Reinforcement Learning-Based Formation Control for Uncrewed Surface Vehicles Under Aperiodic DoS Attacks: A Stackelberg–Nash Game Approach

Jinliang Liu, *Senior Member, IEEE*, Zihan Zhang, Engang Tian, *Senior Member, IEEE*, Chen Peng, *Senior Member, IEEE*, Jinde Cao, *Fellow, IEEE*, and Tingwen Huang, *Fellow, IEEE*

***Abstract*—This article investigates the distributed formation control of uncrewed surface vehicles (USVs) under aperiodic denial-of-service (DoS) attacks within a Stackelberg–Nash game (SNG) framework. An actor–critic (AC) reinforcement learning (RL) algorithm is developed to approximate these policies online, ensuring convergence to the Stackelberg–Nash equilibrium (SNE). To enhance resilience against communication interruptions, a consensus-based estimator is designed to reconstruct missing neighbor data using local information. Rigorous Lyapunov-based analysis guarantees the input-to-state stability (ISS) of the estimator and the semi-globally uniformly ultimately bounded (SGUUB) stability of the closed-loop system. Simulation results verify the framework's effectiveness in achieving accurate trajectory tracking and robustness against frequent DoS attacks.**

***Index Terms*—Actor–critic (AC) learning, denial-of-service (DoS) attacks, formation control, Stackelberg–Nash game (SNG), uncrewed surface vehicles (USVs).**

I. Introduction

UNCREWED surface vehicles (USVs) have gained increasing attention due to their wide range of applications in maritime operations in recent years [1]. Their autonomous capabilities make them highly suitable for missions that are dangerous, repetitive, or long-endurance missions [2], [3]. In particular, USV formations have shown great potential to outperform single-agent systems by enhancing robustness, scalability, and efficiency [4], [5]. Despite these advantages, the implementation of multi-USV formation control systems remains challenging [6], [7].

Game-theoretic approaches have been widely studied for their capability to model competitive and cooperative interactions among multiple agents [8], [9]. Most existing studies assume that agents play symmetric roles, where each agent independently and simultaneously makes decisions to optimize its own objective [10]. However, in many practical scenarios, certain agents may possess distinct advantages that allow them to act first or strategically influence others, while the remaining agents simultaneously seek their own optimal responses [11], [12]. Originating from an economic model involving two participants: a leader and a follower, the Stackelberg framework assumes that the leader selects an optimal action while anticipating the follower's potential response to maximize its own utility. When multiple followers simultaneously make their decisions in response to the leader's action, the resulting framework constitutes a Stackelberg–Nash game (SNG) [13], [14]. In [15], an event-triggered robust adaptive dynamic programming (ADP) algorithm is developed to solve a class of multiplayer SNGs for uncertain nonlinear continuous-time systems. Zhang et al. [16] proposed an ADP-based hierarchical Stackelberg–Nash optimal game control method. Recent studies have further extended SNG to stochastic systems [15] and finite-time control [17], [18]. However, these works primarily rely on ideal communication channels, leaving the system vulnerable to cyber threats. A game-based backstepping design using reinforcement learning (RL) was developed in [19], while a finite-time adaptive fuzzy strategy was proposed in [20] to handle stochastic disturbances. However, the potential of the SNG framework in enhancing system resilience against cyber threats compared to traditional methods remains to be fully explored. While distributed consensus (DC) protocols are effective in ideal environments, they rely heavily on symmetric information exchange, which is vulnerable to denial-of-service (DoS) attacks. To address the hierarchical nature of leader–follower formations and the information

Received 24 January 2026; revised 23 February 2026; accepted 14 March 2026. Date of publication 27 March 2026; date of current version 19 August 2026. This work was supported in part by the National Natural Science Foundation of China under Grant 62373252, Grant 62573122, and Grant 62576098. This article was recommended by Associate Editor C.-T. Lin. *(Corresponding author: Jinliang Liu.)*

Jinliang Liu and Zihan Zhang are with the School of Computer Science, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: liujinliang@vip.163.com; zhangzihan2409@163.com).

Engang Tian is with the School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China (e-mail: tianengang@163.com).

Chen Peng is with the School of Mechatronic Engineering and Automation, Department of Automation, Shanghai University, Shanghai 200444, China (e-mail: c.peng@shu.edu.cn).

Jinde Cao is with Purple Mountain Laboratories, Nanjing 211111, China, also with the School of Mathematics, Southeast University, Nanjing 211189, China, and also with the Department of Mathematics, Luoyang Normal University, Luoyang 471934, China (e-mail: jdcao@seu.edu.cn).

Tingwen Huang is with the Faculty of Computer Science and Control Engineering, Shenzhen University of Advanced Technology, Shenzhen 518055, China (e-mail: huangtw2024@163.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2026.3674952.

Digital Object Identifier 10.1109/TCYB.2026.3674952

asymmetry caused by attacks, we adopt the SNG framework. As demonstrated in recent studies [13], [14], SNG allows the leader to optimize the global formation objective based on the hierarchical structure. This hierarchical guidance provides superior resilience compared to reactive consensus methods, which lack a global coordinator when interfollower links are severed.

The desired Stackelberg–Nash equilibrium (SNE) is determined by the solution of the coupled Hamilton–Jacobi equations of the nonlinear system. RL lies at the core of optimal control theory [21], offering the advantages of real-time learning and optimization [6]. It focuses on how an agent interacts with its environment to maximize the cumulative reward and obtains a near-optimal solution by iteratively solving the Hamilton–Jacobi–Bellman (HJB) equation online [22], [23]. Distributed secure surrounding control based on the Stackelberg game framework was investigated in [24]. The actor–critic (AC) structure enables agents to learn optimal behaviors through interaction with the environment [25], [26]. In marine environments, wireless communication among USVs is subject to latency, packet loss [27], [28], and, more critically, cyber threats such as deception attacks [29], false data injection (FDI) attacks [30], and DoS attacks [31], [32]. DoS attacks are deliberate disruptions of communication channels, resulting in information loss and degraded coordination [33], [34]. Conventional formation control methods often employ passive zero-order hold (ZOH) mechanisms during data loss. However, under aperiodic DoS attacks, this passive approach forces followers to effectively operate open-loop, leading to significant error accumulation and potential instability. To address this, the proposed consensus-based estimator shifts from passive waiting to active reconstruction [9]. By exploiting network redundancy to reconstruct missing neighbor data locally, it ensures formation stability even when direct communication channels are temporarily severed. To mitigate the adverse effects of communication disruptions, researchers have proposed resilient estimation [35], [36] and control mechanisms [37] that enable agents to maintain performance under limited or corrupted information. Significant advancements have been made in enhancing temporal performance and communication efficiency through fixed-time convergent RL [38], [39] and event-triggered mechanisms [40], [41]. Distinct from these studies that primarily focus on convergence speed or bandwidth optimization, our proposed SNG framework is designed to address the hierarchical decision-making challenges within multiagent systems. Furthermore, to complement existing efficiency-oriented designs, we incorporate a consensus-based estimator to ensure active resilience against aperiodic DoS attacks.

To address the aforementioned limitations, we propose a hierarchical control framework. The main contributions are summarized as follows.

1) *Hierarchical SNG Framework:* Unlike symmetric control approaches [10], we establish an SNG framework. This explicitly models the leader–follower hierarchy, enabling the leader to optimize global performance by anticipating followers' responses.

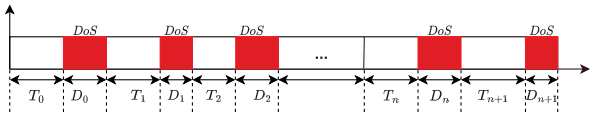

Fig. 1. Illustration of the aperiodic DoS attack intervals.

2) *Model-Free Online Learning:* We propose a novel integration of the AC algorithm within the Stackelberg–Nash framework. Specifically, the leader's actor network is designed to learn a policy that anticipates the followers' Nash equilibrium responses, enabling the online solution of the coupled hierarchical HJB equations without requiring prior knowledge of the nonlinear system dynamics.
3) *Active DoS Resilience:* Distinct from passive ZOH strategies [31], [33], we design a consensus-based active estimator. This mechanism exploits topology redundancy to reconstruct missing data during aperiodic DoS attacks.

II. Preliminaries and Problem Formulation

A. Motion Model of USVs

Consider a formation of $N+1$ networked USVs, and the kinematics and kinetics of the USV$_i$($i=0,1,2,\ldots,N$) are

$$\begin{cases}\dot{\eta}_i = R(\psi_i)\,\nu_i \\ M_i\dot{\nu}_i = f_i(\nu_i) + \tau_i + \omega_i\end{cases} \tag{1}$$

where $\eta_i = [x_i, y_i, \psi_i]^T \in \mathbb{R}^3$ and $\nu_i = [u_i, v_i, r_i]^T \in \mathbb{R}^3$ are the position and velocity vectors, respectively, and $R(\psi_i) = [\cos\psi_i, -\sin\psi_i, 0; \sin\psi_i, \cos\psi_i, 0; 0, 0, 1]$ is the rotation matrix. $M_i \in \mathbb{R}^{3\times3}$ is the inertial matrix; $f_i(\nu_i) \in \mathbb{R}^3$ denotes the combined unknown effect of Coriolis and centripetal forces, hydrodynamic drag, and potentially unknown damping and unmodeled dynamics; $\tau_i \in \mathbb{R}^3$ is the control input; and $\omega_i \in \mathbb{R}^3$ is the bounded environmental disturbance. The model follows the standard 3-DOF USV dynamics [9].

B. Aperiodic DoS Attacks

As shown in Fig. 1, it is assumed that the communication channels between the USVs are subject to aperiodic DoS attacks. Define $T_n = [t_{2n}, t_{2n+1}), (n \in \mathbb{N}, t_0 = 0)$ as the $(n+1)$th secure communication interval during which the DoS attack is inactive, allowing information exchange between USVs. Let $D_n = [t_{2n+1}, t_{2n+2})$ denote the time period of the $(n+1)$th DoS attack occurring. During this interval, the network communication between some USVs is disrupted due to the jamming signals, preventing them from sending and receiving information from neighboring USVs. An example of the aperiodic DoS attacks is illustrated in Fig. 1, where t_{2n+1} denotes the nth DoS attack start instant and $u_n = t_{2n+2} - t_{2n+1}$ is the attack duration. Additionally, considering the energy constraints of DoS attacks in practical scenarios, the following assumptions are proposed to further characterize the nature of DoS attacks.

Assumption 1 (DoS Frequency [9]): Let $n(0,t)$ denote the number of DoS attacks occurring over the interval $[0,t]$. There exist scalars $\Lambda_f, t_f > 0$ such that $n(0,t) \le \Lambda_f + (t/t_f)$.

Assumption 2 (DoS Duration [9]): The total interval of DoS attacks occurring over $[0,t]$ is represented by $\Phi(0,t) = \bigcup_{n\in\mathbb{N}} D_n \cap [0,t]$. There exist scalars $\Lambda_d > 0$ and $t_d > 1$ such that $|\Phi(0,t)| \le \Lambda_d + (t/t_d)$.

Remark 1 (Engineering Rationale): Assumptions 1 and 2 characterize the physical constraints of DoS attacks. Specifically, they ensure average reachability of the leader's commands and reflect practical limitations on attacker energy and stealth, which prevent continuous jamming.

Assumption 3: DoS attacks are assumed to exclusively disrupt interfollower communications. Leaders typically possess high-grade anti-jamming equipment, whereas interfollower links are comparatively more vulnerable.

C. Action Estimators Based on the Consensus Protocol

The communication topology among followers $\mathcal{O} = \{1,\ldots,N\}$ is time-varying due to link-specific aperiodic DoS attacks. This framework considers partial link interruptions. The adjacency weights are defined as

$$b_{i,j}(t) = \begin{cases} 0, & t \in D_n \text{ and the channel } (i,j) \text{ is under attack} \\ a_{ij}, & \text{otherwise} \end{cases} \tag{2}$$

where $i,j = 1,2,\ldots,N$. A DC-based estimator [42] is designed to reconstruct the unknown actions η_j via

$$\dot{\hat{\eta}}_{i,j} = -\mu_{i,j}\left(\sum_{k\in\mathcal{N}_i} b_{i,k}\left(\hat{\eta}_{i,j} - \hat{\eta}_{k,j}\right) + b_{i,j}\left(\hat{\eta}_{i,j} - \eta_j\right)\right) \tag{3}$$

where $i = 1,2,\ldots,N$, $\mu_{i,j} \in \mathbb{R}^{3\times 3}$ is a symmetric positive-definite gain matrix, $\hat{\eta}_{i,j} \in \mathbb{R}^3$ represents the action estimate for the ith player, $\mathcal{N}_i$ denotes its neighbor set, and $b_{i,k}$ and $b_{i,j}$ are adjacency weights.

Define the stacked vectors as $\hat{\eta}_i = \left[\hat{\eta}_{i,1}^{\mathrm{T}},\cdots,\hat{\eta}_{i,N}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{3N}, \eta = \left[\eta_1^{\mathrm{T}},\cdots,\eta_N^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{3N}$. Then, the compact form of estimator dynamics becomes

$$\dot{\hat{\eta}}_i = -\mu_i\left(\sum_{k\in\mathcal{N}_i} b_{i,k}(t)\left(\hat{\eta}_i - \hat{\eta}_k\right) + \left(\Lambda_i \otimes I_3\right)\left(\hat{\eta}_i - \eta\right)\right) \tag{4}$$

where $\mu_i = \mathrm{diag}\{\mu_{i,1},\ldots,\mu_{i,N}\} \in \mathbb{R}^{3N\times 3N}$ and $\Lambda_i = \mathrm{diag}\{b_{i,1},\ldots,b_{i,N}\} \in \mathbb{R}^{N\times N}$. Define the global error vectors $\tilde{\eta} = \hat{\eta} - 1_N \otimes \eta$, with $\hat{\eta} = \left[\hat{\eta}_1^{\mathrm{T}},\cdots,\hat{\eta}_N^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{3NN}$.

Then, the dynamics of $\tilde{\eta}$ are given by

$$\dot{\tilde{\eta}} = -\mu\left(\mathcal{H}\otimes I_3\right)\tilde{\eta} - \mathbf{1}_N \otimes \dot{\eta} \tag{5}$$

where $\mu = \mathrm{diag}(\mu_1,\ldots,\mu_N) \in \mathbb{R}^{3NN\times 3NN}$, $\kappa = \mathrm{diag}(\kappa_1,\ldots,\kappa_N) \in \mathbb{R}^{NN\times NN}$, and $\mathcal{H} = L \otimes I_N + \Lambda$ with $\Lambda = \mathrm{diag}(\Lambda_1,\ldots,\Lambda_N) \in \mathbb{R}^{NN\times NN}$.

Remark 2: The leader is assumed to have exclusive access to the predefined reference trajectory $\eta_r(t)$. This informational asymmetry necessitates a hierarchical SNG framework, with the leader as the primary tracking target.

D. Problem Formulation

Consider a formation of $N+1$ USVs consisting of a leader (USV_0) and N followers, forming a hierarchical decision-making paradigm. Let $e_i(\cdot)$ denote the formation error of USV_i, which is driven to zero by the control input τ_i to achieve formation control. Accordingly, a cost function is constructed for each agent based on its role. For notation, define $\|e\|_Q^2 = e^T Q e$ for any vector e and matrix Q.

Regarding the leader, it is assumed to anticipate the strategies adopted by all followers, and its cost function is defined as follows:

$$J_0\left(e_0(t_0),\tau_0,\tau_{-0}\right) = \int_{t_0}^{\infty} r_0\left(e_0,\tau_0,\tau_{-0}\right)dt \tag{6}$$

$$r_0\left(e_0,\tau_0,\tau_{-0}\right) = \|e_0\|_{Q_0}^2 + \|\tau_0 + \sum_{k=1}^{N} c_k\tau_k\|_{P_0}^2 \tag{7}$$

where $\tau_{-0} = \{\tau_j : j \in \mathcal{O}\}$ represents the policy collection for all USVs excluding τ_0. Q_0 and P_0 are symmetric positive-definite matrices. $c_k > 0$ quantifies follower k's influence magnitude on the leader's dynamics.

For the followers, the associated cost function is

$$J_i\left(e_i(t_0),\tau_0,\tau_i,\tau_{-i}\right) = \int_{t_0}^{\infty} r_i\left(e_i,\tau_0,\tau_i,\tau_{-i}\right)dt \tag{8}$$

$$r_i\left(e_i,\tau_0,\tau_i,\tau_{-i}\right) = \|e_i\|_{Q_i}^2 + \|\tau_i + b_i\tau_0\|_{P_i}^2 \tag{9}$$

in which $\tau_{-i} = \{\tau_j : j \in N_i\}$ represents the policy collection associated with the neighbors of USV i. $Q_i > 0$ and $P_i > 0$ denote two matrices that are positive-definite and symmetric. $b_i \ge 0$ reflects the influence intensity of leader on follower i.

The value function corresponding to the cost function at time is given by the following formula:

$$V_i\left(e_i(t)\right) = \int_t^{\infty} r_i\left(e_i,\tau_i,\tau_{-i}\right)d\tau, \quad i \in \mathcal{V}. \tag{10}$$

The optimal strategies and associated value functions are characterized by the following formula:

$$\tau_i^* = \arg\min_{\tau_i} V_i^*\left(e_i(t)\right), \quad i \in \mathcal{V} \tag{11}$$

$$V_i^*\left(e_i(t)\right) = \begin{cases} \min_{\tau_0}\int_{t_0}^{\infty} r_0\left(e_0,\tau_0,\tau_{-0}^*\right)d\tau, & i = 0 \\ \min_{\tau_i}\int_{t_0}^{\infty} r_i\left(e_i,\tau_0^*,\tau_i,\tau_{-i}^*\right)d\tau, & i \in O. \end{cases} \tag{12}$$

Definition 1 (Admissible Policy [13]): The control policy $\tau_i(t)$ is considered admissible, if it meets the following criteria: $\tau_i(t)$, for each $i \in \mathcal{V}$, maintains continuity, the collection $\{\tau_0(t),\tau_1(t),\ldots,\tau_N(t)\}$ ensures the stability of system, and the associated value function $V_i(e_i(t)) < \infty$.

Definition 2 (SNE [13]): A mapping $\tilde{\mathcal{T}}_i : \tau_0 \to \tau_i$ exists for each i, ensuring that, given any fixed τ_0 and for all τ_i, it has

$$\begin{cases} J_0\left(e_0(t_0),\tilde{\tau}_0,\tilde{\mathcal{T}}_{-0}(\tilde{\tau}_0)\right) \le J_0\left(e_0(t_0),\tau_0,\tilde{\mathcal{T}}_{-0}(\tau_0)\right) \\ J_i\left(e_i(t_0),\tau_0,\tilde{\mathcal{T}}_i(\tau_0),\tilde{\mathcal{T}}_{-i}(\tau_0)\right) \le J_i\left(e_i(t_0),\tau_0,\tau_i,\tilde{\mathcal{T}}_{-i}(\tau_0)\right) \end{cases} \tag{13}$$

then the policy $\{\tilde{\tau}_0,\tilde{\tau}_1,\ldots,\tilde{\tau}_N\}$ with $\tilde{\tau}_i = \mathcal{T}_i(\tilde{\tau}_0)$ is regarded as forming the SNE.

Remark 3: In the Stackelberg framework, the leader anticipates state-dependent best responses (24). To compute the equilibrium, the leader requires real-time states of all followers. Thus, the communication topology assumes that the leader has direct access to all followers as in-degree neighbors to ensure game solvability.

III. Hierarchical AC Control Design Under Stackelberg–Nash Framework

Based on the clarified Stackelberg–Nash interaction mechanism, the AC network is utilized to approximate the value and policy functions and derive optimal control laws for both the leader and followers. Although the control design operates on the nominal structure of (1), it remains independent of precise hydrodynamic parameters. Instead, the AC algorithm adaptively approximates the optimal policy online, compensating for parametric uncertainties without high-precision system identification. To address USV nonlinearities, a hybrid architecture is adopted. Backstepping ensures structural stability by decomposing dynamics, while the integrated AC algorithm minimizes the local cost (8) within the dynamic loop to learn optimal policies online. This synergy guarantees both stability and optimality, effectively overcoming the optimality limits of standard backstepping and the convergence challenges of RL.

A. Optimized Control Design of Follower USVs

After observing the leader's action, the followers maintain a formation with the leader and make optimal decisions to preserve a certain pattern relative to neighboring USVs.

The error variables are defined as

$$e_{\eta_i}(t) = \eta_i(t) - \eta_0(t) - d_{i0} + \sum_{j\in\mathcal{O}_{i-}} \left(\eta_i(t) - \eta_j(t) - d_{ij}\right) \tag{14}$$

$$e_i(t) = \nu_i(t) - \xi_{\eta_i} \tag{15}$$

where $i \in \mathcal{O}$ and $\mathcal{O}_{i-}$ represent the collection of followers excluding i, $d_{ij} \in \mathbb{R}^3$ is the desired displacement between the ith follower and jth neighbor, $d_{i0} \in \mathbb{R}^3$ is the desired displacement of USV$_i$ relative to the leader, and $\xi_{\eta_i} \in \mathbb{R}^3$ is the virtual control input.

The corresponding error dynamics are given by

$$\dot{e}_{\eta_i}(t) = \dot{\eta}_i(t) - \dot{\eta}_0(t) + \sum_{j\in\mathcal{O}_{i-}} \left(\dot{\eta}_i(t) - \dot{\eta}_j(t)\right) \tag{16}$$

$$\dot{e}_i(t) = \dot{\nu}_i(t) - \dot{\xi}_{\eta_i}. \tag{17}$$

The optimized USV control is designed using a two-step backstepping method.

Step 1: Utilizing $e_i(t) = \nu_i(t) - \xi_{\eta_i}$, (16) can be rewritten as

$$\dot{e}_{\eta_i}(t) = R(\psi_i)\xi_{\eta_i} + R(\psi_i)e_i(t) - \dot{\eta}_0 + \sum_{j\in\mathcal{O}_{i-}} \left(R(\psi_i)\xi_{\eta_i} + R(\psi_i)e_i(t) - \dot{\eta}_j(t)\right). \tag{18}$$

To mitigate the impact of intermittent DoS attacks, estimated signals are employed to compensate for potential data loss. According to (3), (18) can be rewritten as

$$\dot{e}_{\eta_i}(t) = R(\psi_i)\xi_{\eta_i} + R(\psi_i)e_i(t) - \dot{\eta}_0(t) + \sum_{j\in\mathcal{O}_{i-}} \left(R(\psi_i)\xi_{\eta_i} + R(\psi_i)e_i(t) - \hat{\dot{\eta}}_{i,j}(t)\right). \tag{19}$$

The virtual control ξ_{η_i} is designed as

$$\xi_{\eta_i} = R(\psi_i)^{-1}\left(A_i^{-1}\left(-k_i e_{\eta_i}(t) + \dot{\eta}_0(t) + \sum_{j\in\mathcal{O}_{i-}} \hat{\dot{\eta}}_{i,j}(t)\right)\right) \tag{20}$$

where $A_i = 1 + |\mathcal{O}_{i-}|$ and $k_i > 1/2$ is the design parameter.

Step 2: The optimal control is obtained via the RL strategy. To commence, formulate the Hamiltonian function for USV $i \in \mathcal{O}$ in the following manner:

$$H_i(e_i, \nabla V_i, \tau_i, \tau_{-i}) = r_i(e_i, \tau_i, \tau_{-i}) + \left(\frac{\partial V_i}{\partial(e_i)}\right)^{\mathrm{T}}\left[M_i^{-1}(f_i(\nu_i) + \tau_i - \omega_i) - \dot{\xi}_{\eta i}\right]. \tag{21}$$

For ease of subsequent analysis, let $\nabla V_i = \partial V_i/\partial(e_i)$. Given an arbitrary policy τ_0, the corresponding optimal response for each follower $i \in \mathcal{O}$ can be formulated as

$$\tau_i^*(\tau_0) = \arg\min_{\tau_i} H_i\left(e_i, \nabla\tilde{V}_i, \tau_i, \tau_{-i}\right) \tag{22}$$

where $\tilde{V}_i$ is the corresponding value function of USV i based on the policy sequence $\{\tau_0, \tau_1^*(\tau_0), \ldots, \tau_N^*(\tau_0)\}$. Since $H_i(e_i, \nabla\tilde{V}_i, \tau_i, \tau_{-i})$ is strongly convex in τ_i, let $(\partial H_i/\partial\tau_i) = 0$, one can gain the following expression:

$$\tau_i^*(\tau_0) = -b_i\tau_0 - \frac{1}{2}P_i^{-1}M_i^{-T}\nabla\tilde{V}_i, \quad i \in \mathcal{O}. \tag{23}$$

Due to $\partial^2 H_i/\partial\tau_i^2 = 2P_i > 0$, $\tau_i^*(\tau_0)$ in (23) can achieve the minimum of the $H_i(e_i, \nabla\tilde{V}_i, \tau_i, \tau_{-i})$. When the leader takes the policy τ_0^*, each follower's best reaction can be expressed as follows:

$$\tau_i^*\left(\tau_0^*\right) = -b_i\tau_0^* - \frac{1}{2}P_i^{-1}M_i^{-T}\nabla V_i^*, \quad i \in \mathcal{O} \tag{24}$$

where V_i^* denotes the follower's optimal game value. To streamline the notation, we denote $\tau_i^*(\tau_0^*)$ simply by τ_i^*, with no risk of ambiguity in the subsequent discussion.

The derivation of the optimal control law τ_i^* utilizes the stationarity condition $(\partial H_i/\partial\tau_i) = 0$, treating neighbors' strategies τ_{-i} as fixed parameters. This follows the standard Nash equilibrium formulation [13]. Crucially, the coupling effects among followers are strictly captured in the subsequent coupled HJB equation (25), where the value function V_i evolves based on the joint actions of the entire neighborhood.

Then, inserting (24) into (21) yields

$$H_i\left(e_i, \nabla V_i^*, \tau_i^*, \tau_{-i}^*\right) = r_i\left(e_i, \tau_i^*, \tau_{-i}^*\right) + \left(\frac{\partial V_i^*}{\partial(e_i)}\right)^{\mathrm{T}}\left[M_i^{-1}\left(f_i(\nu_i) + \tau_i^* - \omega_i\right) - \dot{\xi}_{\eta i}\right] = 0 \tag{25}$$

for $i \in \mathcal{O}$. In summary, the HJB equation (25) for each USV is formed by strengthening the smoothness condition of the Hamiltonian quantity with respect to the control inputs. Decompose the gradient term (∇V_i^*) into two components as

$$\nabla V_i^* = 2z_i e_i(t) + J_i^o(e_i) \tag{26}$$

where $z_i \in \mathbb{R}$ is the positive design parameter.

Substituting into the expression of τ_i^* yields

$$\tau_i^* = -b_i\tau_0^* - z_i P_i^{-1} M_i^{-1} e_i(t) - \frac{1}{2} P_i^{-1} M_i^{-1} J_i^o(e_i). \tag{27}$$

Assumption 4: There exist ideal weights W_i^* and W_{ai}^* such that the optimal value function V_i^* and control policy τ_i^* can be approximated on a compact set Ω_i with bounded errors. Moreover, V_i^* is the unique positive-definite solution among stabilizing admissible control policies [43].

Remark 4: Although nonlinear HJB equations may yield multiple solutions generally, restricting the scope to admissible controls ensures the uniqueness of V_i^* [43]. The proposed update laws minimize the Bellman error to target this specific solution. As proved in Theorem 2, this mechanism guarantees weight convergence to the neighborhood of W^*, ensuring closed-loop stability.

The term $J_i^o(e_i)$ is continuous but uncertain so that an NN can approximate it on the compact set Ω_{τ_i}

$$J_i^o(e_i) = W_i^{\mathrm{T}} S_i(e_i) + \varepsilon_i(e_i) \tag{28}$$

where $W_i^* \in \mathbb{R}^{p_i \times 3}$ is the ideal weight matrix with the neural neuron number p_i, $S_i(e_i) \in \mathbb{R}^{p_i}$ is the basis function vector, and $\varepsilon_i(e_i) \in \mathbb{R}^3$ denotes the approximation error.

Thus, the optimal control and its gradient become

$$\nabla V_i^* = 2z_i e_i(t) + W_i^{\mathrm{T}} S_i(e_i) + \varepsilon_i(e_i) \tag{29}$$

$$\begin{aligned}\tau_i^* = &-b_i\tau_0 - z_i P_i^{-1} M_i^{-1} e_i(t) \\ &- \frac{1}{2} P_i^{-1} M_i^{-1} W_i^{\mathrm{T}} S_i(e_i) - \frac{1}{2} P_i^{-1} M_i^{-1} \varepsilon_i.\end{aligned} \tag{30}$$

Since the ideal weight matrix $W_i^* \in \mathbb{R}^{p_1 \times 3}$ is an unknown constant matrix, the optimal control cannot be implemented. RL is performed by constructing the following critic and actor networks. The critic for evaluating the control performance is

$$\nabla \hat{V}_i^* = 2z_i e_i(t) + \hat{W}_{ci}^{\mathrm{T}}(t) S_i(e_i) \tag{31}$$

where $\nabla \hat{V}_i^*$ is the estimation of ∇V_i^*,$\hat{W}_{ci}(t) \in \mathbb{R}^{p_i \times 3}$ denotes the critic NN weight and is adjusted by the following law:

$$\dot{\hat{W}}_{ci}(t) = -\kappa_{ci} S_i(e_i) S_i^{\mathrm{T}}(e_i) \hat{W}_{ci}(t) \tag{32}$$

where $\kappa_{ci} \in \mathbb{R}$ is the positive critic gain parameter.

The actor for generating the control action is defined as

$$\hat{\tau}_i^* = -b_i\tau_0 - z_i P_i^{-1} M_i^{-1} e_i(t) - \frac{1}{2} P_i^{-1} M_i^{-1} W_{ai}^{\mathrm{T}} S_i(e_i) \tag{33}$$

where $\hat{W}_{ai}(t) \in \mathbb{R}^{p_i \times 3}$ is the actor NN weight, updated by

$$\begin{aligned}\dot{\hat{W}}_{ai}(t) = &-S_i(e_i) S_i^{\mathrm{T}}(e_i) \\ &\times \left(\kappa_{ai}\left(\hat{W}_{ai}(t) - \hat{W}_{ci}(t)\right) + \kappa_{ci}\hat{W}_{ci}(t)\right)\end{aligned} \tag{34}$$

where $\kappa_{ai} \in \mathbb{R}$ is the positive actor gain parameter.

These design parameters z_i, κ_{ci}, and κ_{ai} are selected to satisfy the following conditions:

$$\begin{aligned}&z_i > \frac{\zeta \lambda_{P_i}^{\max}\left(\lambda_{M_i}^{\max}\right)^2}{4}, \quad \kappa_{ai} > \frac{1}{2\left(\lambda_{P_i}^{\min}\right)^2\left(\lambda_{M_i}^{\min}\right)^4}, \\ &\kappa_{ai} > \kappa_{ci} > \tfrac{1}{2}\kappa_{ai}\end{aligned} \tag{35}$$

where $\zeta_i = 9 + 2A_i^2$; $\lambda_{P_i}^{\max}$ and $\lambda_{M_i}^{\max}$ are the largest eigenvalue of P_i, M_i, and $\lambda_{P_i}^{\min}$; and $\lambda_{M_i}^{\min}$ is the smallest eigenvalue of P_i and M_i.

By plugging (31) and (33) into the HJB equation, we obtain

$$\begin{aligned}H_i\left(e_i, \nabla V_i^*, \tau_i^*, \tau_{-i}\right) = &\, e_i^{\mathrm{T}}(t) Q_i e_i(t) \\ &+ \left(-z_i P_i^{-1} M_i^{-1} e_i(t) - \frac{1}{2} P_i^{-1} M_i^{-1} \hat{W}_{ai}^{\mathrm{T}}(t) S_i(e_i)\right)^{\mathrm{T}} \\ &\times P_i \left(-z_i P_i^{-1} M_i^{-1} e_i(t) - \frac{1}{2} P_i^{-1} M_i^{-1} \hat{W}_{ai}^{\mathrm{T}}(t) S_i(e_i)\right) \\ &+ \left(2z_i e_i(t) + \hat{W}_{ci}^{\mathrm{T}}(t) S_i(e_i)\right)^{\mathrm{T}} M_i^{-1} \\ &\times \left(f_i(v_i) - b_i\tau_0 - z_i P_i^{-1} M_i^{-1} e_i(t)\right. \\ &\left. - \frac{1}{2} P_i^{-1} M_i^{-1} \hat{W}_{ai}^{\mathrm{T}}(t) S_i(e_i) + \omega_i - \dot{\xi}_{\eta_i}(t)\right).\end{aligned} \tag{36}$$

Provided that $H_i = 0$ admits a unique solution [44], this condition is tantamount to

$$\frac{\partial H_i}{\partial \hat{W}_{ai}(t)} = S_i(e_i) S_i^{\mathrm{T}}(e_i)\left(\hat{W}_{ai}(t) - \hat{W}_{ci}(t)\right) = 0_{p_i \times 3}. \tag{37}$$

It is evident that $P(t) = 0$ is equal to (37)

$$P(t) = \mathrm{Tr}\left\{\left(\hat{W}_{ai}(t) - \hat{W}_{ci}(t)\right)^{\mathrm{T}}\left(\hat{W}_{ai}(t) - \hat{W}_{ci}(t)\right)\right\} \tag{38}$$

where $\mathrm{Tr}\{\cdot\}$ denotes the trace operator. The time derivative of $P(t)$ is given by

$$\frac{dP(t)}{dt} = \mathrm{Tr}\left\{\frac{\partial P}{\partial \hat{W}_{ci}^{\mathrm{T}}(t)} \dot{\hat{W}}_{ci}(t) + \frac{\partial P}{\partial \hat{W}_{ai}^{\mathrm{T}}(t)} \dot{\hat{W}}_{ai}(t)\right\} \le 0. \tag{39}$$

Inequality (39) guarantees that the RL laws (32) and (34) can ensure asymptotic convergence of $P(t)$ to zero, thereby making sure the satisfaction of condition (37).

B. Optimized Control Design of the Leader

For the leader, after predicting the optimal strategies $\tau_{-0}^*(\tau_0)$ of the followers, the task involves tracking the reference trajectory η_r and maintaining the formation with the followers. The following gives the two-step backstepping AC design.

Step 1: The leader formation error is defined by

$$e_{\eta_0}(t) = \eta_0(t) - \eta_r(t) + \sum_{j \in \mathcal{O}}\left(\eta_0(t) - \eta_j(t) - d_{0j}\right), \tag{40}$$

$$e_0(t) = \nu_0(t) - \xi_{\eta_0}(t) \tag{41}$$

where $d_{0j} \in \mathbb{R}^3$ is the desired displacement of the leader relative to USV$_j$ and $\xi_{\eta_0} \in \mathbb{R}^3$ is a virtual velocity to be designed. The virtual control ξ_{η_0} is designed as

$$\xi_{\eta_0} = R(\psi_0)^{-1}\left(A_0^{-1}\left(-k_0 e_{\eta_0} + \dot{\eta}_r + \dot{\eta}_j\right)\right) \tag{42}$$

where $A_0 = 1 + |\mathcal{O}|$ and $k_0 > 1/2$ is a design parameter.

Step 2: Given the leader's ability to foresee how each follower would optimally react, it makes the decision according to the following expression:

$$\tau_0^* = \arg\min_{\tau_0} H_0\left(e_0, \nabla V_0^*, \tau_0, \tau_{-0}^*(\tau_0)\right) \tag{43}$$

with V_0^* denoting the leader's optimal game value. According to $(\partial H_0/\partial \tau_0) = 0$, we can gain

$$\tau_0^* = K_1 \sum_{k=1}^{N} c_k P_k^{-1} M_k^{-1} \nabla V_k^* - K_1 P_0^{-1} M_0^{-1} \nabla V_0^* \tag{44}$$

where $K_1 = 1/2\left(1 - \sum_{k=1}^{N} c_k b_k\right)$.

Remark 5 (Solvability Condition): To ensure the existence of the leader's optimal control, H_0 must be strictly convex, necessitating the constraint $\sum_{k=1}^{N} c_k b_k < 1$. This is practically satisfied by selecting appropriate weighting matrices in the cost function to bound the aggregate feedback from followers.

Decompose the gradient term (∇V_0^*) into two parts as

$$\nabla V_0^* = 2z_0 e_0(t) + J_0^o(e_0) \tag{45}$$

where $z_0 \in \mathbb{R}$ is a design parameter and $J_0^o(e_0) = -2z_0 e_0(t) + (\mathrm{d}J_0^*/\mathrm{d}e_0) \in \mathbb{R}^3$.

Substituting into the expression of τ_0^* yields

$$\tau_0^* = K_1 \sum_{k=1}^{N} c_k P_k^{-1} M_k^{-1} \nabla V_k^* - z_0 K_1 P_0^{-1} M_0^{-1} e_0(t) - \frac{1}{2} K_1 P_0^{-1} M_0^{-1} J_0^o(e_0). \tag{46}$$

The term $J_0^o(e_0)$ is continuous but uncertain, so that an NN can approximate it on the compact set Ω_{τ_0}

$$J_0^o(e_0) = W_0^{\mathrm{T}} S_0(e_0) + \varepsilon_0(e_0) \tag{47}$$

where $W_0^* \in \mathbb{R}^{p_0 \times 3}$ is the ideal weight matrix with the neural neuron number p_0, $S_0(e_0) \in \mathbb{R}^{p_0}$ is the basis function vector, and $\varepsilon_0(e_0) \in \mathbb{R}^3$ denotes the approximation error. The critic for evaluating the control performance is designed as

$$\nabla \hat{V}_0^* = 2z_0 e_0(t) + \hat{W}_{c0}^{\mathrm{T}}(t) S_0(e_0) \tag{48}$$

where $\nabla \hat{V}_0^*$ is the estimation of ∇V_0^* and $\hat{W}_{c0}(t) \in \mathbb{R}^{p_0 \times 3}$ denotes the critic NN weight and is adjusted by

$$\dot{\hat{W}}_{c0}(t) = -\kappa_{c0} S_0(e_0) S_0^{\mathrm{T}}(e_0) \hat{W}_{c0}(t) \tag{49}$$

where $\kappa_{c0} \in \mathbb{R}$ denotes the positive critic gain parameter.

The actor for generating the control action is defined as

$$\hat{\tau}_0^* = K_1 \sum_{k=1}^{N} c_k P_k^{-1} M_k^{-1} \nabla V_k^* - z_0 K_1 P_0^{-1} M_0^{-1} e_0(t) - \frac{1}{2} K_1 P_0^{-1} M_0^{-1} \hat{W}_{a0}^{\mathrm{T}}(t) S_0(e_0) \tag{50}$$

where $\hat{W}_{a0}(t) \in \mathbb{R}^{p_0 \times 3}$ is the actor NN weight, updated by

$$\dot{\hat{W}}_{a0}(t) = -S_0(e_0) S_0^{\mathrm{T}}(e_0) \times \left(\kappa_{a0}\left(\hat{W}_{a0}(t) - \hat{W}_{c0}(t)\right) + \kappa_{c0} \hat{W}_{c0}(t)\right) \tag{51}$$

where $\kappa_{a0} \in \mathbb{R}$ is the positive actor gain parameter. To implement the Stackelberg prediction, the leader's Actor network incorporates followers' gradient feedback $\nabla \hat{V}_k$. This coupling enables the leader to learn an optimal policy τ_0^* (44) that inherently anticipates the followers' Nash equilibrium responses, ensuring hierarchical coordination without requiring high-precision system identification.

IV. Main Result

Theorem 1: Consider smooth positive-definite solutions $V_i(e_i)$, $i \in \mathcal{V}$ of the coupled HJB equations, satisfying $V_i(0) = 0$. The optimal control τ_i^* is derived from (24) and (44). Consequently, the following results are valid.

1) The dynamics of local formation errors exhibit asymptotic stability.
2) The set $\{\tau_0^*, \tau_1^*, \ldots, \tau_N^*\}$ forms the SNE policy, and the corresponding optimal outcomes are $V_i(e_i(0))$, $i \in \mathcal{V}$.

Proof: First, consider $V_i(e_i)$ as a Lyapunov function

$$\dot{V}_i(e_i) = (\nabla V_i(e_i))^{\mathrm{T}} \left[M_i^{-1}(f_i(\nu_i) + \tau_i + \omega_i) - \dot{\xi}_{\eta i}\right]. \tag{52}$$

With the optimal controls (24) and (44), we have

$$\dot{V}_i(e_i) = \begin{cases} -\|e_0\|_{Q_0}^2 - \|\tau_0^* + \sum_{k=1}^{N} c_k \tau_k^{*T}\|_{P_0}^2, & i = 0, \\ -\|e_i\|_{Q_i}^2 - \|\tau_i^* + b_i \tau_0^{*T}\|_{P_i}^2, & i \in \mathcal{O}. \end{cases} \tag{53}$$

The cost function of USV $i \in \mathcal{V}$ is

$$J_0\left(e_0(t_0), \tau_0^*, \tau_{-0}^*\right) = V_0(e_0(t_0)), \quad i = 0 \tag{54}$$

$$J_i\left(e_i(t_0), \tau_i^*, \tau_{-i}^*\right) = V_i(e_i(t_0)), \quad i \in \mathcal{O}. \tag{55}$$

Following [45], it holds that:

$$\begin{aligned} &H_i\left(e_i, \nabla V_i, \tau_i^*, \tau_{-i}^*\right) \\ &= H_i\left(e_i, \nabla V_i, \tau_i, \tau_{-i}^*\right) + r_i\left(e_i, \tau_i^*, \tau_{-i}^*\right) - r_i\left(e_i, \tau_i, \tau_{-i}^*\right) \\ &\quad + \nabla V_i^T M_i^{-1}\left(\tau_i - \tau_i^*\right). \end{aligned} \tag{56}$$

For the leader, according to (43), for any control policy τ_0, $H_0(e_0, \nabla V_0, \tau_0^*, \tau_{-0}^*) \le H_0(e_0, \nabla V_0, \tau_0, \tau_{-0}^*)$. This implies

$$J_0\left(e_0(t_0), \tau_0, \tau_{-0}^*\right) \ge J_0\left(e_0(t_0), \tau_0^*, \tau_{-0}^*\right). \tag{57}$$

For each follower $i \in \mathcal{O}$, the following holds:

$$J_i\left(e_i(t_0), \tau_i, \tau_{-i}^*\right) = \int_{t_0}^{\infty} \left[\|\tau_i + b_i \tau_0^*\|_{P_i}^2 - \|\tau_i^* + b_i \tau_0^*\|_{P_i}^2 + \nabla V_i^T M_i^{-1}\left(\tau_i - \tau_i^*\right)\right] dt + V_i(e_i(t_0)). \tag{58}$$

In view of (58), it holds that

$$\nabla V_i^T M_i^{-1}\left(\tau_i - \tau_i^*\right) = -2\tau_i^{*T} P_i - 2\tau_i^T P_i \tag{59}$$

which further leads to

$$\begin{aligned} &\|\tau_i + b_i \tau_0^*\|_{P_i}^2 - \|\tau_i^* + b_i \tau_0^*\|_{P_i}^2 + \nabla V_i^T M_i^{-1}\left(\tau_i - \tau_i^*\right) \\ &= \|\tau_i - \tau_i^*\|_{P_i}^2. \end{aligned} \tag{60}$$

From (55) and (60), we can obtain the expression

$$J_i\left(e_i(t_0), \tau_0, \tau_i, \tau_{-i}^*\right) \ge J_i\left(e_i(t_0), \tau_0, \tau_i^*, \tau_{-i}^*\right). \tag{61}$$

Incorporating (57) and (61), the SNE properties are met with $\tilde{\tau}_0 = \tau_0^*$ and $\mathcal{T}_i(\tilde{\tau}_0) = \tau_i^*$, $i \in \mathcal{O}$. Additionally, the corresponding optimal outcomes are $V_i(e_i(t_0))$ for all USVs.

Theorem 2: All error signals of the overall USV system remain semi-globally uniformly ultimately bounded (SGUUB) and the USV tracks the prescribed reference trajectory to the required accuracy if the design parameters satisfy the conditions (32), (34), and (35).

Proof: The Lyapunov candidate is selected as

$$L_1(t) = \frac{1}{2} e_{\eta_i}(t)^{\mathrm{T}} e_{\eta_i}(t). \tag{62}$$

Taking the time derivative along (17) is

$$\dot{L}_1(t) = -k_i e_{\eta_i}^{\mathrm{T}}(t) e_{\eta_i}(t) + A_i e_{\eta_i}^{\mathrm{T}}(t) R(\psi_i) e_i(t). \tag{63}$$

Applying Cauchy–Schwarz and Young's inequalities yields

$$\dot{L}_1(t) \le -\left(k_i - \frac{1}{2}\right) e_{\eta_i}^{\mathrm{T}}(t)\, e_{\eta_i}(t) + \frac{A_i^2}{2} e_i^{\mathrm{T}}(t)\, e_i(t). \tag{64}$$

Define the composite Lyapunov function candidate

$$L_2(t) = L_1(t) + \frac{1}{2} e_i^{\mathrm{T}}(t)\, e_i(t) + \frac{1}{2}\mathrm{Tr}\left[\tilde{W}_{ai}^{\mathrm{T}} \tilde{W}_{ai}\right] + \frac{1}{2}\mathrm{Tr}\left[\tilde{W}_{ci}^{\mathrm{T}} \tilde{W}_{ci}\right] \tag{65}$$

where $\tilde{W}_{ai}(t) = \hat{W}_{ai}(t) - W_i^*$ and $\tilde{W}_{ci}(t) = \hat{W}_{ci}(t) - W_i^*$. Along with system dynamics and adaptive laws, its derivative yields

$$\begin{aligned}\dot{L}_2(t) = {} & \dot{L}_1(t) + e_i^{\mathrm{T}}\left(M_i^{-1}\left(f_i(\nu_i) + \tau_i + \omega_i\right) - \dot{\xi}_{\eta_i}\right) \\ & - \kappa_{ci}\mathrm{Tr}\left[\tilde{W}_{ci}^{\mathrm{T}} S_i(e_i)\, S_i^{\mathrm{T}}(e_i)\, \hat{W}_{ci}\right] \\ & - \mathrm{Tr}\left[\tilde{W}_{ai}^{\mathrm{T}} S_i(e_i)\, S_i^{\mathrm{T}}(e_i)\left(\kappa_{ai}\left(\hat{W}_{ai} - \hat{W}_{ci}\right) + \kappa_{ci}\hat{W}_{ci}\right)\right].\end{aligned} \tag{66}$$

Implementing (33)–(66) and utilizing the Cauchy–Schwarz and Young's inequalities, the following results are obtained:

$$\begin{aligned}\dot{L}_2(t) \le {} & \dot{L}_1(t) - \left(\frac{z_i}{\lambda_{P_i}^{\max}\left(\lambda_{M_i}^{\max}\right)^2} - \frac{9}{4}\right) e_i^{\mathrm{T}}(t)\, e_i(t) \\ & - \frac{\kappa_{ci}}{2}\mathrm{Tr}\left\{\tilde{W}_{ci}^{\mathrm{T}}(t)\, S_i(e_i)\, S_i^{\mathrm{T}}(e_i)\, \tilde{W}_{ci}(t)\right\} \\ & - \frac{\kappa_{ci}}{2}\mathrm{Tr}\left\{\tilde{W}_{ai}^{\mathrm{T}}(t)\, S_i(e_i)\, S_i^{\mathrm{T}}(e_i)\, \tilde{W}_{ai}(t)\right\} \\ & - \left(\kappa_{ci} - \frac{\kappa_{ai}}{2}\right)\mathrm{Tr}\left\{\hat{W}_{ci}^{\mathrm{T}}(t)\, S_i(e_i)\, S_i^{\mathrm{T}}(e_i)\, \hat{W}_{ci}(t)\right\} \\ & - \left(\frac{\kappa_{ai}}{2} - \frac{1}{4\left(\lambda_{P_i}^{\min}\right)^2\left(\lambda_{M_i}^{\min}\right)^4}\right)\mathrm{Tr}\left\{\hat{W}_{ai}^{\mathrm{T}} S_i(e_i)\, S_i^{\mathrm{T}}(e_i)\, \hat{W}_{ai}\right\} \\ & + \frac{\kappa_{ai} + \kappa_{ci}}{2}\mathrm{Tr}\left\{W_i^{*\mathrm{T}} S_i(e_i)\, S_i^{\mathrm{T}}(e_i)\, W_i^*\right\} + \frac{1}{2}\left\|M_i^{-1}\omega\right\|^2 \\ & + \frac{1}{2}\left\|M_i^{-1} f_i(\nu_i)\right\|^2 + \frac{1}{2}\left\|M_i^{-1} b_i \tau_0^*\right\|^2 + \frac{1}{2}\left\|\dot{\xi}_{\eta_i}\right\|^2.\end{aligned} \tag{67}$$

According to condition (35) and (66), (67) becomes

$$\begin{aligned}\dot{L}_2(t) \le {} & -(2k_i - 1) L_1(t) \\ & - \left(\frac{z_i}{\lambda_{P_i}^{\max}\left(\lambda_{M_i}^{\max}\right)^2} - \frac{9 + 2A_i^2}{4}\right) e_i^{\mathrm{T}}(t)\, e_i(t) \\ & - \frac{\kappa_{ci}}{2}\lambda_{S_i}^{\min}\mathrm{Tr}\left\{\tilde{W}_{ci}^{\mathrm{T}}(t)\, \tilde{W}_{ci}(t)\right\} \\ & - \frac{\kappa_{ci}}{2}\lambda_{S_i}^{\min}\mathrm{Tr}\left\{\tilde{W}_{ai}^{\mathrm{T}}(t)\, \tilde{W}_{ai}(t)\right\} + C_1(t)\end{aligned} \tag{68}$$

where $\lambda_{(S_i)}^{\min}$ is the smallest eigenvalue of $S_i(e_i)S_i^{\mathrm{T}}(e_i)$ and $C_1(t) = [(\kappa_{ci} + \kappa_{ai})/2]\|W_i^{*\mathrm{T}} S_i(e_i)\|^2 + (1/2)\|\dot{\xi}_{\eta_i}\|^2 + (1/2)\|M_i^{-1}\omega\|^2 + (1/2)\|M_i^{-1} f_i(\nu_i)\|^2 + (1/2)\|M_i^{-1} b_i \tau_0^*\|^2$. Since every term in $C_1(t)$ is bounded, there is a constant q_1 that makes $|C_1(t)| \le q_1$.

Let $\sigma = (4z_i - \zeta_i \lambda_{P_i}^{\max}(\lambda_{M_i}^{\max})^2)/4\lambda_{P_i}^{\max}(\lambda_{M_i}^{\max})^2$ and $\zeta_i = 9 + 2A_i^2$. Define $a_i = \min\{2k_i - 1, 2\sigma, \kappa_{ci}\lambda_{S_i}^{\min}\}$, and then,

$$\dot{L}_2(t) \le -a_i L_2(t) + q_1. \tag{69}$$

Then, it yields the following inequality:

$$L_2(t) \le e^{-a_i t} L_2(0) + \frac{q_1}{a_i}\left(1 - e^{-a_i t}\right). \tag{70}$$

Intuitively, inequality (69) elucidates the interplay between the stabilizing control forces (the decay term $-a_i L_2$) and the bounded disturbances (the source term q_1). The system dynamics are dominated by the control input when errors are large, forcing the states into a compact set determined by the ratio q_1/a_i, thus guaranteeing uniform ultimate boundedness (UUB). The control laws for the leader can be designed similarly; the proof is omitted here for brevity.

Theorem 3: Considering the estimator dynamics in (3), and under *Assumptions 1 and 2*, the estimation error system is input-to-state stability (ISS) with respect to $\dot{\eta}$.

Proof: Consider the following Lyapunov candidate function:

$$V_a = \frac{1}{2}\tilde{\eta}^{\mathrm{T}}(\mathcal{H} \otimes I_3)\,\tilde{\eta}. \tag{71}$$

Taking the time derivative of V_a along (3) yields

$$\dot{V}_a = -\tilde{\eta}^{\mathrm{T}}(\mathcal{H} \otimes I_3)\,\mu\,(\mathcal{H}(t) \otimes I_3)\,\tilde{\eta} - \tilde{\eta}^{\mathrm{T}}(\mathcal{H} \otimes I_3)\,(\mathbf{1}_N \otimes \dot{\eta}). \tag{72}$$

1) *During Secure Intervals T_n:* When $t \in T_n$, the Laplacian matrix $L(t)$ remains positive-definite with $\lambda_2(L(t)) > 0$ and $\mathcal{H}(t)$ satisfies $\lambda_{\min}(\mathcal{H}(t)) \ge \lambda_2 > 0$. Using the properties of symmetric positive-definite matrices, one obtains

$$\dot{V}_a \le -\lambda_{\min}(\mu)\lambda_2\lambda_{\min}(\mathcal{H})\|\tilde{\eta}\|^2 + \lambda_{\max}(\mathcal{H})\sqrt{N}\|\tilde{\eta}\|\|\dot{\eta}\|.$$

Applying Young's inequality gives

$$\dot{V}_a \le -\frac{q_3}{2}\|\tilde{\eta}\|^2 + \frac{q_4^2}{2q_3}\|\dot{\eta}\|^2 \le -\frac{q_3}{q_2}V_a + \frac{q_4^2}{2q_3}d_{\max}^2 \tag{73}$$

where $q_2 = \lambda_{\max}(\mathcal{H})$, $q_3 = \lambda_{\min}(\mu)\lambda_2\lambda_{\min}(\mathcal{H})$, $q_4 = \lambda_{\max}(\mathcal{H})(N)^{1/2}$, and $d_{\max} = \sup_t \|\dot{\eta}(t)\|$ denotes the maximum variation rate of the estimated signal.

2) *During DoS Intervals D_n:* The convergence term in (72) weakens but remains bounded. Hence,

$$\dot{V}_a \le q_4\|\tilde{\eta}\|\|\dot{\eta}\| \le c\sqrt{V_a} \tag{74}$$

where $c = q_4 d_{\max}(2/\lambda_o)^{1/2}$ and $\lambda_o = \lambda_{\min}(\mathcal{H})$.

The Lyapunov function increment is bounded by

$$\dot{V}_a \le \beta_2, \quad \beta_2 = c\left(\sqrt{V_a(t_s)} + \frac{c}{2}u_D\right). \tag{75}$$

For a unified analysis, the worst case upper bound is

$$\beta_2^{(\mathrm{u})} = q_4 d_{\max}\sqrt{\frac{2}{\lambda_o}}\sqrt{V_a(t_0) + \frac{\lambda_o q_4^2 d_{\max}^2}{2q_3^2}} + \frac{q_4^2 d_{\max}^2}{\lambda_o}u_{D,\max}. \tag{76}$$

Let $T_{sum}(t)$ and $D_{sum}(t)$ denote the total durations of secure and DoS intervals up to time t. The ratio between the two intervals satisfies $T_{sum}(t)/(T_{sum}(t) + D_{sum}(t)) \ge \rho, \quad 0 < \rho < 1$. Combining (73) and (75) yields

$$V_a(t) \le \gamma_1 e^{-\gamma_2(t - t_0)} V_a(t_0) + \gamma_3 \tag{77}$$

where the positive constants $\gamma_1, \gamma_2, \gamma_3 > 0$ depend on the secure/attack time ratio and are defined as $\gamma_1 = 1, \gamma_2 = \rho q_3/q_2, \gamma_3 = (\beta_2^{\mathrm{u}} q_2(1-\rho))/(\rho q_3) + q_2 q_4^2 d_{max}^2/(2q_3)$. ■

The detailed algorithm is summarized in Algorithm 1. The execution flow is iterative. First, the leader computes its policy by anticipating follower responses. Subsequently, followers utilize a consensus-based estimator to reconstruct missing data during DoS attacks. Finally, all agents update their AC networks to minimize Hamiltonians, ensuring convergence to the SNE.

Algorithm 1 Hierarchical Stackelberg–Nash AC Control Under Aperiodic DoS Attacks

Input: Initial states $\eta_{0,i}(0), \nu_{0,i}(0)$; d_{ij}.
Output: Optimal policies $\tau_0^*(t)$ and $\tau_i^*(t)$.
Initialize: Weights $\hat{W}_{a0}, \hat{W}_{c0}, \hat{W}_{ai}, \hat{W}_{ci}$ and $\hat{\eta}_i(0)$.
for $t = 0, \Delta t, \ldots, T$ **do**
  Step 1: Leader Update (Anticipatory Planning)
  Compute $e_0(t)$ and optimal anticipatory control $\tau_0(t)$ via (41)–(50);
  Update $\hat{W}_{c0}$ and $\hat{W}_{a0}$ via (49) and (51);
  Step 2: Follower Update (Active Reconstruction)
  for each follower $i \in \{1, \ldots, N\}$ **do**
    Update $\hat{\eta}_i(t)$ via (3) to compensate for DoS attacks;
    Compute tracking error $e_i(t)$ via (15) and control $\tau_i(t)$ via (31)–(33);
    Update weights $\hat{W}_{ci}$ and $\hat{W}_{ai}$ via (32) and (34);
  end for
end for
Return Optimal control trajectories $\tau_0^*(t)$ and $\tau_i^*(t)$.

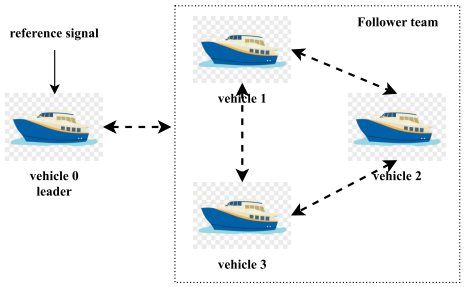

Fig. 2. Communication graph.

V. Simulation

A. Simulation Setup

USV model parameters are adopted from [9]: inertial matrix elements are $m_{11} = 25.8, m_{22} = 33.8, m_{33} = 2.76$, and $m_{23} = m_{32} = 1.0948$ (others zero). Hydrodynamic damping terms D_{mn} are: $D_{11} = 0.7225 + 1.3274|u_i| + 5.8664u_i^2$, $D_{22} = 0.8612 + 36.2823|v_i| + 0.805|r_i|$, $D_{33} = 1.9 - 0.08|v_i| + 0.75|r_i|$, $D_{23} = -0.1079 + 0.845|v_i| + 3.45|r_i|$, and $D_{32} = -0.1052 - 5.0437|v_i| - 0.13|r_i|$. Disturbances are set as $\omega_i(t) = [0.17\sin(0.03t), 0.15\sin(0.12t), 0.06\cos(0.42t)]^{\mathrm{T}}$. Simulation parameters are set as follows. Four USVs follow a reference trajectory $\eta_r(t) = [6 + 15\sin(0.0005t), 6 + 15\cos(0.0005t), 0]^T$ with the graph in Fig. 2. Cost parameters are $c_i = b_i = 0.1$ and $Q_i = P_i = I_3, i \in \mathcal{V}$. For the simulation, we consider a total of $n(0,t) = 15$ DoS attacks occurring over the time interval $[0,t]$, with each attack having a maximum duration of $u_n = 5$ s. Desired offsets are $d_{01} = [-5,5,0]^T, d_{02} = [0,10,0]^T, d_{03} = [5,5,0]^T, d_{12} = [5,5,0]^T, d_{13} = [10,0,0]^T$, and $d_{23} = [5,-5,0]^T$. Initial states are $\eta_0(0) = [6,21,0]^T, \eta_1(0) = [11,16,0]^T, \eta_2(0) = [6,11,0]^T, \eta_3(0) = [1,16,0]^T$ with $\nu_i(0) = [0,0,0]^T$. Optimized RL networks employ 16 nodes with centers ς_i evenly distributed in $[-8,8]$, using radial basis function (RBF) as the activation basis function. For the leader, design parameters are set as $k_0 = 13$, $z_0 = 8$, $\kappa_{a0} = 10$, and $\kappa_{c0} = 8$, with initial weights $\hat{W}_{c0}(0) = [0.58]_{16\times 3}$ and $\hat{W}_{a0}(0) = [0.38]_{16\times 3}$. For followers ($i \in \mathcal{O}$), parameters are $k_i = 8$, $z_i = 4$, $\kappa_{ai} = 15$, and $\kappa_{ci} = 12$. Initial critic weights $\hat{W}_{ci}(0)$ for $i = 1,2,3$ are $[0.45, 0.47, 0.49]_{16\times 3}$, and actor weights $\hat{W}_{ai}(0)$ are $[0.65, 0.67, 0.69]_{16\times 3}$.

Remark 6: The gain matrix μ balances convergence speed against noise sensitivity. We prioritize rapid reconstruction during secure intervals (T_n) to mitigate DoS impacts. While high gains risk noise amplification, robustness is theoretically guaranteed by the ISS property in Theorem 3.

Remark 7 (Parameter Sensitivity): While satisfying (35) ensures stability, higher gains (k_i, z_i) accelerate convergence but are practically constrained by actuator saturation. Simulation results indicate that selecting $k_i \in [5,20]$ and $z_i \in [4,15]$ offers a reliable tradeoff between tracking precision and feasible control effort.

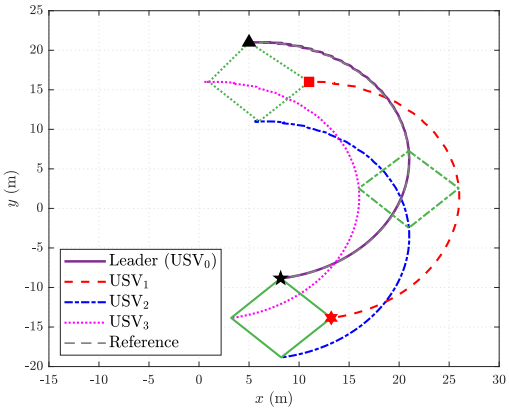

Fig. 3. Actual trajectories of four USVs under aperiodic DoS attacks.

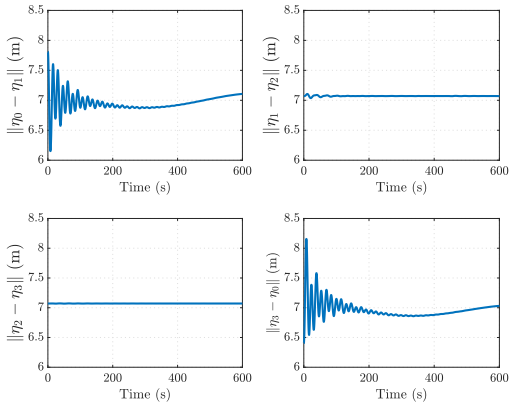

Fig. 4. Distances between the USVs.

B. Simulation Results

By utilizing Algorithm 1, we obtained the simulation results presented in Figs. 3–5. As illustrated in Fig. 3, all USVs successfully achieve the desired circular formation under aperiodic DoS attacks. As illustrated in Fig. 4, the intervehicle distances exhibit distinct transient behaviors driven by initial conditions. The trajectory of d_{23} (between Followers 2 and 3) is notably smooth because their initial positions perfectly satisfy the formation geometry ($\|\eta_2(0) - \eta_3(0)\| \approx \|d_{23}\| \approx 7.07$ m), resulting in zero initial error. Ultimately, the steady

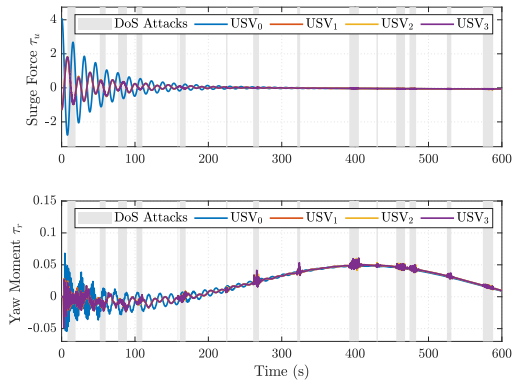

Fig. 5. Time evolution of control inputs (τ_u and τ_r) for the USV formation.

TABLE I
QUANTITATIVE PERFORMANCE METRICS OF FORMATION CONTROL

Agent	Steady-State MSE (m^2)	Max Transient Error (m)
Leader (USV0)	0.0014	1.32
USV1	0.0066	1.41
USV2	0.0066	1.45
USV3	0.0066	1.45

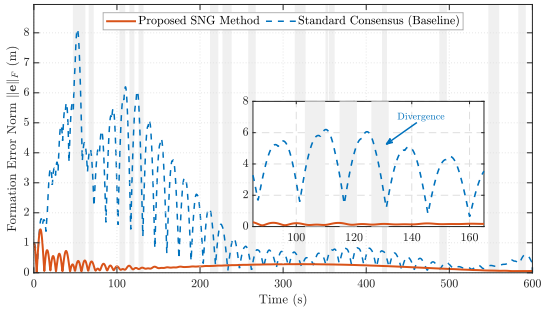

Fig. 6. Comparison of formation error norms $\|\mathbf{e}\|_F$ under DoS attacks.

convergence of all distances validates the achievement of the SNE. Controller feasibility and smoothness are confirmed in Fig. 5. The convergence of both the leader's and followers' policies demonstrates that the system reaches the SNE derived in Theorem 1. To quantitatively validate the method, Table I summarizes key performance metrics. The extremely low steady-state MSE (0.0014–0.0066 m^2) confirms high tracking precision, while the constrained maximum transient errors (1.32–1.45 m) verify system safety and bounded overshoot.

C. Comparative Analysis

To validate the proposed method's superiority, we compare it with a standard DC baseline under identical DoS attacks. As demonstrated in Fig. 6, the DC baseline diverges (error $\approx$ 8 m) as its reactive logic fails without neighbor data. In contrast, the proposed method maintains stability (error <1.5 m), attributed to the synergy of SNG-based hierarchical guidance and the consensus-based estimator, which actively reconstructs missing information to compensate for communication blackouts.

D. Robustness and Stability Boundary Analysis

To verify the system's operational limits, we first conducted stress tests on initial conditions and dynamic tasks. As detailed in Table II, the system exhibits global asymptotic stability, converging to a minimal steady-state error (≈0.06 m) even under large initial deviations (±50 m). Moreover, transient errors during dynamic reconfiguration are strictly suppressed (<0.3 m), ensuring safe maneuvering. Table III validates the constraints of Theorem 3. The system remains stable near the duration bound (u_n = 35 s) but diverges when violating dwell-time conditions (u_n = 80 s). Additionally, excessive frequency (n = 30) marks the operational limit where performance degrades.

TABLE II
QUANTITATIVE ROBUSTNESS METRICS: ERROR SUPPRESSION AND CONTROL EFFICIENCY

Scenario	Detailed Condition Parameter Setting	Max Transient Error [m]	Steady-State Error (SSE) [m]	Avg.Control Input
Initial Deviation	Pos. ±10 m, Yaw ±90°	15.10	0.060	1.55
	Pos. ±30 m, Yaw ±90°	43.00	0.060	2.79
	Pos. ±50 m, Yaw ±90°	71.30	0.060	4.76
Dynamic Spacing	Expansion (7m → 8m)	0.22	0.163	0.45
	Contraction (8m → 7m)	0.19	0.033	0.32

TABLE III
STABILITY BOUNDARY ANALYSIS: DOS STRESS TEST

Test Type	Parameter Setting	Mean RMSE [m]	Peak Error [m]
Attack Duration	Short (u_n = 2 s)	0.030	0.079
	Critical (u_n = 35 s)	0.317	0.955
	Extreme (u_n = 80 s)	0.594	1.702
Attack Frequency	Low (n = 5)	0.156	0.691
	Medium (n = 15)	0.292	0.860
	High (n = 30)	0.424	1.462

VI. CONCLUSION

This article proposes a hierarchical Stackelberg–Nash RL framework for the distributed formation control of USVs subject to aperiodic DoS attacks. Rigorous Lyapunov stability analysis is conducted to verify the SGUUB stability of the closed-loop system and the ISS of the resilient state estimator, which guarantees the system's convergence to the SNE. Simulation results validate the proposed framework's effectiveness in achieving accurate formation tracking, bounded state estimation errors, and strong resilience against aperiodic DoS attacks. However, the continuous online learning of the AC neural networks imposes a higher computational burden. Future work will focus on optimizing computational efficiency, extending the framework to multileader USV systems, and investigating resilience against deception attacks by integrating advanced detection mechanisms.

REFERENCES

[1] S. Xue, W. Zhang, B. Luo, and D. Liu, "Integral reinforcement learning-based dynamic event-triggered nonzero-sum games of USVs," *IEEE Trans. Cybern.*, vol. 55, no. 4, pp. 1706–1716, Apr. 2025.

[2] Z. Zhang, B. Huang, B. Zhou, and J. Miao, "An anti-congestion interaction mechanism based leaderless formation control for unmanned surface vehicles with communication interruptions," *IEEE Trans. Veh. Technol.*, vol. 73, no. 11, pp. 16964–16976, Nov. 2024.

[3] W. Gan, X. Qu, D. Song, and P. Yao, "Multi-USV cooperative chasing strategy based on obstacles assistance and deep reinforcement learning," *IEEE Trans. Autom. Sci. Eng. (from July 2004)*, vol. 21, no. 4, pp. 5895–5910, Oct. 2024.

[4] B. Liu, H.-T. Zhang, H. Meng, D. Fu, and H. Su, "Scanning-chain formation control for multiple unmanned surface vessels to pass through water channels," *IEEE Trans. Cybern.*, vol. 52, no. 3, pp. 1850–1861, Mar. 2022.

[5] G. Lyu, Z. Peng, and J. Wang, "Safety-critical receding-horizon planning and formation control of autonomous surface vehicles via collaborative neurodynamic optimization," *IEEE Trans. Cybern.*, vol. 54, no. 12, pp. 7236–7247, Dec. 2024.

[6] J. Liu, Z. Liu, Y. Li, E. Tian, and C. Peng, "Privacy-protected formation control for unmanned surface vehicles: A neural predictor-based non-cooperative game framework," *IEEE Trans. Veh. Technol.*, early access, 2026, doi: 10.1109/TVT.2026.3651837.

[7] J. Liu, S. Mao, L. Zha, E. Tian, and C. Peng, "Privacy-preserving and collision-free formation control for heterogeneous multiagent systems," *IEEE Trans. Consum. Electron.*, early access, Jan. 24, 2026, doi: 10.1109/TCE.2026.3667772.

[8] C. Qin, X. Qiao, J. Wang, D. Zhang, Y. Hou, and S. Hu, "Barrier-critic adaptive robust control of nonzero-sum differential games for uncertain nonlinear systems with state constraints," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 54, no. 1, pp. 50–63, Jan. 2024.

[9] L. Zha, J. Miao, J. Liu, E. Tian, and C. Peng, "Neural predictor-based formation control for unmanned surface vehicles under aperiodic DoS attacks: A noncooperative game approach," *IEEE Trans. Veh. Technol.*, vol. 74, no. 10, pp. 15013–15025, Oct. 2025.

[10] J. Qin, M. Li, Y. Shi, Q. Ma, and W. X. Zheng, "Optimal synchronization control of multiagent systems with input saturation via off-policy reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 85–96, Jan. 2019.

[11] P. Zhang and Y. Zhang, "Two-step Stackelberg approach for the two weak pursuers and one strong evader closed-loop game," *IEEE Trans. Autom. Control*, vol. 69, no. 2, pp. 1309–1315, Feb. 2024.

[12] T. An, B. Dong, H. Yan, L. Liu, and B. Ma, "Dynamic event-triggered strategy-based optimal control of modular robot manipulator: A multi-player nonzero-sum game perspective," *IEEE Trans. Cybern.*, vol. 54, no. 12, pp. 7514–7526, Dec. 2024.

[13] Q. Kang, D. Yu, and C. L. P. Chen, "Fuzzy weighted regularization for fractional-order nonlinear multiagent systems under Stackelberg–Nash game," *IEEE Trans. Fuzzy Syst.*, vol. 33, no. 9, pp. 3345–3359, Sep. 2025.

[14] H. Song and Z. Wang, "Stackelberg-game-based dynamic event-triggered distributed model predictive control approach for multiagent systems under joint deception attacks," *IEEE Internet Things J.*, vol. 12, no. 13, pp. 25651–25663, Jul. 2025.

[15] M. Lin, B. Zhao, and D. Liu, "Event-triggered robust adaptive dynamic programming for multiplayer Stackelberg–Nash games of uncertain nonlinear systems," *IEEE Trans. Cybern.*, vol. 54, no. 1, pp. 273–286, Jan. 2024.

[16] H. Zhang, Y. Zhang, X. Zhao, and C.-Y. Su, "Distributed adaptive dynamic programming for consensus control of multiagent systems within hierarchical Stackelberg–Nash game framework," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 55, no. 6, pp. 4286–4300, Jun. 2025.

[17] X. Guan, Y. Hu, and K. Peng, "Bayesian-Stackelberg-game-based finite-time sliding mode fault-tolerant secure control for cyber–physical systems under jamming attacks and multiple physical faults," *IEEE Trans. Cybern.*, vol. 55, no. 10, pp. 4710–4723, Oct. 2025.

[18] B.-C. Wang, J. Xu, H. Zhang, and Y. Liang, "Linear quadratic mean field Stackelberg games: Open-loop and feedback solutions," *IEEE Trans. Cybern.*, vol. 55, no. 11, pp. 5318–5331, Nov. 2025.

[19] J. Long, D. Yu, G. Wen, L. Li, Z. Wang, and C. L. P. Chen, "Game-based backstepping design for strict-feedback nonlinear multi-agent systems based on reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 1, pp. 817–830, Jan. 2024.

[20] D. Yu, S. Ma, Y.-J. Liu, Z. Wang, and C. L. P. Chen, "Finite-time adaptive fuzzy backstepping control for quadrotor UAV with stochastic disturbance," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1335–1345, Apr. 2024.

[21] D. Zhang, Q. Yuan, L. Meng, R. Xia, W. Liu, and C. Qin, "Reinforcement learning for single-agent to multi-agent systems: From basic theory to industrial application progress, a survey," *Artif. Intell. Rev.*, vol. 59, no. 2, p. 46, Nov. 2025.

[22] C. Qin, S. Hou, M. Pang, Z. Wang, and D. Zhang, "Reinforcement learning-based secure tracking control for nonlinear interconnected systems: An event-triggered solution approach," *Eng. Appl. Artif. Intell.*, vol. 161, Dec. 2025, Art. no. 112243.

[23] R. Lu et al., "Adaptive optimal surrounding control of multiple unmanned surface vessels via actor-critic reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 7, pp. 12173–12186, Jul. 2025.

[24] K. Yu, Y. Li, M. Lv, and S. Tong, "Distributed optimal formation control of multiple unmanned surface vehicles with Stackelberg differential graphical game," *IEEE Trans. Artif. Intell.*, vol. 5, no. 8, pp. 4058–4073, Aug. 2024.

[25] L. Chen, S.-L. Dai, and C. Dong, "Adaptive optimal tracking control of an underactuated surface vessel using actor–critic reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 6, pp. 7520–7533, Jun. 2024.

[26] P. Ning, L. Duan, and C. Hua, "Optimal tracking control of uncertain nonlinear systems using simplified reinforcement learning," *IEEE Trans. Cybern.*, early access, Jan. 13, 2026, doi: 10.1109/TCYB.2026.3650813.

[27] J. Wu, C. Peng, J. Zhang, and E. Tian, "A sampled-data-based secure control approach for networked control systems under random DoS attacks," *IEEE Trans. Cybern.*, vol. 54, no. 8, pp. 4841–4851, Aug. 2024.

[28] J. Liu, E. Ma, L. Zha, and E. Tian, "Distributed energy management for microgrids: When privacy preservation meets multiple mechanisms," *IEEE Trans. Consum. Electron.*, early access, Jan. 6, 2026, doi: 10.1109/TCE.2026.3651619.

[29] Z. Gu, T. Yin, and Z. Ding, "Path tracking control of autonomous vehicles subject to deception attacks via a learning-based event-triggered mechanism," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 12, pp. 5644–5653, Dec. 2021.

[30] L. Li, Y. Li, J. Lian, and M. Li, "Active security control for switched systems under deception and DoS attacks based on two-tier Stackelberg game," *IEEE Trans. Cybern.*, vol. 55, no. 7, pp. 3251–3261, Jul. 2025.

[31] J. Zhang, D. Yang, W. Li, H. Zhang, G. Li, and P. Gu, "Resilient output control of multiagent systems with DoS attacks and actuator faults: Fully distributed event-triggered approach," *IEEE Trans. Cybern.*, vol. 54, no. 12, pp. 7681–7690, Dec. 2024.

[32] Y. Xu, K. Li, G. Dong, Y. Li, X. Chen, and D. Xie, "Resilient cooperative optimal output regulation control for nonlinear multiagent systems," *IEEE Trans. Cybern.*, early access, Jan. 1, 2026, doi: 10.1109/TCYB.2025.3645097.

[33] F. Yang, J. Liu, Y. Du, and C. Yan, "Predefined-time secure distributed energy management for microgrids against DoS attack based on dynamic event-triggered approach," *IEEE Trans. Autom. Sci. Eng.*, vol. 22, pp. 12791–12801, 2025.

[34] Y. Ma, X. Qi, Z. Li, S. Hu, and M. Á. Sotelo, "Resilient control for networked unmanned surface vehicles with dynamic event-triggered mechanism under aperiodic DoS attacks," *IEEE Trans. Veh. Technol.*, vol. 73, no. 4, pp. 5824–5833, Apr. 2024.

[35] X. Chen, S. Hu, W. Zhang, X. Xie, and D. Yue, "DoS-resilient time varying estimators and controllers co-design for NCSs under sensor and actuator attacks," *IEEE Trans. Inf. Forensics Security*, vol. 20, pp. 7498–7511, 2025, doi: 10.1109/TIFS.2025.3576587.

[36] J. Zhang, Y. Zuo, and S. Tong, "Dynamic event-triggered fuzzy adaptive resilient consensus control for nonlinear MASs under DoS attacks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 54, no. 11, pp. 6683–6693, Nov. 2024.

[37] Y. Su, F. Teng, T. Li, H. Liang, and C. L. Philip Chen, "Fuzzy-based optimal control for an underactuated surface vessel with user-specified performance," *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 5, pp. 7036–7050, May 2025.

[38] F. Yang, Z. Gong, Q. Wei, and Y. Lei, "Secure containment control for multi-UAV systems by fixed-time convergent reinforcement learning," *IEEE Trans. Cybern.*, vol. 55, no. 4, pp. 1981–1994, Apr. 2025.

[39] F. Yang, Y. Fu, and Y. Du, "Event-based fixed-time resilient distributed secondary control in AC microgrids against DoS attacks," *IEEE Trans. Smart Grid*, early access, Dec. 22, 2025, doi: 10.1109/TSG.2025.3646998.

[40] G. Zhang, S. Yin, J. Li, W. Zhang, and W. Zhang, "Game-based event-triggered control for unmanned surface vehicle: Algorithm design and harbor experiment," *IEEE Trans. Cybern.*, vol. 55, no. 6, pp. 2729–2741, Jun. 2025.

[41] C. Qin, M. Pang, Z. Wang, S. Hou, and D. Zhang, "Observer-based fault tolerant control design for saturated nonlinear systems with full state constraints via a novel event-triggered mechanism," *Eng. Appl. Artif. Intell.*, vol. 161, Dec. 2025, Art. no. 112221.

[42] M. Ye and G. Hu, "Distributed Nash equilibrium seeking by a consensus based approach," *IEEE Trans. Autom. Control*, vol. 62, no. 9, pp. 4811–4818, Sep. 2017.

[43] K. G. Vamvoudakis and F. L. Lewis, "Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878–888, 2010.
[44] G. Wen, W. Hao, W. Feng, and K. Gao, "Optimized backstepping tracking control using reinforcement learning for quadrotor unmanned aerial vehicle system," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 8, pp. 5004–5015, Aug. 2022.
[45] Q. Wang, Z. Duan, J. Wang, and G. Chen, "LQ synchronization of discrete-time multiagent systems: A distributed optimization approach," *IEEE Trans. Autom. Control*, vol. 64, no. 12, pp. 5183–5190, Dec. 2019.

Jinliang Liu (Senior Member, IEEE) received the Ph.D. degree in control theory and control engineering from Donghua University, Shanghai, China, in 2011.

He was a Lecturer from June 2011 to August 2013, an Associate Professor from August 2013 to July 2019, and a Professor from July 2019 to May 2023 with the School of Information Engineering, Nanjing University of Finance and Economics, Nanjing, China. He held a post-doctoral position at the School of Automation, Southeast University, Nanjing, from December 2013 to June 2016. He was a Visiting Researcher/Scholar with the Department of Mechanical Engineering, The University of Hong Kong, Hong Kong, from October 2016 to 2017. He was a Visiting Scholar with the Department of Electrical Engineering, Yeungnam University, Gyeongsan, South Korea, from November 2017 to January 2018. Since June 2023, he has been a Professor with the School of Computer Science, Nanjing University of Information Science and Technology, Nanjing. His research interests include cyber-physical systems, autonomous systems, privacy preservation, and game theory.

Zihan Zhang received the B.S. degree in computer science and technology from Huaiyin Normal University, Huaian, China, in 2024. She is currently pursuing the M.S. degree in computer science and technology with the College of Computer Science, Nanjing University of Information Science and Technology, Nanjing, China.

Her research interests include formation control, networked control systems, and neural networks.

Engang Tian (Senior Member, IEEE) received the B.S. degree in mathematics from Shandong Normal University, Jinan, China, in 2002, the M.Sc. degree in operations research and cybernetics from Nanjing Normal University, Nanjing, China, in 2005, and the Ph.D. degree in control theory and control engineering from Donghua University, Shanghai, China, in 2008.

From 2011 to 2012, he was a Post-Doctoral Research Fellow with The Hong Kong Polytechnic University, Hong Kong. From 2015 to 2016, he was a Visiting Scholar with the Department of Information Systems and Computing, Brunel University London, Uxbridge, U.K. From 2008 to 2018, he was an Associate Professor and then a Professor with the School of Electrical and Automation Engineering, Nanjing Normal University. In 2018, he was appointed as an Eastern Scholar at the Municipal Commission of Education, Shanghai, and joined the University of Shanghai for Science and Technology, Shanghai, where he is currently a Professor with the School of Optical-Electrical and Computer Engineering. He has published more than 100 articles in refereed international journals. His research interests include networked control systems, cyberattacks, nonlinear stochastic control, and filtering.

Chen Peng (Senior Member, IEEE) received the B.S. and M.S. degrees in coal preparation and the Ph.D. degree in control theory and control engineering from Chinese University of Mining Technology, Xuzhou, China, in 1996, 1999, and 2002, respectively.

He held research positions at Nanjing Normal University, Nanjing, China, The University of Hong Kong, Hong Kong, Queensland University of Technology, Brisbane City, QLD, Australia, and Central Queensland University, Rockhampton, QLD, Australia, from 2002 to 2012. He is currently a Professor with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China. His research interests include networked control systems, multi-USV systems, power systems, and interconnected systems.

Dr. Peng is an Associate Editor for a number of international journals, including the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, *Information Sciences*, and *Transactions of the Institute of Measurement and Control*. He was named as a Highly Cited Researcher from 2020 to 2022 by Clarivate Analytics.

Jinde Cao (Fellow, IEEE) received the B.S. degree in mathematics from Anhui Normal University, Wuhu, China, in 1986, the M.S. degree from Yunnan University, Kunming, China, in 1989, and the Ph.D. degree in applied mathematics from Sichuan University, Chengdu, China, in 1998.

He was a Post-Doctoral Research Fellow with the Department of Automation and Computer-Aided Engineering, The Chinese University of Hong Kong, Hong Kong, China, from 2001 to 2002. He is the Endowed Chair Professor and the Dean of the School of Mathematics, Southeast University (SEU), Nanjing, China. He directs the SEU Research Center for Complex Systems and Network Sciences, Nanjing, SEU-Jiangsu National Center for Applied Mathematics, Nanjing, and Jiangsu Provincial Key Laboratory of Networked Collective Intelligence Nanjing. He is also an Honorable Professor with the Institute of Mathematics and Mathematical Modeling, Almaty, Kazakhstan.

Prof. Cao is elected as a member of Russian Academy of Sciences, the Academia Europaea (Academy of Europe), European Academy of Sciences and Arts, and the Lithuanian Academy of Sciences and a fellow of African Academy of Sciences and Pakistan Academy of Sciences. He received the National Innovation Award of China, the Obada Prize, and the Clarivate Highly Cited Researcher Award.

Tingwen Huang (Fellow, IEEE) received the B.S. degree in mathematics from Southwest University, Chongqing, China, in 1990, the M.S. degree in mathematics from Sichuan University, Chengdu, China, in 1993, and the Ph.D. degree in mathematics from Texas A&M University, College Station, TX, USA, in 2002.

He has been a Professor with the Faculty of Computer Science and Control Engineering, Shenzhen University of Advanced Technology, Shenzhen, China, since 2024. He has expertise in optimization and control, machine learning, privacy protection, and neural networks.

Prof. Huang is currently serving as the Editor-in-Chief for *IEEE Systems, Man, and Cybernetics Magazine* and an Editorial Board Member for IEEE TRANSACTIONS ON INDUSTRIAL CYBER-PHYSICAL SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE, IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE, and *Neural Networks*.