Stackelberg Game Based Robust Optimal Control of Cyber-Physical System under Hybrid Attack

Junkai Tan, Shuangsi Xue, and Hui Cao

Abstract—**This paper presents a novel framework integrating Stackelberg game theory and reinforcement learning for cyber-physical system (CPS) security. A hierarchical game model is developed, in which defenders and attackers interact through sequential decision-making. The defender-attacker dynamics are formulated as an optimization problem combining H_2 and H_∞ control objectives. Key innovations include a unified game-theoretic approach for modeling hybrid attack-defense mechanisms, online reinforcement learning algorithms for real-time strategy adaptation, and rigorous stability analysis using the Lyapunov theory. Theoretical guarantees of convergence are established for the proposed learning scheme. Comprehensive experiments on a robotic platform validate the effectiveness of the framework in maintaining control performance under diverse attack scenarios.**

***Index Terms*—Cyber-physical system (CPS), Stackelberg game, optimal control, reinforcement learning, adaptive dynamic programming**

I. Introduction

CYBER-physical systems (CPSs) integrate physical processes with computational elements, playing a crucial role in modern society. While widely used in critical infrastructures, such as power grids [1, 2], transportation networks [3, 4], and industrial control systems [5, 6], their increasing connectivity and complexity introduce security vulnerabilities. Various cyber attacks such as denial-of-service (DoS) [7, 8], false data injection (FDI) [9, 10], and malware [11] can severely impact system operations. Therefore, developing effective defense strategies against these threats is essential.

Game theory offers an effective approach for analyzing strategic interactions between adversarial agents in CPS [12, 13]. The Stackelberg game framework, in which defenders act first as leaders followed by attackers' responses, enables systematic modeling of hierarchical security decisions [14, 15]. This sequential structure allows defenders to proactively plan countermeasures by anticipating potential attack strategies [16]. Recent works have explored various aspects of game-theoretic CPS security. In Refs. [17, 18], robust Stackelberg games were formulated to analyze Nash equilibria under hierarchical decision-making. A Hamiltonian-driven approach was proposed in Ref. [19] for deriving optimal stabilization controllers. Researches in Refs. [20, 21] developed single-critic learning algorithms combining H_2 and H_∞ controls for uncertain nonlinear stochastic systems. Data-driven methods were investigated in Refs. [22, 23] to achieve optimal mixed $\mathrm{H}_2/\mathrm{H}_\infty$ performance. For discrete-time linear systems, Ref. [24] proposed a robust Stackelberg game incorporating both control indices. However, existing mixed $\mathrm{H}_2/\mathrm{H}_\infty$ approaches have focused primarily on stabilization without specific performance constraints. The challenging problem of tracking control for nonlinear constrained systems remains largely unexplored in this context.

Reinforcement learning (RL) provides a powerful framework for solving complex decision-making problems in CPS [25, 26]. Through continuous interaction with the environment, RL enables agents to learn and adapt optimal control policies, making it particularly suitable for dynamic attack-defense scenarios in CPS security. Both model-free and model-based RL approaches have been investigated for CPS control. Model-free methods, such as Q-learning [27, 28], can learn stabilizing controllers without prior system knowledge, while integral RL [29, 30] approximates optimal control for partially unknown systems. However, these offline approaches typically require extensive training data. In contrast, model-based RL methods [31, 32] leverage system models for online learning, though they need prior dynamic information. Actor-critic architectures [33, 34] have been explored to simultaneously learn value functions and control policies online. Recent works [35, 36] have extended this to constrained nonlinear tracking control. While existing studies have demonstrated the potential of RL for CPS control [37, 38], few have addressed the critical challenge of hybrid attack-defense mechanisms. This gap motivates our investigation of a game-theoretic RL framework for securing CPS under diverse attack scenarios.

Motivated by these challenges, we propose a novel Stackelberg game framework to analyze hybrid attack-defense interactions in CPS. Unlike existing approaches that focus on single attack types or static defense strategies, we develop a comprehensive model capturing dynamic interactions between multiple attack modes and adaptive defense mechanisms. The problem is formulated as an optimal control scenario where attackers maximize system damage using H_2 performance metrics while defenders minimize impacts through H_∞ control. An online reinforcement learning approach enables real-time strategy adaptation, overcoming limitations of offline methods that require extensive prior training data. Theoretical stability guarantees are established via the

Manuscript received: 2 December 2024; revised: 15 February 2025; accepted: 25 February 2025. *(Corresponding author: Shuangsi Xue.)*

Citation: J. Tan, S. Xue, and H. Cao, Stackelberg game based robust optimal control of cyber-physical system under hybrid attack, *Int. J. Intell. Control Syst.*, 2025, 30(1), 47–54.

Junkai Tan, Shuangsi Xue, and Hui Cao are with Shaanxi Key Laboratory of Smart Grid, Xi'an Jiaotong University, Xi'an 710049, China, and also with School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: tanjk@stu.xjtu.edu.cn; xssxjtu@xjtu.edu.cn; huicao@mail.xjtu.edu.cn).

Digital Object Identifier 10.62678/IJICS202503.10145

Lyapunov analysis. Extensive simulations on a four-wheeled robot platform validate the effectiveness of the framework. The key contributions are as follows:

(1) A unified Stackelberg game framework advances existing works [5, 14, 39] in three aspects: integration of both H_2 and H_∞ performance indices to characterize attack-defense objectives, explicit modeling of stochastic hybrid attacks through Bernoulli switching signals, and consideration of input constraints and system uncertainties in the game formulation.

(2) An efficient actor-critic architecture improves upon traditional methods [7, 12] through: online concurrent learning of value functions and control policies, lower computational complexity without requiring complete system models, and provable convergence guarantees under persistent disturbances.

(3) Comprehensive experimental validation demonstrates clear advantages over baseline approaches [40, 41]: reduction in tracking errors under hybrid attacks, faster convergence to optimal strategies, and enhanced robustness against simultaneous DoS and FDI attacks.

This paper is structured as follows. In Section II, we present the mathematical preliminaries and system modeling. Section III develops the Stackelberg game framework for hybrid attack-defense mechanisms. Section IV provides the theoretical analysis and proposes an online reinforcement learning solution. Section V provides the stability analysis. Section VI validates the framework through comprehensive numerical experiments. Section VII summarizes our findings and discusses future research directions.

II. Preliminary

Consider a continuous-time nonlinear CPS with disturbance

$$\dot{x}(t) = f(x(t)) + \kappa(g(x)u(t), k(x)\omega(t)) \tag{1}$$

where t indicates time, $x(t) \in \mathbb{R}^n$ denotes the system state vector, $u(t) \in \mathbb{R}^m$ represents the control input, $\omega(t) \in \mathbb{R}^m$ indicates external disturbance, $f : \mathbb{R}^n \to \mathbb{R}^n$ characterizes autonomous dynamics, $g : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ defines control distribution, $k : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ captures disturbance coupling, and $\kappa : \mathbb{R}^{n\times m} \times \mathbb{R}^{n\times m} \to \mathbb{R}^n$ models the hybrid attack-defense mechanism.

$$\kappa(g(x)u(t), k(x)\omega(t)) = a(t)g(x)u(t) + b(t)k(x)\omega(t) \tag{2}$$

The switching signals $a(t)$ and $b(t)$ follow Bernoulli distributions with success probabilities α and β respectively, representing the stochastic nature of attack occurrences and defense activations. For reference trajectory tracking, let $x_d(t) \in \mathbb{R}^n$ be the desired state governed by $\dot{x}_d(t) = f_d(x_d(t))$, where $f_d : \mathbb{R}^n \to \mathbb{R}^n$ specifies the reference dynamics. Define tracking error as $e(t) = x(t) - x_d(t)$. The error dynamic is

$$\dot{e}(t) = f(x) - f_d(x_d) + \kappa(g(x)u(t), k(x)\omega(t)) \tag{3}$$

where all functions satisfy local Lipschitz continuity conditions. For analytical purposes, we integrate Eq. (1) and reference dynamics into an augmented form

$$\begin{cases} \dot{X} = F(X) + aG(X)U + bK(X)\omega, \\ Y = H(X, U) \end{cases} \tag{4}$$

where the augmented state $X = [x^{\mathrm{T}}, x_d^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{2\times n}$ combines actual and desired states, control input $U = [u^{\mathrm{T}}, 0_{1\times m}]^{\mathrm{T}} \in \mathbb{R}^{2\times m}$ incorporates system and reference controls, and Y denotes the performance output. The system matrices are given by

$$\begin{aligned} F &= \begin{bmatrix} f(x(t)) \\ f_d(x_d(t)) \end{bmatrix}, \quad K = \begin{bmatrix} k(x(t)) \\ 0_{n\times m} \end{bmatrix}, \\ G &= \begin{bmatrix} g(x(t)) & 0_{n\times n} \\ 0_{n\times m} & 0_{n\times m} \end{bmatrix}, \quad H = \begin{bmatrix} \sqrt{Q}X(t) \\ \sqrt{\alpha R}U(t) \end{bmatrix} \end{aligned} \tag{5}$$

where Q and R are positive definite weighting matrices for state and control costs, respectively. We make Assumption 1.

Assumption 1 For Eq. (4), we assume:

(1) Functions F and G are locally Lipschitz on $X \in \chi \subset \mathbb{R}^n$, with $F(0) = 0$ and $\|G\| \leqslant G_H$ for all $X \in \chi$. χ is the region in n-dimensional Euclidean space and G_H is a constant.

(2) Matrices Q and R satisfy $\underline{\lambda}_Q \mathcal{I} \le Q \le \bar{\lambda}_Q \mathcal{I}$ and $\underline{\lambda}_R \mathcal{I} \le R \le \bar{\lambda}_R \mathcal{I}$, where $0 \leqslant \underline{\lambda}_Q, \underline{\lambda}_R < \bar{\lambda}_Q$, and $\bar{\lambda}_R < \infty$. $\underline{\lambda}_Q$ and $\bar{\lambda}_Q$ are lower and upper bounds on the eigenvalue of Q. $\underline{\lambda}_R$ and $\bar{\lambda}_R$ are lower and upper bounds on the eigenvalue of R.

These preliminaries enable us to formulate the Stackelberg game framework for hybrid attack-defense interactions.

III. Problem Formulation

This paper addresses optimal control design for CPS under hybrid attack-defense scenarios. The attacker aims to maximize system damage using the H_2 control input $U^*(t)$, while the defender minimizes damage through the H_∞ control $\omega^*(t)$. Based on Eq. (4), we model this interaction as a Stackelberg game. For the nonlinear CPS in Eq. (4), the H_2 and H_∞ performance objectives are defined as

$$\begin{aligned} J_D(X_0, U, \omega) = E\left\{\int_t^\infty \|Y\|^2 \mathrm{d}\tau\right\} = \\ E\left\{\int_t^\infty \left(X^{\mathrm{T}}QX + \alpha U^{\mathrm{T}}RU\right)\mathrm{d}\tau\right\} \end{aligned} \tag{6}$$

$$\begin{aligned} J_A(X_0, U, \omega) = E\left\{\int_t^\infty \gamma^2\|\omega\|^2 - \|Y\|^2 \mathrm{d}\tau\right\} = \\ E\left\{\int_t^\infty \left(\beta\gamma^2\|\omega\|^2 - X^{\mathrm{T}}QX - \alpha U^{\mathrm{T}}RU\right)\mathrm{d}\tau\right\} \end{aligned} \tag{7}$$

where subscripts D and A represent defender and attacker, X_0 is the initial value of X, E is the estimate value function, and γ denotes the disturbance attenuation level. J_D measures H_2 performance, while J_A quantifies H_∞ performance. The sequential interaction between attacker and defender is formalized through the following Stackelberg game framework.

Definition 1 Stackelberg game framework Consider a defender D and an attacker A with objectives in Eqs. (6) and (7), respectively. The hierarchical decision process involves:

Level 1 Defense: D determines baseline strategy $U_{\mathrm{L1}} \in \Omega_U$, where Ω_U is the possible space of U

$$J_{D_{\mathrm{L1}}}(X_0, U_{\mathrm{L1}}^*, 0) = \min_{U \in \Omega_U} J_D(X_0, U_{\mathrm{L1}}, 0) \tag{8}$$

Level 2 Attack: Given U_{L1}^*, A optimizes $\omega_{\mathrm{L2}} \in \Omega_W$, where Ω_W is the possible space of W

$$J_{A_{\mathrm{L2}}}(X_0, U_{\mathrm{L1}}^*, \omega_{\mathrm{L2}}^*) = \max_{\omega \in \Omega_W} J_A(X_0, U_{\mathrm{L1}}^*, \omega_{\mathrm{L2}}) \tag{9}$$

Level 3 Defense update: D updates the level 3 defense input $U_{\mathrm{L3}} \in \Omega_U$ given ω_{L2}^*

$$J_{D_{\mathrm{L3}}}(X_0, U^*_{\mathrm{L3}}, \omega^*_{\mathrm{L2}}) = \min_{U \in \Omega_U} J_D(X_0, U_{\mathrm{L3}}, \omega^*_{\mathrm{L2}}) \tag{10}$$

The resulting $U^* \overset{\Delta}{=} U^*_{\mathrm{L3}}$ and $\omega^* \overset{\Delta}{=} \omega^*_{\mathrm{L2}}$ form the Stackelberg equilibrium.

As illustrated in Fig. 1, the defender first establishes an initial control strategy. Then, the attacker optimizes its disturbance input based on the defense policy. Finally, the defender adapts its control to counter the attack. This iterative process converges to the Stackelberg equilibrium strategies U^* and ω^*.

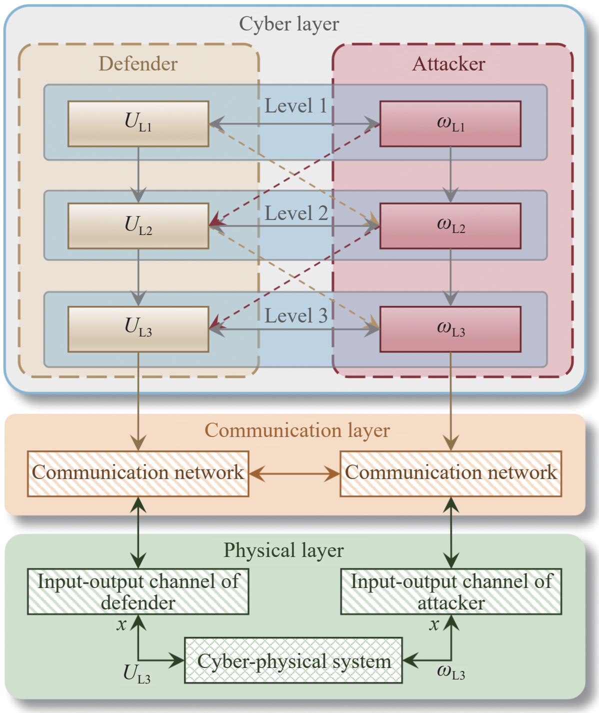


Figure 1 Stackelberg game based hybrid attack-defense interaction.

Problem 1 Stackelberg game framework for hybrid attack-defense Consider the nonlinear CPS in Eq. (4) which is under the hybrid attack-defense mechanisms characterized by stochastic switching signals $a(t)$ and $b(t)$ that follow Bernoulli distributions with success probabilities $P(a=1)=\alpha$ and $P(b=1)=\beta$. The objectives are as follows:

(1) Design optimal defense strategy $U^*(t)$ that maximizes system damage using the H_∞ performance index in Eq. (6).

(2) Develop optimal attack policy $\omega^*(t)$ that minimizes adverse impacts through the H_2 control in Eq. (7).

The interaction could be formulated as the following optimization problem

$$J^*_D(X_0) = \min_{\bar{U} \in \Omega_U} J_D(X_0, \bar{U}, \bar{\omega}^*) \tag{11}$$

$$J^*_A(X_0) = \min_{\bar{\omega} \in \Omega_W} J_A(X_0, \bar{U}, \bar{\omega}) \tag{12}$$

where Ω_U and Ω_W denote the feasible control sets for the defender and attacker, respectively. The optimal control signals are subject to the following constraints

$$\bar{U}^* = \begin{cases} U^*, \ a(t)=1; \\ 0, \ \text{otherwise} \end{cases} \tag{13}$$

$$\bar{\omega}^* = \begin{cases} \omega^*, \ b(t)=1; \\ 0, \ \text{otherwise} \end{cases} \tag{14}$$

where $\bar{U}^*$ and $\bar{\omega}^*$ denote the actual implemented defense and attack control signals under stochastic switching, respectively.

Based on the formulated Stackelberg game framework, we first analyze the optimization problem of the attacker.

$$J^*_A = \max_{\omega} E\left\{\int_t^\infty \left(\beta\gamma^2\|\omega\|^2 - X^{\mathrm{T}}QX - \alpha U^{\mathrm{T}}RU\right)\mathrm{d}\tau\right\} \tag{15}$$

The Hamiltonian function of the attacker is defined as

$$\begin{aligned} H_A(X,U,\omega,\nabla J^*_A) = {} & \nabla J^{*\mathrm{T}}_A(F + \alpha GU + \beta K\omega) + \\ & \beta\gamma^2\|\omega\|^2 - X^{\mathrm{T}}QX - \alpha U^{\mathrm{T}}RU \end{aligned} \tag{16}$$

By minimizing H_A, the optimal attack strategy is obtained as

$$\omega^*(U) = -\frac{K^{\mathrm{T}}}{2\gamma^2}\nabla J^*_A \tag{17}$$

The evolution of the value function of the attacker is captured by costate dynamics.

$$\begin{aligned} \dot{\lambda}_2 = {} & -\left(\frac{\partial F}{\partial x} + \frac{\partial G}{\partial x}U + \frac{\partial K}{\partial x}\omega^* + G\frac{\partial U^{\mathrm{T}}}{\partial x}X\right)^{\mathrm{T}}\nabla J^*_A + \\ & 2QX + 2RU\frac{\partial U}{\partial x} \end{aligned} \tag{18}$$

The optimization of the defender is

$$J^*_D = \min_U E\left\{\int_t^\infty \left(X^{\mathrm{T}}QX + \alpha U^{\mathrm{T}}RU + \eta^{\mathrm{T}}\dot{\lambda}_2\right)\mathrm{d}\tau\right\} \tag{19}$$

where η is the Lagrange multiplier. The Hamiltonian of the defender is

$$\begin{aligned} H_D(X,U,\omega^*,\nabla J^*_D,\eta) = {} & \nabla J^{*\mathrm{T}}_D(F + \alpha GU + \beta K\omega) + \\ & X^{\mathrm{T}}QX + \alpha U^{\mathrm{T}}RU + \eta^{\mathrm{T}}\dot{\lambda}_2 \end{aligned} \tag{20}$$

The optimal defense strategy is derived as

$$U^*(\omega^*) = -\frac{1}{2}R^{-1}\left(G^{\mathrm{T}}\nabla J^*_D - \nabla xG\eta\nabla J^*_A\right) \tag{21}$$

The Lagrange multiplier dynamics are governed by

$$\begin{aligned} \dot{\eta} = {} & \sum_{i=1}^n \eta_i \left(\frac{\partial K}{\partial X_i} \cdot \frac{\partial \omega^*}{\partial \nabla J^*_A}\right)^{\mathrm{T}} \nabla J^*_A + \\ & (\nabla F + \alpha\nabla GU + \beta\nabla K\omega^* + \alpha G\nabla U)\eta - \\ & \left(K \cdot \frac{\partial \omega^*}{\partial \nabla J^*_A}\right)^{\mathrm{T}} \nabla J^*_D \end{aligned} \tag{22}$$

Due to the complexity of solving these nonlinear Hamiltonian optimization problems directly, we propose an actor-critic reinforcement learning approach to efficiently approximate the optimal value functions and control policies online.

IV. Main Result

In this section, we develop an online reinforcement learning solution using actor-critic architecture to solve the formulated Stackelberg game problem. As shown in Fig. 2, the defender and attacker are modeled as hierarchical learning agents that interact through neural networks (NNs). The proposed framework employs dual actor-critic networks for each agent. The actor networks approximate optimal control (AOC) policies while the critic networks evaluate performance by

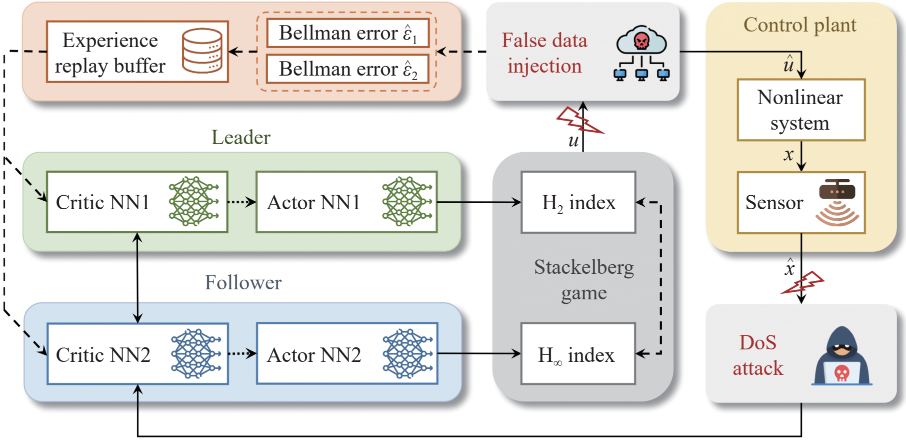


Figure 2 Structure of the proposed Stackelberg game based hybrid attack and defense.

estimating value functions. This concurrent learning scheme enables efficient approximation of both optimal strategies and their corresponding performance metrics through continuous interaction between the adversarial agents.

A. Actor-Critic Architecture for Value Approximation

To derive optimal strategies, we employ parallel actor-critic networks for both agents. The critic networks estimate value functions while actor networks generate control policies. The neural approximation structure is formulated as

$$J_i^*(X) = \mathcal{W}_{ci}^{\mathrm{T}}\phi_{ci}(X) + \epsilon_{ci}(X),\ i = 1,2 \tag{23}$$

$$U^*(X) = -\frac{1}{2}\Big(R^{-1}G^{\mathrm{T}}\left(\nabla\phi_{a1}^{\mathrm{T}}\mathcal{W}_{a1} + \nabla\epsilon_{a1}^{\mathrm{T}}\right) - \left(\mathcal{W}_{a2}^{\mathrm{T}}\nabla\phi_{a2} + \nabla\epsilon_{a2}\right)\nabla xG\eta\Big) \tag{24}$$

$$\omega^*(X) = -\frac{K^{\mathrm{T}}}{2\gamma^2}\left(\nabla\phi_{a2}^{\mathrm{T}}\mathcal{W}_{a2} + \nabla\epsilon_{a2}^{\mathrm{T}}\right) \tag{25}$$

where $\mathcal{W}_{ci}$ and $\mathcal{W}_{ai} \in \mathbb{R}^{n_\phi \times 1}$ denote the target weights for critic and actor networks, with ϵ_{ci} and ϵ_{ai} representing approximation errors. ϕ_{ci} and ϕ_{ai} are basis functions for critic and actor networks, respectively.

Since the ideal weights are unknown, we implement estimated parameters.

$$\hat{J}_i(X) = \hat{\mathcal{W}}_{ci}^{\mathrm{T}}\phi_{ci},\ i = 1,2 \tag{26}$$

$$\hat{U}(X) = -\frac{1}{2}\Big(R^{-1}G^{\mathrm{T}}\nabla\phi_{a1}^{\mathrm{T}}\hat{\mathcal{W}}_{a1} - \hat{\mathcal{W}}_{a2}^{\mathrm{T}}\nabla\phi_{a2}\nabla xG\eta\Big) \tag{27}$$

$$\hat{\omega}(X) = -\frac{K^{\mathrm{T}}}{2\gamma^2}\nabla\phi_{a2}^{\mathrm{T}}\hat{\mathcal{W}}_{a2} \tag{28}$$

By incorporating these approximations into the Hamiltonian functions, we derive the Bellman optimality errors.

$$\delta_1 = \left(\nabla\phi_{c1}^{\mathrm{T}}\hat{\mathcal{W}}_{c1}\right)^{\mathrm{T}}\left(F + \alpha G\hat{U} + \beta K\hat{\omega}\right) + X^{\mathrm{T}}QX + \alpha U^{\mathrm{T}}RU + \eta^{\mathrm{T}}\dot{\lambda}_2 \tag{29}$$

$$\delta_2 = \left(\nabla\phi_{c2}^{\mathrm{T}}\hat{\mathcal{W}}_{c2}\right)^{\mathrm{T}}\left(F + \alpha G\hat{U} + \beta K\hat{\omega}\right) + \beta\gamma^2\|\hat{\omega}\|^2 - X^{\mathrm{T}}QX - \alpha U^{\mathrm{T}}RU \tag{30}$$

For analytical purposes, we make Assumption 2 on network parameters.

Assumption 2 The network weights and activation functions satisfy uniform bounds: $\|\hat{\mathcal{W}}_{ci}\| \leqslant \mathcal{W}_{Hi}$, $\|\sigma_i(X)\| \leqslant \sigma_{Hi}$, $\|\nabla\sigma_i(X)\| \leqslant \sigma'_{Hi}$, $\|\phi_i(X)\| \leqslant \phi_{Hi}$, $\|\nabla\phi_i(X)\| \leqslant \phi'_{Hi}$, $\|\epsilon_i(X)\| \leqslant \epsilon_{Hi}$, and $\|\nabla\epsilon_i(X)\| \leqslant \epsilon'_{Hi}$.

These neural approximation structures enable online learning of optimal strategies through weight updates driven by the Bellman error minimization.

B. Online Learning of Value Function

We present an online learning scheme for actor-critic neural network weights based on minimizing Bellman errors. The defender maintains a historical data stack $[\hat{U}(t), \delta_1(t), [\hat{U}^j(t), \delta_1^j(t)]_{j=1}^N]$, while the attacker stores trajectory data $[\hat{\omega}(t), \delta_2(t), [\hat{\omega}^j(t), \delta_2^j(t)]_{j=1}^N]$, where the superscript j indicates historical samples. Both agents update their neural network weights by minimizing the squared Bellman errors: $E_i = \delta_i^{\mathrm{T}}\delta_i + \sum_{l=1}^{N}\delta_i^{l\mathrm{T}}\delta_i^l$, $i = 1,2$. The critic network weights are updated through gradient descent.

$$\dot{\hat{\mathcal{W}}}_{ci} = -k_{ci,1}\frac{\sigma_i\delta_i}{\rho_i(t)} - \frac{k_{ci,2}}{N}\sum_{l=1}^{N}\frac{\sigma_i^l\delta_i^l}{\rho_i^l(t)},\ i = 1,2 \tag{31}$$

where $k_{ci,j} > 0$ is learning rate, $\rho_i(t) = \left(\sigma_i^{\mathrm{T}}\sigma_i + 1\right)^2$, $\rho_i^l(t) = \left(\sigma_i^{l\mathrm{T}}\sigma_i^l + 1\right)^2$, $\sigma_i = \nabla\phi_{ci}^{\mathrm{T}}\left(F + \alpha G\hat{U} + \beta K\hat{\omega}\right)$, and $\sigma_i^l = \nabla\phi_{ci}^{\mathrm{T}}X^l\left(F + \alpha G\hat{U}^l + \beta K\hat{\omega}^l\right)$. The actor network weights follow a similar gradient-based update.

$$\dot{\hat{\mathcal{W}}}_{ai} = F_i k_{ai}(\hat{\mathcal{W}}_{ci} - \hat{\mathcal{W}}_{ai}),\ i = 1,2 \tag{32}$$

where $k_{ai} > 0$ is actor learning rate and F_i is positive definite matrix. To ensure convergence, we require the following excitation condition.

Assumption 3 Persistent excitation [42, 43] The collected data satisfy

$$\begin{aligned} \Lambda_{1,i}\mathcal{I}_{m,i} &\leqslant \int_t^{t+T}\frac{\sigma_i\sigma_i^{\mathrm{T}}}{\rho_i}\mathrm{d}\tau, \\ \Lambda_{2,i}\mathcal{I}_{m,i} &\leqslant \inf_{t\in\mathbb{R}_{t\geqslant t_0}}\frac{\sigma_i^l\sigma_i^{l\mathrm{T}}}{N\rho_i^l} \end{aligned} \tag{33}$$

where $\mathcal{I}_{m,i}$ is identity matrix, T is the time internal of excitation, $\mathbb{R}_{t\geqslant t_0}$ is the time space after initial time t_0, and constant $\Lambda_{1,i}$ or $\Lambda_{2,i}$ must be positive.

The complete online learning procedure is detailed in Algorithm 1.

Algorithm 1 Online learning algorithm for hybrid attack-defense

```
1: Initialize actor-critic networks:
       Actor weight $\hat{\mathcal{W}}_{ai}$ and critic weight $\hat{\mathcal{W}}_{ci}$
       Learning rates $k_{ci,j}$ and $k_{ai}$
       Projection matrix $F_i,\ i\in\{1,2\}$
2: while $t < T_{\text{end}}$ do
3:     Compute optimal strategies:
           Defense policy $\hat{U}$ via Eq. (27)
           Attack policy $\hat{\omega}$ via Eq. (28)
4:     Evaluate Bellman errors from Eq. (30):
           Defender error $\delta_1(X,\hat{U},\hat{\omega})$
           Attacker error $\delta_2(X,\hat{U},\hat{\omega})$
5:     Update experience replay buffers:
           Defender: $[\hat{U},\delta_1,[\hat{U}^j,\delta_1^j]_{j=1}^N]$
           Attacker: $[\hat{\omega},\delta_2,[\hat{\omega}^j,\delta_2^j]_{j=1}^N]$
6:     Update network parameters:
           Critic weight via Eq. (31)
           Actor weight via Eq. (32)
7:     Execute actions and update system state.
8: end while
```

V. Stability Analysis

In this section, we analyze the stability properties of the closed-loop system using the Lyapunov theory. The main objective is to establish uniform ultimate boundedness (UUB) of both system states and neural network estimation errors under hybrid attacks. Based on the optimal defense policy in Eq. (27) and attack strategy in Eq. (28), we have the following error bounds

$$\|U^*(X)-\hat{U}(X)\|^2 \leqslant \bar{\Gamma}_{u_1}\|\tilde{\mathcal{W}}_{a1}\|^2 + B_{u_1} \tag{34}$$

$$\|\omega^*(X)-\hat{\omega}(X)\|^2 \leqslant \bar{\Gamma}_{u_2}\|\tilde{\mathcal{W}}_{a2}\|^2 + B_{u_2} \tag{35}$$

where $\tilde{\mathcal{W}}_{ai},\ i=1,2$ indicates the error between the optimal and actual weights for actor network, $\bar{\Gamma}_{u_i},\ i=1,2$ is positive constant determined by network activation bound, and $B_{u_i},\ i=1,2$ represents bounded approximation residual.

The key stability results are summarized in Theorems 1 and 2.

Theorem 1 Consider Eq. (4) with the proposed Stackelberg game framework. Under Assumptions 1–3, for networks updated via Eqs. (31) and (32), the augmented error state $Z=[X^{\mathrm{T}},\tilde{\mathcal{W}}_{c1}^{\mathrm{T}},\tilde{\mathcal{W}}_{c2}^{\mathrm{T}},\tilde{\mathcal{W}}_{a1}^{\mathrm{T}},\tilde{\mathcal{W}}_{a2}^{\mathrm{T}}]^{\mathrm{T}}$ remains UUB if

$$\|Z\| \geqslant \sqrt{\Psi_{\text{res}}/(\underline{\lambda}_{\mathcal{H}}\mathcal{I})} \tag{36}$$

where $\tilde{\mathcal{W}}_{ci},\ i=1,2$ indicates the error between the optimal and actual weights for critic network, Ψ_{res} is defined in Eq. (A6), and $\underline{\lambda}_{\mathcal{H}}$ denotes the minimum eigenvalue of $\mathcal{H}$ in Eq. (A7).

Theorem 2 For Eq. (4), the approximate policies $\hat{U}$ in Eq. (27) and $\hat{\omega}$ in Eq. (28) converge to their optimal counterparts U^* and ω^*, respectively, reaching a unique Stackelberg equilibrium.

VI. Numerical Simulation

A. Simulation Setup

To validate the proposed Stackelberg game framework, we conduct numerical experiments on a four-wheeled differential drive robot system. Figure 3 shows the details of four-wheeled system. The kinematic model of the robot follows Eq. (37).

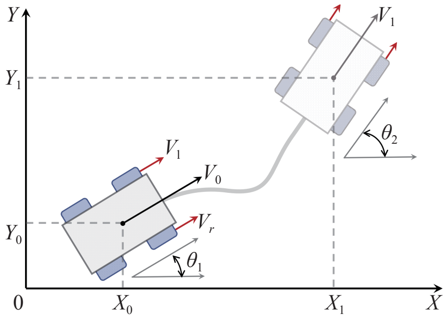


Figure 3 Schematic of the four-wheeled mobile robot.

$$\begin{aligned} f &= 0_{3\times 1}, \\ g &= \begin{bmatrix} \cos(\theta) & 0 \\ \sin(\theta) & 0 \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{37}$$

State $x=[x,y,\theta]^{\mathrm{T}}$ represents position coordinates and heading angle, and control input $u=[v,\Omega]^{\mathrm{T}}$ denotes linear and angular velocities.

The reference trajectory is an elliptical path $(x_d^2/4+y_d^2=1)$ centered at origin with semi-major axis $a=2$ and semi-minor axis $b=1$. The desired heading angle is $\theta_d=\arctan\{(y_d-y)/(x_d-x)\}$. The tracking error is defined as $e=[x-x_d,y-y_d,\theta-\theta_d]^{\mathrm{T}}$.

For neural network implementation, we use basis function $\phi_i=\left[e_1^2,e_2^2,e_1^2+e_2^2,e_3^2,e_1^2+e_3^2,e_2^2+e_3^2\right]^{\mathrm{T}}$, where e_i is the i-th element of e. All network weights are initialized to 5. Key parameters are listed in Table 1. The approximate optimal control method from Ref. [40] serves as baseline for comparison.

Table 1 Parameter of numerical simulation.

Index	Control parameter	Update law
Defender	$R=\text{diag}([1.0,0.1])$	$k_{1,c1}=k_{1,c2}=0.01$
	$Q_1=\mathcal{I}_3$ and $\alpha=0.1$	$k_{1,a}=1$ and $F_1=\mathcal{I}_6$
Attacker	$\gamma=2$	$k_{2,c1}=k_{2,c2}=0.01$
	$Q_2=\mathcal{I}_3$ and $\beta=0.1$	$k_{2,a}=1$ and $F_2=\mathcal{I}_6$

To handle input constraints $v,\ \Omega\in[-5,5]$, we reconstruct the defense penalty function as

$$\Psi(U) = 2R\int_0^U \mu_D \tanh^{-1}\left(\frac{\zeta_U}{\mu_D}\right) \mathrm{d}\zeta_U \tag{38}$$

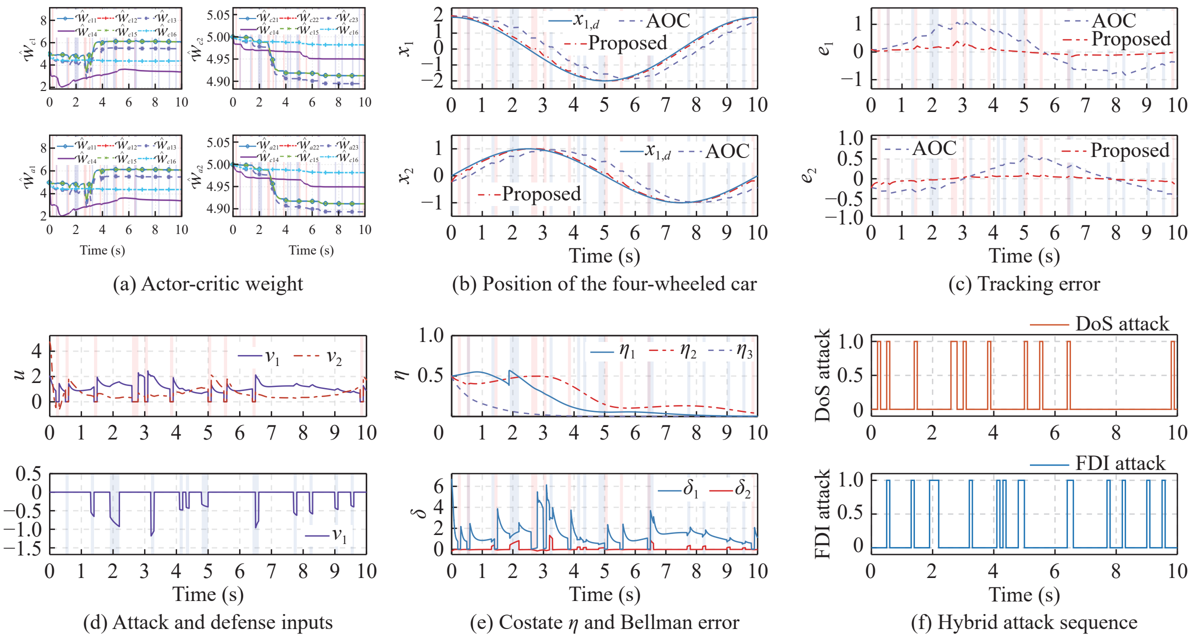
The constrained defense input becomes

$$\hat{U} = -\mu_D \tanh\left\{\frac{R^{-1}G^{\mathrm{T}}\nabla\phi_{a1}^{\mathrm{T}}\hat{W}_{a1} - \hat{W}_{a2}^{\mathrm{T}}\nabla\phi_{a2}\nabla xG\eta}{2\mu_D}\right\} \tag{39}$$

where $\mu_D = 5$ denotes the input saturation bound.

B. Simulation Result

The numerical results demonstrate the effectiveness of the framework through various performance metrics. Figure 4(a) shows the evolution of neural network parameters. The rapid convergence of both critic and actor weights within 10 s validates the learning efficiency of the proposed algorithm. Figure 4(b) illustrates the system state trajectories. Despite persistent attacks, the states closely track their reference values, indicating strong robustness of the controller design. Figure 4(c) presents the tracking error dynamics. The errors remain bounded within ±0.2 and exhibit convergent behavior, providing empirical support for the theoretical stability analysis in Theorem 1. Figure 4(d) depicts the control signals from both agents. The defender generates smooth control inputs while effectively counteracting the disturbances of the attacker, demonstrating the ability of the framework to balance performance and energy efficiency. Figure 4(e) evaluates the learning performance through Bellman errors. Their asymptotic convergence towards zero confirms successful approximation of optimal policies and attainment of game equilibrium. Figure 4(f) visualizes the stochastic hybrid attack sequence. The binary switching pattern with probability α captures the random nature of cyber attacks in practical systems. Figure 5 demonstrates the path of robot following capability. The minimal deviation from desired trajectories under attack validates the resilience of the framework in maintaining control objectives. These comprehensive results verify two key theoretical claims: (1) uniform ultimate boundedness of closed-loop stability, and (2) convergence to Stackelberg equilibrium between competing agents.



(a) Actor-critic weight

(b) Position of the four-wheeled car

(c) Tracking error

(d) Attack and defense inputs

(e) Costate η and Bellman error

(f) Hybrid attack sequence

Figure 4 Simulation result of the proposed Stackelberg game based hybrid attack-defense framework.

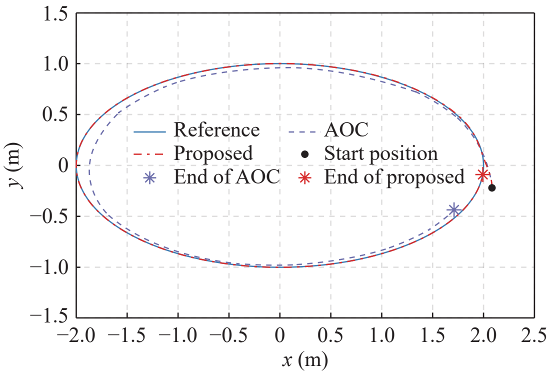


Figure 5 Trajectory of the system state.

C. Application to Critical Infrastructure System

Our framework demonstrates significant potential for real-world critical infrastructure protection, which can be applied in several critical domains.

Power grid application: The designed framework can be applied in power grid, including wide-area monitoring against coordinated cyber-attacks, real-time detection of false data injection attacks, adaptive defense against DoS attacks, and optimal protection of critical grid nodes.

Autonomous vehicle system: The designed framework can be applied in autonomous vehicle system, including secure vehicle to vehicle (V2V) communication protocols, robust trajectory tracking under global positioning system (GPS) spoofing, dynamic defense for platooning safety, and multi-layer protection for vehicle networks.

Industrial control system: The designed framework can also be applied in industrial control system, including protection of supervisory control and data acquisition (SCADA) networks, resilient control of robotic systems, adaptive security for smart factories, and real-time attack detection in process control.

Our ongoing research focuses on hardware-in-the-loop validation and pilot implementations across these domains. Initial results demonstrate the effectiveness of the framework in maintaining system stability and performance under various attack scenarios.

VII. Conclusion

This paper develops a Stackelberg game based framework to analyze hybrid attack-defense dynamics in CPS. A novel leader-follower structure models the sequential decision-making process between attackers and defenders. The objectives of attacker and defender are formulated using H_2 and H_∞ indices. An efficient reinforcement learning approach is proposed to learn optimal strategies online. Theoretical analysis establishes uniform ultimate boundedness of the closed-loop system. Numerical experiments on a four-wheeled robot demonstrate the capability of the framework to maintain control performance under attacks. Future work will explore extensions to cooperative multi-agent systems and practical implementations.

Appendix
Proof of Theorems 1 and 2

The proof of Theorem 1 follows from the Lyapunov stability analysis.

Proof Consider the Lyapunov function candidate

$$\mathcal{V}=J_1^*+J_2^*+\sum_{i=1}^{2}\frac{1}{2}\left(\tilde{\mathcal{W}}_{ci}^{\mathrm{T}}\tilde{\mathcal{W}}_{ci}+\tilde{\mathcal{W}}_{ai}^{\mathrm{T}}\tilde{\mathcal{W}}_{ai}\right) \tag{A1}$$

Define the Bellman errors δ_i and δ_i^l

$$\delta_i=-\sigma_i^{\mathrm{T}}\tilde{\mathcal{W}}_{ci}+\varPhi_i(\tilde{\mathcal{W}}_{a1},\tilde{\mathcal{W}}_{a2})+\varDelta_i \tag{A2}$$

$$\delta_i^l=-(\sigma_i^l)^{\mathrm{T}}\tilde{\mathcal{W}}_{ci}+\varPhi_i^l(\tilde{\mathcal{W}}_{a1},\tilde{\mathcal{W}}_{a2})+\varDelta_i^l \tag{A3}$$

where $\varPhi_i$ and $\varPhi_i^l$ contain quadratic terms, and $\varDelta_i$ and $\varDelta_i^l$ are bounded residuals.

Taking the derivative of $\mathcal{V}$ along system trajectories yields

$$\dot{\mathcal{V}}=\sum_{i=1}^{2}\left(\nabla J_i^*\dot{X}+\tilde{\mathcal{W}}_{ci}^{\mathrm{T}}\dot{\hat{\mathcal{W}}}_{ci}+\tilde{\mathcal{W}}_{ai}^{\mathrm{T}}\dot{\hat{\mathcal{W}}}_{ai}\right) \tag{A4}$$

Substituting the weight update laws and applying the Young's inequality, one has

$$\dot{\mathcal{V}}\leqslant -Z^{\mathrm{T}}\mathcal{H}Z+\varPsi_{\mathrm{res}} \tag{A5}$$

where the Hamiltonian matrix $\mathcal{H}$ is positive definite and

$$\begin{aligned}\varPsi_{\mathrm{res}}=&\sum_{i=1}^{2}\Big(\frac{1}{2}k_{ci,1}\|\varDelta_{\mathcal{W}i}\|^2+\frac{1}{2}k_{ci,2}\|\varDelta_{\mathcal{W}i}^l\|^2\Big)+\\&\gamma^2B_{u_2}+\bar{\lambda}_RB_{u_1}\end{aligned} \tag{A6}$$

$$\mathcal{H}=\begin{bmatrix}\mu_1&0&0&0&0\\0&\mu_2&0&0&0\\0&\mu_3&\mu_4&0&0\\0&\mu_5&0&\mu_6&0\\0&0&\mu_7&0&\mu_8\end{bmatrix} \tag{A7}$$

where matrix coefficients μ_i are defined as

$$\begin{cases}\mu_1=\underline{\lambda}_Q,\\ \mu_2=\dfrac{1}{2}(k_{c1,1}\sigma_1\sigma_1^{\mathrm{T}}+k_{c1,2}\varLambda_{2,1}\mathcal{I}_{m,1}),\\ \mu_3=(k_{c1,1}+k_{c2,1})\sigma_1\sigma_2^{\mathrm{T}},\\ \mu_4=\dfrac{1}{2}(k_{c2,1}\sigma_2\sigma_2^{\mathrm{T}}+k_{c2,2}\varLambda_{2,2}\mathcal{I}_{m,2}),\\ \mu_5=-F_1\mathcal{I}_{m,1},\\ \mu_6=F_1\mathcal{I}_{m,1}-\bar{\lambda}_{R,1}\varGamma_{u_1}\mathcal{I}_{m,1},\\ \mu_7=-F_2\mathcal{I}_{m,2},\\ \mu_8=F_2\mathcal{I}_{m,2}+\gamma^2\varGamma_{u_2}\mathcal{I}_{m,2}\end{cases} \tag{A8}$$

where $\varDelta_{\mathcal{W}1}=0.25\tilde{\mathcal{W}}_{a1}^{\mathrm{T}}\sigma_1\tilde{\mathcal{W}}_{a1}+\varDelta_1+\xi_{H1}$, $\varDelta_{\mathcal{W}2}=0.25\tilde{\mathcal{W}}_{a2}^{\mathrm{T}}\sigma_2\tilde{\mathcal{W}}_{a2}-0.25\tilde{\mathcal{W}}_{a1}^{\mathrm{T}}\sigma_1\tilde{\mathcal{W}}_{a1}+\varDelta_2$, $\varDelta_{\mathcal{W}1}^l=0.25\tilde{\mathcal{W}}_{a1}^{\mathrm{T}}\sigma_1^l\tilde{\mathcal{W}}_{a1}+\varDelta_1^l$, and $\varDelta_{\mathcal{W}2}^l=0.25\tilde{\mathcal{W}}_{a2}^{\mathrm{T}}\sigma_2^l\tilde{\mathcal{W}}_{a2}-0.25\tilde{\mathcal{W}}_{a1}^{\mathrm{T}}\sigma_1^l\tilde{\mathcal{W}}_{a1}+\varDelta_2^l$. Therefore, the augmented error state Z converges to the compact set defined by Eq. (A6), establishing UUB. ■

The detailed proof of Theorem 2 is referred to Theorem 2 in Refs. [17, 18].

Acknowledgment

This work was supported by the Postdoctoral Science Foundation of China (No. 2024M762602).

References

[1] H. Li and Q. Wei, A multi-observer based optimal control method for nonlinear systems under sensor attacks, *IEEE Trans. Autom. Sci. Eng.*, 2025, 22, 7632–7641.

[2] S. Sridhar, A. Hahn, and M. Govindarasu, Cyber-physical system security for the electric power grid, *Proc. IEEE*, 2012, 100(1), 210–224.

[3] J. Dong, Z. Ye, and D. Zhang, Finite-time security control of networked unmanned marine vehicle systems subject to DoS attack, *IEEE Trans. Intell. Veh.*, 2024, 9(2), 3464–3477.

[4] N. Anwar, G. Xiong, W. Lu, P. Ye, H. Zhao, and Q. Wei, Cyber-physical-social systems for smart cities: An overview, in *Proc. 2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence*, Beijing, China, 2021, 348–353.

[5] Y. Li, S. Liu, and L. Zhu, A stochastic Bayesian game for securing secondary frequency control of microgrids against spoofing attacks with incomplete information, *IEEE Trans. Ind. Cyber Phys. Syst.*, 2024, 2, 118–129.

[6] Q. Wei, H. Li, and F. Wang, Parallel control for continuous-time linear systems: A case study, *IEEE/CAA J. Autom. Sin.*, 2020, 7(4), 919–928.

[7] H. Liu, SINR-based multi-channel power schedule under DoS attacks: A Stackelberg game approach with incomplete information, *Automatica*, 2019, 100, 274–280.

[8] X. Chen, L. Xiao, W. Feng, N. Ge, and X. Wang, DDoS defense for IoT: A Stackelberg game model-enabled collaborative framework, *IEEE Internet Things J.*, 2022, 9(12), 9659–9674.

[9] Y. Wu, M. Chen, H. Li, and M. Chadli, Event-triggered distributed intelligent learning control of six-rotor UAVs under FDI attacks, *IEEE Trans. Artif. Intell.*, 2024, 5(7), 3299–3312.

[10] Y. Xu, T. Li, Y. Yang, S. Tong, and C. Chen, Simplified ADP for event-triggered control of multiagent systems against FDI attacks, *IEEE Trans. Syst. Man Cybern. Syst.*, 2023, 53(8), 4672–4683.

[11] C. Fei, J. Shen, H. Qiu, and Z. Zhang, Data driven secure control for cyber-physical systems under hybrid attacks: A Stackelberg game approach, *J. Franklin Inst.*, 2024, 361(6), 106715.

[12] W. Xing, X. Zhao, Y. Li, and L. Liu, Denial-of-service attacks on cyber-physical systems against linear quadratic control: A Stackelberg-game analysis, *IEEE Trans. Autom. Control*, 2025, 70(1), 595–602.

[13] J. Tan, S. Xue, H. Cao, and H. Li, Nash equilibrium solution based on safety-guarding reinforcement learning in nonzero-sum game, in *Proc. 2023 International Conference on Advanced Robotics and Mechatronics*, Sanya, China, 2023, 630–635.

[14] J. Lian, P. Jia, F. Wu, and X. Huang, A Stackelberg game approach to the stability of networked switched systems under DoS attacks, *IEEE Trans. Netw. Sci. Eng.*, 2023, 10(4), 2086–2097.

[15] P. Shukla, L. An, A. Chakrabortty, and A. Duel-Hallen, A robust Stackelberg game for cyber-security investment in networked control systems, *IEEE Trans. Control Syst. Technol.*, 2023, 31(2), 856–871.

[16] Z. Wang, H. Shen, H. Zhang, S. Gao, and H. Yan, Optimal DoS attack strategy for cyber-physical systems: A Stackelberg game-theoretical approach, *Inf. Sci.*, 2023, 642, 119134.

[17] Z. Jing, X. Li, P. Ju, and H. Zhang, Optimal control and filtering for hierarchical decision problems with H_∞ constraint based on Stackelberg strategy, *IEEE Trans. Autom. Control*, 2024, 69(9), 6238–6245.

[18] X. Li, Z. Jing, and P. Ju, Leader-follower based online reinforcement learning algorithm in problem with hierarchy decision makers, *Int. J. Intell. Control Syst.*, 2024, 29(1), 48–58.

[19] Y. Yang, M. Mazouchi, and H. Modares, Hamiltonian-driven adaptive dynamic programming for mixed H_2/H_∞ performance using sum-of-squares, *Int. J. Robust Nonlinear Control*, 2021, 31(6), 1941–1963.

[20] Z. Ming, H. Zhang, X. Tong, and Y. Yan, Mixed H_2/H_∞ control with dynamic event-triggered mechanism for partially unknown nonlinear stochastic systems, *IEEE Trans. Autom. Sci. Eng.*, 2023, 20(3), 1934–1944.

[21] Z. Ming, H. Zhang, Q. Li, and X. Tong, Mixed H_2/H_∞ control for nonlinear stochastic systems with cooperative and non-cooperative differential game, *IEEE Trans. Circuits Syst. II Express Briefs*, 2022, 69(12), 4874–4878.

[22] Z. Ming, H. Zhang, Y. Li, and Y. Liang, Mixed H_2/H_∞ control for nonlinear closed-loop Stackelberg games with application to power systems, *IEEE Trans. Autom. Sci. Eng.*, 2024, 21(1), 69–77.

[23] S. Yu, H. Zhang, Z. Ming, and J. Sun, Adaptive optimal control via continuous-time Q-learning for Stackelberg-Nash games of uncertain nonlinear systems, *IEEE Trans. Syst. Man Cybern. Syst.*, 2024, 54(7), 4461–4470.

[24] Y. Ren, Q. Wang, and Z. Duan, Output-feedback Q-learning for discrete-time linear H_∞ tracking control: A Stackelberg game approach, *Int. J. Robust Nonlinear Control*, 2022, 32(12), 6805–6828.

[25] 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 Trans. Cybern.*, 2021, 51(9), 4648–4660.

[26] W. Tushar, C. Yuen, T. K. Saha, S. Nizami, M. R. Alam, D. B. Smith, and H. V. Poor, A survey of cyber-physical systems from a game-theoretic perspective, *IEEE Access*, 2023, 11, 9799–9834.

[27] Y. Liu, L. Tang, S. Tong, C. Chen, and D. Li, Reinforcement learning design-based adaptive tracking control with less learning parameters for nonlinear discrete-time MIMO systems, *IEEE Trans. Neural Netw. Learn. Syst.*, 2015, 26(1), 165–176.

[28] W. Shi, S. Song, C. Wu, and C. Chen, Multi pseudo Q-learning-based deterministic policy gradient for tracking control of autonomous underwater vehicles, *IEEE Trans. Neural Netw. Learn. Syst.*, 2019, 30(12), 3534–3546.

[29] R. Song, F. L. Lewis, and Q. Wei, Off-policy integral reinforcement learning method to solve nonlinear continuous-time multiplayer nonzero-sum games, *IEEE Trans. Neural Netw. Learn. Syst.*, 2017, 28(3), 704–713.

[30] J. Lu, Q. Wei, T. Zhou, Z. Wang, and F. Wang, Event-triggered near-optimal control for unknown discrete-time nonlinear systems using parallel control, *IEEE Trans. Cybern.*, 2023, 53(3), 1890–1904.

[31] F. Wang, N. Jin, D. Liu, and Q. Wei, Adaptive dynamic programming for finite-horizon optimal control of discrete-time nonlinear systems with ε-error bound, *IEEE Trans. Neural Netw.*, 2011, 22(1), 24–36.

[32] J. Lu, X. Wang, Q. Wei, and F. Wang, Nearly optimal stabilization of unknown continuous-time nonlinear systems: A new parallel control approach, *Neurocomputing*, 2024, 578, 127421.

[33] K. Zhang, H. Zhang, Y. Mu, and C. Liu, Decentralized tracking optimization control for partially unknown fuzzy interconnected systems via reinforcement learning method, *IEEE Trans. Fuzzy Syst.*, 2021, 29(4), 917–926.

[34] H. Su, H. Zhang, H. Jiang, and Y. Wen, Decentralized event-triggered adaptive control of discrete-time nonzero-sum games over wireless sensor-actuator networks with input constraints, *IEEE Trans. Neural Netw. Learn. Syst.*, 2020, 31(10), 4254–4266.

[35] L. Xia, Q. Li, and R. Song, Adaptive event-triggered average tracking control with activable event-triggering mechanisms, *IEEE Trans. Syst. Man Cybern. Syst.*, 2023, 53(10), 6067–6079.

[36] R. Song, L. Liu, and B. Hu, Aperiodic sampling artificial-actual H_∞ optimal control for interconnected constrained systems, *IEEE Trans. Autom. Sci. Eng*., to be published.

[37] J. Tan, S. Xue, Z. Guo, H. Li, H. Cao, and B. Chen, Data-driven optimal shared control of unmanned aerial vehicles, *Neurocomputing*, 2025, 622, 129428.

[38] J. Tan, S. Xue, H. Cao, and S. Ge, Human-AI interactive optimized shared control, *J. Autom. Intell*., to be published.

[39] J. Tan, S. Xue, H. Li, Z. Guo, H. Cao, and D. Li, Prescribed performance robust approximate optimal tracking control via Stackelberg game, *IEEE Trans. Automat. Sci. Eng.*, 2025, 22, 12871–12883.

[40] J. Tan, S. Xue, H. Li, H. Cao, and D. Li, Safe stabilization control for interconnected virtual-real systems via model-based reinforcement learning, in *Proc. 2024 14th Asian Control Conference*, Dalian, China, 2024, 605–610.

[41] J. Tan, S. Xue, H. Cao, and H. Li, Safe human-machine cooperative game with level-k rationality modeled human impact, in *Proc. 2023 IEEE International Conference on Development and Learning*, Macao, China, 2023, 188–193.

[42] Q. Wei, Z. Zhu, J. Zhang, and F. Wang, A parallel control method for zero-sum game with unknown time-varying system, *Int. J. Intell. Control Syst.*, 2024, 29(1), 37–40.

[43] Q. Wang and Z. Wang, Adaptive bipartite consensus of multi-agent systems with parameter uncertainty and leader of nonzero input under signed digraph, *Int. J. Intell. Control Syst*., to be published.

Junkai Tan received the BE degree in electrical engineering from School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China, in 2023. He is currently pursuing the ME degree in electrical engineering at School of Electrical Engineering, Xi' an Jiaotong University, China. His research interests include adaptive dynamic programming and inverse reinforcement learning.

Shuangsi Xue received the BE degree in electrical engineering and automation from Hunan University, Changsha, China, in 2014, and the ME and PhD degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2018 and 2023, respectively. He is currently an assistant professor at School of Electrical Engineering, Xi'an Jiaotong University, China. His research interests include adaptive control and data-driven control of networked systems.

Hui Cao received the BE, ME, and PhD degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2000, 2004, and 2009, respectively. He is currently a professor at School of Electrical Engineering, Xi'an Jiaotong University, China. He was a postdoctoral research fellow at Department of Electrical and Computer Engineering, National University of Singapore, Singapore, from 2014 to 2015. He has authored or coauthored over 30 scientific and technical papers in recent years. He was a recipient of the Second Prize of National Technical Invention Award. His research interests include knowledge representation and discovery.