ISA Transactions xxx (xxxx) xxx

Contents lists available at ScienceDirect

ISA Transactions

journal homepage: www.elsevier.com/locate/isatrans

Research article

Critic-only adaptive dynamic programming algorithms' applications to the secure control of cyber–physical systems

He Jiang [a], Huaguang Zhang [a,*], Xiangpeng Xie [b]

[a] College of Information Science and Engineering, Northeastern University, Box 134, 110819, Shenyang, PR China
[b] Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, 210003, Nanjing, PR China

HIGHLIGHTS

- Secure control problems are converted into zero-sum game issues.
- Two mainstream ADP methods are introduced to solve HJI equations.
- Tuning conditions of secure control parameters are derived.
- The proposed control scheme is tested on the Quanser helicopter.

ARTICLE INFO

Article history:
Received 26 September 2018
Received in revised form 22 January 2019
Accepted 14 February 2019
Available online xxxx

Keywords:
Cyber–physical systems
Adaptive dynamic programming
Reinforcement learning
Neural networks

ABSTRACT

Industrial cyber–physical systems generally suffer from the malicious attacks and unmatched perturbation, and thus the security issue is always the core research topic in the related fields. This paper proposes a novel intelligent secure control scheme, which integrates optimal control theory, zero-sum game theory, reinforcement learning and neural networks. First, the secure control problem of the compromised system is converted into the zero-sum game issue of the nominal auxiliary system, and then both policy-iteration-based and value-iteration-based adaptive dynamic programming methods are introduced to solve the Hamilton–Jacobi–Isaacs equations. The proposed secure control scheme can mitigate the effects of actuator attacks and unmatched perturbation, and stabilize the compromised cyber–physical systems by tuning the system performance parameters, which is proved through the Lyapunov stability theory. Finally, the proposed approach is applied to the Quanser helicopter to verify the effectiveness.

1. Introduction

With the developments of computational intelligence and network communication, cyber–physical systems (CPSs) [1,2] build an intelligent bridge between the cyber-world and the physical world. Recently, CPSs emerge in many industrial applications including power systems [3], communication networks, transportation systems [4], aerospace systems and health-care systems.

In the fields of control theory and control engineering, the security issue is one of the hot topics for CPSs. That is because the integration of the industrial control systems with advanced communication technologies may cause system perturbation, component failures and adversarial attacks, which affect the system stability and degrade the control performance. Up to now, there have been several significant works concerning the security issues of CPSs. For the large-scale systems, the signal transmissions usually suffer from time-delays and packet losses due to the physical limitation of communication. Therefore, for the CPSs with incomplete measurements, the topic of the state estimation deserves much attention. In [5], based on the locally received information, a novel distributed filtering approach was proposed for the fuzzy systems with sensor saturation and packet dropouts. In [6], the asynchronous state estimation for switched complex networks with communication constraints was addressed, and the associated estimator gain was obtained by solving a convex optimization problem. In some occasions, the CPSs may undergo the threats from attackers, such as sensor attacks and actuator attacks. In [7], a new state observer was designed by means of the adaptive switching mechanism for the CPSs with sensor attacks. In [8], the time-varying sensor and actuator attacks were both considered by using an adaptive control law which ensured the stability results were uniformly ultimately bounded. Based on the framework of [8], an improved adaptive resilient control policy was presented against the cyber-attacks in [9]. For the

* Corresponding author.
E-mail addresses: jianghescholar@163.com (H. Jiang), hgzhang@ieee.org (H. Zhang), xiexiangpeng1953@163.com (X. Xie).

https://doi.org/10.1016/j.isatra.2019.02.012
0019-0578/

issues of component faults or failures, the fault-tolerant control techniques [10–12] were regarded as the effective tools. In addition to secure control, the optimal control problem of CPSs also receives much attention, because optimal control can not only guarantee the system stability but also minimize the control energy and reduce control costs. The existing works regarding the secure control generally neglect the optimal control design, and thus the control performance may not be guaranteed. To the best of our knowledge, there are still few works considering the security issues of CPSs through the optimal control design, which motivates the research of this paper.

Adaptive dynamic programming (ADP) [13–15], an important branch of reinforcement learning (RL) [16–18], is a powerful tool in solving various optimal control problems such as robust optimal control [19–21], constrained optimal control [22–24], optimal tracking control [25–27], zero-sum game [28–30] and non-zero-sum game [31,32]. As is known, H_∞ robust control issues can be converted into zero-sum games [28,33,34]. The solutions of zero-sum games rely on Hamilton–Jacobi–Isaacs (HJI) equations. It is generally difficult or even impossible to solve HJI equations, especially for the nonlinear cases. In [33], an iterative simultaneous policy update algorithm was proposed to deal with HJI equations. Afterwards, the associated discrete-time version was developed in [29], where the convergence of the proposed algorithm was well proved. The aforementioned works [29,33] both require the knowledge of system models during the algorithm iteration process. In [30], a recurrent neural network (NN) was utilized to identify the unknown system first, and then put the obtained identification results into the online learning algorithm. Although the real system models were replaced by the identification results, the NN approximation errors were inevitable. It should be also pointed out that most of the aforementioned works belong to the policy iteration (PI) methods, which require the initial condition of admissible control policies. However, for some large-scale CPSs, it is impractical to obtain the initial admissible control. Therefore, it will be interesting to investigate the security issues of CPSs with value iteration (VI) methods, which can start without the initial admissible condition.

In this paper, we build a relationship between the zero-sum game and the secure control of CPSs by using optimal control theory and RL methods. The contributions and main works of this paper can be summarized as below. First, the secure control problem of the compromised system is converted into the zero-sum game issue of the nominal auxiliary system. Second, two mainstream ADP methods including PI and VI are reviewed and introduced to solve the HJI equations. Third, based on the solution of the associated zero-sum game, the proposed secure control scheme can mitigate the effects of actuator attacks and unmatched perturbation, and stabilize the compromised CPSs by tuning the system performance parameters. Furthermore, how to set the values of parameters is derived through the Lyapunov stability theory. Finally, the proposed control scheme is tested on the Quanser helicopter to demonstrate the effectiveness.

2. Problem formulation

Consider the following linear CPSs with unmatched perturbation [21,35,36]:

$$\dot{x}(t) = Ax(t) + Bu(t) + C\Delta w(x(t)) \tag{1}$$

where $x \in \mathbb{R}^n$ denotes the system state; $u \in \mathbb{R}^m$ is the control input; $\Delta w(x) \in \mathbb{R}^p$ represents the unmatched perturbation with $\|\Delta w(x)\| \le k_w \|x\|$ on the given compact set; $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ and $C \in \mathbb{R}^{n\times p}$ denote the constant system matrices.

If the system (1) is compromised by the actuator attack, then the system dynamics can be described by [8,9,37,38]

$$\dot{x}(t) = Ax(t) + B(\bar{u}(t) + \rho_a(t, x(t))) + C\Delta w(x(t)) \tag{2}$$

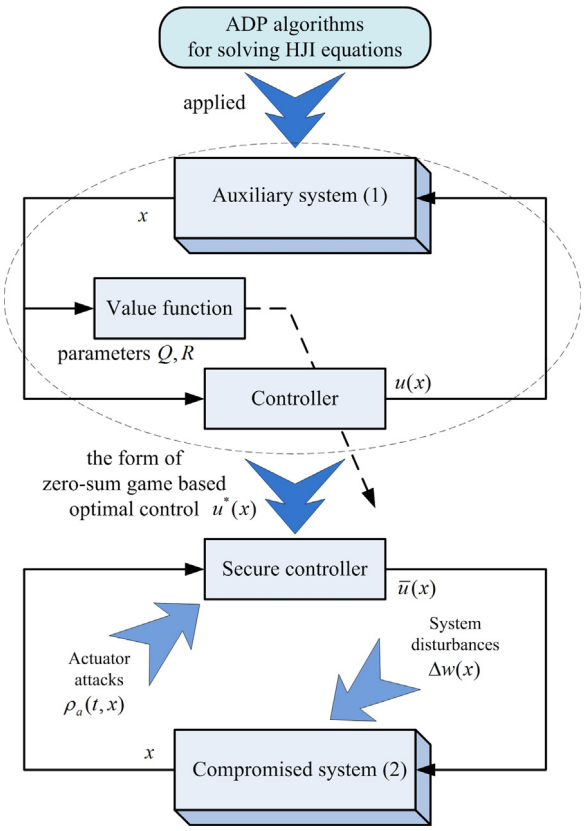

Fig. 1. Schematic diagram of the secure control design in this paper.

where the actuator attack $\rho_a(t, x)$ can be parameterized as $\rho_a(t, x) = \delta_a(t)\sigma_a(x)$ with $\|\delta_a(t)\| \le \delta_{am}$ and $\|\sigma_a(x)\| \le \varpi_a \|x\|$ on the given compact set [8,9,37,38]; $\bar{u}$ is the secure controller to be designed later.

Remark 1. It is impractical to investigate the compromised system (2) directly due to the existence of the actuator attack. The main idea of this paper is to design a secure control policy for the system (2) based on the auxiliary system (1). For the auxiliary system (1), the zero-sum game based control scheme is a suitable approach to attenuate unmatched disturbances. Furthermore, the effect caused by the actuator attack can be eliminated through tuning the parameters for the zero-sum game, which will be shown in Section 4. The schematic diagram of the secure control design is shown in Fig. 1.

The zero-sum game based control design aims to find out a control policy such that the system (1) is asymptotically stable and has the $\mathcal{L}_2$-gain no larger than γ, that is,

$$\int_0^\infty (x^T(\tau)Qx(\tau) + u^T(\tau)Ru(\tau))d\tau \le \gamma^2 \int_0^\infty \Delta w^T(\tau)\Delta w(\tau)d\tau \tag{3}$$

where Q and R are positive definite symmetric matrices.

The solution of the zero-sum game is a saddle point $(u^*, \Delta w^*)$, where u^* is the optimal control and Δw^* is viewed as the worst-case disturbance. Therefore, we define the performance index function as

$$J(x(0), u, \Delta w) = \int_0^\infty r(x(\tau), u(\tau), \Delta w(\tau))d\tau \tag{4}$$

where $r(x, u, \Delta w) = x^T Qx + u^T Ru - \gamma^2 \Delta w^T \Delta w$.

Please cite this article as: H. Jiang, H. Zhang and X. Xie, Critic-only adaptive dynamic programming algorithms' applications to the secure control of cyber–physical systems. ISA Transactions (2019), https://doi.org/10.1016/j.isatra.2019.02.012.

Given the admissible policies $u(x)$ and $\Delta w(x)$, the value function is expressed as

$$V(x(t)) = \int_t^\infty r(x(\tau), u(x(\tau)), \Delta w(x(\tau)))d\tau. \tag{5}$$

If the saddle point solution exists, the Nash equilibrium satisfies the following condition:

$$V^*(x) \triangleq \min_u \max_{\Delta w} V(x) = \max_{\Delta w} \min_u V(x). \tag{6}$$

According to the stationarity condition [30,33,39,40], the optimal control policy and the worst-case disturbance policy can be derived as

$$u^*(x) = -\frac{1}{2}R^{-1}B^T\nabla V^*(x), \tag{7}$$

$$\Delta w^*(x) = \frac{1}{2\gamma^2}C^T\nabla V^*(x) \tag{8}$$

where $\nabla V^*(x) = \partial V^*(x)/\partial x$ and $V^*(x)$ satisfies the HJI equation:

$$0 = r(x, u^*(x), \Delta w^*(x)) + \nabla V^{*T}(x)(Ax + Bu^*(x) + C\Delta w^*(x)). \tag{9}$$

3. ADP algorithms for solving HJI equations

From Section 2, it can be observed that the zero-sum game based optimal control relies on solving the associated HJI equations. However, HJI equations are generally difficult or even impossible to be solved, especially for the complex nonlinear systems. In this section, both PI-based and VI-based ADP methods will be reviewed and introduced to deal with this issue.

3.1. PI-based ADP algorithm

Inspired by the classical ADP works [34,39,41], a PI-based method is presented in the following Algorithm 1. By using Algorithm 1, one can obtain $V^*(x)$, $u^*(x)$ and $\Delta w^*(x)$ as the iteration index $i \to \infty$.

Algorithm 1 PI method

Step 1: (Initialization) Let the iteration index $i = 0$. Select a small enough computation precision ϵ. Choose initial admissible policies $u^{(0)}(x)$ and $\Delta w^{(0)}(x)$.

Step 2: (Policy Evaluation) With $u^{(i)}(x)$ and $\Delta w^{(i)}(x)$, compute $V^{(i+1)}(x)$ by

$$0 = r(x, u^{(i)}(x), \Delta w^{(i)}(x)) + (\nabla V^{(i+1)}(x))^T(Ax + Bu^{(i)}(x) + C\Delta w^{(i)}(x)). \tag{10}$$

Step 3: (Policy Improvement) With $V^{(i+1)}(x)$, update the control and disturbance policies, respectively, by

$$u^{(i+1)}(x) = -\frac{1}{2}R^{-1}B^T\nabla V^{(i+1)}(x), \tag{11}$$

$$\Delta w^{(i+1)}(x) = \frac{1}{2\gamma^2}C^T\nabla V^{(i+1)}(x). \tag{12}$$

Step 4: If $\left\|V^{(i+1)}(x) - V^{(i)}(x)\right\| \le \epsilon$ on the given compact set, stop and the optimal value function is acquired; Else, let $i = i + 1$ and go back to Step 2.

Remark 2. The PI algorithm is popular in the field of ADP. Under the help of initial admissible control, the PI-based method will soon find out the optimal solution. This is the main advantage of the PI-based method. However, for some large-scale CPSs, it is generally difficult or impractical to obtain the initial admissible control condition. Therefore, an iterative learning method without the requirement of initial admissible control is desired.

3.2. VI-based ADP algorithm

Different from PI-based methods, VI-based methods can start without the initial admissible condition. Motivated by the significant works [24,33,42,43], the VI method is presented in the following Algorithm 2.

Algorithm 2 VI method

Step 1: (Initialization) Let the iteration index $i = 0$. Set a computation precision ϵ. Choose an initial value function $V^{(0)}(x)$.

Step 2: (Policy Improvement) With $V^{(i)}(x)$, update the control and disturbance policies, respectively, by

$$u^{(i)}(x) = -\frac{1}{2}R^{-1}B^T\nabla V^{(i)}(x), \tag{13}$$

$$\Delta w^{(i)}(x) = \frac{1}{2\gamma^2}C^T\nabla V^{(i)}(x). \tag{14}$$

Step 3: (Policy Evaluation) With $u^{(i)}(x)$ and $\Delta w^{(i)}(x)$, compute $V^{(i+1)}(x)$ by

$$V^{(i+1)}(x(t)) = \int_t^{t+\Delta t} r(x(\tau), u^{(i)}(x(\tau)), \Delta w^{(i)}(x(\tau)))d\tau + V^{(i)}(x(t+\Delta t)). \tag{15}$$

Step 4: If $\left\|V^{(i+1)}(x) - V^{(i)}(x)\right\| \le \epsilon$ on the given compact set, stop and the optimal value function is acquired; Else, let $i = i + 1$ and go back to Step 2.

3.3. Critic-only NN implementation

According to the universal approximation property, the optimal value function has a NN representation:

$$V^*(x) = \varphi_c^T(x)W_c^* \tag{16}$$

where W_c^* is the ideal NN weight with the associated NN activation function $\varphi_c(x)$.

To implement Algorithm 2, we construct a critic NN $\hat{V}^{(i)}$ to approximate the iterative value function $V^{(i)}$:

$$\hat{V}^{(i)}(x) = \varphi_c^T(x)W_c^{(i)} \tag{17}$$

where $W_c^{(i)}$ is the iterative NN weight. The critic NN implies the iterative control and disturbance policies, i.e., $u^{(i)}$ and $\Delta w^{(i)}$, can be approximated by $\hat{u}^{(i)}$ and $\Delta\hat{w}^{(i)}$:

$$\hat{u}^{(i)}(x) = -\frac{1}{2}R^{-1}B^T\nabla\varphi_c^T(x)W_c^{(i)}, \tag{18}$$

$$\Delta\hat{w}^{(i)}(x) = \frac{1}{2\gamma^2}C^T\nabla\varphi_c^T(x)W_c^{(i)}. \tag{19}$$

Then, let the NN error function be

$$e^{(i)} = \hat{V}^{(i+1)}(x(t)) - \int_t^{t+\Delta t} r(x(\tau), \hat{u}^{(i)}(x(\tau)), \Delta\hat{w}^{(i)}(x(\tau)))d\tau - \hat{V}^{(i)}(x(t+\Delta t)). \tag{20}$$

To minimize $E^{(i)} = \frac{1}{2}e^{(i)T}e^{(i)}$, the gradient descent based updating law is derived by

$$W_c^{(i+1),k+1} = W_c^{(i+1),k} - \alpha_c\frac{\partial E^{(i)}(x)}{\partial e^{(i)}(x)}\frac{\partial e^{(i)}(x)}{\partial\hat{V}^{(i+1)}(x)}\frac{\partial\hat{V}^{(i+1)}(x)}{\partial W_c^{(i+1),k}} \tag{21}$$

where α_c is the learning rate and k is the number of calculation times for the gradient descent method. The schematic diagram of critic-only implementation is clearly shown in Fig. 2.

Please cite this article as: H. Jiang, H. Zhang and X. Xie, Critic-only adaptive dynamic programming algorithms' applications to the secure control of cyber–physical systems. ISA Transactions (2019), https://doi.org/10.1016/j.isatra.2019.02.012.

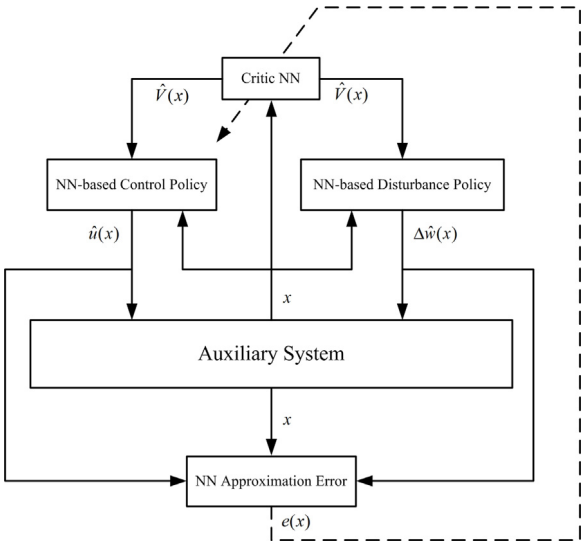

Fig. 2. Schematic diagram of critic-only implementation.

Remark 3. Compared with PI-based Algorithm 1, VI-based Algorithm 2 has the following advantages: (1) The initial condition of Algorithm 2 is not strict as that of the PI method, and the VI algorithm can start without admissible control. This is more practical for the real-world applications. (2) Algorithm 2 does not involve the information of system matrix A, while Algorithm 1 requires all the knowledge of system matrices A, B and C. That is, Algorithm 2 is partially model-free while Algorithm 1 is completely model-based. (3) In the policy evaluation step (15), $V^{(i+1)}$ can be easily attained by adding two terms. However, in the step (10) of Algorithm 1, $V^{(i+1)}$ seems more difficult to be solved. Therefore, Algorithm 2 is more easy-to-realize than Algorithm 1. Furthermore, the critic-only implementation is simpler than the actor–critic structure [38,44] or multiple-network architecture [34,43], and reduces computational burden.

4. Secure control design and stability analysis

In this section, the secure control scheme is designed based on the solution of the zero-sum game, and the tuning laws of parameters are derived through Lyapunov stability theory.

4.1. Secure control design

The following Theorem 1 introduces the secure control design in details.

Theorem 1. *If the form of the zero-sum game based optimal controller (7) is applied to the compromised system (2), that is, $\bar{u}(x) = u^*(x)$, then the system (2) can be asymptotically stable by tuning the parameters Q and R.*

Proof. Choose the Lyapunov function candidate as $V^*(x)$. Then, according to (7)–(9), it can be acquired that

$$\begin{aligned}\dot{V}^*(x) =& \nabla V^{*T}(x)(Ax + B(u^*(x) + \rho_a(t,x)) + C\Delta w(x))\\ =& \nabla V^{*T}(x)(Ax + Bu^*(x) + C\Delta w^*(x))\\ &+ \nabla V^{*T}(x)(B\rho_a(t,x) - C\Delta w^*(x) + C\Delta w(x))\\ \le& -x^TQx - u^{*T}(x)Ru^*(x) + \gamma^2\Delta w^{*T}(x)\Delta w^*(x)\\ &+ \frac{1}{2}\nabla V^{*T}(x)BB^T\nabla V^*(x) + \frac{1}{2}\rho_a^T(t,x)\rho_a(t,x)\\ &+ \nabla V^{*T}(x)CC^T\nabla V^*(x) + \frac{1}{2}\Delta w^{*T}(x)\Delta w^*(x)\\ &+ \frac{1}{2}\Delta w^T(x)\Delta w(x). \end{aligned} \tag{22}$$

Based on (22), it can be further derived that

$$\begin{aligned}\dot{V}^*(x) \le& -(\lambda_{\min}(Q) - \frac{1}{2}k_w^2 - \frac{1}{2}\delta_{am}^2\varpi_a^2)\|x\|^2\\ &- (\frac{1}{4}\lambda_{\min}(R^{-1}) - \frac{1}{2})\left\|B^T\nabla V^*(x)\right\|^2\\ &+ (1 + \frac{1}{4\gamma^2} + \frac{1}{8\gamma^4})\left\|C^T\nabla V^*(x)\right\|^2\\ \le& -(\lambda_{\min}(Q) - \frac{1}{2}k_w^2 - \frac{1}{2}\delta_{am}^2\varpi_a^2\\ &- (1 + \frac{1}{4\gamma^2} + \frac{1}{8\gamma^4})C_m^2k_v^2)\|x\|^2\\ &- (\frac{1}{4}\lambda_{\min}(R^{-1}) - \frac{1}{2})\left\|B^T\nabla V^*(x)\right\|^2 \end{aligned} \tag{23}$$

where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of a matrix. Let $\|\nabla V^*(x)\|^2 \le k_v^2\|x\|^2$ and $\|C\| \le C_m$. Since the system (1) considered in this paper is linear, the optimal solution $V^*(x)$ is the quadratic form. Therefore, it is reasonable to attain $\|\nabla V^*(x)\|^2 \le k_v^2\|x\|^2$. The purpose of $\dot{V}^*(x) \le 0$ can be achieved, if we set the parameters Q and R to satisfy the following condition:

$$\begin{cases}\lambda_{\min}(Q) \ge \frac{1}{2}k_w^2 + \frac{1}{2}\delta_{am}^2\varpi_a^2 + (1 + \frac{1}{4\gamma^2} + \frac{1}{8\gamma^4})C_m^2k_v^2\\ \lambda_{\min}(R^{-1}) \ge 2\end{cases}. \tag{24}$$

Note that, using the Lyapunov stability theory may bring conservativeness for selecting the parameters Q and R. Actually, for some given Q and R, even though they do not satisfy the condition (24), they may also make the system stable. This will be shown in the simulation part.

Furthermore, Eq. (23) can be also derived by

$$\begin{aligned}\dot{V}^*(x) \le& -(\lambda_{\min}(Q) - \frac{1}{2}B_m^2k_v^2 - \frac{1}{2}k_w^2 - \frac{1}{2}\delta_{am}^2\varpi_a^2\\ &- (1 + \frac{1}{4\gamma^2} + \frac{1}{8\gamma^4})C_m^2k_v^2)\|x\|^2\\ &- u^{*T}(x)Ru^*(x) \end{aligned} \tag{25}$$

where $\|B\| \le B_m$.

Hence, to ensure $\dot{V}^*(x) \le 0$, one can also choose Q to satisfy the following relation:

$$\lambda_{\min}(Q) \ge \frac{1}{2}B_m^2k_v^2 + \frac{1}{2}k_w^2 + \frac{1}{2}\delta_{am}^2\varpi_a^2 + (1 + \frac{1}{4\gamma^2} + \frac{1}{8\gamma^4})C_m^2k_v^2. \tag{26}$$

From (25) and (26), it can be observed that the matrix Q plays an important role in the system stability. In addition, the matrix Q also determines the system control performance, which will be also demonstrated in the simulation result. The proof is completed. ■

4.2. Stability analysis of NN approximation error

It is known that ADP algorithms are generally implemented by the universal approximators, such as NNs. However, NN approximation errors are inevitable, and may affect the system stability. Among the existing ADP works, NN approximation errors are rarely discussed. In this paper, we attempt to consider this issue, and present the following stability analysis.

After the NN learning procedure, the NN weight will converge to a constant value $\hat{W}_c$. Then, the NN-based approximate optimal control policy can be expressed as

$$\hat{u}(x) = -\frac{1}{2}R^{-1}B^T\nabla\varphi_c^T(x)\hat{W}_c. \tag{27}$$

By means of (7) and (16), one has

$$u^*(x) = -\frac{1}{2}R^{-1}B^T\nabla\varphi_c^T(x)W_c^*. \tag{28}$$

Let the NN weight approximation error be $\tilde{W}_c = W_c^* - \hat{W}_c$ with $\left\|\tilde{W}_c\right\| \le \tilde{W}_{cm}$ and $\theta(x) = \frac{1}{2}R^{-1}B^T\nabla\varphi_c^T(x)$. Using (27) and (28) yields

$$\begin{aligned}\hat{u}(x) =& u^*(x) - \frac{1}{2}R^{-1}B^T\nabla\varphi_c^T(x)\hat{W}_c + \frac{1}{2}R^{-1}B^T\nabla\varphi_c^T(x)W_c^*\\ =& u^*(x) + \theta(x)\tilde{W}_c.\end{aligned} \tag{29}$$

Since the optimal solution $V^*(x)$ is quadratic, the NN activation function $\varphi_c(x)$ should be also given by the quadratic form of x. Therefore, it is reasonable to attain $\|\theta(x)\| \le \frac{1}{2}\left\|R^{-1}B^T\right\|\left\|\nabla\varphi_c^T(x)\right\| \le k_\theta\|x\|$.

Theorem 2. *If the NN-based approximate optimal controller (27) is applied to the compromised system (2), that is, $\bar{u}(x) = \hat{u}(x)$, then the system (2) can be asymptotically stable by tuning the parameters Q and R.*

Proof. Choose the Lyapunov function candidate as $V^*(x)$, which implies

$$\begin{aligned}\dot{V}^*(x) =& \nabla V^{*T}(x)(Ax + B(u^*(x) + \theta(x)\tilde{W}_c + \rho_a(t,x)) + C\Delta w(x))\\ =& \nabla V^{*T}(x)(Ax + B(u^*(x) + \rho_a(t,x)) + C\Delta w(x)) + \nabla V^{*T}(x)B\theta(x)\tilde{W}_c\\ \le& -(\lambda_{\min}(Q) - \frac{1}{2}k_w^2 - \frac{1}{2}\delta_{am}^2\varpi_a^2 - (1 + \frac{1}{4\gamma^2} + \frac{1}{8\gamma^4})C_m^2k_v^2)\|x\|^2\\ & - (\frac{1}{4}\lambda_{\min}(R^{-1}) - \frac{1}{2})\left\|B^T\nabla V^*(x)\right\|^2 + \frac{1}{2}\nabla V^{*T}(x)BB^T\nabla V^*(x)\\ & + \frac{1}{2}\tilde{W}_c^T\theta^T(x)\theta(x)\tilde{W}_c\\ \le& -(\lambda_{\min}(Q) - \frac{1}{2}k_w^2 - \frac{1}{2}\delta_{am}^2\varpi_a^2 - (1 + \frac{1}{4\gamma^2} + \frac{1}{8\gamma^4})C_m^2k_v^2\\ & - \frac{1}{2}\tilde{W}_{cm}^2k_\theta^2)\|x\|^2\\ & - (\frac{1}{4}\lambda_{\min}(R^{-1}) - 1)\left\|B^T\nabla V^*(x)\right\|^2.\end{aligned} \tag{30}$$

If the NN weight approximation error is not large, the following condition is easy to be realized:

$$\begin{cases}\lambda_{\min}(Q) \ge \frac{1}{2}k_w^2 + \frac{1}{2}\delta_{am}^2\varpi_a^2 + (1 + \frac{1}{4\gamma^2} + \frac{1}{8\gamma^4})C_m^2k_v^2 + \frac{1}{2}\tilde{W}_{cm}^2k_\theta^2\\ \lambda_{\min}(R^{-1}) \ge 4\end{cases}. \tag{31}$$

By the same way as (25), one can derive that

$$\begin{aligned}\dot{V}^*(x) \le& -(\lambda_{\min}(Q) - B_m^2k_v^2 - \frac{1}{2}k_w^2 - \frac{1}{2}\delta_{am}^2\varpi_a^2\\ & - (1 + \frac{1}{4\gamma^2} + \frac{1}{8\gamma^4})C_m^2k_v^2 - \frac{1}{2}\tilde{W}_{cm}^2k_\theta^2)\|x\|^2\\ & - u^{*T}(x)Ru^*(x).\end{aligned} \tag{32}$$

To guarantee $\dot{V}^*(x) \le 0$, one can directly choose the matrix Q to satisfy the following condition:

$$\lambda_{\min}(Q) \ge B_m^2k_v^2 + \frac{1}{2}k_w^2 + \frac{1}{2}\delta_{am}^2\varpi_a^2 + (1 + \frac{1}{4\gamma^2} + \frac{1}{8\gamma^4})C_m^2k_v^2 + \frac{1}{2}\tilde{W}_{cm}^2k_\theta^2. \tag{33}$$

The aforementioned derivation indicates the NN-based approximate optimal controller with admissible NN approximation error can guarantee the system stability. The proof is completed. ■

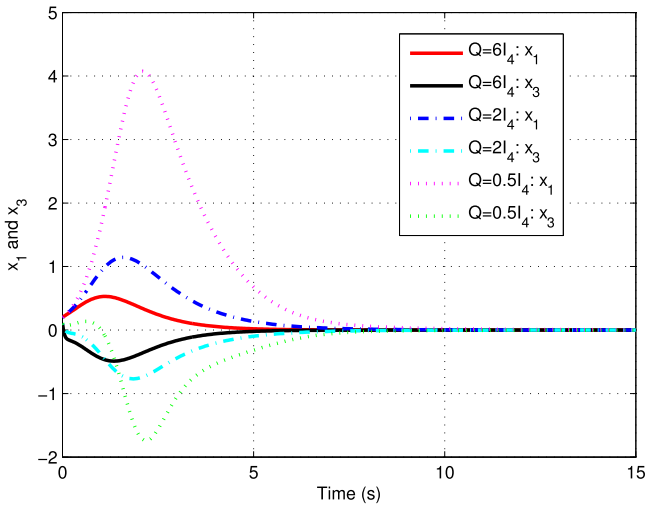

Fig. 3. Dynamics of the system states x_1 and x_3.

5. Simulation results

In this section, the secure control scheme is applied to the Quanser helicopter. By utilizing the Euler–Lagrange formula, the system model is described by [45]

$$\begin{aligned}\ddot{\theta} =& -\frac{B_p}{J_{T_p}}\dot{\theta} + \frac{K_{pp}}{J_{T_p}}F_p + \frac{K_{py}}{J_{T_p}}F_y\\ \ddot{\psi} =& -\frac{B_y}{J_{T_y}}\dot{\psi} + \frac{K_{yp}}{J_{T_y}}F_p + \frac{K_{yy}}{J_{T_y}}F_y\end{aligned} \tag{34}$$

where B_p and B_y denote the equivalent viscous damping about the pitch axis and yaw axis, respectively; K_{pp} is the thrust force constant of the yaw motor; K_{yy}, K_{py} and K_{yp} are the thrust torque constants acting on the yaw axis from the yaw motor, the pitch axis from the yaw motor and the yaw axis from the pitch motor, respectively; J_p and J_y denote the moments of inertia about pitch axis and yaw axis, respectively; θ and ψ represent pitch angle and yaw angle, respectively; F_p and F_y denote the voltage of the pitch motor and the yaw motor, respectively; l and m are used to describe the center of mass length along the helicopter body from the pitch axis and the moving mass of the helicopter, respectively. Let $J_{T_p} = J_p + ml^2$ and $J_{T_y} = J_y + ml^2$.

Define the system states $x = [\theta, \psi, \dot{\theta}, \dot{\psi}]^T$ and the control inputs $u = [F_p, F_y]^T$. Let the perturbation input be $\Delta w = 3\sin(x_2)x_1$ and the actuator attack be $\rho_a(t,x) = [4\cos(t)\sin(x_2)x_1, 5\sin(t)\sin(x_4)x_2]^T$. The helicopter model with attacks and perturbation in the state-space form can be given by

$$\dot{x} = Ax + B(\bar{u} + \rho_a(t,x)) + C\Delta w \tag{35}$$

where $A = \begin{bmatrix}0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & -\frac{B_p}{J_{T_p}} & 0\\ 0 & 0 & 0 & -\frac{B_y}{J_{T_y}}\end{bmatrix}$, $B = \begin{bmatrix}0 & 0\\ 0 & 0\\ \frac{K_{pp}}{J_{T_p}} & \frac{K_{py}}{J_{T_p}}\\ \frac{K_{yp}}{J_{T_y}} & \frac{K_{yy}}{J_{T_y}}\end{bmatrix}$ and $C = \begin{bmatrix}1\\0\\1\\0\end{bmatrix}$.

The parameters are provided by $B_p = 0.8$ N/V, $B_y = 0.318$ N/V, $K_{pp} = 0.204$ N · m/V, $K_{yy} = 0.072$ N · m/V, $K_{py} = 0.0068$ N · m/V, $K_{yp} = 0.0219$ N · m/V, $J_p = 0.0178$ kg · m^2, $J_y = 0.0084$ kg · m^2, $l = 0.186$ m and $m = 1.3872$ kg.

Please cite this article as: H. Jiang, H. Zhang and X. Xie, Critic-only adaptive dynamic programming algorithms' applications to the secure control of cyber–physical systems. ISA Transactions (2019), https://doi.org/10.1016/j.isatra.2019.02.012.

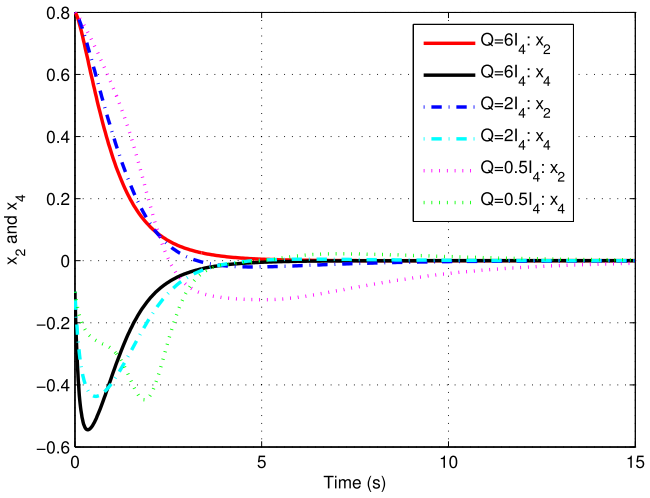

Fig. 4. Dynamics of the system states x_2 and x_4.

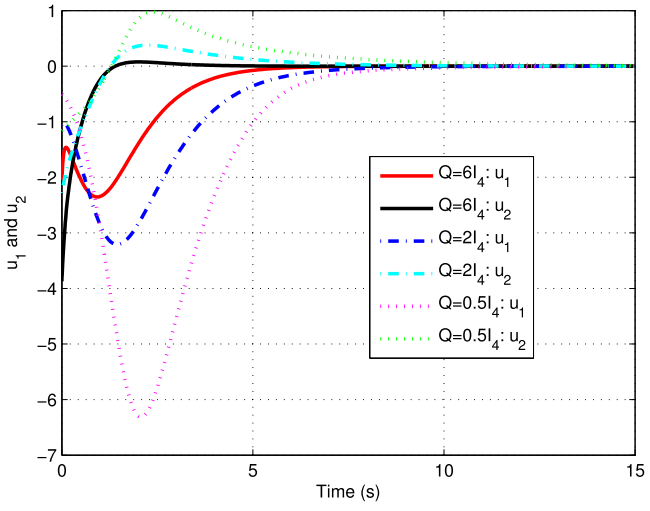

Fig. 5. Dynamics of the control inputs u_1 and u_2.

Set $R = 0.2I_2$ and $\gamma = 5$. In order to further investigate the effect of the matrix Q on system performance, we choose three different cases including $Q = 6I_4$, $Q = 2I_4$ and $Q = 0.5I_4$. Simulation results are shown in Figs. 3–5. In Fig. 3, under $Q = 6I_4$, the system states x_1 and x_3 achieve the minimum amplitude of oscillation, and become stable at about 5 s; when $Q = 0.5I_4$, the system states x_1 and x_3 get the maximum amplitude of oscillation and the slowest convergence. In Fig. 4, the system states x_2 and x_4 under $Q = 6I_4$ achieve convergence faster than them under $Q = 2I_4$ and $Q = 0.5I_4$. In Fig. 5, the control inputs u_1 and u_2 under $Q = 6I_4$ use less control energy than them under $Q = 2I_4$ and $Q = 0.5I_4$. Therefore, it can be observed that the matrix Q plays an important role in both system stability and control performance.

6. Conclusion

This paper has integrated optimal control theory, zero-sum game theory, RL methods and NNs to deal with the secure control issue of the CPSs with actuator attacks and unmatched perturbation. Both PI-based and VI-based ADP algorithms have been reviewed and introduced to solve the HJI equation. The tuning conditions of parameters for the secure controller have been derived through the Lyapunov stability theory. It has been proved that this secure control scheme can mitigate the effects of actuator attacks and unmatched perturbation, and stabilize the compromised CPSs. In the future work, it is expected that our proposed scheme can be applied to other CPSs.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61433004, 61627809, 61621004).

References

[1] Zhang D, Song H, Yu L. Robust fuzzy-model-based filtering for nonlinear cyber-physical systems with multiple stochastic incomplete measurements. IEEE Trans Syst Man Cybern A 2017;47(8):1826–38.
[2] Wang F-Y. The emergence of intelligent enterprises: From CPS to CPSS. IEEE Intell Syst 2010;25(4):85–8.
[3] Wei Q, Liu D, Liu Y, Song R. Optimal constrained self-learning battery sequential management in microgrid via adaptive dynamic programming. IEEE/CAA J Autom Sinica 2017;4(2):168–76.
[4] Wang F-Y. Scanning the issue and beyond: Computational transportation and transportation 5.0. IEEE Trans Intell Transp Syst 2014;15(5):1861–8.
[5] Zhang D, Nguang SK, Srinivasan D, Yu L. Distributed filtering for discrete-time T-S fuzzy systems with incomplete measurements. IEEE Trans Fuzzy Syst 2018;26(3):1459–71.
[6] Zhang D, Wang QG, Srinivasan D, Li H, Yu L. Asynchronous state estimation for discrete-time switched complex networks with communication constraints. IEEE Trans Neural Netw Learn Syst 2018;29(5):1732–46.
[7] An L, Yang GH. Secure state estimation against sparse sensor attacks with adaptive switching mechanism. IEEE Trans Automat Control 2018;63(8):2596–603.
[8] Jin X, Haddad WM, Yucelen T. An adaptive control architecture for mitigating sensor and actuator attacks in cyber-physical systems. IEEE Trans Automat Control 2017;62(11):6058–64.
[9] An L, Yang GH. Improved adaptive resilient control against sensor and actuator attacks. Inform Sci 2018;423:145–56.
[10] Han J, Zhang H, Wang Y, Liu Y. Disturbance observer based fault estimation and dynamic output feedback fault tolerant control for fuzzy systems with local nonlinear models. ISA Trans 2015;59:114–24.
[11] Sun S, Zhang H, Wang Y, Cai Y. Dynamic output feedback-based fault-tolerant control design for T-S fuzzy systems with model uncertainties. ISA Trans. http://dx.doi.org/10.1016/j.isatra.2018.07.022.
[12] Xie X, Yue D, Zhang H, Xue Y. Fault estimation observer design for discrete-time Takagi–Sugeno fuzzy systems based on homogenous polynomially parameter-dependent Lyapunov functions. IEEE Trans Cybern 2017;47(9):2504–13.
[13] Zhong X, He H. An event-triggered ADP control approach for continuous-time system with unknown internal states. IEEE Trans Cybern 2017;47(3):683–94.
[14] Luo B, Liu D, Wu H, Wang D, Lewis FL. Policy gradient adaptive dynamic programming for data-based optimal control. IEEE Trans Cybern 2017;47(10):3341–54.
[15] Zhong X, He H, Zhang H, Wang Z. Optimal control for unknown discrete-time nonlinear Markov jump systems using adaptive dynamic programming. IEEE Trans Neural Netw Learn Syst 2014;25(12):2141–55.
[16] Lewis FL, Vrabie D. Reinforcement learning and adaptive dynamic programming for feedback control. IEEE Circuits Syst Mag 2009;9(3):32–50.
[17] Luo B, Yang Y, Liu D. Adaptive Q-learning for data-based optimal output regulation with experience replay. IEEE Trans Cybern 2018;48(12):3337–48.
[18] He H, Zhong X. Learning without external reward [research frontier]. IEEE Comput Intell Mag 2018;13(3):48–54.
[19] Wang D, Liu D, Zhang Q, Zhao D. Data-based adaptive critic designs for nonlinear robust optimal control with uncertain dynamics. IEEE Trans Syst Man Cybern A 2016;46(11):1544–55.
[20] Wang D, Liu D, Li H, Ma H. Neural-network-based robust optimal control design for a class of uncertain nonlinear systems via adaptive dynamic programming. Inform Sci 2014;282:167–79.
[21] Yang X, He H. Self-learning robust optimal control for continuous-time nonlinear systems with mismatched disturbances. Neural Netw 2018;99:19–30.
[22] Luo B, Liu D, Wu H-N. Adaptive constrained optimal control design for data-based nonlinear discrete-time systems with critic-only structure. IEEE Trans Neural Netw Learn Syst 2017;29(6):2099–111.

[23] Wang D, Mu C, Yang X, Liu D. Event-based constrained robust control of affine systems incorporating an adaptive critic mechanism. IEEE Trans Syst Man Cybern A 2017;47(7):1602–12.
[24] Xiao G, Zhang H, Qu Q, Jiang H. General value iteration based single network approach for constrained optimal controller design of partially-unknown continuous-time nonlinear systems. J Franklin Inst B 2018;355(5):2610–30.
[25] Luo B, Liu D, Huang T, Wang D. Model-free optimal tracking control via critic-only Q-learning. IEEE Trans Neural Netw Learn Syst 2016;27(10):2134–44.
[26] Luo B, Liu D, Huang T, Liu J. Output tracking control based on adaptive dynamic programming with multistep policy evaluation. IEEE Trans Syst Man Cybern A 2017;PP(99):1–11.
[27] Mu C, Ni Z, Sun C, He H. Data-driven tracking control with adaptive dynamic programming for a class of continuous-time nonlinear systems. IEEE Trans Cybern 2017;47(6):1460–70.
[28] Zhong X, He H, Wang D, Ni Z. Model-free adaptive control for unknown nonlinear zero-sum differential game. IEEE Trans Cybern 2018;48(5):1633–46.
[29] Wei Q, Liu D, Lin Q, Song R. Adaptive dynamic programming for discrete-time zero-sum games. IEEE Trans Neural Netw Learn Syst 2018;29(4):957–69.
[30] Wei Q, Song R, Yan P. Data-driven zero-sum neuro-optimal control for a class of continuous-time unknown nonlinear systems with disturbance using ADP. IEEE Trans Neural Netw Learn Syst 2016;27(2):444–58.
[31] Zhao D, Zhang Q, Wang D, Zhu Y. Experience replay for optimal control of nonzero-sum game systems with unknown dynamics. IEEE Trans Cybern 2016;46(3):854–65.
[32] Song R, Lewis FL, Wei Q. Off-policy integral reinforcement learning method to solve nonlinear continuous-time multiplayer nonzero-sum games. IEEE Trans Neural Netw Learn Syst 2017;28(3):704–13.
[33] Wu H-N, Luo B. Neural network based online simultaneous policy update algorithm for solving the HJI equation in nonlinear H_∞ control. IEEE Trans Neural Netw Learn Syst 2012;23(12):1884–95.
[34] Modares H, Lewis FL, Jiang Z-P. H_∞ tracking control of completely unknown continuous-time systems via off-policy reinforcement learning. IEEE Trans Neural Netw Learn Syst 2015;26(10):2550–62.
[35] Yang X, He H, Zhong X. Adaptive dynamic programming for robust regulation and its application to power systems. IEEE Trans Ind Electron 2018;65(7):5722–32.
[36] Yang X, He H, Wei Q, Luo B. Reinforcement learning for robust adaptive control of partially unknown nonlinear systems subject to unmatched uncertainties. Inform Sci 2018;463–464:307–22.
[37] Zhang H, Qu Q, Xiao G, Cui Y. Optimal guaranteed cost sliding mode control for constrained-input nonlinear systems with matched and unmatched disturbances. IEEE Trans Neural Netw Learn Syst 2018;29(6):2112–26.
[38] Fan Q, Yang G. Adaptive actor-critic design-based integral sliding-mode control for partially unknown nonlinear systems with input disturbances. IEEE Trans Neural Netw Learn Syst 2016;27(1):165–77.
[39] Luo B, Wu H-N, Huang T. Off-policy reinforcement learning for H_∞ control design. IEEE Trans Cybern 2015;45(1):65–76.
[40] Wang D, He H, Mu C, Liu D. Intelligent critic control with disturbance attenuation for affine dynamics including an application to a microgrid system. IEEE Trans Ind Electron 2017;64(6):4935–44.
[41] Luo B, Wu H-N. Computationally efficient simultaneous policy update algorithm for nonlinear H_∞ state feedback control with Galerkin's method. Internat J Robust Nonlinear Control 2013;23(9):991–1012.
[42] Lewis FL, Vrabie D, Vamvoudakis KG. Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers. IEEE Control Syst 2012;32(6):76–105.
[43] Xiao G, Zhang H, Zhang K, Wen Y. Value iteration based integral reinforcement learning approach for H_∞ controller design of continuous-time nonlinear systems. Neurocomputing 2018;285:51–9.
[44] Vamvoudakis KG, Lewis FL. Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. Automatica 2010;46(5):878–88.
[45] Luo B, Wu H-N, Huang T. Optimal output regulation for model-free Quanser helicopter with multi-step Q-learning. IEEE Trans Ind Electron 2018;65(6):4953–61.

Please cite this article as: H. Jiang, H. Zhang and X. Xie, Critic-only adaptive dynamic programming algorithms' applications to the secure control of cyber–physical systems. ISA Transactions (2019), https://doi.org/10.1016/j.isatra.2019.02.012.