ln(\sigma)}} = \sigma^{(1+N_0)} e^{(\xi-\eta)(T-t_p^A)}.$$

Since $\eta < \xi$, one obtains $\int_{t_p}^{t_p^A} \upsilon e^{\xi t} dt \leqslant e^{(\xi-\eta)t_p^A} \int_{t_p}^{t_p^A} \upsilon e^{\eta t} dt$. From (33), one has

$$\mathcal{P}(T^-) \leqslant \sigma^{(1+N_0)} e^{(\xi-\eta)T} \int_0^T \upsilon e^{\eta t} dt + \sigma^{N_\varsigma(0,T)} \mathcal{P}(0). \tag{36}$$

It follows from (33) and (36), one obtains

$$\mathcal{L}^{\varsigma(T^-)}(T^-) \leqslant e^{-\xi t}(\sigma^{(1+N_0)} e^{(\xi-\eta)T} \int_0^T \upsilon e^{\eta t} dt + \sigma^{N_\varsigma(0,T)} \mathcal{P}(0)) \leqslant \sigma^{N_0} e^{-\eta T} \mathcal{L}^{\varsigma(t)}(0) + \frac{\upsilon(1-e^{-\eta T})}{\eta} \sigma^{(N_0+1)}.$$

Then, we obtain

$$\lim_{t\to\infty} \mathcal{L}^{\varsigma(t)}(t) \leqslant \sigma^{(N_0+1)} \frac{\upsilon}{\eta} = \bar{\mathbb{G}}.$$

Finally, referring to the analysis in [12] and choosing suitable parameters, we can show that all signals in the MASs remain bounded, and the MASs achieve stability.

This completes the proof.

Remark 5. For the dormant periods $[t_p, t_p^A)$, based on the inequality (27), one has $|\varpi_{i,1}| \leqslant [2(\frac{\upsilon_{i,1}}{\xi_{i,1}} + \mathcal{L}_i^{\varsigma(t)}(0))]^{\frac{1}{2}}$. During the active periods $[t_p^A, t_{p+1})$, invoking (30) and (31), it is obtained that $|x_i| \leqslant [2(\frac{\upsilon_{i,2}}{\xi_{i,2}} + \mathcal{L}_i^{\varsigma(t)}(0))]^{\frac{1}{2}}$ and $|\varpi_{i,2}| \leqslant [2(\frac{\upsilon_{i,3}}{\xi_{i,3}} + \mathcal{L}_i^{\varsigma(t)}(0))]^{\frac{1}{2}}$, respectively. In general, the variables $\varpi_{i,1}, \varpi_{i,2}$ and x_i converge to the set $[2(\frac{\upsilon}{\xi} + \mathcal{L}^{\varsigma(t)}(0))]^{\frac{1}{2}}$. This implies that the variables can be decreased by increasing ξ or decreasing υ. In addition, for the cost function, the value of Q_{ii} can reflect the speed of convergence of the states, while parameters α_i and β_i characterize the asymmetric input constraints. Choosing the appropriate Q_{ii}, α_i, and β_i can achieve a better control effect.

4 Simulation

4.1 Application example

A nonlinear MAS consisting of five single-link robot arms is provided to verify the effectiveness of the proposed control scheme. The agent model is adopted from [6], where the physical parameters g, M_i, D_i, J_i and d_i can be found. We select $J_i = D_i = 1, M_i g d_i = 10$ in this paper. Simulations are conducted in MATLAB R2025a to validate the efficacy of the proposed strategy. The initial states of the followers and the leader are given as $x_{i,1}(0) = x_{i,2}(0) = x_{0,1}^s(0) = x_{0,2}^s(0) = -2, i = 1, \cdots, 4$. For the optimal controller, choose $Q_{ii} = \mathbb{I}_2$, $u_{i,\min} = -4, u_{i,\max} = 2$ for the cost functions. Select the initial critic NN weight vectors $\theta_i^{(0)} = [0,0,0,0,0]^{\mathrm{T}}$. The activation functions are given as $\phi_i(\cdot) = [\tanh(\cdot(1))^2; \tanh(\cdot(1)\cdot(2)); \tanh(\cdot(2))^2; \tanh(\cdot(1)^2\cdot(2)); \tanh(\cdot(2)^2\cdot(1))]$, where $\cdot = \varpi_{i,1}, x_i, \varpi_{i,2}$. The learning rates are chosen as $\mu_{i,1} = \mu_{i,2} = \mu_{i,3} = 0.01$. The leader's input u_0^s is defined as

$$u_0^s = \begin{cases} \cos\left(\dfrac{\pi}{5}t\right), & \text{if } 1\,\text{s} \leqslant t < 10\,\text{s}, \\ \sin\left(\dfrac{\pi(t-20)}{5}\right), & \text{if } 20\,\text{s} \leqslant t < 35\,\text{s}, \\ 0, & \text{otherwise.} \end{cases}$$

The switching signal $\varsigma(t)$, which represents the different topology modes, satisfies the following condition:

$$\varsigma(t) = \begin{cases} 0, & t \in [0\,\text{s}, 10\,\text{s}) \cup [15\,\text{s}, 35\,\text{s}) \cup [45\,\text{s}, 50\,\text{s}], \\ 1, & t \in [10\,\text{s}, 15\,\text{s}), \\ 2, & t \in [35\,\text{s}, 45\,\text{s}). \end{cases} \tag{37}$$

As illustrated in (37), the attack occurs at 10 s, 35 s and disappears at 15 s, 45 s. Specifically, (1) $\varsigma(t) = 0$ during 0–10 s, 15–35 s and 45–50 s, representing the topology during dormant periods, which corresponds to $\tilde{\mathcal{G}}_0$ in Figure

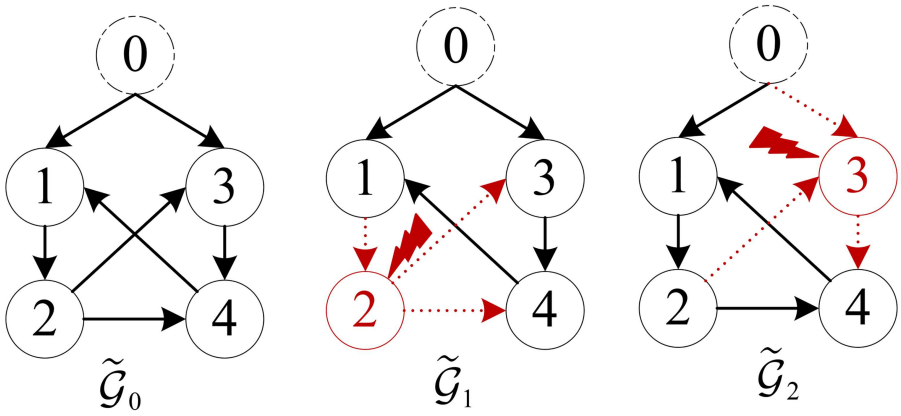

Figure 2 (Color online) Communication graph under DoS attacks.

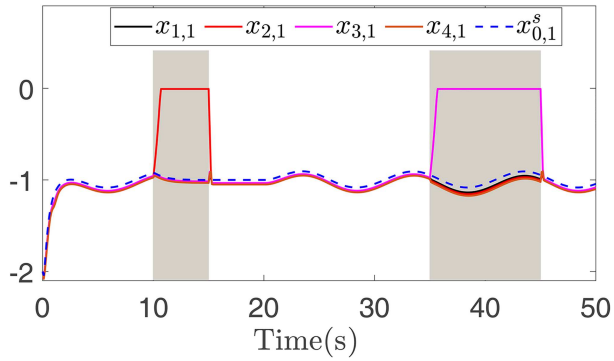

Figure 3 (Color online) Curves of the states $x_{i,1}$ and $x^s_{0,1}$ under DoS attacks.

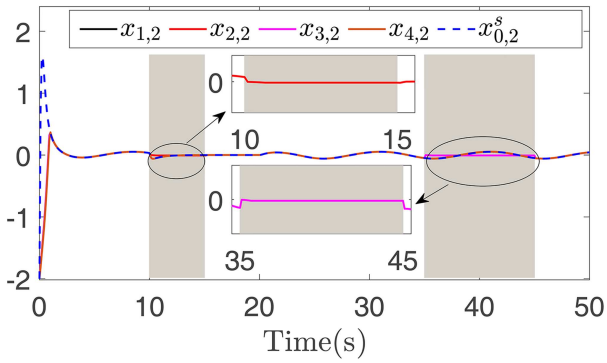

Figure 4 (Color online) Curves of the states $x_{i,2}$ and $x^s_{0,2}$ under DoS attacks.

2; (2) $\varsigma(t) = 1$ during 10–15 s, representing the topological configuration $\tilde{\mathcal{G}}_1$ in Figure 2; (3) $\varsigma(t) = 2$ during 5–45 s, corresponding to the topological relationship $\tilde{\mathcal{G}}_2$ in Figure 2.

Figures 3 and 4 display the state trajectories of the supervisor-guided leader and the followers. Corresponding to the attacked nodes designated in Figure 2, agent 2 maintains stabilizing control during the interval of 10 s to 15 s, while agent 3 performs stabilizing control between 35 and 45 s, as indicated by the gray shaded regions. During these attack periods, the remaining unaffected agents continue to achieve consensus control. Once the DoS attacks cease and the communication topology is restored, all agents recover their consensus performance. Figure 5 illustrates the convergence process of the consensus errors. In particular, transient spikes appear in the error curves at the switching moments of the attack modes, which are also reflected in the control input curves in Figure 6. These spikes result from abrupt changes in topological connectivity that lead to discontinuous variations in consensus errors. For example, when agent 2 is subject to the DoS attack from 10 to 15 s, its communication with neighbors is interrupted. By applying a stabilizing control strategy, the control law becomes solely dependent on its local states, causing a sudden shift in the error dynamics. Figure 6 presents the curves of the constrained optimal control input, which remain strictly within the asymmetric limits $[-4, 2]$. The boundedness of the critic NN weights is confirmed in Figure 7. Finally, the convergence curves of the cost functions during the training process are shown in Figure 8, showing that they converge to the optimal values.

4.2 Comparison experiment

Motivated by practical engineering requirements, this work considers asymmetric input constraints, which are addressed through a nonquadratic cost function. This approach offers a higher degree of generality compared to the symmetric constraints investigated in [22]. To provide a baseline for comparison with [22], the optimal control input under symmetric constraints during dormant periods is formulated as

$$u_i^*(\varpi_{i,1}) = -\alpha_i \tanh\left(\frac{d_i + b_i}{2\alpha_i} g_i^{\mathrm{T}}(x_i) \nabla \mathcal{V}_i^*(\varpi_{i,1})\right). \tag{38}$$

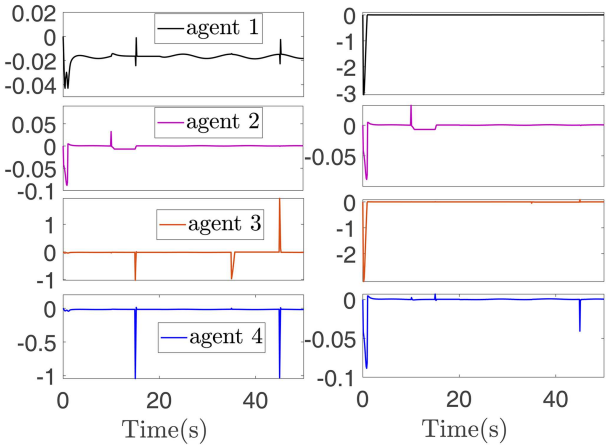

Figure 5 (Color online) Curves of consensus error under DoS attacks.

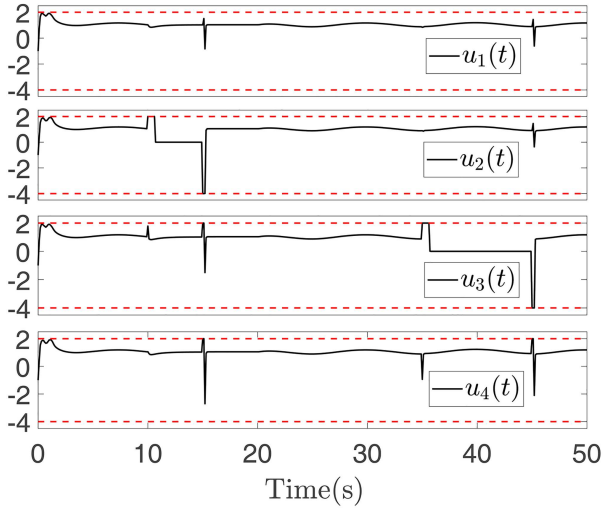

Figure 6 (Color online) Curves of constrained optimal control input u_i.

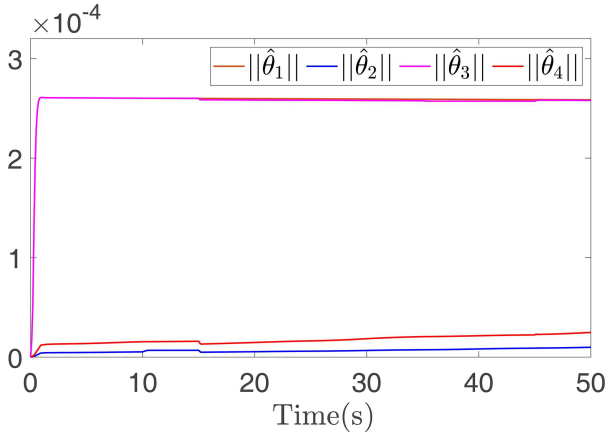

Figure 7 (Color online) Convergence curves of the norm of NN weights $\hat{\theta}_i(t)$.

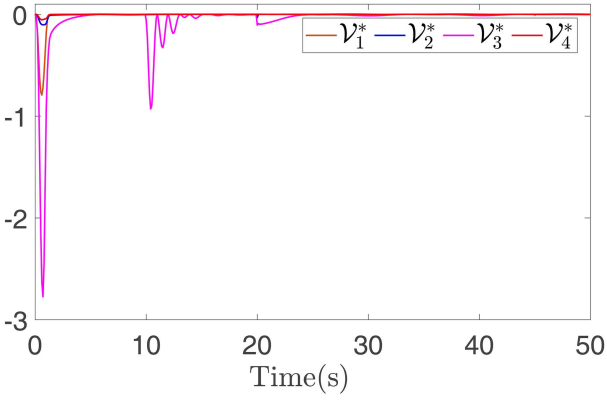

Figure 8 (Color online) Convergence curves of cost functions $\mathcal{V}_i^*$.

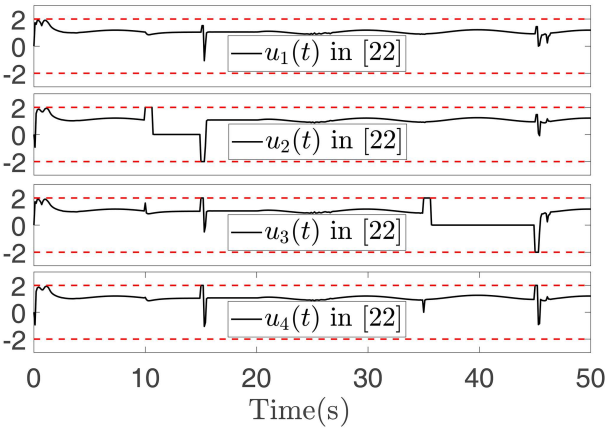

Figure 9 (Color online) Curves of symmetric constrained optimal control input u_i [22].

According to (38), the control input is constrained within the symmetric range $[-\alpha_i, \alpha_i]$. Setting $u_{i,\min} = -2, u_{i,\max} = 2, \alpha_i = 2$ and conducting the simulation under conditions as previously described, the result is

illustrated in Figure 9. A comparison with the symmetric constrained optimal control in [22] demonstrates that setting $|u_{i,\min}| = |u_{i,\max}|$, the proposed u_i^* recovers the symmetric form given in (38). Furthermore, asymmetric constraints characterize practical engineering systems more accurately, such as aero-engines [33] and mobile robots [34], where actuator limitations are inherently non-uniform.

5 Conclusion

The resilient HiTL optimal cooperative control problem for constrained-input MASs under DoS attacks has been addressed in this paper. First, the leader has been designed to be non-autonomous. Then, a nonquadratic cost function has been constructed to tackle the asymmetric constraints, and an HJB equation has been defined to obtain the optimal control strategy. Afterward, the solution to the HJB equation was learned using the critic NN. Using Lyapunov stability theory, it has been shown that all signals in closed-loop systems remain bounded. Finally, the effectiveness and superiority of the proposed control scheme have been demonstrated through simulation. In future work, we will extend the results to MASs with hybrid cyber attacks.

Acknowledgements This work was supported in part by National Natural Science Foundation of China (Grant Nos. 62533006, 52471376, 62273072, 62322307), Fundamental Research Funds for the Central Universities (Grant No. ZYGX2024Z018), and Sichuan Science and Technology Program (Grant No. 2024JDHJ0037).

References

1 Okubo A. Dynamical aspects of animal grouping: swarms, schools, flocks, and herds. Adv Biophys, 1986, 22: 1–94
2 Ren W. Formation keeping and attitude alignment for multiple spacecraft through local interactions. J Guidance Control Dyn, 2007, 30: 633–638
3 Olfati-Saber R. Flocking for multi-agent dynamic systems: algorithms and theory. IEEE Trans Automat Contr, 2006, 51: 401–420
4 Kiumarsi B, Basar T. Human-in-the-loop control of distributed multi-agent systems: a relative input-output approach. In: Proceedings of 2018 IEEE Conference on Decision and Control (CDC), 2018. 3343–3348
5 Ma L, Zhu F. Human-in-the-loop formation control for multi-agent systems with asynchronous edge-based event-triggered communications. Automatica, 2024, 167: 111744
6 Huang Z, Li T, Long Y, et al. Observer-based human-in-the-loop optimal output cluster synchronization control for multiagent systems: a model-free reinforcement learning method. IEEE Trans Cybern, 2025, 55: 649–660
7 Lin G, Ren H, Zhou Q, et al. Fuzzy dynamic event-triggered containment control for human-in-the-loop MASs with error constraints. IEEE Trans Fuzzy Syst, 2024, 32: 2496–2508
8 Liu P M, Guo X G, Wang J L, et al. Preset-time and preset-accuracy human-in-the-loop cluster consensus control for MASs under stochastic actuation attacks. IEEE Trans Automat Contr, 2024, 69: 1675–1688
9 Gong X, Gui J, Chen Y, et al. Resilient human-in-the-loop formation-tracking of multi-UAV systems against Byzantine attacks. IEEE Trans Automat Sci Eng, 2025, 22: 3797–3809
10 Lv M, Gu N, Wang D, et al. Human-in-the-loop coordinated path following of marine vehicles based on continuous twisting control. IEEE Trans Ind Inf, 2025, 21: 465–474
11 Werbos P J. Reinforcement learning and approximate dynamic programming-foundations, common misconceptions, and the challenges ahead. In: Reinforcement Learning and Approximate Dynamic Programming for Feedback Control. Hoboken: John Wiley & Sons, Inc, 2012. 1–30
12 Wu Y, Chen M, Li H, et al. Mixed-zero-sum-game-based memory event-triggered cooperative control of heterogeneous MASs against DoS attacks. IEEE Trans Cybern, 2024, 54: 5733–5745
13 Werbos P. Advanced forecasting methods for global crisis warning and models of intelligence. In: General System Yearbook. Louisville: Society for General Systems Research, 1977. 25–38
14 Zhang C J, Ji L H, Yang S S, et al. Distributed optimal consensus control for multiagent systems based on event-triggered and prioritized experience replay strategies. Sci China Inf Sci, 2025, 68: 112206
15 Xu W K, Wang L, Sun S W, et al. A novel policy iteration algorithm for solving the optimal consensus control problem of a discrete-time multiagent system with unknown dynamics. Sci China Inf Sci, 2023, 66: 189204
16 Padhi R, Unnikrishnan N, Wang X, et al. A single network adaptive critic (SNAC) architecture for optimal control synthesis for a class of nonlinear systems. Neural Netws, 2006, 19: 1648–1660
17 Yang F, Gong Z, Wei Q, et al. Secure containment control for multi-UAV systems by fixed-time convergent reinforcement learning. IEEE Trans Cybern, 2025, 55: 1981–1994
18 Wei Q, Jiang H. Event-/self-triggered adaptive optimal consensus control for nonlinear multiagent system with unknown dynamics and disturbances. IEEE Trans Cybern, 2025, 55: 1476–1485
19 Gong Z, Yang F, Yuan Y, et al. Secure formation control of multiagent system against FDI attack using fixed-time convergent reinforcement learning. IEEE Trans Control Netw Syst, 2025, 12: 1203–1214
20 Abu-Khalaf M, Lewis F L. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. Automatica, 2005, 41: 779–791
21 Liu D, Yang X, Wang D, et al. Reinforcement-learning-based robust controller design for continuous-time uncertain nonlinear systems subject to input constraints. IEEE Trans Cybern, 2015, 45: 1372–1385
22 Xue S, Luo B, Liu D, et al. Event-triggered ADP for tracking control of partially unknown constrained uncertain systems. IEEE Trans Cybern, 2022, 52: 9001–9012
23 Yang X, Wei Q. Adaptive dynamic programming for robust event-driven tracking control of nonlinear systems with asymmetric input constraints. IEEE Trans Cybern, 2024, 54: 6333–6344
24 He W, Xu W, Ge X, et al. Secure control of multiagent systems against malicious attacks: a brief survey. IEEE Trans Ind Inf, 2022, 18: 3595–3608
25 Zhang Y, Wu Z G, Shi P. Resilient event-/self-triggering leader-following consensus control of multiagent systems against DoS attacks. IEEE Trans Ind Inf, 2023, 19: 5925–5934
26 Tan W, Hou Z, Li Y X. Data-driven containment control for unknown MIMO nonlinear MASs under aperiodic DoS attacks. IEEE Trans Automat Sci Eng, 2025, 22: 7762–7772
27 Su W, Mu C X, Zhu S, et al. Event-triggered leader-follower bipartite consensus control for nonlinear multi-agent systems under DoS attacks. Sci China Inf Sci, 2025, 68: 132206

28 Zhao R, Zuo Z, Wang Y. Event-triggered control for switched systems with denial-of-service attack. IEEE Trans Automat Contr, 2022, 67: 4077–4090
29 Sun Q Y, Wang B Y, Feng X M, et al. Small-signal stability and robustness analysis for microgrids under time-constrained DoS attacks and a mitigation adaptive secondary control method. Sci China Inf Sci, 2022, 65: 162202
30 Yang X, Zhao B. Optimal neuro-control strategy for nonlinear systems with asymmetric input constraints. IEEE CAA J Autom Sin, 2020, 7: 575–583
31 Sun J, Long T. Event-triggered distributed zero-sum differential game for nonlinear multi-agent systems using adaptive dynamic programming. ISA Trans, 2021, 110: 39–52
32 Liu T, Wang S, Huang J. An adaptive distributed observer for a class of uncertain linear leader systems over jointly connected switching networks and its application. IEEE Trans Automat Contr, 2024, 69: 7340–7355
33 Wang K, Wu D, Li P, et al. Output-constrained switching anti-windup compensation for aero-engines with asymmetric input saturation. IEEE Control Syst Lett, 2024, 8: 3075–3080
34 Tian L, Wang X, Chen H, et al. Time-varying formation tracking for nonholonomic mobile robots with asymmetric input constraints. IEEE Trans Syst Man Cybern Syst, 2025, 55: 8195–8209