Stackelberg Game-based Robust Optimal Control of Cyber-Physical System Under Hybrid Attacks

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. We develop a hierarchical game model where 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: 1) A unified game-theoretic approach for modeling hybrid attack-defense mechanisms, 2) Online reinforcement learning algorithms for real-time strategy adaptation, and 3) Rigorous stability analysis using Lyapunov theory. Theoretical guarantees of convergence are established for the proposed learning scheme. Comprehensive experiments on a robotic platform validate the framework's effectiveness in maintaining control performance under diverse attack scenarios.**

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

I. Introduction

CYBER-PHYSICAL systems (CPS) integrate physical processes with computational elements, playing a crucial role in modern society. While widely used in critical infrastructures like 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, where 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 work has explored various aspects of game-theoretic CPS security. In [17, 18], robust Stackelberg games were formulated to analyze Nash equilibria under hierarchical decision-making. A Hamiltonian-driven approach was proposed in [19] for deriving optimal stabilization controllers. Research in [20, 21] developed single-critic learning algorithms combining H_2 and H_∞ control for uncertain nonlinear stochastic systems. Data-driven methods were investigated in [22, 23] to achieve optimal mixed H_2/H_∞ performance. For discrete-time linear systems, [24] proposed a robust Stackelberg game incorporating both control indices. However, existing mixed H_2/H_∞ 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 like 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 demonstrate RL's potential 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, this paper proposes 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 Lyapunov analysis. Extensive simulations on a four-wheeled robot platform validate the framework's effectiveness. The key contributions are:

Junkai Tan, Shuangsi Xue, and Hui Cao are with the Shaanxi Key Laboratory of Smart Grid, Xi'an Jiaotong University, Xi'an 710049, China, and also with the State Key Laboratory of Electrical Insulation and Power Equipment, 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) (*Corresponding author: Shuangsi Xue*).

1) A unified Stackelberg game framework that advances existing work [5, 14, 39] in three aspects: (1) Integration of both H_2 and H_∞ performance indices to characterize attack-defense objectives (2) Explicit modeling of stochastic hybrid attacks through Bernoulli switching signals (3) Consideration of input constraints and system uncertainties in the game formulation
2) An efficient actor-critic architecture that improves upon traditional methods [7, 12] through: (1) Online concurrent learning of value functions and control policies (2) Lower computational complexity without requiring complete system models (3) Provable convergence guarantees under persistent disturbances
3) Comprehensive experimental validation demonstrating clear advantages over baseline approaches [40, 41]: (1) Reduction in tracking errors under hybrid attacks (2) Faster convergence to optimal strategies (3) 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 VI validates the framework through comprehensive numerical experiments. Section VII summarizes our findings and discusses future research directions.

II. Preliminaries

Consider a continuous-time nonlinear CPS with disturbances:

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

where $x(t) \in \mathbb{R}^n$ denotes the system state vector, $u(t) \in \mathbb{R}^m$ represents the control input, $\omega(t) \in \mathbb{R}^m$ indicates external disturbances, $f : \mathbb{R}^n \to \mathbb{R}^n$ 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 dynamics are:

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

where all functions satisfy local Lipschitz continuity conditions. For analytical purposes, we integrate the system (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 $X = [x^\top, x_d^\top]^\top \in \mathbb{R}^{2\times n}$ combines actual and desired states, control input $U = [u^\top, 0_{1\times m}]^\top \in \mathbb{R}^{2\times m}$ incorporates system and reference controls, and Y denotes the performance output. The system matrices are given by:

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

where Q and R are positive definite weighting matrices for state and control costs. We make the following assumptions:

Assumption 1. *For system (4), we assume:*

1) *Functions F and G are locally Lipschitz on $X \in \chi \subset \mathbb{R}^n$, with $F(0) = 0$ and $\|G\| \le G_H$ for all $X \in \chi$.*
2) *Matrices Q and R satisfy $\underline{\lambda}_Q\mathcal{I} \preceq Q \preceq \bar{\lambda}_Q\mathcal{I}$ and $\underline{\lambda}_R\mathcal{I} \preceq R \preceq \bar{\lambda}_R\mathcal{I}$, where $0 \le \underline{\lambda}_Q, \underline{\lambda}_R < \bar{\lambda}_Q, \bar{\lambda}_R < \infty$.*

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 H_2 control input $U^*(t)$, while the defender minimizes damage through H_∞ control $\omega^*(t)$. Based on system (4), we model this interaction as a Stackelberg game. For the nonlinear CPS (4), the H_2 and H_∞ performance objectives are defined as:

$$\begin{aligned} J_{\mathbb{D}}(X_0, U, \omega) &= \mathbf{E}\left\{\int_t^\infty \|Y\|^2 d\tau\right\} \\ &= \mathbf{E}\left\{\int_t^\infty \left(X^\top QX + \alpha U^\top RU\right) d\tau\right\} \end{aligned} \tag{5}$$

$$\begin{aligned} J_{\mathbb{A}}(X_0, U, \omega) &= \mathbf{E}\left\{\int_t^\infty \gamma^2\|\omega\|^2 - \|Y\|^2 d\tau\right\} \\ &= \mathbf{E}\left\{\int_t^\infty \left(\beta\gamma^2\|\omega\|^2 - X^\top QX - \alpha U^\top RU\right) d\tau\right\} \end{aligned} \tag{6}$$

where γ denotes the disturbance attenuation level. $J_{\mathbb{D}}$ measures H_2 performance while $J_{\mathbb{A}}$ 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 $\mathbb{D}$ and attacker $\mathbb{A}$ with objectives (5) and (6). The hierarchical decision process involves:*

Level 1 (Defense): *$\mathbb{D}$ determines baseline strategy $U_{L1} \in \Omega_U$:*

$$J_{\mathbb{D}_{L1}}(X_0, U_{L1}^*, 0) = \min_{U\in\Omega_U} J_{\mathbb{D}}(X_0, U_{L1}, 0)$$

Level 2 (Attack): *Given U_{L1}^*, $\mathbb{A}$ optimizes $\omega_{L2} \in \Omega_W$:*

$$J_{\mathbb{A}_{L2}}(X_0, U_{L1}^*, \omega_{L2}^*) = \max_{\omega\in\Omega_W} J_{\mathbb{A}}(X_0, U_{L1}^*, \omega_{L2})$$

Level 3 (Defense Update): *$\mathbb{D}$ updates $U_{L3} \in \Omega_U$ given ω_{L2}^*:*

$$J_{\mathbb{D}_{L3}}(X_0, U_{L3}^*, \omega_{L2}^*) = \min_{U\in\Omega_U} J_{\mathbb{D}}(X_0, U_{L3}, \omega_{L2}^*)$$

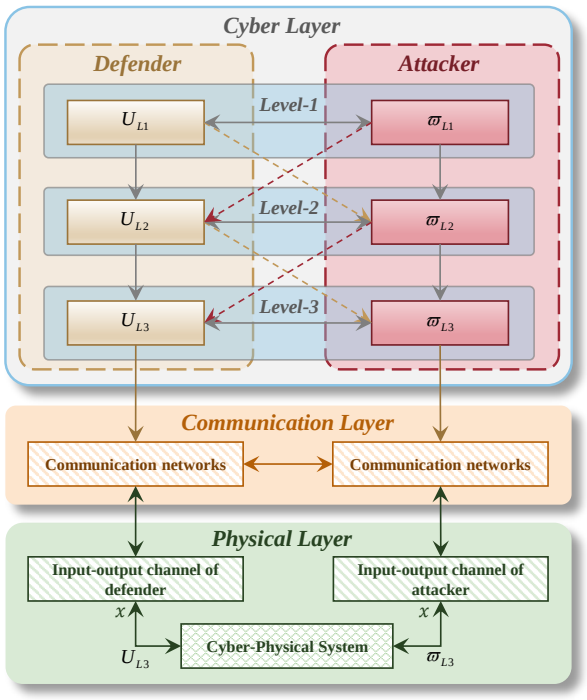


Fig. 1. Stackelberg game-based hybrid attack-defense interaction.

The resulting $U^* \triangleq U^*_{L3}$ *and* $\omega^* \triangleq \omega^*_{L2}$ *form the Stackelberg equilibrium.*

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

Problem 1. *(Stackelberg Game Framework for Hybrid Attack-Defense) Consider the nonlinear CPS (4) under hybrid attack-defense mechanisms characterized by stochastic switching signals* $a(t)$ *and* $b(t)$ *following Bernoulli distributions with success probabilities* $P(a=1)=\alpha$ *and* $P(b=1)=\beta$*. The objective is to:*

1) *Design optimal defense strategy* $U^*(t)$ *that maximizes system damage using* H_∞ *performance index (5)*
2) *Develop optimal attack policy* $\omega^*(t)$ *that minimizes adverse impacts through* H_2 *control (6)*

The interaction could be formulated as the following optimization problem:

$$J^*_{\mathbb{D}}(X_0) = \min_{\bar{U}\in\Omega_U} J_{\mathbb{D}}(X_0,\bar{U},\bar{\omega}^*)$$
$$J^*_{\mathbb{A}}(X_0) = \min_{\bar{\omega}\in\Omega_W} J_{\mathbb{A}}(X_0,\bar{U},\bar{\omega})$$

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

$$\bar{U}^* = \begin{cases} U^*, & a(t)=1 \\ 0, & otherwise \end{cases}$$
$$\bar{\omega}^* = \begin{cases} \omega^*, & b(t)=1 \\ 0, & otherwise \end{cases} \tag{7}$$

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

Based on the formulated Stackelberg game framework, we first analyze the attacker's optimization problem:

$$J^*_{\mathbb{A}} = \max_{\omega} \mathbf{E}\left\{\int_t^\infty \left(\beta\gamma^2\|\omega\|^2 - X^\top QX - \alpha U^\top RU\right) d\tau\right\} \tag{8}$$

The attacker's Hamiltonian function is defined as:

$$H_{\mathbb{A}}(X,U,\omega,\nabla J^*_{\mathbb{A}}) = \nabla J^{*\top}_{\mathbb{A}}(F+\alpha GU+\beta K\omega) + \beta\gamma^2\|\omega\|^2 - X^\top QX - \alpha U^\top RU \tag{9}$$

By minimizing $H_{\mathbb{A}}$, the optimal attack strategy is obtained as:

$$\omega^*(U) = -\frac{K^\top}{2\gamma^2}\nabla J^*_{\mathbb{A}} \tag{10}$$

The evolution of the attacker's value function is captured by costate dynamics:

$$\dot{\lambda}_2 = -\left(\frac{\partial F}{\partial x} + \frac{\partial G}{\partial x}U + \frac{\partial K}{\partial x}\omega^* + G\frac{\partial U^\top}{\partial x}X\right)^\top \nabla J^*_{\mathbb{A}} + 2QX + 2RU\frac{\partial U}{\partial x} \tag{11}$$

For the defender's optimization:

$$J^*_{\mathbb{D}} = \min_{U} \mathbf{E}\left\{\int_t^\infty \left(X^\top QX + \alpha U^\top RU + \eta^\top\dot{\lambda}_2\right) d\tau\right\} \tag{12}$$

where η is the Lagrange multiplier. The defender's Hamiltonian is:

$$H_{\mathbb{D}}(X,U,\omega^*,\nabla J^*_{\mathbb{D}},\eta) = \nabla J^{*\top}_{\mathbb{D}}(F+\alpha GU+\beta K\omega) + X^\top QX + \alpha U^\top RU + \eta^\top\dot{\lambda}_2 \tag{13}$$

The optimal defense strategy is derived as:

$$U^*(\omega^*) = -\frac{1}{2}R^{-1}\left(G^\top\nabla J^*_{\mathbb{D}} - \nabla_x G\eta\nabla J^*_{\mathbb{A}}\right) \tag{14}$$

The Lagrange multiplier dynamics are governed by:

$$\dot{\eta} = \sum_{i=1}^n \eta_i\left(\frac{\partial K}{\partial X_i}\cdot\frac{\partial\omega^*}{\partial\nabla J^*_{\mathbb{A}}}\right)^\top\nabla J^*_{\mathbb{A}} + (\nabla F + \alpha\nabla GU + \beta\nabla K\omega^* + \alpha G\nabla U)\,\eta - \left(K\cdot\frac{\partial\omega^*}{\partial\nabla J^*_{\mathbb{A}}}\right)^\top\nabla J^*_{\mathbb{D}} \tag{15}$$

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.

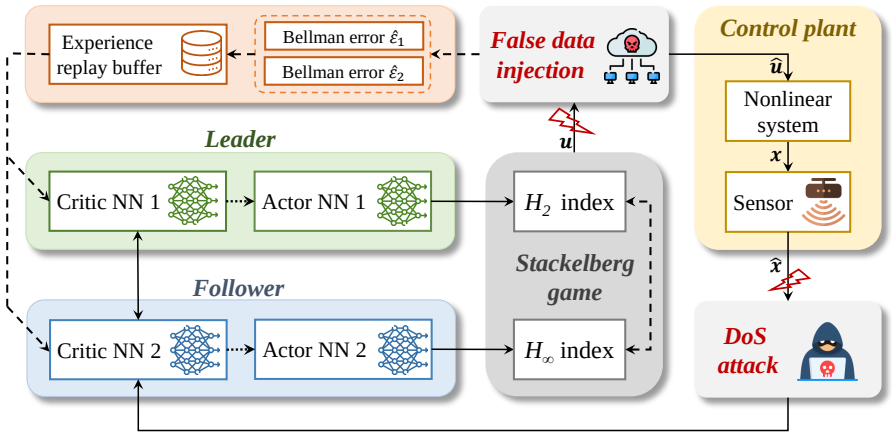


Fig. 2. Structure of the proposed Stackelberg game-based hybrid attack and defense.

IV. Main Results

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 policies while the critic networks evaluate performance by 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:

$$J_i^*(X) = \mathcal{W}_{ci}^\top \phi_{ci}(X) + \epsilon_{ci}(X),\ i = 1, 2 \tag{16}$$

$$U^*(X) = -\frac{1}{2}\Big(R^{-1}G^\top\left(\nabla\phi_{a1}^\top \mathcal{W}_{a1} + \nabla\epsilon_{a1}^\top\right) - \left(\mathcal{W}_{a2}^\top\nabla\phi_{a2} + \nabla\epsilon_{a2}\right)\nabla_x G\eta\Big) \tag{17}$$

$$\omega^*(X) = -\frac{K^\top}{2\gamma^2}\left(\nabla\phi_{a2}^\top\mathcal{W}_{a2} + \nabla\epsilon_{a2}^\top\right) \tag{18}$$

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

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

$$\hat{J}_i(X) = \hat{\mathcal{W}}_{ci}^\top\phi_{ci},\ i = 1, 2 \tag{19}$$

$$\hat{U}(X) = -\frac{1}{2}\Big(R^{-1}G^\top\nabla\phi_{a1}^\top\hat{\mathcal{W}}_{a1} - \hat{\mathcal{W}}_{a2}^\top\nabla\phi_{a2}\nabla_x G\eta\Big) \tag{20}$$

$$\hat{\omega}(X) = -\frac{K^\top}{2\gamma^2}\nabla\phi_{a2}^\top\hat{\mathcal{W}}_{a2} \tag{21}$$

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

$$\delta_1 = \left(\nabla\phi_{c1}^\top\hat{\mathcal{W}}_{c1}\right)^\top\left(F + \alpha G\hat{U} + \beta K\hat{\omega}\right) + X^\top QX + \alpha U^\top RU + \eta^\top\dot{\lambda}_2 \tag{22}$$

$$\delta_2 = \left(\nabla\phi_{c2}^\top\hat{\mathcal{W}}_{c2}\right)^\top\left(F + \alpha G\hat{U} + \beta K\hat{\omega}\right) + \beta\gamma^2\|\hat{\omega}\|^2 - X^\top QX - \alpha U^\top RU \tag{23}$$

For analytical purposes, we make the following assumption on network parameters:

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

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

B. Online learning of value functions

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: $E_i = \delta_i^\top\delta_i + \sum_{k=1\cdots N}\delta_i^{k\top}\delta_i^k, i = 1, 2$. The critic network weights are updated through gradient descent:

$$\dot{\hat{\mathcal{W}}}_{ci} = -k_{ci,1}\frac{\sigma_i\delta_i}{\rho_i(t)} - \frac{k_{ci,2}}{N}\sum_{k=1}^{N}\frac{\sigma_i^k\delta_i^k}{\rho_i^k(t)},\ i = 1, 2 \tag{24}$$

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

$$\dot{\hat{\mathcal{W}}}_{ai} = F_i k_{ai}\left(\hat{\mathcal{W}}_{ci} - \hat{\mathcal{W}}_{ai}\right),\ i = 1, 2 \tag{25}$$

where $k_{ai} > 0$ are actor learning rates and F_i are positive definite matrices. To ensure convergence, we require the following excitation condition:

Assumption 3. *(Persistent excitation) [42, 43] The collected data satisfies:*

$$\Lambda_{1,i}\mathcal{I}_{m,i} \leqslant \int_t^{t+T} \frac{\sigma_i \sigma_i^\top}{\rho_i} \mathrm{d}\tau, \quad \Lambda_{2,i}\mathcal{I}_{m,i} \leqslant \inf_{t\in\mathbf{R}_{t\geq t_0}} \frac{\sigma_i^k \sigma_i^{k\top}}{N\rho_i^k}$$

where $\mathcal{I}_{m,i}$ is identity matrix and either $\Lambda_{1,i}$ 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 weights $\hat{\mathcal{W}}_{ai}$ and critic weights $\hat{\mathcal{W}}_{ci}$
- Learning rates $k_{ci,j}, k_{ai}$
- Projection matrices F_i, $i, j \in \{1, 2\}$

2: **while** $t < T_{end}$ **do**

3: Compute optimal strategies:
- Defense policy $\hat{U}$ via (20)
- Attack policy $\hat{\omega}$ via (21)

4: Evaluate Bellman errors from (23):
- 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 weights via (24)
- Actor weights via (25)

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 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 (20) and attack strategy in (21), we have the following error bounds:

$$\|U^*(X) - \hat{U}(X)\|^2 \leq \bar{\Sigma}_{u_1}\|\tilde{W}_{a1}\|^2 + \Pi_{u_1} \tag{26}$$

$$\|\omega^*(X) - \hat{\omega}(X)\|^2 \leq \bar{\Sigma}_{u_2}\|\tilde{W}_{a2}\|^2 + \Pi_{u_2} \tag{27}$$

where $\bar{\Sigma}_{u_i}$ are positive constants determined by network activation bounds, and Π_{u_i} represent bounded approximation residuals.

The key stability results are summarized in the following theorems:

Theorem 1. *Consider system (4) with the proposed Stackelberg game framework. Under Assumptions 1-3, for networks updated via (24)-(25), the augmented error state $Z = [X^\top, \tilde{W}_{c1}^\top, \tilde{W}_{c2}^\top, \tilde{W}_{a1}^\top, \tilde{W}_{a2}^\top]^\top$ remains UUB if:*

$$\|Z\| \geq \sqrt{\Psi_{\text{res}}/(\underline{\lambda}_{\mathcal{H}}\mathcal{I})} \tag{28}$$

where Ψ_{res} is defined in (36), and $\underline{\lambda}_{\mathcal{H}}$ denotes the minimum eigenvalue of $\mathcal{H}$ in (37).

Theorem 2. *For system (4), the approximate policies $\hat{U}$ (20) and $\hat{\omega}$ (21) converge to their optimal counterparts U^* and ω^*, reaching a unique Stackelberg equilibrium.*

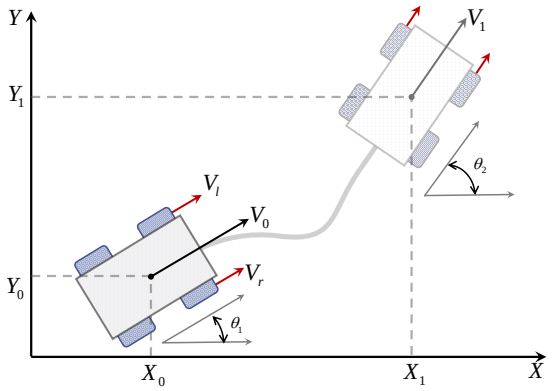


Fig. 3. Schematic of the four-wheeled mobile robot.

VI. Numerical Simulations

A. Simulation setup

To validate the proposed Stackelberg game framework, we conduct numerical experiments on a four-wheeled differential drive robot system. The robot's kinematic model follows:

$$f = \mathbf{0}_{3\times 1}, \quad g = \begin{bmatrix} \cos(\theta) & 0 \\ \sin(\theta) & 0 \\ 0 & 1 \end{bmatrix} \tag{29}$$

where states $x = [x, y, \theta]^\top$ represent position coordinates and heading angle, and control inputs $u = [v, \Omega]^\top$ denote 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 $e = [x - x_d, y - y_d, \theta - \theta_d]^\top$.

For neural network implementation, we use basis functions: $\phi_i = \left[e_1^2, e_2^2, e_1^2 + e_2^2, e_3^2, e_1^2 + e_3^2, e_2^2 + e_3^2\right]^\top$. All network weights are initialized to 5. Key parameters are listed in Table I. The approximate optimal control (AOC) method from [40] serves as baseline for comparison.

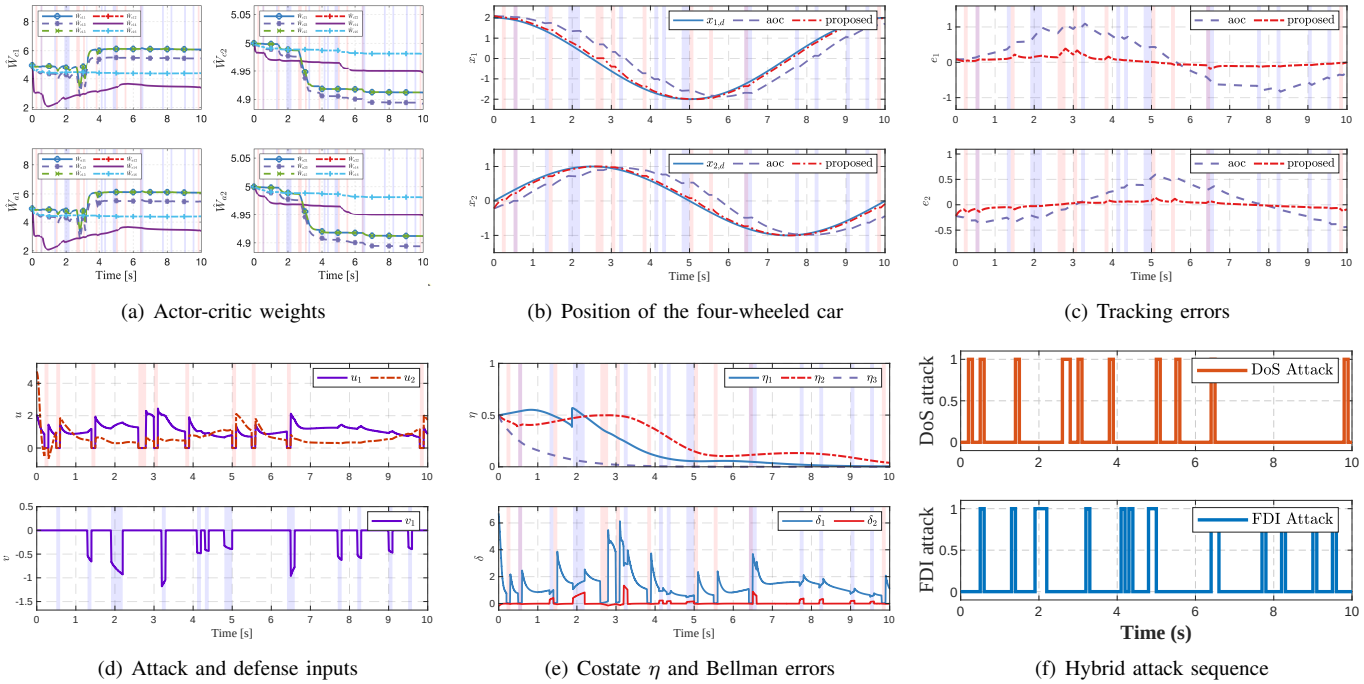
To handle input constraints $v, \Omega \in [-5, 5]$, we reconstruct the defense penalty function as:

$$\Psi(U) = 2R \int_0^U \mu_{\mathbb{D}} \tanh^{-1}\left(\frac{\zeta_U}{\mu_{\mathbb{D}}}\right) \mathrm{d}\zeta_U \tag{30}$$

The constrained defense input becomes:

$$\hat{U} = -\mu_{\mathbb{D}} \tanh\left\{\frac{R^{-1}G^\top \nabla\phi_{a1}^\top \hat{\mathcal{W}}_{a1} - \hat{\mathcal{W}}_{a2}^\top \nabla\phi_{a2}\nabla_x G\eta}{2\mu_{\mathbb{D}}}\right\}$$

where $\mu_{\mathbb{D}} = 5$ denotes the input saturation bound.



(a) Actor-critic weights (b) Position of the four-wheeled car (c) Tracking errors

(d) Attack and defense inputs (e) Costate η and Bellman errors (f) Hybrid attack sequence

Fig. 4. Simulation results of the proposed Stackelberg game-based hybrid attack-defense framework.

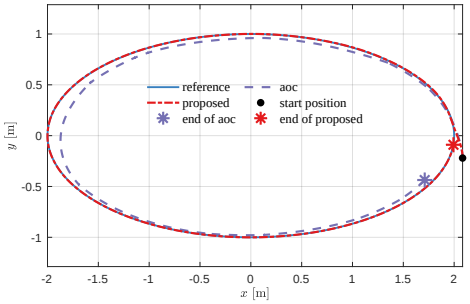


Fig. 5. Trajectory of the system states.

TABLE I
PARAMETERS OF THE NUMERICAL SIMULATION

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

B. Simulation results

The numerical results demonstrate the framework's effectiveness through various performance metrics: Fig. 4(a) shows the evolution of neural network parameters. The rapid convergence of both critic and actor weights within 10s validates the learning efficiency of the proposed algorithm. Fig. 4(b) illustrates the system state trajectories. Despite persistent attacks, the states closely track their reference values, indicating strong robustness of the controller design. Fig. 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. Fig. 4(d) depicts the control signals from both agents. The defender generates smooth control inputs while effectively counteracting the attacker's disturbances, demonstrating the framework's ability to balance performance and energy efficiency. Fig. 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. Fig. 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. Fig. 5 demonstrates the robot's path following capability. The minimal deviation from desired trajectories under attack validates the framework's resilience 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.

C. Application to Critical Infrastructure Systems

Our framework demonstrates significant potential for real-world critical infrastructure protection:
Our framework can be applied in several critical domains: **Power Grid Applications:** 1) Wide-area monitoring against coordinated cyber-attacks; 2) Real-time detection of false data injection attacks; 3) Adaptive defense against DoS attacks; 4) Optimal protection of critical grid nodes. **Autonomous Vehicle Systems:** 1) Secure V2V communication protocols; 2) Robust trajectory tracking under GPS spoofing; 3) Dynamic defense for platooning safety; 4) Multi-layer protection for vehicle networks. **Industrial Control Systems:** 1) Protection of SCADA networks; 2) Resilient control of

robotic systems; 3) Adaptive security for smart factories; 4) 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 framework's effectiveness in maintaining system stability and performance under various attack scenarios.

VII. Conclusions

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 attacker's and defender's objectives 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 framework's capability to maintain control performance under attacks. Future work will explore extensions to cooperative multi-agent systems and practical implementations.

Acknowledgements

This research is supported by China Postdoctoral Science Foundation (Project ID: 2024M762602).

Appendix: Proof of Theorem 1-2

The proof of Theorem 1 follows from Lyapunov stability analysis.

Proof. Consider the Lyapunov candidate:

$$\mathcal{V} = J_1^* + J_2^* + \sum_{i=1}^{2} \frac{1}{2}(\tilde{W}_{ci}^\top \tilde{W}_{ci} + \tilde{W}_{ai}^\top \tilde{W}_{ai}) \tag{31}$$

Define the Bellman errors δ_i and δ_i^k:

$$\delta_i = -\sigma_i^\top \tilde{W}_{ci} + \Phi_i(\tilde{W}_{a1}, \tilde{W}_{a2}) + \Delta_i \tag{32}$$

$$\delta_i^k = -(\sigma_i^k)^\top \tilde{W}_{ci} + \Phi_i^k(\tilde{W}_{a1}, \tilde{W}_{a2}) + \Delta_i^k \tag{33}$$

where Φ_i and Φ_i^k contain quadratic terms, and Δ_i, Δ_i^k are bounded residuals.

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

$$\dot{\mathcal{V}} = \sum_{i=1}^{2} \left(\nabla J_i^* \dot{X} + \tilde{W}_{ci}^\top \dot{\tilde{\mathcal{W}}}_{ci} + \tilde{W}_{ai}^\top \dot{\tilde{\mathcal{W}}}_{ai} \right) \tag{34}$$

Substituting the weight update laws and applying Young's inequality:

$$\dot{\mathcal{V}} \leq -Z^T \mathcal{H} Z + \Psi_{\text{res}} \tag{35}$$

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

$$\Psi_{\text{res}} = \sum_{i=1}^{2} \left(\frac{1}{2} k_{ci,1} \|\Delta_{\mathcal{W}i}\|^2 + \frac{1}{2} k_{ci,2} \|\Delta_{\mathcal{W}i}^k\|^2 \right) + \gamma^2 \Pi_{u_2} + \bar{\lambda}_{R,1} \Pi_{u_1} \tag{36}$$

$$\mathcal{H} = \begin{bmatrix} \mu_1 & 0 & 0 & 0 & 0 \\ 0 & \mu_2 & 0 & 0 & 0 \\ 0 & \mu_3 & \mu_4 & 0 & 0 \\ 0 & \mu_5 & 0 & \mu_6 & 0 \\ 0 & 0 & \mu_7 & 0 & \mu_8 \end{bmatrix} \tag{37}$$

where matrix coefficients μ_i are defined as:

$$\begin{cases} \mu_1 = \underline{\lambda}_{Q1} - \underline{\lambda}_{Q2} \\ \mu_2 = \frac{1}{2}(k_{c1,1}\sigma_1\sigma_1^T + k_{c1,2}\Lambda_{2,1}\mathcal{I}_{m,1}) \\ \mu_3 = (k_{c1,1} + k_{c2,1})\sigma_1\sigma_2^T \\ \mu_4 = \frac{1}{2}(k_{c2,1}\sigma_2\sigma_2^T + k_{c2,2}\Lambda_{2,2}\mathcal{I}_{m,2}) \\ \mu_5 = -F_1\mathcal{I}_{m,1} \\ \mu_6 = F_1\mathcal{I}_{m,1} - \bar{\lambda}_{R,1}\Sigma_{u_1}\mathcal{I}_{m,1} \\ \mu_7 = -F_2\mathcal{I}_{m,2} \\ \mu_8 = F_2\mathcal{I}_{m,2} + \gamma^2\Sigma_{u_2}\mathcal{I}_{m,2} \end{cases} \tag{38}$$

where $\Delta_{\mathcal{W}1} = 0.25\tilde{W}_{a1}^T G_\sigma \tilde{W}_{a1} + \Delta_1 + \xi_{H1}$, $\Delta_{\mathcal{W}2} = 0.25\tilde{W}_{a2}^T K_\sigma \tilde{W}_{a2} - 0.25\tilde{W}_{a1}^T G_\sigma \tilde{W}_{a1} + \Delta_2$, $\Delta_{\mathcal{W}1}^k = 0.25\tilde{W}_{a1}^T G_{\sigma,k} \tilde{W}_{a1} + \Delta_1^k$, $\Delta_{\mathcal{W}2}^k = 0.25\tilde{W}_{a2}^T K_{\sigma,k} \tilde{W}_{a2} - 0.25\tilde{W}_{a1}^T G_{\sigma,k} \tilde{W}_{a1} + \Delta_2^k$. Therefore, the augmented error state Z converges to the compact set defined by (28), establishing UUB. □

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

References

[1] H. Li and Q. Wei, "A Multi-Observer Based Optimal Control Method for Nonlinear Systems Under Sensor Attacks," *IEEE Transactions on Automation Science and Engineering*, pp. 1–10, 2024.

[2] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber–Physical System Security for the Electric Power Grid," *Proceedings of the IEEE*, vol. 100, pp. 210–224, Jan. 2012.

[3] J. Dong, Z. Ye, and D. Zhang, "Finite-Time Security Control of Networked Unmanned Marine Vehicle Systems Subject to DoS Attack," *IEEE Transactions on Intelligent Vehicles*, vol. 9, pp. 3464–3477, Feb. 2024.

[4] N. Anwar, G. Xiong, W. Lu, P. Ye, H. Zhao, and Q. Wei, "Cyber-Physical -Social Systems for Smart Cities: An Overview," in *2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI)*, pp. 348–353, July 2021.

[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 Transactions on Industrial Cyber-Physical Systems*, vol. 2, pp. 118–129, 2024.

[6] Q. Wei, H. Li, and F.-Y. Wang, "Parallel control for continuous-time linear systems: A case study," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, pp. 919–928, July 2020.

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

[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 of Things Journal*, vol. 9, pp. 9659–9674, June 2022.

[9] Y. Wu, M. Chen, H. Li, and M. Chadli, "Event-Triggered Distributed Intelligent Learning Control of Six-Rotor UAVs Under FDI Attacks," *IEEE Transactions on Artificial Intelligence*, vol. 5, pp. 3299–3312, July 2024.

[10] Y. Xu, T. Li, Y. Yang, S. Tong, and C. L. P. Chen, "Simplified ADP for Event-Triggered Control of Multiagent Systems Against FDI Attacks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, pp. 4672–4683, Aug. 2023.

[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," *Journal of the Franklin Institute*, vol. 361, p. 106715, Apr. 2024.
[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 Transactions on Automatic Control*, pp. 1–8, 2024.
[13] J. Tan, S. Xue, H. Cao, and H. Li, "Nash Equilibrium Solution Based on Safety-Guarding Reinforcement Learning in Nonzero-Sum Game," in *2023 International Conference on Advanced Robotics and Mechatronics (ICARM)*, pp. 630–635, July 2023.
[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 Transactions on Network Science and Engineering*, vol. 10, pp. 2086–2097, July 2023.
[15] P. Shukla, L. An, A. Chakrabortty, and A. Duel-Hallen, "A Robust Stackelberg Game for Cyber-Security Investment in Networked Control Systems," *IEEE Transactions on Control Systems Technology*, vol. 31, pp. 856–871, Mar. 2023.
[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," *Information Sciences*, vol. 642, p. 119134, Sept. 2023.
[17] Z. Jing, X. Li, P. Ju, and H. Zhang, "Optimal Control and Filtering for Hierarchical Decision Problems With H_ınfty Constraint Based on Stackelberg Strategy," *IEEE Transactions on Automatic Control*, vol. 69, pp. 6238–6245, Sept. 2024.
[18] Z. Jing, P. Ju, and X. Li, "Leader-Follower Based Online Reinforcement Learning Algorithm in Problem with Hierarchy Decision Makers," *The International Journal of INTELLIGENT CONTROL AND SYSTEMS*, Mar. 2024.
[19] Y. Yang, M. Mazouchi, and H. Modares, "Hamiltonian-driven adaptive dynamic programming for mixed H_2 / H_∞ performance using sum-of-squares," *International Journal of Robust and Nonlinear Control*, vol. 31, pp. 1941–1963, Apr. 2021.
[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 Transactions on Automation Science and Engineering*, vol. 20, pp. 1934–1944, July 2023.
[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 Transactions on Circuits and Systems II: Express Briefs*, vol. 69, pp. 4874–4878, Dec. 2022.
[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 Transactions on Automation Science and Engineering*, vol. 21, pp. 69–77, Jan. 2024.
[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 Transactions on Systems, Man, and Cybernetics: Systems*, vol. 54, pp. 4461–4470, July 2024.
[24] Y. Ren, Q. Wang, and Z. Duan, "Output-feedback Q-learning for discrete-time linear H∞ tracking control: A Stackelberg game approach," *International Journal of Robust and Nonlinear Control*, vol. 32, no. 12, pp. 6805–6828, 2022.
[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 Transactions on Cybernetics*, vol. 51, pp. 4648–4660, Sept. 2021.
[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*, vol. 11, pp. 9799–9834, 2023.
[27] Y.-J. Liu, L. Tang, S. Tong, C. L. P. Chen, and D.-J. Li, "Reinforcement Learning Design-Based Adaptive Tracking Control With Less Learning Parameters for Nonlinear Discrete-Time MIMO Systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, pp. 165–176, Jan. 2015.
[28] W. Shi, S. Song, C. Wu, and C. L. P. Chen, "Multi Pseudo Q-Learning-Based Deterministic Policy Gradient for Tracking Control of Autonomous Underwater Vehicles," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, pp. 3534–3546, Dec. 2019.
[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 Transactions on Neural Networks and Learning Systems*, vol. 28, pp. 704–713, Mar. 2017.
[30] J. Lu, Q. Wei, T. Zhou, Z. Wang, and F.-Y. Wang, "Event-Triggered Near-Optimal Control for Unknown Discrete-Time Nonlinear Systems Using Parallel Control," *IEEE Transactions on Cybernetics*, vol. 53, pp. 1890–1904, Mar. 2023.
[31] F.-Y. 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 Transactions on Neural Networks*, vol. 22, pp. 24–36, Jan. 2011.
[32] J. Lu, X. Wang, Q. Wei, and F.-Y. Wang, "Nearly optimal stabilization of unknown continuous-time nonlinear systems: A new parallel control approach," *Neurocomputing*, vol. 578, p. 127421, Apr. 2024.
[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 Transactions on Fuzzy Systems*, vol. 29, pp. 917–926, Apr. 2021.
[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 Transactions on Neural Networks and Learning Systems*, vol. 31, pp. 4254–4266, Oct. 2020.
[35] L. Xia, Q. Li, and R. Song, "Adaptive Event-Triggered Average Tracking Control With Activable Event-Triggering Mechanisms," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, pp. 6067–6079, Oct. 2023.
[36] R. Song, L. Liu, and B. Hu, "Aperiodic Sampling Artificial-Actual $H_{\text{ınfty}}$ Optimal Control for Interconnected Constrained Systems," *IEEE Transactions on Automation Science and Engineering*, pp. 1–11, 2023.
[37] J. Tan, S. Xue, Z. Guo, H. Li, H. Cao, and B. Chen, "Data-driven optimal shared control of unmanned aerial vehicles," *Neurocomputing*, p. 129428, Jan. 2025.
[38] J. Tan, S. Xue, H. Cao, and S. S. Ge, "Human-AI interactive optimized shared control," *Journal of Automation and Intelligence*, Jan. 2025.
[39] J. Dong, Z. Ye, and D. Zhang, "Finite-Time Security Control of Networked Unmanned Marine Vehicle Systems Subject to DoS Attack," *IEEE Transactions on Intelligent Vehicles*, vol. 9, pp. 3464–3477, Feb. 2024.
[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 *2024 14th Asian Control Conference (ASCC)*, pp. 605–610, July 2024.
[41] J. Tan, S. Xue, H. Cao, and H. Li, "Safe Human-Machine Cooperative Game with Level-k Rationality Modeled Human Impact," in *2023 IEEE International Conference on Development and Learning (ICDL)*, pp. 188–193, Nov. 2023.
[42] Q. Wei, Z. Zhu, J. Zhang, and F.-Y. Wang, "A Parallel Control Method For Zero-Sum Games With Unknown Time-varying System," *The International Journal of INTELLIGENT CONTROL AND SYSTEMS*, Mar. 2024.
[43] Q. Wang and Z. Wang, "Adaptive Bipartite Consensus of Multi-agent Systems with Parameter Uncertainty and Leader of Nonzero Input under Signed Digraph," *The International Journal of INTELLIGENT CONTROL AND SYSTEMS*, Mar. 2024.

Junkai Tan received the B.E. degree in electrical engineering at the School of Electrical Engineering in Xi'an Jiaotong University, Xi'an, China. He is currently working toward the M.E. degree in electrical engineering at the School of Electrical Engineering, Xi'an Jiaotong University.

His current research interest includes adaptive dynamic programming and inverse reinforcement learning.

Shuangsi Xue (M'24) received the B.E. degree in electrical engineering and automation from Hunan University, Changsha, China, in 2014, and the M.E. and Ph.D. degrees in electrical engineering from Xian Jiaotong University, Xian, China, in 2018 and 2023, respectively. He is currently an Assistant Professor at the School of Electrical Engineering, Xian Jiaotong University.

His current research interest includes adaptive control and data-driven control of networked systems.

Hui Cao (M'11) received the B.E., M.E., and Ph.D. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2000, 2004, and 2009, respectively.

He is a Professor at the School of Electrical Engineering, Xi'an Jiaotong University. He was a Postdoctoral Research Fellow at the 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. His current research interest includes knowledge representation and discovery. Dr. Cao was a recipient of the Second Prize of National Technical Invention Award.