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

Zero-Sum Game Based Secure Tracking Control of UAV against FDI Attacks Using Fixed-Time Convergent Reinforcement Learning

Zhenyu Gong[1], Feisheng Yang[1] and Dongrui Wu[2]

Abstract— **In this paper, a fixed-time convergent reinforcement learning (RL) algorithm is developed to realize the secure tracking control of the unmanned aerial vehicle (UAV) via the zero-sum game for the first time. To mitigate FDI attack on actuators that may cause the UAV to deviate from the reference trajectory, a zero-sum differential game framework is built in which the secure controller tries to minimize the common performance function, yet the attacker plays a contrary role. Obtaining the optimal secure tracking controller depends on solving the Hamilton-Jacobi-Isaacs (HJI) equation related to the zero-sum game. Therefore, a critic-only online RL algorithm is proposed that can converge in a fixed time interval, with the corresponding convergence proof provided. A simulation example is given to show the effectiveness of the raised method.**

I. INTRODUCTION

Moving towards becoming more intelligent and more autonomous, the unmanned aerial vehicle (UAV) has received enormous attention from researchers [1]. The UAV has been broadly employed in military and civilian fields due to the advantages of rapid response, flexible deployment and low cost. To carry out the assigned task, such as tracking reference trajectory, the UAV needs to accurately execute control commands launched by the ground control station (GCS). It is critical to guarantee reliable communication between the UAV and the GCS. However, once hostile adversaries intrude on vulnerable communication and inject false data into the control signal, it will lead to system performance degradation and mission failure. Therefore, it is significant to design a secure control scheme to defend against false data injection (FDI) attack on the UAV.

There are mainly two kinds of secure control methods against FDI attacks. One refers to fault-tolerant control [2]. The other includes the detection mechanism that uses attack detectors and secure controllers to compensate for compromised systems [3], [4]. Some researchers have applied these techniques to UAVs. [5] employed the attack detector and the linear quadratic Gaussian controller against FDI attack. An attack detector based on modified sliding innovation sequences was proposed for the UAV in [6]. In reference [7], authors investigated the event-based intermittent secure formation control for the multi-UAV system.

[1]Zhenyu Gong and Feisheng Yang are with Research & Development Institute of Northwestern Polytechnical University in Shenzhen and School of Automation, Northwestern Polytechnical University, Xi'an 710129, China gongzhenyu@mail.nwpu.edu.cn, yangfeisheng@nwpu.edu.cn

[2]Dongrui Wu is with the Key Laboratory of the Ministry of Education for Image Processing and Intelligent Control and the School of Artificial and Automation, Huazhong University of Science and Technology, Wuhan 430074, China drwu@hust.edu.cn

Most existing references only concentrated on the secure control problem from attackers' or defenders' perspectives. Nevertheless, attackers and defenders have limited resources to carry out their policies. Both of them need to evaluate the impact of policies to be implemented on themselves and their adversaries. Game theory provides a unified framework that allows them to study the interaction [8]. [9] developed the Stackelberg game for describing the attack-defense process in the switched control system and provided a sufficient condition for seeking the Stackelberg-Nash equilibrium. [10] designed an optimal resilient controller for the multi-agent system in a hybrid game framework, including the zero-sum and nonzero-sum game.

The reinforcement learning (RL) technique provides a learning scheme to solve the optimization-based problem by designing adaptive policies [11]. It also refers to a feasible scheme for solving the game in the secure control problem [12]. [13] investigated the relationship between FDI attack probability and the existence of the solution to the Hamilton-Jacobi-Isaacs (HJI) equation and designed an optimal secure controller using Q-learning. Authors in [14] proposed a novel FDI attack detection mechanism based on a threat-detection level function and developed an attack mitigation strategy by the off-policy algorithm. A model-free algorithm was applied to design the secure controller for the Markov jump system in [15]. In reference [16], authors proposed a resilient controller for the leader-follower multi-agent system, where the off-policy algorithm was developed to solve the game Riccati equation.

Although existing references investigated secure control by game theory and RL approach, few focused on improving the convergence of RL algorithms. It inspires us to conduct this work. We propose a fixed-time convergent RL algorithm based on the critic-only neural network (NN) structure. Convergence and stability are discussed successively. Moreover, to our best knowledge, it is the first time that the fixed-time convergent RL algorithm has been applied to solve the secure control problem.

The rest of this paper is arranged as follows. Section II introduces the UAV model under FDI attack. Section III formulates the attack-defense process between the attacker and the secure controller as a zero-sum game. Section IV presents a fixed-time convergent RL algorithm based on the critic-only NN structure and analyzes the algorithm convergence and system stability. Section V provides a simulation example. Section VI draws a conclusion.

Notations: The $m \times n$ dimensional zero matrix and $m \times m$ dimensional identity matrix are denoted by $0_{m\times n}$ and I_m,

respectively. The minimum singular value and the minimum eigenvalue are denoted by $\sigma_{\min}(\cdot)$ and $\lambda_{\min}(\cdot)$, respectively. The operator $\lceil\cdot\rfloor$ represents $\mathrm{sgn}(\cdot)|\cdot|$, where $\mathrm{sgn}(\cdot)$ is the sign function and $|\cdot|$ is the absolute value function. The operator ∇ represents gradient computation. The Euclidean norm is denoted by $\|\cdot\|_2$. We define the set $\mathcal{B}_\varsigma[x_0]$ as the closed-ball with radius ς and center x_0. The Cartesian product of two sets $\mathcal{S}_1$ and $\mathcal{S}_2$ is denoted by $\mathcal{S}_1 \times \mathcal{S}_2$.

II. PROBLEM FORMULATION

The dynamics of the UAV can be described by

$$\begin{aligned}\dot{p} &= v\\ \dot{v} &= -cg + u\end{aligned} \tag{1}$$

where $p = [p_x \quad p_y \quad p_z]^{\mathrm{T}}$ and $v = [v_x \quad v_y \quad v_z]^{\mathrm{T}}$ represent the position vector and velocity vector, respectively. g is gravity acceleration, $c = [0 \quad 0 \quad 1]^{\mathrm{T}}$, and u is the controller. For simplicity, define the system state as $x = [p^{\mathrm{T}} \quad v^{\mathrm{T}}]^{\mathrm{T}}$. The UAV dynamics becomes

$$\dot{x} = Ax + Bu + Cg \tag{2}$$

where $A = \begin{bmatrix} 0_{3\times3} & I_3 \\ 0_{3\times3} & 0_{3\times3} \end{bmatrix}$, $B = \begin{bmatrix} 0_{3\times3} \\ I_3 \end{bmatrix}$, $C = [0_{1\times5} \quad -1]^{\mathrm{T}}$. Tracking control aims to design a controller to make the UAV reach the reference trajectory. Given the reference trajectory function p_r, the position and velocity tracking error can be defined by $e_p = p - p_r$, $e_v = v - \dot{p}_r$, respectively. The error dynamics can be written as

$$\dot{e} = Ae + Bu + Cg + f \tag{3}$$

where $e = [e_p^{\mathrm{T}} \quad e_v^{\mathrm{T}}]^{\mathrm{T}}$ is the tracking error state and $f = [0_{1\times3} \quad -\ddot{p}_r^{\mathrm{T}}]^{\mathrm{T}}$. Obviously, the UAV reaches the desired trajectory if $e \to 0$. Consider that the adversarial attacker tries to inject false data into the control signal to deteriorate the system performance, resulting in the UAV deviating from the desired trajectory. For the attacker, we have the following assumption.

Assumption 1: The attacker knows the system dynamics and the control objective.

According to Assumption 1, the compromised control signal can be rewritten as

$$u = u_s + u_a \tag{4}$$

where u_a is the FDI attack signal, and u_s is the secure tracking controller to be designed.

Note that there exist the constant term g and the time-dependent term f in the error system. We can eliminate them by introducing their opposite terms. Thus, design the following secure tracking controller

$$u_s = u_c + u_e \tag{5}$$

where $u_e = cg + \ddot{p}_r$, u_c is the controller to be further designed to mitigate FDI attack and stabilize the error system. Substituting (4) and (5) into (3), one has

$$\dot{e} = Ae + Bu_c + Bu_a. \tag{6}$$

Furthermore, define the following performance function for the error system (6)

$$\begin{aligned}J(e(0), u_c, u_a) &= \int_0^\infty U(e(t), u_c(t), u_a(t))dt\\ &= \int_0^\infty (e^{\mathrm{T}}Qe + u_c^{\mathrm{T}}Ru_c - \gamma^2 u_a^{\mathrm{T}}Tu_a)dt\end{aligned} \tag{7}$$

where symmetric weight matrices $Q > 0$, $R > 0$, $T > 0$, and the attack attenuation level $\gamma > 0$.

III. SECURE ZERO-SUM GAME

Regarding the secure controller u_c and the FDI attacker u_a as zero-sum game players, in which the secure controller intends to minimize the performance function (7) but the attacker aims to maximize it. Hence, we can obtain the game value as follows

$$V(e(0)) = \min_{u_c}\max_{u_a} J(e(0), u_c, u_a). \tag{8}$$

In addition, the saddle point (u_c^*, u_a^*) of the game is the Nash equilibrium if the following condition is satisfied

$$J(e(0), u_c, u_a^*) \geq J(e(0), u_c^*, u_a^*) \geq J(e(0), u_c^*, u_a). \tag{9}$$

With the value function $V(e)$, define the Hamiltonian function as

$$H(e, u_c, u_a) = U(e(t), u_c(t), u_a(t)) + \nabla V^{\mathrm{T}}(e)\dot{e} \tag{10}$$

where $\nabla V(e) = \partial V(e)/\partial e$. Employing the stationary conditions in (10), we can obtain the optimal secure control policy and the optimal attack policy

$$\frac{\partial H}{\partial u_c} = 0 \Rightarrow u_c^* = -\frac{1}{2}R^{-1}B^{\mathrm{T}}\nabla V(e) \tag{11}$$

$$\frac{\partial H}{\partial u_a} = 0 \Rightarrow u_a^* = \frac{1}{2\gamma^2}T^{-1}B^{\mathrm{T}}\nabla V(e). \tag{12}$$

Substituting (11) and (12) into (10), one has the HJI equation below

$$\begin{aligned}&\frac{1}{4}\nabla V^{\mathrm{T}}(e)BR^{-1}B^{\mathrm{T}}\nabla V(e) - \frac{1}{4\gamma^2}\nabla V^{\mathrm{T}}(e)BT^{-1}B^{\mathrm{T}}\nabla V(e)\\ &+ e^{\mathrm{T}}Qe + \nabla V^{\mathrm{T}}(e)(Ae + Bu_c + Bu_a) = 0.\end{aligned} \tag{13}$$

We can acquire the following theorem for ensuring fixed-time stability with existence of the Nash equilibrium.

Theorem 1: Consider the error system (6) with the performance function (7) and the given attack attenuation level γ. Assume that there exist a radially unbounded, continuously differentiable, positive function $V(e) : \mathbb{R}^6 \to \mathbb{R}$, $V(0) = 0$ and real numbers $c_1 > 0, c_2 > 0, 0 < \bar{q} < 1, \bar{r} > 1$ satisfying

$$\nabla V^{\mathrm{T}}(e)\dot{e} \leq -c_1(V(e))^{\bar{q}} - c_2(V(e))^{\bar{r}}. \tag{14}$$

Then the system (6) with optimal policies $u_c = u_c^*$, $u_a = u_a^*$ is globally fixed-time stable, and the settling time is upper bounded by

$$\tilde{\mathcal{T}} \leq \frac{1}{c_1(1-\bar{q})} + \frac{1}{c_2(\bar{r}-1)}. \tag{15}$$

Furthermore, the saddle point (u_c^*, u_a^*) is the Nash equilibrium and the game value is $V(e(0))$.

Proof: By virtue of [17], we can obtain the settling-time function (15) directly.

The performance function (7) can be rewritten as

$$J(e(0),u_c,u_a)=\int_0^\infty \big(U(e,u_c,u_a)+\nabla V^{\mathrm{T}}(e)\dot{e}\big)dt + V(e(0))-V(e(\infty)). \tag{16}$$

Since $\lim_{t\to\bar{\mathcal{T}}} e(t)=\lim_{t\to\infty} e(t)=0$, one has $V(e(\infty))=V(0)=0$.

Now consider that the secure control policy and the attack policy are given by (11) and (12), respectively. By completing the squares for (16), one yields

$$J(e(0),u_c,u_a)=V(e(0))+\int_0^\infty (u_c-u_c^*)^{\mathrm{T}}R(u_c-u_c^*)dt - \int_0^\infty \gamma^2(u_a-u_a^*)^{\mathrm{T}}T(u_a-u_a^*)dt. \tag{17}$$

If $u_c=u_c^*$, one has

$$J(e(0),u_c^*,u_a)=-\int_0^\infty \gamma^2(u_a-u_a^*)^{\mathrm{T}}T(u_a-u_a^*)dt + V(e(0)). \tag{18}$$

Similarly, if $u_a=u_a^*$, we can obtain

$$J(e(0),u_c,u_a^*)=\int_0^\infty (u_c-u_c^*)^{\mathrm{T}}R(u_c-u_c^*)dt + V(e(0)). \tag{19}$$

Combining (18) and (19), it yields

$$J(e(0),u_c,u_a^*)\geq J(e(0),u_c^*,u_a^*)\geq J(e(0),u_c^*,u_a).$$

Note that the Nash equilibrium condition (9) holds. Furthermore, one has

$$J(e(0),u_c^*,u_a^*)=V(e(0)) \tag{20}$$

which gives the zero-sum game value. ■

Remark 1: Once the attacker and secure controller reach the Nash equilibrium, the attacker can not gain more income by adjusting its policy if the secure control policy remains unchanged. The same is true for the secure controller.

IV. FIXED-TIME CONVERGENT RL ALGORITHM

According to the Weierstrass higher-order approximation theorem, the value function can be approximated by the NN. Consider the following critic NN to approximate the value function

$$V(e)=W^{\mathrm{T}}\phi(e)+\varepsilon(e),\quad e\in\mathcal{E} \tag{21}$$

where $W\in\mathbb{R}^N$, $\phi(\cdot)$, $\varepsilon(\cdot)$, $\mathcal{E}\subseteq\mathbb{R}^6$ represent the expected NN weights, activation function, approximation error of the critic NN and error state set, respectively. Thereby, the optimal policies can be approximated by

$$u_c^*=-\frac{1}{2}R^{-1}B^{\mathrm{T}}(\nabla\phi^{\mathrm{T}}(e)W+\nabla\varepsilon(e)) \tag{22}$$

$$u_a^*=\frac{1}{2\gamma^2}T^{-1}B^{\mathrm{T}}(\nabla\phi^{\mathrm{T}}(e)W+\nabla\varepsilon(e)). \tag{23}$$

Remark 2: The secure game can be viewed as a deterministic Markov decision process. The controller and attacker implement policies on the UAV system, which can be regarded as an interaction with the environment. Combined with the response from the system, the value function is utilized to evaluate the current policies. Then, the controller and attacker can improve their policies. It indicates that we can use RL to study the secure game problem.

Remark 3: Since the derivative of the activation function is required, the differentiable activation function can be selected as polynomial, sigmoid, hyperbolic tangent functions, etc.

Because the expected NN weights are unknown, the estimation critic NN weights are provided for estimating W. One has

$$\hat{V}(e)=\hat{W}^{\mathrm{T}}\phi(e). \tag{24}$$

Then the estimated policies become

$$\hat{u}_c=-\frac{1}{2}R^{-1}B^{\mathrm{T}}\nabla\phi^{\mathrm{T}}(e)\hat{W} \tag{25}$$

$$\hat{u}_a=\frac{1}{2\gamma^2}T^{-1}B^{\mathrm{T}}\nabla\phi^{\mathrm{T}}(e)\hat{W}. \tag{26}$$

Substituting (24), (25) and (26) into Hamiltonian function (10), one has

$$H(e,\hat{u}_c,\hat{u}_a,\hat{W})\equiv\xi=\hat{W}^{\mathrm{T}}\nabla\phi(e)(Ae+B\hat{u}_c+B\hat{u}_a) + U(e,\hat{u}_c,\hat{u}_a). \tag{27}$$

Define the following loss function

$$E(\hat{W})=\frac{1}{q+1}\left|\frac{\xi}{1+\psi^{\mathrm{T}}\psi}\right|^{q+1}+\frac{1}{r+1}\left|\frac{\xi}{1+\psi^{\mathrm{T}}\psi}\right|^{r+1} \tag{28}$$

where $\psi=\nabla\phi(e)(Ae+B\hat{u}_c+B\hat{u}_a)$, $0<q<1$ and $r>1$.

The aim of selecting $\hat{W}$ is to minimize $E(\hat{W})$. According to the gradient descent principle, we can derive the NN weight update law, which guarantees that NN weights converge in a fixed time. The weight update law is given by

$$\dot{\hat{W}}=-\alpha\frac{\partial E(\hat{W})}{\partial\hat{W}} =-\alpha\frac{\psi}{1+\psi^{\mathrm{T}}\psi}\left(\left\lceil\frac{\xi}{1+\psi^{\mathrm{T}}\psi}\right\rfloor^q+\left\lceil\frac{\xi}{1+\psi^{\mathrm{T}}\psi}\right\rfloor^r\right) \tag{29}$$

where α is the learning rate of the critic NN.

Define the weight error as $\tilde{W}=\hat{W}-W$. The weight error dynamics is

$$\dot{\tilde{W}}=-\alpha\frac{\psi}{1+\psi^{\mathrm{T}}\psi}\left(\left\lceil\frac{\psi^{\mathrm{T}}\tilde{W}+\epsilon}{1+\psi^{\mathrm{T}}\psi}\right\rfloor^q+\left\lceil\frac{\psi^{\mathrm{T}}\tilde{W}+\epsilon}{1+\psi^{\mathrm{T}}\psi}\right\rfloor^r\right) \tag{30}$$

where $\epsilon=W^{\mathrm{T}}\nabla\phi(e)(Ae+B\hat{u}_c+B\hat{u}_a)+U(e,\hat{u}_c,\hat{u}_a)$.

Next, the fixed-time convergence of critic weights will be discussed by the following theorem. Before that, the following lemmas are necessary.

Lemma 1 [18] : For any $\hat{\epsilon} > 0$, there exist functions $L(\hat{\epsilon}) > 0$ and $N_0(\hat{\epsilon}) > 0$ such that $\sup|\epsilon| < L(\hat{\epsilon})$, $N_0(\hat{\epsilon}) \leq N$ and $\epsilon \equiv 0$ if $N \to \infty$.

Lemma 2 [19] : For $\eta_1, \eta_2, \cdots, \eta_n \geq 0$, $0 < q < 1$ and $r > 1$, then

$$\begin{cases} \left(\sum_{i=1}^n \eta_i\right)^q \leq \left(\sum_{i=1}^n \eta_i^q\right) \leq n^{1-q}\left(\sum_{i=1}^n \eta_i\right)^q, \\ n^{1-r}\left(\sum_{i=1}^n \eta_i\right)^r \leq \left(\sum_{i=1}^n \eta_i{}^r\right) \leq \left(\sum_{i=1}^n \eta_i\right)^r. \end{cases}$$

Theorem 2: Let $\bar{\psi} = \frac{\psi}{1+\psi^{\mathrm{T}}\psi}$, $\Psi = \bar{\psi}\bar{\psi}^{\mathrm{T}}$, $0 < \theta < 1$, $h = (1-\theta)\sigma_{\min}^{\frac{q+1}{2}}(\Psi)$ and $\bar{h} = 2^{1-r}\sigma_{\min}^{\frac{r+1}{2}}(\Psi)$. Define $\mu = \left(\frac{(\bar{\epsilon}_m^q + \bar{\epsilon}_m^r)\bar{\psi}_m}{\theta\sigma_{\min}^{(q+1)/2}(\Psi)}\right)^{\frac{1}{q}}$ with $\bar{\epsilon}_m > 0$, $\bar{\psi}_m > 0$. Then, the solution to the weight error dynamics (30) is 1) globally fixed-time stable with the settling time

$$\mathcal{T} \leq \frac{2}{\sigma_{\min}^{\frac{q+1}{2}}(\Psi)(2\alpha)^{\frac{q+1}{2}}(1-q)} + \frac{2}{\sigma_{\min}^{\frac{r+1}{2}}(\Psi)(2\alpha)^{\frac{r+1}{2}}(r-1)}$$

if $\epsilon \equiv 0$, and 2) globally fixed-time uniformly ultimately bounded (UUB) with the settling time

$$\mathcal{T} \leq \frac{2^{\frac{1-q}{2}}\alpha^{-\frac{q+1}{2}} - \alpha^{-1}\mu^{1-q}}{h(1-q)} + \frac{2^{\frac{1-r}{2}}\alpha^{-\frac{r+1}{2}}}{\bar{h}(r-1)}$$

if $\epsilon \not\equiv 0$.

Proof: Select the following Lyapunov function for the weight error

$$\tilde{V}(\tilde{W}) = \frac{1}{2\alpha}\tilde{W}^{\mathrm{T}}\tilde{W}. \tag{31}$$

Calculating the time derivative of (31) along (30), one has

$$\dot{\tilde{V}}(\tilde{W}) = -\frac{\tilde{W}^{\mathrm{T}}\psi}{1+\psi^{\mathrm{T}}\psi}\left(\left\lceil\frac{\psi^{\mathrm{T}}\tilde{W}+\epsilon}{1+\psi^{\mathrm{T}}\psi}\right\rfloor^q + \left\lceil\frac{\psi^{\mathrm{T}}\tilde{W}+\epsilon}{1+\psi^{\mathrm{T}}\psi}\right\rfloor^r\right). \tag{32}$$

1) If $\epsilon \equiv 0$, suppose that the persistence of excitation condition is met, then one has

$$\begin{aligned}\dot{\tilde{V}}(\tilde{W}) &= -\frac{\tilde{W}^{\mathrm{T}}\psi}{1+\psi^{\mathrm{T}}\psi}\left(\left\lceil\frac{\psi^{\mathrm{T}}\tilde{W}}{1+\psi^{\mathrm{T}}\psi}\right\rfloor^q + \left\lceil\frac{\psi^{\mathrm{T}}\tilde{W}}{1+\psi^{\mathrm{T}}\psi}\right\rfloor^r\right) \\ &= -\left|\frac{\psi^{\mathrm{T}}\tilde{W}}{1+\psi^{\mathrm{T}}\psi}\right|^{q+1} - \left|\frac{\psi^{\mathrm{T}}\tilde{W}}{1+\psi^{\mathrm{T}}\psi}\right|^{r+1} \\ &= -\left(\left\|\frac{\psi^{\mathrm{T}}\tilde{W}}{1+\psi^{\mathrm{T}}\psi}\right\|_2^2\right)^{\frac{q+1}{2}} - \left(\left\|\frac{\psi^{\mathrm{T}}\tilde{W}}{1+\psi^{\mathrm{T}}\psi}\right\|_2^2\right)^{\frac{r+1}{2}} \\ &\leq -\sigma_{\min}^{\frac{q+1}{2}}(\Psi)\|\tilde{W}\|_2^{q+1} - \sigma_{\min}^{\frac{r+1}{2}}(\Psi)\|\tilde{W}\|_2^{r+1}. \end{aligned} \tag{33}$$

Moreover, (33) can be rewritten as

$$\begin{aligned}\dot{\tilde{V}}(\tilde{W}) \leq &-\sigma_{\min}^{\frac{q+1}{2}}(\Psi)(2\alpha)^{\frac{q+1}{2}}(\tilde{V}(\tilde{W}))^{\frac{q+1}{2}} \\ &-\sigma_{\min}^{\frac{r+1}{2}}(\Psi)(2\alpha)^{\frac{r+1}{2}}(\tilde{V}(\tilde{W}))^{\frac{r+1}{2}}.\end{aligned}$$

It shows that the solution $\tilde{W} \equiv 0$ to (30) is globally fixed-time stable. The corresponding settling time $\mathcal{T}$ satisfies

$$\mathcal{T} \leq \frac{2}{\sigma_{\min}^{\frac{q+1}{2}}(\Psi)(2\alpha)^{\frac{q+1}{2}}(1-q)} + \frac{2}{\sigma_{\min}^{\frac{r+1}{2}}(\Psi)(2\alpha)^{\frac{r+1}{2}}(r-1)}. \tag{34}$$

2) If $\epsilon \not\equiv 0$, recalling that the time derivative of the Lyapunov function (31) is given by (32). According to Lemma 1, one has $|\psi^{\mathrm{T}}\tilde{W}| \leq |\epsilon|$. Then, using Lemma 2, one obtains

$$\begin{aligned}&|\bar{\psi}^{\mathrm{T}}\tilde{W}|^q \leq |\bar{\psi}^{\mathrm{T}}\tilde{W} + \bar{\epsilon}|^q + |\bar{\epsilon}|^q \\ &|\bar{\psi}^{\mathrm{T}}\tilde{W}|^r \leq 2^{r-1}(|\bar{\psi}^{\mathrm{T}}\tilde{W} + \bar{\epsilon}|^r + |\bar{\epsilon}|^r).\end{aligned}$$

Hence, one attains

$$|\bar{\psi}^{\mathrm{T}}\tilde{W}|^q - |\bar{\epsilon}|^q \leq |\bar{\psi}^{\mathrm{T}}\tilde{W} + \bar{\epsilon}|^q \tag{35}$$

$$2^{1-r}|\bar{\psi}^{\mathrm{T}}\tilde{W}|^r - |\bar{\epsilon}|^r \leq |\bar{\psi}^{\mathrm{T}}\tilde{W} + \bar{\epsilon}|^r. \tag{36}$$

Based on (35) and (36), one can acquire

$$\begin{aligned}\dot{\tilde{V}}(\tilde{W}) \leq &-\sigma_{\min}^{\frac{q+1}{2}}(\Psi)\|\tilde{W}\|_2^{q+1} - 2^{1-r}\sigma_{\min}^{\frac{r+1}{2}}(\Psi)\|\tilde{W}\|_2^{r+1} \\ &+ |\bar{\psi}^{\mathrm{T}}\tilde{W}||\bar{\epsilon}|^q + |\bar{\psi}^{\mathrm{T}}\tilde{W}||\bar{\epsilon}|^r.\end{aligned} \tag{37}$$

Note that there exist $\bar{\epsilon}_m$ and $N_0 > 0$ satisfying $\sup|\epsilon| < \bar{\epsilon}_m$ for any $N > N_0$ by Lemma 1. Thereby, we can directly know $|\bar{\epsilon}| < \bar{\epsilon}_m$. Due to bounded $\bar{\psi}$, there exists $\bar{\psi}_m > 0$ that satisfies $\|\bar{\psi}\|_2 \leq \bar{\psi}_m$. Then (37) becomes

$$\begin{aligned}\dot{\tilde{V}}(\tilde{W}) \leq &-\sigma_{\min}^{\frac{q+1}{2}}(\Psi)\|\tilde{W}\|_2^{q+1} - 2^{1-r}\sigma_{\min}^{\frac{r+1}{2}}(\Psi)\|\tilde{W}\|_2^{r+1} \\ &+ (\bar{\epsilon}_m^q + \bar{\epsilon}_m^r)\bar{\psi}_m\|\tilde{W}\|_2 \\ \leq &-(1-\theta)\sigma_{\min}^{\frac{q+1}{2}}(\Psi)\|\tilde{W}\|_2^{q+1} - \theta\sigma_{\min}^{\frac{q+1}{2}}(\Psi)\|\tilde{W}\|_2^{q+1} \\ &- 2^{1-r}\sigma_{\min}^{\frac{r+1}{2}}(\Psi)\|\tilde{W}\|_2^{r+1} + (\bar{\epsilon}_m^q + \bar{\epsilon}_m^r)\bar{\psi}_m\|\tilde{W}\|_2.\end{aligned} \tag{38}$$

We can further obtain

$$\begin{aligned}\dot{\tilde{V}}(\tilde{W}) \leq &-(1-\theta)\sigma_{\min}^{\frac{q+1}{2}}(\Psi)\|\tilde{W}\|_2^{q+1} \\ &-2^{1-r}\sigma_{\min}^{\frac{r+1}{2}}(\Psi)\|\tilde{W}\|_2^{r+1}, \quad \|\tilde{W}\|_2 \geq \mu.\end{aligned} \tag{39}$$

Hence, the solution to (30) is globally fixed-time UUB with the bound μ.

Next, we can derive the settling time of (30). Rewrite (39) as

$$\begin{aligned}\frac{dt}{d\tilde{V}(\tilde{W})} \geq &\frac{1}{-h(2\alpha)^{\frac{q+1}{2}}(\tilde{V}(\tilde{W}))^{\frac{q+1}{2}} - \bar{h}(2\alpha)^{\frac{r+1}{2}}(\tilde{V}(\tilde{W}))^{\frac{r+1}{2}}}, \\ &\|\tilde{W}\|_2 \geq \mu.\end{aligned} \tag{40}$$

If $\tilde{V}(\tilde{W}(0)) \leq 1$, integrating both sides of (40), it yields

$$\begin{aligned}\mathcal{T} &\leq -\frac{1}{h(2\alpha)^{\frac{q+1}{2}}}\int_1^{\frac{\mu^2}{2\alpha}} z^{-\frac{q+1}{2}}dz \\ &\leq \frac{2^{\frac{1-q}{2}}\alpha^{-\frac{q+1}{2}} - \alpha^{-1}\mu^{1-q}}{h(1-q)} \triangleq \mathcal{T}_a.\end{aligned} \tag{41}$$

If $\tilde{V}(\tilde{W}(0)) > 1$, one obtains

$$\mathcal{T} \leq \int_{\frac{\mu^2}{2\alpha}}^1 \frac{1}{h(2\alpha)^{\frac{q+1}{2}}z^{\frac{q+1}{2}}}dz + \int_1^{\tilde{V}(\tilde{W}(0))} \frac{1}{\bar{h}(2\alpha)^{\frac{r+1}{2}}z^{\frac{r+1}{2}}}dz.$$

It is obvious that $\int_1^{\tilde{V}(\tilde{W}(0))}(\cdot)dz \leq \int_1^{+\infty}(\cdot)dz$. Let $\bar{z} = z^{1-\frac{r+1}{2}}$. One has $d\bar{z} = (1-\frac{r+1}{2})z^{-\frac{r+1}{2}}dz$. We can further obtain

$$\frac{1}{\bar{h}(2\alpha)^{\frac{r+1}{2}}}\int_1^{+\infty}\frac{1}{z^{\frac{r+1}{2}}}dz = \frac{1}{\bar{h}(2\alpha)^{\frac{r+1}{2}}}\int_1^0 \frac{1}{1-\frac{r+1}{2}}d\bar{z}$$

$$=\frac{2^{\frac{1-r}{2}}\alpha^{-\frac{r+1}{2}}}{\bar{h}(r-1)}$$

and it follows that

$$\mathcal{T} \leq \mathcal{T}_a + \frac{2^{\frac{1-r}{2}}\alpha^{-\frac{r+1}{2}}}{\bar{h}(r-1)}. \tag{42}$$

Finally, combining (41) and (42), the settling time is

$$\mathcal{T} \leq \frac{2^{\frac{1-q}{2}}\alpha^{-\frac{q+1}{2}} - \alpha^{-1}\mu^{1-q}}{h(1-q)} + \frac{2^{\frac{1-r}{2}}\alpha^{-\frac{r+1}{2}}}{\bar{h}(r-1)}. \tag{43}$$

■

Substituting policies (25) and (26) into (6), we can analyze the stability of the closed-loop system with the tracking error and weight error by the following theorem.

Theorem 3: Consider the error closed-loop system (6) using policies (25) and (26). Let $\epsilon \not\equiv 0$. Given $\tilde{\mu} = \left(\frac{\iota+(\bar{\epsilon}_m^q+\bar{\epsilon}_m^r)\bar{\psi}_m}{\theta\sigma_{\min}^{(q+1)/2}(\Psi)}\right)^{\frac{1}{q}}$ with proper $0<\theta<1$ and $\iota>0$, the augmented state $\Gamma = [\tilde{W}^{\mathrm{T}}, e^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^N \times \mathcal{E}$ is fixed-time UUB with the settling time

$$\bar{\mathcal{T}} \leq \frac{2^{\frac{1-q}{2}}\alpha^{-\frac{q+1}{2}} - \alpha^{-1}\tilde{\mu}^{1-q}}{h(1-q)} + \frac{2^{\frac{1-r}{2}}\alpha^{-\frac{r+1}{2}}}{\bar{h}(r-1)}.$$

Proof: Select the Lyapunov function as

$$\bar{V}(\Gamma) = V(e) + \tilde{V}(\tilde{W}). \tag{44}$$

Using the optimal policies (22), (23) and the estimated policies (25), (26), the time derivative of $V(e)$ satisfies

$$\begin{aligned}\dot{V}(e) =& \nabla V^{\mathrm{T}}(e)(Ae + B\hat{u}_c + B\hat{u}_a)\\ =& \nabla V^{\mathrm{T}}(e)(Ae + Bu_c^* + Bu_a^* + B(\hat{u}_c - u_c^*)\\ &+ B(\hat{u}_a - u_a^*)\\ \leq& -c_1(V(e))^{\bar{q}} - c_2(V(e))^{\bar{r}}\\ &+ \frac{1}{2}\nabla V^{\mathrm{T}}(e)BR^{-1}B^T(-\nabla\phi^{\mathrm{T}}(e)\tilde{W} + \nabla\varepsilon(e))\\ &+ \frac{1}{2\gamma^2}\nabla V^{\mathrm{T}}(e)BT^{-1}B^T(\nabla\phi^{\mathrm{T}}(e)\tilde{W} - \nabla\varepsilon(e)).\end{aligned}$$

Suppose that W, $\nabla\phi(e)$ and $\nabla\varepsilon(e)$ have upper bounds W_m, $\nabla\phi_m$ and $\nabla\varepsilon_m$, respectively. Then, it follows that

$$\begin{aligned}\dot{V}(e) \leq& -c_1(V(e))^{\bar{q}} - c_2(V(e))^{\bar{r}} + (W_m\nabla\phi_m + \nabla\varepsilon_m)\\ &\cdot\left(\frac{1}{2}\|B\|_2^2\lambda_{\min}(R) + \frac{1}{2\gamma^2}\|B\|_2^2\lambda_{\min}(T)\right)\\ =& -c_1(V(e))^{\bar{q}} - c_2(V(e))^{\bar{r}} + \bar{m}(\nabla\phi_m\|\tilde{W}\|_2 + \nabla\varepsilon_m)\end{aligned}$$

where $\bar{m} = (W_m\nabla\phi_m + \nabla\varepsilon_m)(\frac{1}{2}\|B\|_2^2\lambda_{\min}(R) + \frac{1}{2\gamma^2}\|B\|_2^2\lambda_{\min}(T))$. Select $\iota>0$ such that $\bar{m}(\nabla\phi_m\|\tilde{W}\|_2 + \nabla\varepsilon_m) \leq \iota\|\tilde{W}\|_2$. Let $\bar{\mu} = \frac{\nabla\varepsilon_m}{\frac{\iota}{\bar{m}}-\phi_m}$. One has

$$\begin{aligned}\dot{V}(e) \leq& -c_1(V(e))^{\bar{q}} - c_2(V(e))^{\bar{r}} + \iota\|\tilde{W}\|_2,\\ &\tilde{W} \in \mathcal{B}_{\bar{\mu}}[0], \quad e \in \mathcal{E}.\end{aligned} \tag{45}$$

Taking into account (45) and Theorem 2, we can acquire

$$\begin{aligned}\dot{\bar{V}}(\Gamma) \leq& -c_1(V(e))^{\bar{q}} - c_2(V(e))^{\bar{r}} - h\|\tilde{W}\|_2^{q+1}\\ &- \bar{h}\|\tilde{W}\|_2^{r+1} - \theta\sigma_{\min}^{\frac{q+1}{2}}(\Psi)\|\tilde{W}\|_2^{q+1}\\ &+ (\iota + (\bar{\epsilon}_m^q + \bar{\epsilon}_m^r)\bar{\psi}_m)\|\tilde{W}\|_2, \quad \tilde{W} \in \mathcal{B}_{\tilde{\mu}}[0],\ e \in \mathcal{E}.\end{aligned}$$

Choosing the parameter θ allows

$$\begin{aligned}\dot{\bar{V}}(\Gamma) \leq& -c_1(V(e))^{\bar{q}} - c_2(V(e))^{\bar{r}} - h\|\tilde{W}\|_2^{q+1} - \bar{h}\|\tilde{W}\|_2^{r+1},\\ &\tilde{W} \in \mathcal{B}_{\tilde{\mu}}[0],\ e \in \mathcal{E}\end{aligned} \tag{46}$$

where $\mathcal{B}_{\bar{\mu}}[0] \subseteq \mathcal{B}_{\tilde{\mu}}[0]$ for sufficiently small θ. Performing the similar operation as in Theorem 2, we can derive the settling time as follows

$$\bar{\mathcal{T}} \leq \frac{2^{\frac{1-q}{2}}\alpha^{-\frac{q+1}{2}} - \alpha^{-1}\tilde{\mu}^{1-q}}{h(1-q)} + \frac{2^{\frac{1-r}{2}}\alpha^{-\frac{r+1}{2}}}{\bar{h}(r-1)}. \tag{47}$$

Therefore, Γ is fixed-time UUB with an upper bound $\mathcal{B}_{\tilde{\mu}}[0]\times\mathcal{E}$. ■

V. SIMULATION EXAMPLE

The reference trajectory is given by $p_r = [15\sin(\frac{\pi}{5}t)\ 15\cos(\frac{\pi}{5}t)\ 2+0.2t]^{\mathrm{T}}$. The gravity acceleration is $g = 9.8\mathrm{m/s}^2$. The attack attenuation level is $\gamma = 8$. The weight matrices Q, R, T are selected as identity matrices with proper dimensions. The weight update law's corresponding parameters are selected as $\alpha = 0.1$, $q = 0.1$, $r = 2$. Choose the polynomial function as the activation function. Suppose that the initial state is $x_0 = [5\ 12\ 4\ 10\ 2\ 1]^{\mathrm{T}}$. Simulation results are shown as Fig. 1-Fig. 4.

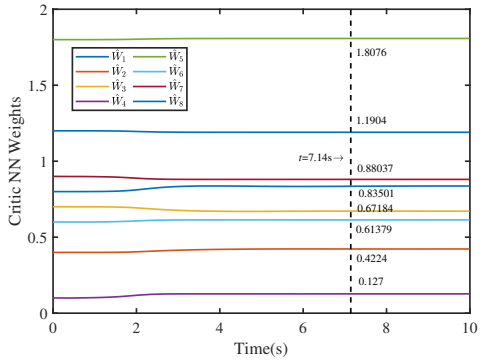

Fig. 1. Time evolution of critic NN weights by fixed-time convergent RL.

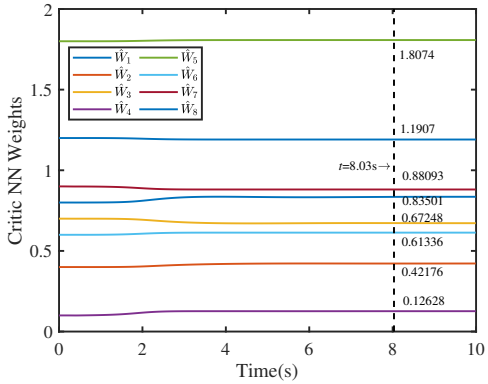

Fig. 2. Time evolution of critic NN weights by finite-time convergent RL.

Fig. 1 and Fig. 2 depict the time evolution of critic NN weights by fixed-time convergent RL and finite-time convergent RL, respectively. In the finite-time convergent RL, the critic NN update law is given by

$$\dot{\hat{W}} = -\alpha\frac{\psi}{1+\psi^{\mathrm{T}}\psi}\left[\frac{\xi}{1+\psi^{\mathrm{T}}\psi}\right]^q, 0<q<1.$$

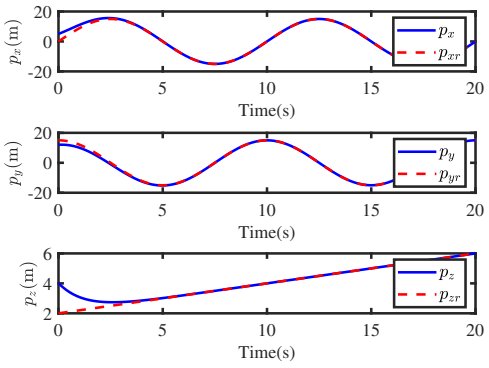

Fig. 3. The trajectories of the UAV and reference.

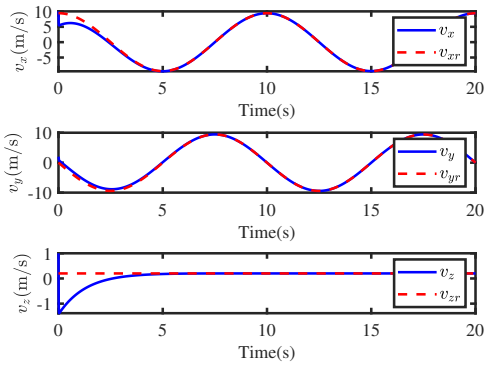

Fig. 4. The velocities of the UAV and reference.

It shows that the proposed approach (t=7.14s) has faster convergence rate than finite-time convergent RL (t=8.03s). Fig. 3-Fig. 4 demonstrate the designed optimal controller can guarantee that the UAV reaches the reference trajectory and the desired velocity under the optimal FDI attack. It shows that the UAV can reach the desired trajectory and velocity in the time of $t = 6$s with trained NN weights.

VI. CONCLUSION

In this article, we study the secure tracking control in the presence of FDI attacks for the UAV under the zero-sum game framework. A fixed-time convergent RL algorithm is proposed for solving the game and obtaining the optimal secure tracking controller. Future research will focus on exploiting the secure control scheme in multi-UAV systems.

ACKNOWLEDGEMENT

This work was supported partially by National Natural Science Foundation of China (62073269), Key Research and Development Program of Shaanxi (2022GY-244), Natural Science Foundation of Chongqing, China (CSTB2022NSCQ-MSX0963), Guangdong Basic and Applied Basic Research Foundation (2023A1515011220), Aeronautical Science Foundation of China (2020Z034053002, 2018ZE53052), and Basic Scientific Research Program of China (JCKY2020203C025).

REFERENCES

[1] H. Wang, H. Zhao, J. Zhang, D. Ma, J. Li, and J. Wei, "Survey on unmanned aerial vehicle networks: A cyber physical system perspective," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1027–1070, 2020.

[2] X. M. Li, Q. Zhou, P. Li, H. Li, and R. Lu, "Event-triggered consensus control for multi-agent systems against false data-injection attacks," *IEEE Transactions on Cybernetics*, vol. 50, no. 5, pp. 1856–1866, 2020.

[3] A. Sargolzaei, K. Yazdani, A. R. Abbaspour, C. D. Crane, and W. E. Dixon, "Detection and mitigation of false data injection attacks in networked control systems," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4281–4292, 2020.

[4] Y. Tan, Q. Liu, J. Liu, X. Xie, and S. Fei, "Observer-based security control for interconnected semi-Markovian jump systems with unknown transition probabilities," *IEEE Transactions on Cybernetics*, vol. 52, no. 9, pp. 9013–9025, 2022.

[5] H. Lin, P. Sun, C. Cai, S. Lu, and H. Liu, "Secure LQG control for a quadrotor under false data injection attacks," *IET Control Theory & Applications*, vol. 16, no. 9, pp. 925–934, 2022.

[6] J. Xiao and M. Feroskhan, "Cyber attack detection and isolation for a quadrotor UAV with modified sliding innovation sequences," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 7, pp. 7202–7214, 2022.

[7] T. Yin, Z. Gu, and J. H. Park, "Event-based intermittent formation control of multi-UAV systems under deception attacks," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12, 2022, doi: 10.1109/TNNLS.2022.3227101.

[8] H. Ishii and Q. Zhu, *Security and Resilience of Control Systems: Theory and Applications*. Switzerland: Springer Cham, 2022.

[9] Y. Huang and J. Zhao, "Switching defence for switched systems under malicious attacks: A Stackelberg game approach," *Nonlinear Analysis: Hybrid Systems*, vol. 42, p. 101092, 2021.

[10] J. Shen, X. Ye, and D. Feng, "A game-theoretic method for resilient control design in industrial multi-agent CPSs with Markovian and coupled dynamics," *International Journal of Control*, vol. 94, no. 11, pp. 3079–3090, 2021.

[11] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers," *IEEE Control Systems Magazine*, vol. 32, no. 6, pp. 76–105, 2012.

[12] S. R. Etesami and T. Başar, "Dynamic games in cyber-physical security: An overview," *Dynamic Games and Applications*, vol. 9, no. 4, pp. 884–913, 2019.

[13] C. Wu, X. Li, W. Pan, J. Liu, and L. Wu, "Zero-sum game based optimal secure control under actuator attacks," *IEEE Transactions on Automatic Control*, vol. 66, no. 8, pp. 3773–3780, 2021.

[14] Y. Zhou, K. G. Vamvoudakis, W. M. Haddad, and Z. P. Jiang, "A secure control learning framework for cyber-physical systems under sensor and actuator attacks," *IEEE Transactions on Cybernetics*, vol. 51, no. 9, pp. 4648–4660, 2021.

[15] X. Gao, F. Deng, and P. Zeng, "Zero-sum game-based security control of unknown nonlinear Markov jump systems under false data injection attacks," *International Journal of Robust and Nonlinear Control*, pp. 1–11, 2022, doi: 10.1002/rnc.6418.

[16] M. Rohollah and M. Hamidreza, "Resilient autonomous control of distributed multiagent systems in contested environments." *IEEE Transactions on Cybernetics*, vol. 49, no. 11, pp. 3957–3967, 2018.

[17] A. Polyakov, "Nonlinear feedback design for fixed-time stabilization of linear control systems," *IEEE Transactions on Automatic Control*, vol. 57, no. 8, pp. 2106–2110, 2012.

[18] N.-M. T. Kokolakis and K. G. Vamvoudakis, "Safety-aware pursuit-evasion games in unknown environments using Gaussian processes and finite-time convergent reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2022, doi: 10.1109/TNNLS.2022.3203977.

[19] Z. Zuo, J. Song, B. Tian, and M. Basin, "Robust fixed-time stabilization control of generic linear systems with mismatched disturbances," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 2, pp. 759–768, 2022.