An Intelligent and Secure Control Approach for Nonlinear Systems under Attacks

Xiangnan Zhong
Department of Electrical Engineering and Computer Science
Florida Atlantic University
Boca Raton, FL, USA
xzhong@fau.edu

Zhen Ni
Department of Electrical Engineering and Computer Science
Florida Atlantic University
Boca Raton, FL, USA
zhenni@fau.edu

2021 IEEE Symposium Series on Computational Intelligence (SSCI) | 978-1-7281-9048-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/SSCI50451.2021.9659857

***Abstract*—This paper designs an intelligent and secure control approach based on adaptive dynamic programming for a class of nonlinear systems under the actuator attacks. The designed method can monitor and identify the attacks by an established state estimator based detector. When an attack is triggered, the control process automatically switches to a game-theoretical architecture for attack mitigation. The intelligent learning process is developed for both nominal and attack conditions without the requirement of system dynamics and attack information. Neural network techniques are applied to implement the proposed method with two critic networks and the control signals are calculated accordingly. Therefore, the designed intelligent control method can reduce the computation complexity. Simulation studies and results demonstrate the necessary of attack detection and mitigation during the learning process, and also verify the effectiveness of the developed method.**

***Index Terms*—Adaptive dynamic programming, reinforcement learning, attack mitigation, intelligent control, attack detection and identification.**

I. Introduction

Networked control systems have attracted significant increasing attention over the past decades due to the development of more decentralized control applications and the rise of cyber-physical systems [1]–[6]. Generally, networked control systems integrate sensing, control and actuation components through a communication network [7], which is usually vulnerable to malicious attacks. Therefore, security of such systems is one of the critical requirements to guarantee the operation of various infrastructure and control systems without leading to catastrophic failures.

Recently, extensive efforts and studies have been dedicated on attack detection, prevention and resilient control designs from both theoretical research and real-world applications [8]–[12]. For instance, a polynomial fuzzy detection filter was designed in [13] to safeguard the system against faults and guarantee the stability and control performance. In [14], the dynamic response of a system under optimal data injection attacks was analyzed. The authors developed a switching condition to obtain the optimal attack sequence and a closed-form switching policy for data injection attacks. The attack detection and identification problem was considered in [15] for cyber-physical systems. Considering the fundamental limitations of a class of monitors, they designed the centralized and distributed monitors to completely detect and identify the attacks applied on the systems. In [16], the security of networked control systems was considered and the attack scenarios corresponding to denial-of-service, replay, zero-dynamics, and bias injection attacks were analyzed using the networked control framework. A hidden moving target defense approach was developed in [17] to improve the stealthiness which cannot be detected by the attackers. In [18], a false data detection mechanism was developed based on the separation of nominal power grid state and anomalies. Two methods were considered as the nuclear norm minimization and low rank matrix factorization to solve the false data injection problem. However, most of the above literature have focused on the setting that the system dynamics and attack information are known by the designers with varying degree of availability. With the ever increasing complexity and dimensionality of the control systems and communication networks, the explicit information of the system models is usually difficult or even unfeasible to achieve.

Fortunately, recent studies on reinforcement learning (RL) and adaptive dynamic programming (ADP) have made it possible to solve the feedback control problem with partially known or fully unknown system dynamics. By attempting to obtain the approximate solutions of the Hamilton–Jacobi–Bellman (HJB) equations, RL and ADP have been widely recognized as one of the core methodologies to achieve optimal control in stochastic process [19]–[23]. Extensive efforts and promising results have been achieved over the past decade [24]–[30], which demonstrate the effectiveness and performance of RL- and ADP-based learning methods without the requirement of explicit information of system models. In recent years, such techniques have also been studied in the game theory to estimate the solution of Hamilton-Jacobi-Isaacs (HJI) equation [31]–[36]. One of the popular problems is the two-player zero-sum game with the two players acting as the defensive and adversarial agents respectively [37]–[41]. This idea was further extended to the systems under external attacks to handle the adversary environment in the control and learning process [42]. Besides, to address the cyber-physical security in the net-

This work was supported in part by the National Science Foundation under Grant CNS-2047010, CNS-1947418, ECCS-1947419 and ECCS 2047064.

worked systems, RL and ADP techniques have been integrated into the attack detection problem to develop secure learning architectures. Particularly, a learning-based secure method was designed in [7] for cyber-physical systems under sensor and actuator attacks and a threat-detection level function was also developed to characterize the estimators which were used to detect the attacks. In [43], the compositional attacks such as the eavesdropping and covert attack were considered and solved by a proposed attack-resilient RL algorithm. For the attacks targeting the communication links, a networked attack detection residual was developed in [44] to determine the existence of attacks. The attack detection method with the event-triggered learning design was provided in [45] with explicit stability analysis.

Motivated by the above observations and literature studies, this paper develops an ADP-based intelligent and secure control method for a class of nonlinear systems under the attacks. The major contributions of this paper are as follows: First, this paper designs a state estimator-based detector to monitor the system and identify the attacks along the learning process. Second, we consider a secure switching mechanism to automatically switch the control strategy between the nominal and attack models. Since the impacts of the adversaries are only considered in the attack model, the computation complexity will be significantly reduced. Finally, we formulate the system in the attack model in a game-theoretical problem for attack mitigation, with one agent to minimize the performance index and one adversary to maximize it. In addition, the ADP techniques are designed to intelligently solve the switching problem. This design is suitable for both partially known and fully unknown system dynamics. Then, the neural network techniques are applied to implement the developed control method and the rigorous theoretical analysis is also provided to guarantee the control performance and safety. Comparing with [7], this paper considers a class of nonlinear systems which is more widely used in the real-world applications. Furthermore, comparing with [44] and [45], this paper develops an ADP control method with the critic networks design only in the learning process which reduces the computation complexity.

The rest of this paper is organized as follows. In Section II, we formulate the nonlinear system in both nominal and attack models. In order to automatically mitigate the attacks during the learning process, Section III designs an attack detector based on the state estimation techniques to monitor and identify the attacks applied on the system. In Section IV, we develop the ADP-based controller for secure learning and attack mitigation. When the measured state estimation error exceeds a threshold, an attack is alert which triggers the secure control process into a game-theoretical architecture. The neural network techniques are applied in Section V to implement the designed method. In Section VI, a numerical example is provided to show the necessary of the attack detector design and the effectiveness of the proposed ADP-based secure control method.

II. Problem Formulation

Consider the following continuous-time nonlinear system in the nominal condition

$$\dot{x}(t) = f(x(t)) + Bu(t) \tag{1}$$

where $x(t)$ is the state vector with the initial state as $x(0)$, $u(t)$ is the control input applied to the system, and B is the input matrix which is assumed known. The nonlinear function $f(x(t))$ with $f(0) = 0$ denotes the actual system function and is considered unknown in this paper.

Under the nominal condition, it is desired to find the control action $u(t) = \mathcal{K}(x(t))$ by optimizing the following cost function

$$J_n(x(t), u(t)) = \int_t^\infty \left[x^T(\tau)\mathcal{Q}_n x(\tau) + u^T(\tau)\mathcal{R}_n u(\tau)\right] d\tau \tag{2}$$

where $\mathcal{Q}_n$ and $\mathcal{R}_n$ are the symmetric and positive definite matrices in the nominal condition.

In this paper, we consider that the system and the controller are sending data through a vulnerable communication channel, which may make the data exchange altered. Most attacks and faults can be modeled by additive inputs on system actuator and sensor measurement. This paper will focus on the actuator attacks. Hence, the altered system dynamics becomes

$$\begin{cases} \dot{x}(t) & = f(x(t)) + Bu_a(t) \\ u_a(t) & = \mathcal{K}(x(t)) + D_a w(t) \end{cases} \tag{3}$$

where $u_a(t)$ is the measured control signal under attacks, D_a is the attack matrix, and $w(t)$ is the attack input. Hence, we can rewrite (3) as

$$\begin{aligned} \dot{x}(t) &= f(x(t)) + B\Big(\mathcal{K}(x(t)) + D_a w(t)\Big) \\ &= f(x(t)) + Bu(t) + BD_a w(t). \end{aligned} \tag{4}$$

Since the nonlinear function $f(x(t))$ is assumed unknown in this paper, we establish a multi-layer neural network to reconstruct the function as,

$$f(x(t)) = \zeta^{*T}\phi(x(t)) + \gamma(t) \tag{5}$$

where ζ^* denotes the ideal weights of the neural network which is bounded by $\|\zeta^*\| \le \zeta_M$, $\phi(x(t))$ is the bounded polynomial basis function with $\|\phi(x(t))\| \le \phi_M$, and $\gamma(t)$ is the neural network construction error which is bounded by $\|\gamma(t)\| \le \gamma_M$. Generally, the ideal weights ζ^* are difficult to achieve. Therefore, we consider the neural network output with the estimated weights $\hat{\zeta}(t)$, such that

$$\hat{f}(x(t)) = \hat{\zeta}^T(t)\phi(x(t)). \tag{6}$$

III. Attack Detector Design

In this section, a state estimator-based detector is designed to monitor the learning system and identify the attacks applied on the system. The detector is designed as

$$\dot{\hat{x}}(t) = \hat{\zeta}^T(t)\phi(x(t)) + Bu(t) - L\Big(x(t) - \hat{x}(t)\Big) \tag{7}$$

where $\hat{x}(t)$ is the estimated system state and L is the detector feedback gain matrix.

Define the state estimation error as $\tilde{x}(t) = x(t) - \hat{x}(t)$. By substituting (7) from (4), we have

$$\dot{\tilde{x}}(t) = L\tilde{x}(t) + \tilde{\zeta}^T(t)\phi(x(t)) + BD_a w(t) + \gamma(t) \tag{8}$$

where $\tilde{\zeta}(t) = \zeta^* - \hat{\zeta}(t)$ is the weights estimation error.

Theorem 1: Consider the nonlinear system (1) without any attacks. If the updating law for the detector (7) is given as

$$\dot{\hat{\zeta}}(t) = -\beta\left(\tilde{x}^T(t)L^{-1}\right)^T\phi(x(t)) \tag{9}$$

then the state estimation error $\tilde{x}(t)$ and the weights estimation error $\tilde{\zeta}(t)$ are uniformly ultimately bounded (UUB).

Proof: Consider matrix M as positive definite and satisfying $L^TM + ML = -\mathcal{U}$, where $\mathcal{U}$ is a symmetric positive definite matrix. Hence, define the Lyapunov function as

$$L_D = \frac{1}{2}\tilde{x}^T(t)M\tilde{x}(t) + \mathrm{tr}\left\{\tilde{\zeta}^T(t)\tilde{\zeta}(t)\right\} \tag{10}$$

where $\mathrm{tr}\{\cdot\}$ describes the matrix trace.

Considering (8) with $w(t) = 0$, we have the first derivative of (10) with respect to the system trajectory as

$$\begin{aligned}\dot{L}_D =& \frac{1}{2}\tilde{x}^T(t)\left(L^TM + ML\right)\tilde{x}(t) + \tilde{x}^T(t)M\left(\tilde{\zeta}^T(t)\phi(x(t))\right.\\ &\left.+ \gamma(t)\right) + \mathrm{tr}\left\{\tilde{\zeta}^T(t)\beta L^{-T}\tilde{x}(t)\phi(x(t))\right\}\\ \le& -\frac{1}{2}\lambda_{\min}(\mathcal{U})\|\tilde{x}(t)\|^2 + \|\tilde{x}(t)\|\|M\|\gamma_M\\ &+ \|\mathcal{L}_a\|\|\tilde{x}(t)\|\zeta_M\phi_M\end{aligned} \tag{11}$$

where $\mathcal{L}_a = \beta L^{-T} + M$. Define $\lambda_{\min}(\cdot)$ as the minimal eigenvalue of matrix. We have $\dot{L}_D < 0$ as long as the following condition is satisfied,

$$\|\tilde{x}(t)\| > \frac{2\left(\|M\|\gamma_M + \|\mathcal{L}_a\|\zeta_M\phi_M\right)}{\lambda_{\min}(\mathcal{U})} \doteq \Omega_x. \tag{12}$$

This completes the proof. ∎

Remark 1: Theorem 1 shows that without the network attacks, the state estimation error is UUB and the bound is Ω_x. This fact can be used in attack detection by considering the estimation error as the detection residual [45].

Remark 2: If the communication network is under attacks, the system dynamics become (3). Therefore, the state estimation error (8) in this situation contains the attack input $w(t) \neq 0$. Assume $\hat{x}(0) = x(0)$, i.e., $\tilde{x}(0) = 0$. Taking the integral on both sides of (8), we have

$$\tilde{x}(t) = \int_0^t \left(L\tilde{x}(\tau) + \tilde{\zeta}^T\phi(x(\tau)) + BD_a w(\tau) + \gamma(\tau)\right)d\tau. \tag{13}$$

According to the reverse triangle inequality, it becomes

$$\|\tilde{x}(t)\| \ge \left\|BD_a\int_0^t w(\tau)d\tau\right\| - \mathcal{G}_a \tag{14}$$

where

$$\mathcal{G}_a = \left\|\int_0^t \left(L\tilde{x}(\tau) + \tilde{\zeta}^T\phi(x(\tau)) + \gamma(\tau)\right)d\tau\right\|. \tag{15}$$

Considering (12), if the injected input satisfies

$$\begin{aligned}\left\|BD_a\int_0^t w(\tau)d\tau\right\| >& \Omega_x\\ &+ \left\|\int_0^t \left(L\tilde{x}(\tau) + \tilde{\zeta}^T\phi(x(\tau)) + \gamma(\tau)\right)d\tau\right\|\end{aligned} \tag{16}$$

we have

$$\|\tilde{x}(t)\| > \underbrace{\Omega_x + \left\|\int_0^t \left(L\tilde{x}(\tau) + \tilde{\zeta}^T\phi(x(\tau)) + \gamma(\tau)\right)d\tau\right\| - \mathcal{G}_a}_{\Omega_x} \tag{17}$$

which means a class of attacks can be detected by the designed detector. Therefore, the detector will constantly monitor the condition (12). Once the state estimation error is larger than the bound Ω_x, we consider the system is under attacks.

IV. Controller Design and Attack Mitigation

Based on Theorem 1, our designed detector can automatically identify whether the system is under attacks. Specifically, we consider the system in the nominal condition ($w(t) = 0$) at the beginning until the state estimation error exceeds the bound Ω_x. At this moment, the control mechanism considers that the system is under attacks ($w(t) \neq 0$) and switches to a zero-sum differential game problem with two players, i.e., $u(t)$ as the agent input and $w(t)$ as the adversary input. Therefore, instead of just solving $u(t)$ itself, the controller is designed to approximate both the optimal control signal and the worst-case attack input under this condition.

Nominal Condition: Let us start with the nominal condition. The control signal is provided to optimize the cost function (2). Hence, the optimal performance index will be defined as

$$V_n^*(x(t)) = \min_{u(t)}\int_t^\infty \left(x^T(\tau)\mathcal{Q}_n x(\tau) + u^T(\tau)\mathcal{R}_n u(\tau)\right)d\tau. \tag{18}$$

Assuming (18) is continuously differentiable and considering the system dynamics (1), we have the Hamiltonian function in the nominal condition as

$$\begin{aligned}\mathcal{H}_n\left(\frac{\partial V_n}{\partial x}, x, u\right) =& \frac{\partial V_n^{*T}(x(t))}{\partial x(t)}\left(f(x(t)) + Bu(t)\right)\\ &+ x^T(t)\mathcal{Q}_n x(t) + u^T(t)\mathcal{R}_n u(t).\end{aligned} \tag{19}$$

Then, the optimal control input $u^*(t)$ can be obtained based on the stationary condition for optimality $\partial\mathcal{H}_n/\partial u = 0$ and we have

$$u^*(t) = -\frac{1}{2}\mathcal{R}_n^{-1}B^T\frac{\partial V_n^*(x(t))}{\partial x(t)}. \tag{20}$$

Note that, in this paper, we consider the input matrix B is known. However, even if B is unknown in the design phase, it can be easily obtained from the detector design by calculating the derivative of (7) with respect to $u(t)$.

Attack Condition: When the state estimation error identified by the designed detector (7) exceeds the bound Ω_x, an attack alarm is triggered, at which moment, the control mechanism

switches to the game-theoretical architecture for attack mitigation.

Under this condition, the system can be provided in (3) with $w(t) \neq 0$. Hence, the cost function is augmented as

$$J_a(x,u,w) = \int_t^\infty \big[x^T(\tau)\mathcal{Q}_a x(\tau) + u^T(\tau)\mathcal{R}_a u(\tau) - \rho^2 w^T(\tau)w(\tau)\big]d\tau \quad (21)$$

where $\mathcal{Q}_a > 0$ and $\mathcal{R}_a > 0$ are the symmetric matrices, and ρ is the amount of attenuation of the attack input to the defined performance. In this way, we formulate the problem into a differential game between two players $u(t)$ and $w(t)$. Specifically, we assume the goal of player $u(t)$ is to regulate the system state to the origin and the player $w(t)$ is just the opposite. The term $\rho^2 w^T(t)w(t)$ represents the consideration of the subsequent energy consumption which is directly determined by $w(t)$. Note that even if the attacker can introduce arbitrary attack input to the system, in order to guarantee the control performance, we consider the worst-case attack input it can bring into rather than the actual one. Therefore, we assume the attacker intends to maximize the state deviation from the origin while minimizing $\rho^2 w^T(t)w(t)$, which means consuming as less energy as possible to cause a maximum damage. This is because an attacker cannot inject arbitrarily large false data due to the physical limitations of electronic instruments in the real world.

Hence, the optimal performance index in the attack condition becomes

$$V_a^*(x(t)) = \min_{u(t)} \max_{w(t)} J_a(x,u,w). \quad (22)$$

Assuming (22) is continuously differentiable, we have the Hamiltonian function in the attack condition as

$$\mathcal{H}_a\left(\frac{\partial V_a}{\partial x}, x, u, w\right) = \frac{\partial V_a^{*T}(x(t))}{\partial x(t)}\Big(f(x(t)) + Bu(t) + \mathcal{D}w(t)\Big) + \mathcal{F}(x(t),u(t),w(t)) \quad (23)$$

where $\mathcal{D} = BD_a$ and $\mathcal{F}(x(t),u(t),w(t)) = x^T(t)\mathcal{Q}_a x(t) + u^T(t)\mathcal{R}_a u(t) - \rho^2 w^T(t)w(t)$. The optimal solution satisfies the first order necessary condition. Hence, we obtain the optimal control input as

$$\frac{\partial \mathcal{H}_a}{\partial u} = 0 \longrightarrow u^*(t) = -\frac{1}{2}\mathcal{R}_a^{-1}B^T\frac{\partial V_a^*(x(t))}{\partial x(t)} \quad (24)$$

and the worst-case attack input as

$$\frac{\partial \mathcal{H}_a}{\partial w} = 0 \longrightarrow w^*(t) = \frac{1}{2\rho^2}\mathcal{D}^T\frac{\partial V_a^*(x(t))}{\partial x(t)}. \quad (25)$$

The solution $(u^*(t), w^*(t))$ given in (24) and (25) generates a saddle point solution, such that

$$J_a(x(0),u^*,w) \leq J_a(x(0),u^*,w^*) \leq J_a(x(0),u,w^*) \quad (26)$$

where $J_a(x(0),u^*,w^*) = V_a^*(x(0))$.

Now we will show that the system can achieve Nash equilibrium with the solution $(u^*(t), w^*(t))$. Since the equilibrium point of the system is at the origin, we have $x(t) = 0$ when $t \to \infty$. Therefore, the cost function satisfies $J_a(0,0,0) = 0$. The time derivative of the optimal cost function becomes

$$\dot{J}_a(x(0),u^*,w^*) = \frac{\partial J_a(x(0),u^*,w^*)}{\partial x(t)}\dot{x}(t). \quad (27)$$

Considering the optimality $\mathcal{H}_a(\frac{\partial V_a}{\partial x}, x, u^*, w^*) = 0$ and $J_a(x(0),u^*,w^*) = V_a^*(x(0))$, we have

$$\dot{J}_a(x(0),u^*,w^*) + x^T(t)\mathcal{Q}_a x(t) + u^{*T}(t)\mathcal{R}_a u^*(t) - \rho^2 w^{*T}(t)w^*(t) = 0. \quad (28)$$

Taking the integral on both sides of (28), it becomes

$$J_a(x(0),u^*,w^*) + \int_0^\infty \Big[x^T(\tau)\mathcal{Q}_a x(\tau) + u^{*T}(\tau)\mathcal{R}_a u^*(\tau) - \rho^2 w^{*T}(\tau)w^*(\tau)\Big]d\tau = 0. \quad (29)$$

Hence, we add (29) on the cost function (21) and consider the starting point as $t = 0$. It becomes

$$J_a(x(0),u,w) = \int_0^\infty \mathcal{F}(x(t),u(t),w(t))d\tau + J_a(x(0),u^*,w^*) + \int_0^\infty \Big[x^T(\tau)\mathcal{Q}_a x(\tau) + u^*(\tau)\mathcal{R}_a u^*(\tau) - \rho^2 w^*(\tau)w^*(\tau)\Big]d\tau. \quad (30)$$

Considering (23), (24) and (25), we can further rewrite

$$J_a(x(0),u,w) = J_a(x(0),u^*,w^*) + \int_0^\infty \Big[x^T(\tau)\mathcal{Q}_a x(\tau) + u^T(\tau)\mathcal{R}_a u(\tau) - \rho^2 w^T(\tau)w(\tau) + \frac{\partial V^*(x(\tau))}{\partial x(\tau)} \cdot \big(f(x(\tau)) + Bu^*(\tau) + \mathcal{D}w^*(\tau)\big)\Big]d\tau \quad (31)$$

$$= J_a(x(0),u^*,w^*) + \int_0^\infty \Big[\big(u(t)-u^*(t)\big)^T\mathcal{R}_a\big(u(t) - u^*(t)\big) - \rho^2\big(w(t)-w^*(t)\big)^T\big(w(t)-w^T(t)\big)\Big]d\tau. \quad (32)$$

Therefore, condition (26) is satisfied and the Nash equilibrium is achieved.

V. Online Learning and Stability Analysis

We design two critic networks (i.e., nominal and attack critic networks) to estimate the performance index $V_n(x(t))$ and $V_a(x(t))$ respectively,

$$V_n(x(t)) = \theta_n^{*T}\psi(x,u) + \epsilon(x(t)) \quad (33)$$

$$V_a(x(t)) = \theta_a^{*T}\varphi(x,u,w) + \varepsilon(x(t)) \quad (34)$$

where θ_n^* and θ_a^* are the ideal neural network weights, $\psi(x,u)$ and $\varphi(x,u,w)$ are the activation functions, and $\epsilon(\cdot)$ and $\varepsilon(\cdot)$ are the bounded neural network errors for the nominal and attack critic networks, respectively. The controller will switch between these two conditions based on the designed detector (7). When the system is considered in the nominal condition, we have $\frac{\partial V_n(x(t))}{\partial x(t)} = \psi_{\mathcal{X}}^T(x,u)\theta_n^* + \epsilon_{\mathcal{X}}(x(t))$, where $\psi_{\mathcal{X}}(x,u) = \frac{\partial \psi(x,u)}{\partial x(t)}$ and $\epsilon_{\mathcal{X}}(x(t)) = \frac{\epsilon(x(t))}{\partial x(t)}$. Because the ideal weights θ_n^* is difficult to achieve, we estimate the value as

$\hat{\theta}_n(t)$ and consider the corresponding estimated output of the nominal critic network as

$$\hat{V}_n(x(t)) = \hat{\theta}_n^T(t)\psi(x,u) \tag{35}$$

and its derivative as

$$\frac{\partial \hat{V}_n(x(t))}{\partial x(t)} = \psi_{\mathcal{X}}^T(x,u)\hat{\theta}_n(t). \tag{36}$$

Define the error function for the nominal critic network as $e_{c1} = \mathcal{H}\big(\frac{\partial V_n}{\partial x}, x, u\big) - \mathcal{H}\big(\frac{\partial V_n^*}{x}, x, u^*\big)$. Since the latter term is zero, we have the updating law as

$$\dot{\hat{\theta}}_n(t) = -\beta_n \Upsilon(t)\Big(\psi_{\mathcal{X}}^T(x,u)\dot{x}^T(t)\hat{\theta}_n + x^T(t)\mathcal{Q}_n x(t) + u^T(t)\mathcal{R}_n u(t)\Big) \tag{37}$$

where $\Upsilon(t) = \frac{\upsilon(t)}{(\upsilon^T(t)\upsilon(t)+1)^2}$, $\upsilon(t) = \psi_{\mathcal{X}}(x,u)\dot{x}(t)$ and $\beta_n > 0$ is the learning rate of the nominal critic network. Hence, we have the estimated control input when the system is in the nominal condition as

$$u(t) = -\frac{1}{2}\mathcal{R}_n^{-1}B^T\psi_{\mathcal{X}}^T(x,u)\hat{\theta}_n(t). \tag{38}$$

The attack detector will constantly monitor the system dynamics. When the state estimation error exceeds the bound Ω_x, we consider the system is under attacks. At this moment, the controller will automatically switch to the game-theoretical architecture and consider the zero-sum control policy. Therefore, the attack critic network will be applied and the derivative of the performance index becomes

$$\frac{\partial \hat{V}_a(x(t))}{\partial x(t)} = \varphi_{\mathcal{X}}^T(x,u,w)\hat{\theta}_a(t) \tag{39}$$

which is the estimated version of derivative $\frac{\partial V_a(x(t))}{\partial x(t)} = \varphi_{\mathcal{X}}^T(x,u,w)\theta_a^* + \varepsilon_{\mathcal{X}}(x(t))$ with the estimated neural network weights $\hat{\theta}_a(t)$, where $\varphi_{\mathcal{X}}(x,u,w) = \frac{\partial \varphi(x,u,w)}{\partial x(t)}$ and $\varepsilon_{\mathcal{X}}(x(t)) = \frac{\varepsilon(x(t))}{\partial x(t)}$. Consider the error function $e_{c2} = \mathcal{H}\big(\frac{\partial V_n}{\partial x}, x, u, w\big) - \mathcal{H}\big(\frac{\partial V_n^*}{x}, x, u^*, w^*\big)$, the attack critic network weights will be updated as

$$\dot{\hat{\theta}}_a(t) = -\beta_a \Delta(t)\Big(\varphi_{\mathcal{X}}^T(x,u,w)\dot{x}^T(t)\hat{\theta}_a + x^T(t)\mathcal{Q}_a x(t) + u^T(t)\mathcal{R}_a u(t) - \rho^2 w^T(t)w(t)\Big) \tag{40}$$

where $\Delta(t) = \frac{\delta(t)}{(\delta^T(t)\delta(t)+1)^2}$, $\delta(t) = \varphi_{\mathcal{X}}(x,u,w)\dot{x}(t)$ and $\beta_a > 0$ is the learning rate of the attack critic network. Under this condition, the estimated optimal control input becomes

$$u(t) = -\frac{1}{2}\mathcal{R}_a^{-1}B^T\varphi_{\mathcal{X}}^T(x,u,w)\hat{\theta}_a(t) \tag{41}$$

and the estimated worst-case attack input can also be obtained as

$$w(t) = \frac{1}{2\rho^2}\mathcal{D}^T\varphi_{\mathcal{X}}^T(x,u,w)\hat{\theta}_a(t). \tag{42}$$

Note that upon detection of the attack, the dynamics of the designed detector becomes

$$\dot{\hat{x}}(t) = \hat{\zeta}^T(t)\phi(x(t)) + Bu(t) - L\big(x(t) - \hat{x}(t)\big) + \frac{1}{2\rho^2}\mathcal{D}\mathcal{D}^T\varphi_{\mathcal{X}}^T(x,u,w)\hat{\theta}_a(t) \tag{43}$$

which means the attack input is compensated to the dynamics. The detector will maintain the dynamics (43) until the state estimation error exceeds Ω_x again, after which moment, the controller will reduce to the nominal optimization design.

The following theorem provides the stability of the closed-loop design.

Theorem 2: Consider the system (1) with an attack detector (7). Two critic networks (33) and (34) are established with one operating in the nominal condition updated as (37) and another one operating in the attack condition with the updating law (40). Then the control and attack inputs are estimated as (38), (41) and (42) for nominal and attack conditions, respectively. Then, all the signals for the closed-loop system are UUB.

Proof: Define the following Lyapunov function:

$$L_{\mathcal{C}} = V_n(x(t)) + V_a(x(t)) + \beta_n^{-1}\,\mathrm{tr}\left\{\tilde{\theta}_n^T(t)\tilde{\theta}_n(t)\right\} + \beta_a^{-1}\,\mathrm{tr}\left\{\tilde{\theta}_a^T(t)\tilde{\theta}_a(t)\right\} \tag{44}$$

where $\tilde{\theta}_n(t) = \theta_n^* - \hat{\theta}_n(t)$ and $\tilde{\theta}_a(t) = \theta_a^* - \hat{\theta}_a(t)$ are the errors of the nominal and attack critic network weights, respectively. From (37) and (40), we have $\dot{\tilde{\theta}}_n(t) = \beta_n\Upsilon(t)\big(\upsilon^T(t)\hat{\theta}_n(t) + x^T(t)\mathcal{Q}_n x(t) + u^T(t)\mathcal{R}_n u(t)\big)$ and $\dot{\tilde{\theta}}_a(t) = \beta_a\Delta(t)\big(\delta^T(t)\hat{\theta}_a(t) + x^T(t)\mathcal{Q}_a x(t) + u^T(t)\mathcal{R}_a u(t) - \rho^2 w^T(t)w(t)\big)$.

This boundedness proof is carried out in two parts for the nominal and attack dynamics. We will show that both dynamics of the closed-loop design are UUB. Let us start with the nominal dynamics without any attack inputs. We take the time derivative of (44) and obtain

$$\begin{aligned}\dot{L}_{\mathcal{C}} =& \frac{\partial V_n^T(x(t))}{\partial x(t)}\dot{x}(t) + \beta_n^{-1}\,\mathrm{tr}\left\{\tilde{\theta}_n^T(t)\dot{\tilde{\theta}}_n(t)\right\} \\ \le& -\lambda_{\min}(\mathcal{Q}_n)\|x(t)\|^2 - \lambda_{\min}(\mathcal{R}_n)\|u(t)\|^2 \\ &+ \beta_n^{-1}\,\mathrm{tr}\left\{\beta_n\tilde{\theta}_n^T(t)\frac{\upsilon(t)}{(\upsilon^T(t)\upsilon(t)+1)^2}\right. \\ &\left.\cdot\Big(\upsilon^T(t)\hat{\theta}_n(t) + x^T(t)\mathcal{Q}_n x(t) + u^T(t)\mathcal{R}_n u(t)\Big)\right\}.\end{aligned} \tag{45}$$

where $\lambda_{\min}(\cdot)$ denotes the minimal eigenvalue of matrix. This deviation is achieved by considering the Hamiltonian function $\mathcal{H}(\frac{\partial V_n^*}{\partial x}, x, u^*) = 0$. Since $\hat{\theta}_n(t) = \theta_n^* - \tilde{\theta}_n(t)$, we have

$$\begin{aligned}\dot{L}_{\mathcal{C}} \le& -\lambda_{\min}(\mathcal{Q}_n)\|x(t)\|^2 - \lambda_{\min}(\mathcal{R}_n)\|u(t)\|^2 - \big\|\mathscr{V}(t)\big\|^2 \\ &\cdot\big\|\tilde{\theta}_n(t)\big\|^2 + \frac{1}{2\beta_n}\left(\beta_n^2\big\|\mathscr{V}(t)\big\|^2\big\|\tilde{\theta}_n(t)\big\|^2\right. \\ &\left.+\left\|\frac{1}{\upsilon^T(t)\upsilon(t)+1}\right\|^2\|\mathscr{M}(t)\|^2\right)\end{aligned} \tag{46}$$

where $\mathscr{V}(t) = \frac{\upsilon(t)}{\upsilon^T(t)\upsilon(t)+1}$ and $\mathscr{M}(t) = \upsilon^T(t)\theta_n^* + x^T(t)\mathcal{Q}_n x(t) + u^T(t)\mathcal{R}_n u(t)$. Therefore, we can further rewrite (46) as

$$\begin{aligned}\dot{L}_{\mathcal{C}} \leq & -\lambda_{\min}(\mathcal{Q}_n)\|x(t)\|^2 - \lambda_{\min}(\mathcal{R}_n)\|u(t)\|^2 \\ & -\left(1-\frac{\beta_n}{2}\right)\|\mathscr{V}(t)\|^2\|\tilde{\theta}_n(t)\|^2 \\ & +\frac{\|\mathscr{M}(t)\|^2}{2\beta}\left\|\frac{1}{\upsilon^T(t)\upsilon(t)+1}\right\|^2 \end{aligned} \tag{47}$$

Therefore, if the following conditions are satisfied

$$\beta_n < 2, \quad \|\tilde{\theta}_n(t)\|^2 > \beta_n^{\mathscr{P}}\frac{\|\mathscr{M}(t)\|^2}{\|\upsilon(t)\|^2} \tag{48}$$

where $\beta_n^{\mathscr{P}} = \frac{1}{2\beta_n(1-\beta_n/2)}$, we have $\dot{L} < 0$. This means the nominal dynamics of the closed-loop design are UUB.

When an attack alert is triggered based on the detector design in Theorem 1, the system operates in attack dynamics. At this moment, the attack critic network is applied with the updating law (40). Therefore, the time derivative of the Lyapunov function becomes

$$\begin{aligned}\dot{L}_{\mathcal{C}} =& \frac{\partial V_a(x(t))}{\partial x(t)}\dot{x}(t) + \beta_a^{-1}\,\mathrm{tr}\left\{\tilde{\theta}_a^T(t)\dot{\tilde{\theta}}_a(t)\right\} \\ =& -\left(x^T(t)\mathcal{Q}_a x(t) + u^T(t)\mathcal{R}_a u(t) - \rho^2 w^T(t)w(t)\right) \\ & +\beta_a^{-1}\,\mathrm{tr}\Big\{\beta_a\tilde{\theta}_a^T(t)\Delta(t)\Big(\delta^T(t)\hat{\theta}_a(t) + x^T(t)\mathcal{Q}_a x(t) \\ & +u^T(t)\mathcal{R}_a u(t) - \rho^2 w^T(t)w(t)\Big)\Big\}. \end{aligned} \tag{49}$$

Assuming the condition $B\mathcal{R}_a^{-1}B^T \succ \mathcal{D}^T\mathcal{D}$ is satisfied, we have

$$\begin{aligned}\dot{L}_{\mathcal{C}} \leq & -x^T(t)\mathcal{Q}_a x(t) + \beta_a^{-1}\,\mathrm{tr}\Big\{-\beta_a\tilde{\theta}_a^T(t)\Delta(t)\delta^T(t)\tilde{\theta}_a(t) \\ & +\beta_a\tilde{\theta}_a^T(t)\Delta(t)\Big(\delta^T(t)\theta_a^* + x^T(t)\mathcal{Q}_a x(t) \\ & +u^T(t)\mathcal{R}_a u(t) - \rho^2 w^T(t)w(t)\Big)\Big\} \\ \leq & -\lambda_{\min}(\mathcal{Q}_a)\|x(t)\|^2 - \left(1-\frac{\beta_a}{2}\right)\|\mathscr{W}(t)\|^2\|\tilde{\theta}_a\|^2 \\ & +\frac{1}{2\beta_a}\|\mathscr{W}(t)\|^2\frac{\|\mathscr{N}(t)\|^2}{\|\delta(t)\|^2} \end{aligned} \tag{50}$$

where $\mathscr{N}(t) = \delta^T(t)\theta_a^* + x^T(t)\mathcal{Q}_a x(t) + u^T(t)\mathcal{R}_a u(t) - \rho^2 w^T(t)w(t)$ and $\mathscr{W}(t) = \Delta(t)(\delta^T(t)\delta(t)+1)$. Hence, we have $\dot{L}_{\mathcal{C}} < 0$ if the following conditions are satisfied,

$$\beta_a < 2, \quad \|\tilde{\theta}_a(t)\|^2 > \beta_a^{\mathscr{P}}\frac{\|\mathscr{N}(t)\|^2}{\|\delta(t)\|^2} \tag{51}$$

where $\beta_a^{\mathscr{P}} = \frac{1}{2\beta_a(1-\beta_a/2)}$. Therefore, the attack dynamics of the designed closed-loop system is also UUB, which completes the proof. ■

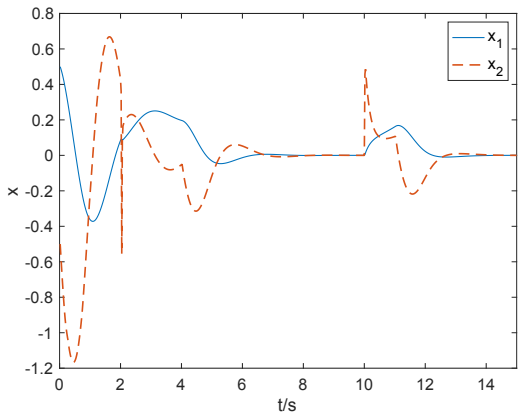

Fig. 1. The trajectories of system states with $x_1(t) = \eta(t)$ and $x_2(t) = \alpha(t)$ under the proposed ADP-based secure control approach.

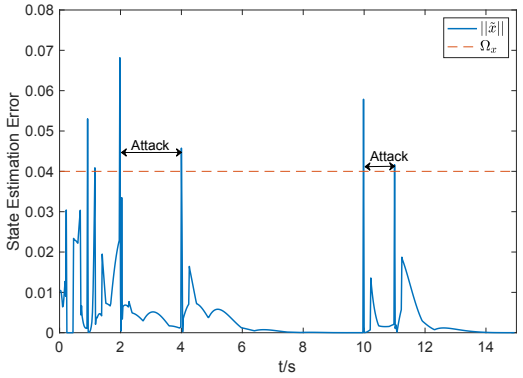

Fig. 2. The comparison of state estimation error $\|\tilde{x}(t)\|$ and threshold Ω_x.

VI. Simulation Results

This section considers a nonlinear continuous-time system to verify the performance of the designed ADP-based secure control method with attack detection. The system dynamics are described as follows:

$$\begin{cases} \dot{\eta}(t) = \alpha(t) \\ J\dot{\alpha}(t) = -Mgl\sin\eta(t) - f_d\dot{\eta}(t) + u(t). \end{cases} \tag{52}$$

This is a torsional pendulum system [46] with the system states $x(t) = [x_1(t), x_2(t)] = [\eta(t), \alpha(t)]$ which describe the angle position and the angular velocity of the pendulum, receptively. The input signal is denoted as $u(t)$. Hence, the matrix $B = [0; 1]$. Other parameters include that $M = 1/3$kg and $l = 2/3$m are the mass and the length of the pendulum, respectively, $J = 4/3$kg·m^2 is the rotary inertia, $g = 9.8$m/s^2 is the acceleration of gravity, and $f_d = 0.2$N·m·s/rad is the frictional factor. Suppose that over the time intervals $2s \leq t \leq 4s$ and $10s \leq t \leq 11s$, an attacker has access to the actuator and launches an attack signal. During the attack, the input signal becomes $u_a(t) = u(t) + w(t)$ which means $D_a = 1$.

A state estimator-based detector is designed to identify the attacks. Based on Theorem 1, we design the attack threshold as $\Omega_x = 0.04$. When an attack alarm is triggered, the control

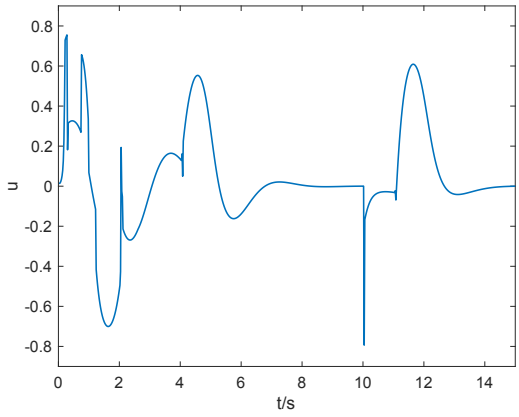

Fig. 3. The trajectory of control input $u(t)$ under the proposed ADP-based secure control approach.

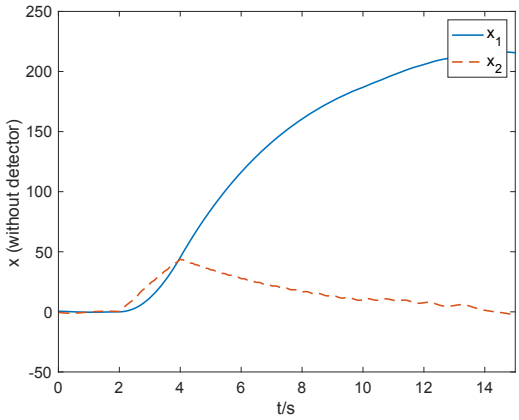

Fig. 4. The trajectories of system states under the conventional ADP-based approach without the detector design.

mechanism will automatically switch to the game-theoretical architecture and the dynamics of the detector will become (43). The control will continue under this architecture until the state estimation error exceeds Ω_x again, at which moment the optimization will reduce back to the nominal condition. Two critic networks are established. The inputs for these two neural networks are designed as $[x_1(t), x_2(t), u(t)]$ and $[x_1(t), x_2(t), u(t), w(t)]$, respectively. The initial weights are chosen randomly within $[-0.5, 0.5]$. The control signals and adversary input are estimated based on (38), (41) and (42).

Assume the initial system state as $x(0) = [-0.5, 0.5]$. The trajectories of the system states during this learning process are shown in Fig. 1. We observe that there is a sharp drop/increase at $t = 2s$ and $t = 10s$, respectively, due to the attacks. Based on the designed ADP-based secure control method, the state can be quickly stabilized to the equilibrium point. The comparison of the state estimation errors $\|\tilde{x}(t)\|$ and the attack threshold Ω_x is provided in Fig. 2. It is shown that at time $t = 2s$ and $t = 10s$, the state estimation error exceeds the threshold Ω_x which trigger the attack alarm. Therefore, the controller switches to the game-theoretical architecture for attack mitigation until $\|\tilde{x}(t)\|$ exceeds Ω_x again which are at $t = 4s$ and $t = 11s$, respectively. Note that during the learning process, there exists the situation that $\|\tilde{x}(t)\|$ exceeds Ω_x, but there is no attack applied on the system, e.g. at $t = 0.92s$. This is because the detector is learning online in this method and may cause some noise during the learning process. However, even though this happens, the developed controller can quickly correct itself to switch back to the nominal situation. The trajectory of the control input in this process is shown in Fig. 3. In addition, to further show the effectiveness of the developed ADP-based secure control method, we also apply the conventional ADP control process without the attack detector on the system and provide the results in Fig. 4. It is shown that only applying the nominal controller cannot stabilize the system due to the attacks. This comparison further demonstrate the necessary of the detector design for attack mitigation and the effectiveness of the developed secure control method.

VII. Conclusion

In this paper, we developed an ADP-based secure control method with attack detection for a class of nonlinear systems. An attack detector was designed based on the state estimation approaches to monitor and identify the attacks on the system. When the state estimation error exceeded the threshold, an attack was alert which triggers the control process to a game-theoretical architecture for attack mitigation. Two critic networks were established for both nominal and attack conditions to implement the control method with explicit stability analysis. The numerical example verified the performance of the designed attack detector and demonstrated the effectiveness of the proposed method.

References

[1] L. Sedghi, Z. Ijaz, M. Noor-A-Rahim, K. Witheephanich, and D. Pesch, "Machine learning in event-triggered control: Recent advances and open issues," 2020.

[2] Y. Cao, W. Yu, W. Ren, and G. Chen, "An overview of recent progress in the study of distributed multi-agent coordination," *IEEE Transactions on Industrial informatics*, vol. 9, no. 1, pp. 427–438, 2013.

[3] J. Lin, A. S. Morse, and B. D. Anderson, "The multi-agent rendezvous problem-the asynchronous case," in *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, vol. 2, pp. 1926–1931, IEEE, 2004.

[4] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on automatic control*, vol. 49, no. 9, pp. 1520–1533, 2004.

[5] C. P. Chen, G.-X. Wen, Y.-J. Liu, and F.-Y. Wang, "Adaptive consensus control for a class of nonlinear multiagent time-delay systems using neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 6, pp. 1217–1226, 2014.

[6] G. Wen, C. P. Chen, Y.-J. Liu, and Z. Liu, "Neural network-based adaptive leader-following consensus control for a class of nonlinear multiagent state-delay systems," *IEEE transactions on cybernetics*, vol. 47, no. 8, pp. 2151–2160, 2017.

[7] 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*, 2020, in press.

[8] F. Pasqualetti, F. Dorfler, and F. Bullo, "Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 110–127, 2015.

[9] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, "A systems and control perspective of cps security," *Annual Reviews in Control*, vol. 47, pp. 394–411, 2019.

[10] L. An and G.-H. Yang, "Secure state estimation against sparse sensor attacks with adaptive switching mechanism," *IEEE Transactions on Automatic Control*, vol. 63, no. 8, pp. 2596–2603, 2017.

[11] E. Mousavinejad, F. Yang, Q.-L. Han, and L. Vlacic, "A novel cyber attack detection method in networked control systems," *IEEE transactions on cybernetics*, vol. 48, no. 11, pp. 3254–3264, 2018.

[12] A. Ameli, A. Hooshyar, E. F. El-Saadany, and A. M. Youssef, "Attack detection and identification for automatic generation control systems," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4760–4774, 2018.

[13] H. Li, Z. Chen, L. Wu, H.-K. Lam, and H. Du, "Event-triggered fault detection of nonlinear networked systems," *IEEE transactions on cybernetics*, vol. 47, no. 4, pp. 1041–1052, 2016.

[14] G. Wu, J. Sun, and J. Chen, "Optimal data injection attacks in cyber-physical systems," *IEEE transactions on cybernetics*, vol. 48, no. 12, pp. 3302–3312, 2018.

[15] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE transactions on automatic control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[16] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, 2015.

[17] J. Tian, R. Tan, X. Guan, and T. Liu, "Enhanced hidden moving target defense in smart grids," *IEEE transactions on smart grid*, vol. 10, no. 2, pp. 2208–2223, 2018.

[18] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 612–621, 2014.

[19] P. J. Werbos, "Intelligence in the brain: A theory of how it works and how to build it," *Neural Networks*, vol. 22, no. 3, pp. 200–212, 2009.

[20] P. J. Werbos, "Using adp to understand and replicate brain intelligence: The next level design?," *Neurodynamics of cognition and consciousness*, pp. 109–123, 2007.

[21] B. R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A. A. Al Sallab, S. Yogamani, and P. Pérez, "Deep reinforcement learning for autonomous driving: A survey," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[22] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, *et al.*, "Mastering the game of go with deep neural networks and tree search," *nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[23] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, *et al.*, "Mastering the game of go without human knowledge," *nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[24] R. A. Brooks, "Intelligence without reason," in *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, Sydney, New South Wales, Australia, 1991.

[25] R. Pfeifer and C. Scheier, *Understanding Intelligence*. MIT Press, Cambridge, MA, 1999.

[26] J. Si, A. G. Barto, W. B. Powell, and D. W. II, eds., *Handbook of Learning and Approximate Dynamic Programming*. Wiley-IEEE, 2004.

[27] F. L. Lewis and D. Liu, eds., *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*. Wiley-IEEE, 2012.

[28] H. Zhang, D. Liu, Y. Luo, and D. Wang, *Adaptive Dynamic Programming for Control: Algorithms and Stability* . London: Springer, 2013.

[29] Y. Jiang and Z. Jiang, "Robust adaptive dynamic programming and feedback stabilization of nonlinear systems," *Neural Networks and Learning Systems, IEEE Transactions on*, vol. 25, no. 5, pp. 882–893, 2014.

[30] X. Zhong, H. He, and D. V. Prokhorov, "Robust controller design of continuous-time nonlinear system using neural network," in *Proc. Int. Joint Conf. Neural Networks*, pp. 1–8, 2013.

[31] T. Dierks and S. Jagannathan, "Optimal control of affine nonlinear continuous-time systems using an online hamilton-jacobi-isaacs formulation," in *49th IEEE Conference on Decision and Control (CDC)*, pp. 3048–3053, IEEE, 2010.

[32] K. G. Vamvoudakis and F. L. Lewis, "Online solution of nonlinear two-player zero-sum games using synchronous policy iteration," *International Journal of Robust and Nonlinear Control*, vol. 22, no. 13, pp. 1460–1483, 2012.

[33] X. Zhong, H. He, D. Wang, and Z. Ni, "Model-free adaptive control for unknown nonlinear zero-sum differential game," *IEEE transactions on cybernetics*, vol. 48, no. 5, pp. 1633–1646, 2017.

[34] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal Control*. Wiley, Hoboken, NJ, USA, 2012.

[35] K. G. Vamvoudakis, "Game-theoretic tracking control for actuator attack attenuation in cyber-physical systems," in *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 4233–4240, IEEE, 2016.

[36] C. Sun and K. G. Vamvoudakis, "Continuous-time safe learning with temporal logic constraints in adversarial environments," in *2020 American Control Conference (ACC)*, pp. 4786–4791, IEEE, 2020.

[37] L. Koçkesen and E. A. Ok, "An introduction to game theory," *University Efe A. Ok New York University July*, vol. 8, 2007.

[38] T. Başar and P. Bernhard, *H_∞ optimal control and related minimax design problems: a dynamic game approach*. Springer Science & Business Media, 2008.

[39] L. Wei and Z. Wu, "Recursive zero-sum stochastic differential game," in *2008 International Conference on Intelligent Computation Technology and Automation (ICICTA)*, vol. 2, pp. 998–1001, IEEE, 2008.

[40] C. Qin, H. Zhang, and Y. Luo, "Model-free adaptive dynamic programming for online optimal solution of the unknown nonlinear zero-sum differential game," in *2014 International Joint Conference on Neural Networks (IJCNN)*, pp. 3815–3820, IEEE, 2014.

[41] K. G. Vamvoudakis and F. R. P. Safaei, "Stochastic zero-sum nash games for uncertain nonlinear markovian jump systems," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pp. 5582–5589, IEEE, 2017.

[42] K. G. Vamvoudakis and J. P. Hespanha, "Cooperative q-learning for rejection of persistent adversarial inputs in networked linear quadratic systems," *IEEE Transactions on Automatic Control*, vol. 63, no. 4, pp. 1018–1031, 2017.

[43] S. Mukherjee and V. Adetola, "A secure learning control strategy via dynamic camouflaging for unknown dynamical systems under attacks," *arXiv preprint arXiv:2102.00573*, 2021.

[44] H. Niu and S. Jagannathan, "Neural network-based attack detection in nonlinear networked control systems," in *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 4249–4254, IEEE, 2016.

[45] H. Niu, C. Bhowmick, and S. Jagannathan, "Attack detection and approximation in nonlinear networked control systems using neural networks," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 1, pp. 235–245, 2019.

[46] D. Liu and Q. Wei, "Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 3, pp. 621–634, 2013.