Event-Driven Intelligent Dynamic Positioning for Networked Unmanned Marine Vehicles Within Reinforcement Learning Framework

Jian Liu, *Member, IEEE*, Jiachen Ke, Jinliang Liu, *Member, IEEE*, Xiangpeng Xie, *Senior Member, IEEE*, and Engang Tian, *Member, IEEE*

***Abstract*—This correspondence develops an intelligent dynamic positioning regulation strategy for nonlinear networked unmanned marine vehicle (NUMV) with unmatched interference. Distinct from the existing Takagi-Sugeno fuzzy approach, reinforcement learning (RL) framework with simplified parameters is utilized to pursue the regulation optimality. On basis of adjustable weight and threshold function, we provide a novel adaptive weight event-driven scheme (AWEDS) to achieve the efficient information transmission. Furthermore, an AWEDS-boosted critic-sole iterative RL algorithm is proposed for implementing the approximately optimal control policy. Afterwards, the nonlinear NUMV and the critic weight estimation error are ensured to be uniformly ultimately bounded. The feasibility and validity of the proposed learning algorithm are eventually illustrated by a comparative experiment.**

***Index Terms*—Reinforcement learning, networked unmanned marine vehicles, adaptive weight event-driven scheme, adaptive critic learning, unmatched interference.**

I. Introduction

Owing to the conjoint deployment of internet medium and autonomous intelligent technology, networked unmanned marine vehicle (NUMV) has been utilized in emergency aid, hydrological detection and other assignments [1], [2]. In realistic implementations, complicated ocean circumstance unpredictably affects the motion regulation of nonlinear NUMV. Towards this end, plentiful effective control schemes for NUMV have been proposed at the current research phase. For instance, Wang et al. [3] devised the fault-tolerant tracking control scheme for NUMV on basis of integral sliding mode method. In [4], an event-driven (ED) secure positioning control policy was presented for nonlinear NUMV against deception attacks and denial of service attacks. In the aforementioned researches, the nonlinearities of NUMV were tackled via Takagi-Sugeno (T-S) fuzzy model [5], which involves the larger number of computational parameters. Undoubtedly, it will increase the complexity of control performance analysis and operation. Thus, there is an urgent requirement for the advanced techniques to efficiently handle the nonlinearities.

As a target-oriented instrument to treat nonlinear dynamical systems, reinforcement learning (RL) presents a unified framework to obtain the approximate solution of Hamilton-Jacobi-Isaacs equation (HJIE). Under RL framework, the numerical solutions of HJIE for different control issue have been acquired in extensive researches [6], [7], [8]. To mention a few, the authors in [6] investigated the solution of ED-assisted HJIE for unknown nonlinear systems by utilizing RL approach. In [7], Zhao et al. derived the consensus-related HJIE for multiagent systems and obtained the corresponding approximate solution. From the aforementioned observations, the existing RL technique mainly employs the actor-critic dual networks (ACDNs) and the critic-sole network. Originally, the ACDNs [9], [10], [11] were exploited to iteratively acquire the near-optimal control strategy. For sake of enhancing the computational efficiency, the structure of ACDNs can be tactfully simplified as a single critic network [12]. It not only diminishes the approximation error caused by the actor network, but also reduces the computational complexity in weight training process. Due to the merits of critic-sole network, a series of efficacious control policies have been provided for various dynamics, such as interconnected systems [13], multi-agent systems [14], robotic systems [15], aircraft systems [16] and so on. However, the critic network-assisted RL framework for nonlinear NUMV is still lacking. In order to pursue the control optimality and lessen the analysis complexity, it is vital to investigate the associated control strategy under the critic-sole network.

Attributed to the compact integration of wireless network and remote control technology, the topic of restricted communication capacity has obtained extensive concerns. In response to communication and computation burden, a significant quantity of ED mechanisms [17], [18] were proposed on basis of the different triggered conditions. Comprehensively taking control optimality into consideration, several RL-based ED control algorithms have been presented in [19], [20]. It is worth pointing out that these triggered conditions mainly concentrate on ensuring the stabilization of nonlinear systems. Furthermore, Liu et al. [21] put forward a decentralized RL algorithm by virtue of adaptive ED scheme (AEDS), which dedicates to reduce the superfluous information dispatch. Nevertheless, the excellent operation of AEDS requires the appropriate predetermination for the weight matrix in triggered condition. On the contrary, it is uncomplicated to cause the imbalance between the control and triggered performance. Hence, we endeavor to develop a novel ED scheme to conquer the mentioned deficiency.

Inspired by the aforesaid contents, a RL-boosted ED dynamic positioning control strategy is presented for nonlinear NUMV with unmatched interference. The central innovations of this correspondence are embodied as follows:

1) A novel adaptive weight ED scheme (AWEDS) is proposed to efficiently regulate the signal transmission. Distinguishing from AEDS adopted in [21], the weight matrix in AWEDS can adaptively change according to the contribution of each state component. Under this mechanism, the effective balance of control and triggered performance can be attained.
2) In comparison with T-S fuzzy model [3], [4], the nonlinearities of NUMV subject to unmatched interference are appropriately

Manuscript received 16 January 2024; revised 4 April 2024; accepted 24 May 2024. Date of publication 28 May 2024; date of current version 17 October 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62001210, Grant 62373252, Grant 61973152, and Grant 62373196, in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20211290, and in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant KYCX23_1887. The review of this article was coordinated by Prof. Claudia Campolo. *(Corresponding author: Jinliang Liu.)*

Jian Liu and Jiachen Ke are with the College of Information Engineering, Nanjing University of Finance and Economics, Nanjing 210023, China (e-mail: by.liujian@gmail.com; 2212602143@qq.com).

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

Xiangpeng Xie is with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China (e-mail: xiexiangpeng1953@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).

Digital Object Identifier 10.1109/TVT.2024.3406347

0018-9545 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Nanjing University of Information Science and Technology. Downloaded on October 19,2024 at 09:12:33 UTC from IEEE Xplore. Restrictions apply.

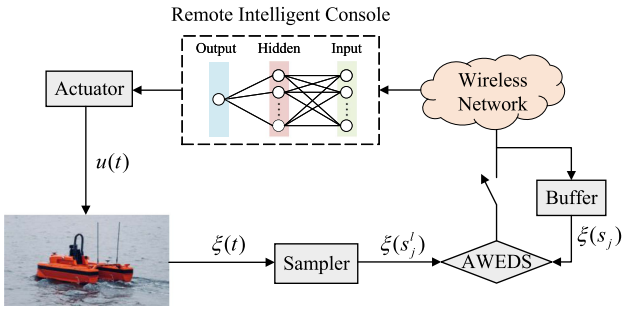

Fig. 1. The NUMV with remote intelligent console.

tackled within RL framework. Since the utilization of intricate parameters in T-S fuzzy model has been avoided, the complexity of control synthesis and analysis can be further relieved.

3) In [9], the intelligent control problem for NUMV were addressed on basis of ACDNs and time-driven scheme, which requires a significant amount of computing resources. To boost the operational efficiency, an AWEDS-enhanced critic-sole iterative RL (CIRL) algorithm is provided to obtain the numerical solution of HJIE.

II. Problem Formulation and Conversion

A. The Nonlinear NUMV

In Fig. 1, the configuration of NUMV with remote intelligent console is depicted. Focused on the motion in surge, sway and yaw, the following body-fixed NUMV dynamics [22] is presented:

$$\begin{cases} \dot{\eta}(t) = \mathcal{R}(\lambda(t))\varsigma(t) \\ U\dot{\varsigma}(t) + V\varsigma(t) + L\eta(t) = u(t) \end{cases} \tag{1}$$

where $\varsigma(t) = [\varsigma_1^T(t)\ \varsigma_2^T(t)\ \varsigma_3^T(t)]^T$ with the components $\varsigma_1^T(t), \varsigma_2^T(t)$, and $\varsigma_3^T(t)$ representing the velocity of surge, sway and yaw, respectively. $\eta(t) = [\bar{x}^T(t)\ \bar{y}^T(t)\ \lambda^T(t)]^T$ denotes the position and yaw angle. $u(t) = [u_1^T(t)\ u_2^T(t)\ u_3^T(t)]^T$ stands for the forces and moment. $U = U^T \in \mathbb{R}^{3\times3}$, $V \in \mathbb{R}^{3\times3}$ and $L \in \mathbb{R}^{3\times3}$ are the matrices of inertia, damping and mooring forces, respectively. $\mathcal{R}(\lambda(t)) = \text{diag}\{\mathcal{R}_1(\lambda(t)), 1\}$ with

$$\mathcal{R}_1(\lambda(t)) = \begin{bmatrix} \cos(\lambda(t)) & -\sin(\lambda(t)) \\ \sin(\lambda(t)) & \cos(\lambda(t)) \end{bmatrix}.$$

For brevity, the notation t will be omitted unless emphasized otherwise. Define $\xi = [\xi_1^T, \ldots, \xi_6^T]^T$ with $\xi_1 = \bar{x}$, $\xi_2 = \bar{y}$, $\xi_3 = \lambda$, $\xi_4 = \varsigma_1$, $\xi_5 = \varsigma_2$, and $\xi_6 = \varsigma_3$. Taking account of the unmatched ocean perturbation δ, the nonlinear NUMV (1) is transformed as

$$\dot{\xi} = h(\xi) + Gu + F\delta \tag{2}$$

where $F \in \mathbb{R}^{6\times3}$ denotes the interference dynamics and $F \neq G$. $h(\xi) \in \mathbb{R}^{6\times1}$ and $G \in \mathbb{R}^{6\times3}$ are described by

$$h(\xi) = \begin{bmatrix} \mathcal{R}(\lambda)\varsigma \\ -U^{-1}V\varsigma - U^{-1}L\eta \end{bmatrix},\ G = \begin{bmatrix} 0 \\ U^{-1} \end{bmatrix}.$$

B. Dynamic Positioning Control Problem

This correspondence aims to provide the dynamic positioning control policy, which assures the uniformly ultimately bounded (UUB) stability of nonlinear NUMV (2). Nevertheless, it is tough to promptly devise the corresponding controller due to the existence of nonlinearity and perturbation. To attain the aforementioned objective, the dynamic positioning control problem will be converted by constructing the performance function as follows:

$$\mathcal{V}(\xi) = \int_t^\infty r(\xi(\tau), u(\tau), \delta(\tau))d\tau \tag{3}$$

where $r(\xi, u, \delta) = \xi^T Q\xi + u^T Pu - \varpi^2\delta^T\delta$ is the utility function. $Q \in \mathbb{R}^{6\times6}$ and $P \in \mathbb{R}^{3\times3}$ denote symmetric positive definite matrices. $\varpi > 0$ serves as the predetermined interference attenuation level.

In order to assist the subsequent descriptions, the Hamiltonian can be stated as

$$\mathcal{H}(\mathcal{V}_\xi, \xi, u, \delta) = \mathcal{V}_\xi^T(h(\xi) + Gu + F\delta) + r(\xi, u, \delta) \tag{4}$$

in which $\mathcal{V}_\xi = \partial\mathcal{V}(\xi)/\partial\xi$ is the partial derivative of $\mathcal{V}(\xi)$.

In light of Bellman optimality principle, the associated regulation issue is transformed into acquiring the solution of HJIE: $\min_u \max_\delta \mathcal{H}(\mathcal{V}_\xi^*, \xi, u, \delta) = 0$, from which $\mathcal{V}_\xi^*$ denotes the partial derivative of optimal performance function $\mathcal{V}^*(\xi)$. As stated in [23], it is apparent to obtain the stationary conditions $\partial\mathcal{H}(\mathcal{V}_\xi^*, \xi, u, \delta)/\partial u = 0$ and $\partial\mathcal{H}(\mathcal{V}_\xi^*, \xi, u, \delta)/\partial\delta = 0$. On this basis, the optimal control strategy $u^*(\xi)$ and the worst perturbation $\delta^*(\xi)$ can be derived:

$$u^*(\xi) = -\frac{1}{2}P^{-1}G^T\mathcal{V}_\xi^*,\ \delta^*(\xi) = \frac{1}{2\varpi^2}F^T\mathcal{V}_\xi^*. \tag{5}$$

On account of $u^*(\xi)$ and $\delta^*(\xi)$, the aforementioned HJIE for NUMV (2) is acquired as

$$\mathcal{V}_\xi^{*T}(h(\xi) + Gu^*(\xi) + F\delta^*(\xi)) + r(\xi, u^*(\xi), \delta^*(\xi)) = 0 \tag{6}$$

where $\mathcal{V}_\xi^{*T}$ stands for $(\mathcal{V}_\xi^*)^T$, and the similar symbol will still be utilized in later analysis.

C. Description of AWEDS

To effectively decrease the network utilization, a novel AWEDS is proposed to comprehensively regulate the transmitted signals and weight matrix. To be concrete, the sampled information will be dispatched as long as the following condition is violated:

$$e_j^T\Psi(s_j^l)e_j \leq \mu(s_j^l)\xi^T(s_j^l)\Psi(s_j^l)\xi(s_j^l) \tag{7}$$

where $s_j^l = s_j + lT$ and s_j $(j = 0, 1, \ldots; l = 1, 2, \ldots)$ denote the current sampled time and the previous triggered time, respectively. T serves as the sampling period. $e_j = \xi(s_j^l) - \xi(s_j)$ represents the triggered error. For the convenience of subsequent demonstration, $\Psi(s_j^l)$ and $\xi(s_j)$ are replaced by $\Psi_{j,l}$ and ξ_{s_j}, respectively. $\Psi_{j,l} = \text{diag}\{\Psi_{j,l}^1, \ldots, \Psi_{j,l}^6\} \in \mathbb{R}^{6\times6}$ is the positive-definite adaptive weight to determine the contribution of the different state elements.

In addition, the adaptive threshold function $\mu(s_j^l)$ is obtained according to

$$\mu(s_j^l) = -(\underline{\mu} - \overline{\mu})e^{-\varepsilon e_j^T\Psi_{j,l}e_j} + \underline{\mu} \tag{8}$$

where $\underline{\mu}$ and $\overline{\mu}$ are the lower and upper bounds (LUBs) of $\mu(s_j^l)$. $\varepsilon > 0$ denotes the given constant to regulate the sensitivity of $\mu(s_j^l)$.

Moreover, the α-th component $\Psi_{j,l}^\alpha$ in the adaptive weight matrix $\Psi_{j,l}$ is designed as

$$\Psi_{j,l}^\alpha = -(\underline{\Psi}^\alpha - \overline{\Psi}^\alpha)\tanh\left(\frac{\rho_\alpha||\xi_\alpha(s_j^l)||}{\sum_{\varrho=1}^6||\xi_\varrho(s_j^l)||}\right) + \underline{\Psi}^\alpha \tag{9}$$

in which $\underline{\Psi}^\alpha$ and $\overline{\Psi}^\alpha$ denote the LUBs of $\Psi_{j,l}^\alpha$, $\alpha \in \{1, \ldots, 6\}$. ρ_α is a positive constant to govern the sensitivity of $\Psi_{j,l}^\alpha$.

Authorized licensed use limited to: Nanjing University of Information Science and Technology. Downloaded on October 19,2024 at 09:12:33 UTC from IEEE Xplore. Restrictions apply.

Then, the updating law of the triggered time s_{j+1} can be described as follows:

$$s_{j+1} = s_j + \min_{l\geq 1}\{lT| \text{ the condition (7) violates}\}. \tag{10}$$

On basis of the proposed AWEDS and zero-order holder technique, the optimal ED control policy $u^*(\xi_{s_j})$ can be elicited as

$$u^*(\xi(t)) = u^*(\xi_{s_j}) = -\frac{1}{2}P^{-1}G^T\mathcal{V}^*_{\xi_{s_j}},\ t\in[s_j, s_{j+1}). \tag{11}$$

D. Stability Analysis

As illustrated in [13], [21], the following essential assumptions are outlined to promote this analysis.

Assumption 1: There exist positive scalars G_M, F_M and $\mathcal{V}_{eM}$ such that $||G||\leq G_M$, $||F||\leq F_M$ and $||\mathcal{V}^*_\xi||\leq \mathcal{V}_{eM}$.

Assumption 2: $u^*(\xi)$ is presumed to be Lipschitz continuous, i.e., $||u^*(\xi)-u^*(\xi_{s_j})||\leq \mathcal{L}_{u^*}||e_j||$ with $\mathcal{L}_{u^*}>0$ being the given scalar.

Theorem 1: Under the proposed AWEDS (7), Assumptions 1 and 2, the optimal ED control law $u^*(\xi_{s_j})$ in (11) can drive the NUMV (2) with the worst perturbation $\delta^*(\xi)$ in (5) to be UUB stability when the following inequality satisfies:

$$\Pi = \lambda_{\min}(Q) - \overline{\mu}\mathcal{L}^2_{u^*}\lambda_{\max}(P)\frac{\max\{\overline{\Psi}^1,\ldots,\overline{\Psi}^6\}}{\min\{\underline{\Psi}^1,\ldots,\underline{\Psi}^6\}} > 0. \tag{12}$$

Proof: Select the Lyapunov function as $L=\mathcal{V}^*(\xi)$. According to (2), (5) and (6), the derivative of L is obtained:

$$\begin{aligned}
\dot{L} &= \mathcal{V}^{*T}_\xi(h(\xi)+Gu^*(\xi_{s_j})+F\delta^*(\xi))\\
&= \mathcal{V}^{*T}_\xi(h(\xi)+Gu^*(\xi)+F\delta^*(\xi))\\
&\quad + \mathcal{V}^{*T}_\xi G(u^*(\xi_{s_j})-u^*(\xi))\\
&= -\xi^TQ\xi - u^{*T}(\xi)Pu^*(\xi)+\varpi^2\delta^{*T}(\xi)\delta^*(\xi)\\
&\quad + 2u^{*T}(\xi)P(u^*(\xi)-u^*(\xi_{s_j}))\\
&= -\xi^TQ\xi + (u^*(\xi)-u^*(\xi_{s_j}))^TP(u^*(\xi)-u^*(\xi_{s_j}))\\
&\quad - u^{*T}(\xi_{s_j})Pu^*(\xi_{s_j}) + \varpi^2||\frac{1}{2\varpi^2}F^T\mathcal{V}^*_\xi||^2.
\end{aligned} \tag{13}$$

By adopting Assumptions 1 and 2, the expression (13) is derived as follows:

$$\dot{L}\leq -\lambda_{\min}(Q)||\xi||^2 + \lambda_{\max}(P)\mathcal{L}^2_{u^*}||e_j||^2 + \frac{F^2_M\mathcal{V}^2_{eM}}{4\varpi^2}. \tag{14}$$

When $t\in[s_j, s_{j+1})$, it is readily elicited that the inequality (7) will hold, which yields

$$||e_j||^2\leq \overline{\mu}\frac{\max\{\overline{\Psi}^1,\ldots,\overline{\Psi}^6\}}{\min\{\underline{\Psi}^1,\ldots,\underline{\Psi}^6\}}||\xi||^2. \tag{15}$$

Substituting (15) into (14), one has

$$\dot{L}\leq -\Pi||\xi||^2 + \frac{F^2_M\mathcal{V}^2_{eM}}{4\varpi^2}. \tag{16}$$

Under the condition (12), it is obvious that $\dot{L}<0$ can be satisfied if $||\xi||>\sqrt{(F^2_M\mathcal{V}^2_{eM})/(4\varpi^2\Pi)}$. Thus, the UUB stability of NUMV (2) subject to the worst disturbance $\delta^*(\xi)$ is assured by adopting the optimal ED control policy $u^*(\xi_{s_j})$. The proof of Theorem 1 has been completed. ■

III. AWEDS-Enhanced Learning Control Strategy

A. The CIRL Algorithm

With the aim of economizing the computational resources, a unique critic network is employed to approximate $\mathcal{V}^*(\xi)$. After that, the following formulas can be acquired:

$$\begin{cases}\mathcal{V}^*(\xi) = \Theta^T_c\vartheta_c(\xi)+\varpi_c(\xi)\\ \mathcal{V}^*_\xi = \nabla\vartheta^T_c(\xi)\Theta_c+\nabla\varpi_c(\xi)\end{cases} \tag{17}$$

in which $\Theta_c\in\mathbb{R}^{m_c}$ stands for the optimal critic weight. $\vartheta_c(\xi)\in\mathbb{R}^{m_c}$ represents the activation function. m_c denotes the quantity of neurons. $\varpi_c(\xi)$ is the residual error.

On the ground of (17), the ED control policy $u^*(\xi_{s_j})$ in (11) and the perturbation $\delta^*(\xi)$ in (5) are deduced as

$$\begin{cases}u^*(\xi_{s_j}) = -\frac{1}{2}P^{-1}G^T(\nabla\vartheta^T_c(\xi_{s_j})\Theta_c+\nabla\varpi_c(\xi_{s_j}))\\ \delta^*(\xi) = \frac{1}{2\varpi^2}F^T(\nabla\vartheta^T_c(\xi)\Theta_c+\nabla\varpi_c(\xi)).\end{cases} \tag{18}$$

Due to the unavailability of critic weight Θ_c in engineering practice, the derived strategies (18) cannot be instantly applied. To conquer this obstacle, the approximation $\hat{\Theta}_c$ is introduced such that $\mathcal{V}^*(\xi)$ and $\mathcal{V}^*_\xi$ can be estimated by

$$\hat{\mathcal{V}}(\xi) = \hat{\Theta}^T_c\vartheta_c(\xi),\ \hat{\mathcal{V}}_\xi = \nabla\vartheta^T_c(\xi)\hat{\Theta}_c. \tag{19}$$

Subsequently, the approximations of $u^*(\xi_{s_j})$ and $\delta^*(\xi)$ are expressed as follows:

$$\begin{cases}\hat{u}(\xi_{s_j}) = -\frac{1}{2}P^{-1}G^T\nabla\vartheta^T_c(\xi_{s_j})\hat{\Theta}_c\\ \hat{\delta}(\xi) = \frac{1}{2\varpi^2}F^T\nabla\vartheta^T_c(\xi)\hat{\Theta}_c.\end{cases} \tag{20}$$

Denote the estimation error $\epsilon_c = \hat{\mathcal{H}}(\hat{\mathcal{V}}_\xi,\xi,\hat{u}(\xi_{s_j}),\hat{\delta}(\xi)) - \mathcal{H}(\mathcal{V}^*_\xi,\xi,u^*(\xi_{s_j}),\delta^*(\xi))$, where $\hat{\mathcal{H}}(\hat{\mathcal{V}}_\xi,\xi,\hat{u}(\xi_{s_j}),\hat{\delta}(\xi)) = \hat{\Theta}^T_c\nabla\vartheta_c(\xi)(h(\xi)+G\hat{u}(\xi_{s_j})+F\hat{\delta}(\xi))+r(\xi,\hat{u}(\xi_{s_j}),\hat{\delta}(\xi))$ is the approximated Hamiltonian. Based on the aforesaid HJIE, it can be observed that $\mathcal{H}(\mathcal{V}^*_\xi,\xi,u^*(\xi_{s_j}),\delta^*(\xi))=0$. Then, the following expression can be acquired:

$$\epsilon_c = \hat{\Theta}^T_c\varphi_c + r(\xi,\hat{u}(\xi_{s_j}),\hat{\delta}(\xi)) \tag{21}$$

where $\varphi_c = \nabla\vartheta_c(\xi)(h(\xi)+G\hat{u}(\xi_{s_j})+F\hat{\delta}(\xi))$.

To efficaciously approximate the optimal weight Θ_c, the steepest descent approach is employed to minimize $\mathcal{E}_c = \frac{\epsilon^T_c\epsilon_c}{2(1+\varphi^T_c\varphi_c)^2}$, and the following tuning policy can be deduced:

$$\dot{\hat{\Theta}}_c = -\gamma_c\frac{\partial\mathcal{E}_c}{\partial\hat{\Theta}_c} = -\gamma_c\frac{\varphi_c\epsilon_c}{(1+\varphi^T_c\varphi_c)^2} \tag{22}$$

with γ_c representing the positive learning parameter.

Defining $\tilde{\Theta}_c = \Theta_c - \hat{\Theta}_c$ as the critic weight approximation error, we can derive

$$\dot{\tilde{\Theta}}_c = -\gamma_c\frac{\varphi_c\varphi^T_c\tilde{\Theta}_c}{(1+\varphi^T_c\varphi_c)^2} + \gamma_c\frac{\varphi_c\sigma_c}{(1+\varphi^T_c\varphi_c)^2} \tag{23}$$

where $\sigma_c = -\nabla\varpi^T_c(\xi)(h(\xi)+G\hat{u}(\xi_{s_j})+F\hat{\delta}(\xi))$.

B. Validity Analysis

Before proceeding, a commonly adopted assumption [12], [20], [21] is presented to facilitate the next analysis process.

Assumption 3: $||\nabla\vartheta_c(\xi)||\leq\nabla\vartheta_{cM}$, $||\sigma_c||\leq\sigma_{cM}$ and $||\nabla\varpi_c(\xi)||\leq\nabla\varpi_{cM}$, where $\nabla\vartheta_{cM}$, σ_{cM} and $\nabla\varpi_{cM}$ denote the positive scalars.

Theorem 2: For the AWEDS-assisted control policy $\hat{u}(\xi_{s_j})$, the near-worst perturbation $\hat{\delta}(\xi)$ in (20) and the critic weight tuning rule in

Authorized licensed use limited to: Nanjing University of Information Science and Technology. Downloaded on October 19,2024 at 09:12:33 UTC from IEEE Xplore. Restrictions apply.

(22), the UUB stability of NUMV (2) and the critic weight approximation error $\tilde{\Theta}_c$ can be ensured under Assumptions 1-3 when the following conditions hold:

$$\begin{cases} \Omega_1 = \lambda_{\min}(Q) - 2\overline{\mu}\mathcal{L}_{u^*}^2\lambda_{\max}(P)\frac{\max\{\overline{\Psi}^1,\ldots,\overline{\Psi}^6\}}{\min\{\underline{\Psi}^1,\ldots,\underline{\Psi}^6\}} > 0 \\ \Omega_2 = \frac{1}{2}\gamma_c\lambda_{\min}(\Sigma_c) - \frac{1}{2\varpi^2}F_M^2\nabla\vartheta_{cM}^2 \\ \quad -\lambda_{\max}(P)\lambda_{\min}^{-2}(P)G_M^2\nabla\vartheta_{cM}^2 > 0 \end{cases} \tag{24}$$

where $\Sigma_c = \varphi_c\varphi_c^T/(1+\varphi_c^T\varphi_c)^2$.

Proof: The Lyapunov function is constructed as $L_s(t) = \mathcal{V}^*(\xi_{s_j}) + \mathcal{V}^*(\xi) + \frac{1}{2}\tilde{\Theta}_c^T\tilde{\Theta}_c$. Afterwards, the overall analysis will conduct from the following two situations.

Situation I: $t \in [s_j, s_{j+1})$, thus $\dot{L}_s(t) = d(\frac{1}{2}\tilde{\Theta}_c^T\tilde{\Theta}_c)/dt + \dot{\mathcal{V}}^*(\xi)$. By virtue of (23), Assumption 3 and the inequality $2\tilde{m}^T\tilde{n} \le \tilde{m}^T\tilde{m} + \tilde{n}^T\tilde{n}$, it implies

$$\begin{aligned} \frac{d(\frac{1}{2}\tilde{\Theta}_c^T\tilde{\Theta}_c)}{dt} =& -\gamma_c\tilde{\Theta}_c^T\frac{\varphi_c\varphi_c^T\tilde{\Theta}_c - \varphi_c\sigma_c}{(1+\varphi_c^T\varphi_c)^2} \\ \le& -\frac{\gamma_c}{2}\tilde{\Theta}_c^T\frac{\varphi_c\varphi_c^T}{(1+\varphi_c^T\varphi_c)^2}\tilde{\Theta}_c + \frac{\gamma_c}{2}||\sigma_c||^2 \\ \le& -\frac{\gamma_c}{2}\lambda_{\min}(\Sigma_c)||\tilde{\Theta}_c||^2 + \frac{\gamma_c}{2}\sigma_{cM}^2. \end{aligned} \tag{25}$$

From (5) and (6), it is readily obtained that

$$\begin{aligned} \dot{\mathcal{V}}^*(\xi) =& -\xi^TQ\xi + 2\varpi^2\delta^{*T}(\xi)(\hat{\delta}(\xi) - \delta^*(\xi)) \\ &+ \varpi^2||\delta^*(\xi)||^2 - \hat{u}^T(\xi_{s_j})P\hat{u}(\xi_{s_j}) \\ &+ (u^*(\xi) - \hat{u}(\xi_{s_j}))^TP(u^*(\xi) - \hat{u}(\xi_{s_j})) \\ \le& \lambda_{\max}(P)||u^*(\xi) - \hat{u}(\xi_{s_j})||^2 - \lambda_{\min}(Q)||\xi||^2 \\ &+ 2\varpi^2||\delta^*(\xi)||^2 + \varpi^2||\hat{\delta}(\xi) - \delta^*(\xi)||^2. \end{aligned} \tag{26}$$

Utilizing Assumptions 1–3, the formulas (18), (20) and the inequality $||\tilde{m}+\tilde{n}||^2 \le 2||\tilde{m}||^2 + 2||\tilde{n}||^2$ [23], one has

$$\begin{aligned} &||u^*(\xi) - \hat{u}(\xi_{s_j})||^2 \\ &= ||u^*(\xi) - u^*(\xi_{s_j}) + u^*(\xi_{s_j}) - \hat{u}(\xi_{s_j})||^2 \\ &\le 2||u^*(\xi) - u^*(\xi_{s_j})||^2 + 2||u^*(\xi_{s_j}) - \hat{u}(\xi_{s_j})||^2 \\ &\le 2\mathcal{L}_{u^*}^2||e_j||^2 + 2||\frac{1}{2}P^{-1}G^T(\nabla\vartheta_c^T\tilde{\Theta}_c + \nabla\varpi_c(\xi))||^2 \\ &\le 2\mathcal{L}_{u^*}^2||e_j||^2 + \lambda_{\min}^{-2}(P)G_M^2(\nabla\vartheta_{cM}^2||\tilde{\Theta}_c||^2 + \nabla\varpi_{cM}^2). \end{aligned} \tag{27}$$

Similarly,

$$\begin{aligned} &||\hat{\delta}(\xi) - \delta^*(\xi)||^2 \\ &= ||\frac{1}{2\varpi^2}F^T(\nabla\vartheta_c(\xi)\tilde{\Theta}_c + \nabla\varpi_c(\xi))||^2 \\ &\le \frac{1}{2\varpi^4}F_M^2\nabla\vartheta_{cM}^2||\tilde{\Theta}_c||^2 + \frac{1}{2\varpi^4}F_M^2\nabla\varpi_{cM}^2. \end{aligned} \tag{28}$$

$$2\varpi^2||\delta^*(\xi)||^2 = 2\varpi^2||\frac{1}{2\varpi^2}F^T\mathcal{V}_\xi^*||^2 \le \frac{1}{2\varpi^2}F_M^2\mathcal{V}_{em}^2. \tag{29}$$

Under the combination of (15) and (25)–(29), one has

$$\dot{L}_s \le -\Omega_1||\xi||^2 - \Omega_2||\tilde{\Theta}_c||^2 + \Omega_3 \tag{30}$$

where $\Omega_3 = \lambda_{\max}(P)\lambda_{\min}^{-2}G_M^2\nabla\varpi_{cM}^2 + \frac{1}{2\varpi^2}F_M^2(\nabla\varpi_{cM}^2 + \mathcal{V}_{eM}^2) + \frac{\gamma_c}{2}\sigma_{cM}^2$.

Evidently, when $||\xi|| > \sqrt{\Omega_3/\Omega_1}$ or $||\tilde{\Theta}_c|| > \sqrt{\Omega_3/\Omega_2}$, $\dot{L}_s(t) < 0$ holds if the condition (24) is satisfied.

Situation II: $t = s_{j+1}$, then the difference of $L_s(t)$ is obtained as follows:

$$\Delta L_s(t) = \mathcal{V}^*(\xi_{s_{j+1}}) - \mathcal{V}^*(\xi_{s_j}) + \Delta\Upsilon \tag{31}$$

where $\Delta\Upsilon = \frac{1}{2}\tilde{\Theta}_c^T(s_{j+1})\tilde{\Theta}_c(s_{j+1}) - \frac{1}{2}\tilde{\Theta}_c^T(s_{j+1}^-)\tilde{\Theta}_c(s_{j+1}^-) + \mathcal{V}^*(\xi(s_{j+1})) - \mathcal{V}^*(\xi(s_{j+1}^-))$. $\xi(s_{j+1}^-) = \lim_{\kappa\to 0^+}\xi(s_{j+1}-\kappa)$ and $\tilde{\Theta}_c(s_{j+1}^-) = \lim_{\kappa\to 0^+}\tilde{\Theta}_c(s_{j+1}-\kappa)$ with $\kappa \in (0, s_{j+1}-s_j)$.

As proved in *Situation I*, $d(\frac{1}{2}\tilde{\Theta}_c^T\tilde{\Theta}_c)/dt + \dot{\mathcal{V}}^*(\xi) < 0$ can be obtained, which illustrates that the term $\frac{1}{2}\tilde{\Theta}_c^T\tilde{\Theta}_c + \mathcal{V}^*(\xi)$ is monotonically decreasing over $[s_j, s_{j+1})$. Besides, there is no doubt that the term $\frac{1}{2}\tilde{\Theta}_c^T\tilde{\Theta}_c + \mathcal{V}^*(\xi)$ is continuous over $[s_j, s_{j+1}]$. Thus, it is deduced for $\kappa \in (0, s_{j+1}-s_j)$ that

$$\begin{aligned} &\frac{1}{2}\tilde{\Theta}_c^T(s_{j+1})\tilde{\Theta}_c(s_{j+1}) + \mathcal{V}^*(\xi(s_{j+1})) \\ &\le \frac{1}{2}\tilde{\Theta}_c^T(s_{j+1}-\kappa)\tilde{\Theta}_c(s_{j+1}^-) + \mathcal{V}^*(\xi(s_{j+1}-\kappa)). \end{aligned} \tag{32}$$

When $\kappa \to 0^+$, one can conclude that

$$\begin{aligned} &\frac{1}{2}\tilde{\Theta}_c^T(s_{j+1})\tilde{\Theta}_c(s_{j+1}) + \mathcal{V}^*(\xi(s_{j+1})) \\ &\le -\frac{1}{2}\tilde{\Theta}_c^T(s_{j+1}^-)\tilde{\Theta}_c(s_{j+1}^-) - \mathcal{V}^*(\xi(s_{j+1}^-)). \end{aligned} \tag{33}$$

Based on the aforementioned analysis, $\Delta\Upsilon < 0$ in (31) can be satisfied. Furthermore, the boundedness of system state ξ has also been elaborated in *Situation I*, which implies $\mathcal{V}^*(\xi_{s_{j+1}}) \le \mathcal{V}^*(\xi_{s_j})$. Then, we can obtain $\Delta L_s(t) < 0$. Taking account of the aforementioned situations comprehensively, the NUMV (2) and weight estimation error $\tilde{\Theta}_c$ are guaranteed to be UUB. This completes the proof of Theorem 2.■

IV. Experiment and Discussions

In this section, the proposed AWEDS-assisted dynamic positioning control strategy will be validated. Inspired by [22], the relevant matrices in NUMV (2) are listed as

$$U = \begin{bmatrix} 1.0852 & 0 & 0 \\ 0 & 2.0575 & -0.4087 \\ 0 & -0.4087 & 0.2153 \end{bmatrix},\ L = \begin{bmatrix} \bar{L} & 0 \\ 0 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} 0.0865 & 0 & 0 \\ 0 & 0.0762 & 0.0151 \\ 0 & 0.0151 & 0.0310 \end{bmatrix},\ \bar{L} = \begin{bmatrix} 0.0389 & 0 \\ 0 & 0.0266 \end{bmatrix}$$

$$F = \begin{bmatrix} 0 \\ \bar{F} \end{bmatrix},\ \bar{F} = \begin{bmatrix} 0.8715 & 0 & 0 \\ 0 & 0.8126 & 1.5156 \\ 0 & 1.5156 & 6.5243 \end{bmatrix}.$$

In performance function (3), $Q = 10I_6$, $P = 0.1I_3$ and $\varpi = 5$. Moreover, the involved LUBs in AWEDS are ascertained as $\underline{\mu} = 0.02$, $\overline{\mu} = 0.08$, $\underline{\Psi}^\alpha = 1$ and $\overline{\Psi}^\alpha = 50$, $\alpha = 1,\ldots,6$. The sampling period T is set to be 0.04. Additionally, we determine the approximate critic

Authorized licensed use limited to: Nanjing University of Information Science and Technology. Downloaded on October 19,2024 at 09:12:33 UTC from IEEE Xplore. Restrictions apply.

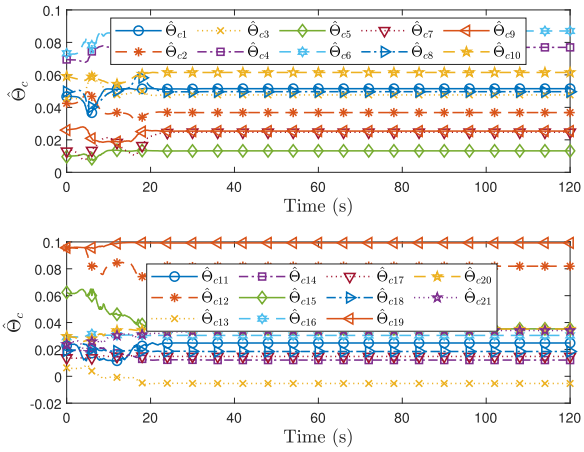

Fig. 2. Response of critic weight $\hat{\Theta}_c$.

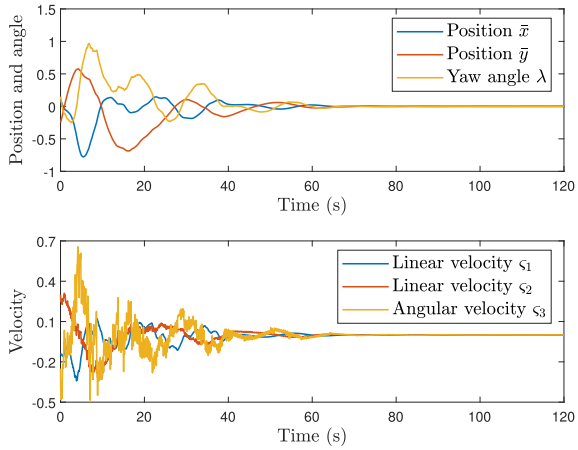

Fig. 3. Evolution of position, angle and velocity.

weight $\hat{\Theta}_c = [\hat{\Theta}_{c1}, \ldots, \hat{\Theta}_{c21}]^T$ and the activation function $\vartheta_c(\xi) = [\xi_1^2, \xi_1\xi_2, \xi_1\xi_3, \xi_1\xi_4, \xi_1\xi_5, \xi_1\xi_6, \xi_2^2, \xi_2\xi_3, \xi_2\xi_4, \xi_2\xi_5, \xi_2\xi_6, \xi_3^2, \xi_3\xi_4, \xi_3\xi_5, \xi_3\xi_6, \xi_4^2, \xi_4\xi_5, \xi_4\xi_6, \xi_5^2, \xi_5\xi_6, \xi_6^2]^T$. To sufficiently promote the weight learning in critic-sole network, a probing noise is added into $\hat{u}(\xi_{s_j})$ before 40 s, which is assigned as $v_e(t) = -0.1e^{-0.05t}[\sin^2(2t)\cos(2t) + \sin^2(1.2t)\cos(10t) + \sin^2(-1.8t)\cos(2.5t) + \sin^2(1.2t) + \sin^3(2.4t)\cos(4t)]$. On basis of trial and error approach, the initial critic weight $\hat{\Theta}_c(0)$ is selected as $[0.0466, 0.0425, 0.0455, 0.0695, 0.0096, 0.0733, 0.0130, 0.0497, 0.026, 0.0592, 0.0228, 0.0955, 0.0063, 0.0183, 0.0624, 0.0291, 0.0137, 0.0189, 0.0957, 0.0297, 0.0237]^T$. By implementing the designed CIRL algorithm, Fig. 2 plots the evolution of critic weight $\hat{\Theta}_c$, from which the critic weight can ultimately tend to the stable status. With the initial state $\xi(0) = [0.1, -0.25, 0.15, -0.15, 0.3, -0.25]^T$, the state responses of NUMV (2) are depicted in Fig. 3. It is apparent that the system states eventually converge to zero, which indicates the validity of the proposed control policy.

For further demonstrating the merits of AWEDS, the performance comparison consequences are emerged in Table I by adopting the predefined weight matrices in AEDS [21] and the adaptive weight $\Psi_{j,l}$ in (9). Due to $\Psi_{j,l}^{\alpha} \in (1, 50)$ involved in AWEDS, we randomly select the weight matrices for AEDS within (1,50) to guarantee the fairness in comparison: $\Psi_1 = \text{diag}\{13.0574, 29.4931, 46.9462, 3.3416, 3.6449, 2.0103\}$, $\Psi_2 = \text{diag}\{11.4780, 2.6969, 23.1051, 1.6759, 24.2118, 47.6087\}$. It

TABLE I
PERFORMANCE COMPARISONS UNDER DIFFERENT WEIGHT MATRICES

Weight Ψ	I_6	$50I_6$	Ψ_1	Ψ_2	$\Psi_{j,l}$
Convergence	$\times$	$\times$	68.84s	80.02s	72.36s
Triggered ratio	36.07%	38.03%	39.3%	23.37%	26.40%

is worth noticing that the notation $\times$ in Table I represents the divergent system states. It is evidently observed from Table I that the state convergence of NUMV (2) and the triggered ratio cannot be balanced well when the weight matrix in AEDS is inappropriately chosen. In contrast, the proposed AWEDS with adaptive weight can effectively balance the control and triggered performance. From the aforementioned discussion, the practicability and advantages of AWEDS-enhanced CIRL algorithm has been further verified.

V. CONCLUSION

In this correspondence, a dynamic positioning regulation strategy is presented for nonlinear NUMV subject to unmatched interference. Facilitated by RL approach, the dynamic positioning control problem is converted into attaining the solution of HJIE. In order to obviate the communication constraints, an adaptive weight-boosted AWEDS is put forward to manage the networked transmission of sampled state information. Furthermore, we provide a CIRL algorithm to acquire the near-optimal control policy. Ultimately, a simulation experiment is conducted to elaborate the practicability and advantages of the designed control scheme. In future research, it is relatively desired to extend the proposed control scheme for unknown networked systems under the data-driven technique. Moreover, the effective defense strategy within RL framework will be investigated under the insecure communication environment.

REFERENCES

[1] Z. Ye, D. Zhang, J. Cheng, and Z.-G. Wu, "Event-triggering and quantized sliding mode control of UMV systems under DoS attack," *IEEE Trans. Veh. Technol.*, vol. 71, no. 8, pp. 8199–8211, Aug. 2022.

[2] H. Sun, J. Shi, and L. Hou, "Dynamic event-triggered output feedback fault-tolerant control for dynamic positioning of unmanned surface vehicles with multiple estimators," *IEEE Trans. Veh. Technol.*, vol. 73, no. 3, pp. 3218–3229, Mar. 2024.

[3] Y. Wang, B. Jiang, Z.-G. Wu, S. Xie, and Y. Peng, "Adaptive sliding mode fault-tolerant fuzzy tracking control with application to unmanned marine vehicles," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 11, pp. 6691–6700, Nov. 2021.

[4] Z. Ye, D. Zhang, Z.-G. Wu, and H. Yan, "A3C-based intelligent event-triggering control of networked nonlinear unmanned marine vehicles subject to hybrid attacks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 12921–12934, Aug. 2022.

[5] J. Liu, J. Ke, J. Liu, X. Xie, and E. Tian, "Secure event-triggered control for IT-2 fuzzy networked systems with stochastic communication protocol and FDI attacks," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 3, pp. 1167–1180, Mar. 2024.

[6] H. Zhang, Z. Ming, Y. Yan, and W. Wang, "Data-driven finite-horizon H_∞ tracking control with event-triggered mechanism for the continuous-time nonlinear systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 8, pp. 4687–4701, Aug. 2023.

[7] Y. Zhao, H. Liang, G. Zong, and H. Wang, "Event-based distributed finite-horizon H_∞ consensus control for constrained nonlinear multiagent systems," *IEEE Syst. J.*, vol. 17, no. 4, pp. 5369–5380, Dec. 2023.

[8] Z. Guo, H. Ren, H. Li, and Q. Zhou, "Adaptive-critic-based event-triggered intelligent cooperative control for a class of second-order constrained multiagent systems," *IEEE Trans. Artif. Intell.*, vol. 4, no. 6, pp. 1654–1665, Dec. 2023.

Authorized licensed use limited to: Nanjing University of Information Science and Technology. Downloaded on October 19,2024 at 09:12:33 UTC from IEEE Xplore. Restrictions apply.

[9] N. Wang, Y. Gao, H. Zhao, and C. K. Ahn, "Reinforcement learning-based optimal tracking control of an unknown unmanned surface vehicle," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 3034–3045, Jul. 2021.

[10] D. Yang, T. Li, H. Zhang, and X. Xie, "Event-trigger-based robust control for nonlinear constrained-input systems using reinforcement learning method," *Neurocomputing*, vol. 340, pp. 158–170, 2019.

[11] H. A. Shah, L. Zhao, and I.-M. Kim, "Joint network control and resource allocation for space-terrestrial integrated network through hierarchal deep actor-critic reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4943–4954, May 2021.

[12] Z. Ming, H. Zhang, W. Li, and Y. Luo, "Neurodynamic programming and tracking control for nonlinear stochastic systems by PI algorithm," *IEEE Trans. Circuits Syst. II: Exp. Briefs*, vol. 69, no. 6, pp. 2892–2896, Jun. 2022.

[13] J. Wang, J. Wu, H. Shen, J. Cao, and L. Rutkowski, "A decentralized learning control scheme for constrained nonlinear interconnected systems based on dynamic event-triggered mechanism," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 53, no. 8, pp. 4934–4943, Aug. 2023.

[14] Z. Wang, X. Wang, and C. Zhao, "Event-triggered containment control for nonlinear multiagent systems via reinforcement learning," *IEEE Trans. Circuits Syst. II: Exp. Briefs*, vol. 70, no. 8, pp. 2904–2908, Aug. 2023.

[15] Z. Peng, W. Yan, R. Huang, H. Cheng, K. Shi, and B. K. Ghosh, "Event-triggered learning robust tracking control of robotic systems with unknown uncertainties," *IEEE Trans. Circuits Syst. II: Exp. Briefs*, vol. 70, no. 7, pp. 2540–2544, Jul. 2023.

[16] Q. Wei, Z. Yang, H. Su, and L. Wang, "Online adaptive dynamic programming for optimal self-learning control of VTOL aircraft systems with disturbances," *IEEE Trans. Automat. Sci. Eng.*, vol. 21, no. 1, pp. 343–352, Jan. 2024.

[17] Z. Gu, X. Huang, X. Sun, X. Xie, and J. H. Park, "Memory-event-triggered tracking control for intelligent vehicle transportation systems: A leader-following approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 5, pp. 4021–4031, May 2024.

[18] H.-T. Sun and C. Peng, "Event-triggered adaptive security path following control for unmanned ground vehicles under sensor attacks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 7, pp. 8500–8509, Jun. 2023.

[19] X. Huo, H. R. Karimi, X. Zhao, B. Wang, and G. Zong, "Adaptive-critic design for decentralized event-triggered control of constrained nonlinear interconnected systems within an identifier-critic framework," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 7478–7491, Aug. 2022.

[20] H. Zhang, Y. Liang, H. Su, and C. Liu, "Event-driven guaranteed cost control design for nonlinear systems with actuator faults via reinforcement learning algorithm," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 50, no. 11, pp. 4135–4150, Nov. 2020.

[21] J. Liu, N. Zhang, L. Zha, X. Xie, and E. Tian, "Reinforcement learning-based decentralized control for networked interconnected systems with communication and control constraints," *IEEE Trans. Automat. Sci. Eng.*, early access, Nov. 2023, doi: 10.1109/TASE.2023.3300917.

[22] Y.-L. Wang and Q.-L. Han, "Network-based modelling and dynamic output feedback control for unmanned marine vehicles in network environments," *Automatica*, vol. 91, pp. 43–53, 2018.

[23] X. Yang and H. He, "Event-driven H_∞-constrained control using adaptive critic learning," *IEEE Trans. Cybern.*, vol. 51, no. 10, pp. 4860–4872, Oct. 2021.

Authorized licensed use limited to: Nanjing University of Information Science and Technology. Downloaded on October 19,2024 at 09:12:33 UTC from IEEE Xplore. Restrictions apply.