Resilient Control Under Denial-of-Service and Uncertainty: An Adaptive Dynamic Programming Approach

Weinan Gao, *Senior Member, IEEE*, Zhong-Ping Jiang, *Fellow, IEEE*, and Tianyou Chai, *Life Fellow, IEEE*

***Abstract*—In this article, a new framework for the resilient control of continuous-time linear systems under denial-of-service (DoS) attacks and system uncertainty is presented. Integrating techniques from reinforcement learning and output regulation theory, it is shown that resilient optimal controllers can be learned directly from real-time state and input data collected from the systems subjected to attacks. Sufficient conditions are given under which the closed-loop system remains stable given any upper bound of DoS attack duration. Simulation results are used to demonstrate the efficacy of the proposed learning-based framework for resilient control under DoS attacks and model uncertainty.**

***Index Terms*—Adaptive dynamic programming (ADP), denial-of-service (DoS) attack, output regulation, resilient optimal control.**

I. Introduction

As modern control engineering systems become more and more complex and uncertain, it is important to learn controllers from online data that avoids the reliance on the accurate model knowledge. A plausible strategy is to learn the dynamics of control systems first, and then, design controllers based on the learned dynamics [1], [34]. The success of this strategy usually requires the discrepancy between learned and actual dynamics to be small enough. The second strategy is to learn control policy directly from online data instead of learning dynamics [15], [16], [24], [25]. Following this strategy, adaptive dynamic programming (ADP) has been used to search for the solution to an optimal decision making problem based on active agent–environment interactions [9], [17], [21], [23], [29], [31], [33]. Its efficiency has been proved especially in situations where the exact system model is difficult or even impossible to be acquired. Apart from the unknown system dynamics, modern control systems with ADP algorithms must also deal with nonvanishing disturbance, dynamic uncertainty, and cyberattacks.

By the integration of the differential game theory and ADP, one can learn the Nash policy to attenuate the effect from nonvanishing disturbances [22], [30]. In order to further reject disturbance while asymptotically tracking some prescribed reference signal, ADP has been leveraged to achieve output regulation in an adaptive optimal sense [3], [7], [14]. However, these methods usually consider static uncertainties in the system based on an assumption that the system order is known exactly. In order to relax this assumption, robust ADP (for short, RADP) has been proposed via a combination of ADP and nonlinear small-gain theory [13]. Under some mild small-gain condition, one can learn a robust optimal control policy via RADP for unknown control systems in the presence of dynamic uncertainties. Nevertheless, most of existing (robust) ADP methods ignore the resiliency of control systems when the control system is subject to malicious attacks.

In order to overcome the effect of denial-of-service (DoS) attacks, safety-constrained optimal control [2], resilient model predictive control [27], and distributed resilient control with setpoint optimization [19] have been studied for networked systems. As a superset of ADP, reinforcement learning (RL) has been studied to improve the resilience of control systems while preserving the optimality or at least suboptimality; see [11] and [12] and references therein. However, most of these RL studies fail to guarantee the asymptotic stability of the closed-loop system. In our previous work, we have proposed a model-free method to analyze the resiliency and the stability of control systems in closed loop with controllers learned by ADP [8]. Without knowing exactly the system dynamics, one can provide an estimate on the upper bound of cyberattacks for which the control system remains operational. However, this method of [8] may fail to work whenever the cyberattacks go beyond our estimation. As cyberattacks become increasingly frequent in modern engineering systems, it is urgent to develop a novel learning-based controller design framework that addresses resiliency besides the common issues of stability and optimality.

A. Contributions

The contributions of this article are summarized as follows. First, we develop an original learning-based resilient optimal controller design framework via ADP for a class of continuous-time linear systems affected by disturbances and DoS attacks. Interestingly, one can rely on this framework to design a resilient optimal control policy to achieve output regulation (i.e., asymptotic tracking with disturbance rejection) given any upper bound of DoS attack duration. Second, comparing with existing resilient control design approaches [4], [5], [6], [28], the overall control design process under the developed framework does not require the precise knowledge of system dynamics in the state equation. Third, this framework is a data-based generalization of the average dwell-time approach in the model-based switched system theory [20], [35].

B. Structure

The rest of this article is organized as follows. Section II formulates an optimal output regulation problem with desired convergence rate and presents a model-based solution to this problem. In Section III, we

Received 10 August 2024; accepted 5 January 2025. Date of publication 8 January 2025; date of current version 30 May 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62373090, in part by the National Key R&D Program of China under Grant 2024YFA1012700, in part by the Key Research and Development Program of Liaoning Province under Grant 2023JH26/10200011, in part by the Fundamental Research Funds for the Central Universities under Grant N2424004, in part by the Research Program of the Liaoning Liaohe Laboratory under Grant LLL23ZZ-05-01, and in part by the U.S. National Science Foundation under Grant CNS-2227153. Recommended by Associate Editor S. Amin. *(Corresponding author: Tianyou Chai.)*

Weinan Gao and Tianyou Chai are with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China (e-mail: weinan.gao@nyu.edu; tychai@mail.neu.edu.cn).

Zhong-Ping Jiang is with the Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, Brooklyn, NY 11201 USA (e-mail: zjiang@nyu.edu).

Digital Object Identifier 10.1109/TAC.2025.3527305

1558-2523 © 2025 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: New York University. Downloaded on August 07,2025 at 19:57:16 UTC from IEEE Xplore. Restrictions apply.

propose a learning-based resilient optimal controller design framework and an algorithm based on the proposed framework. We have also rigorously analyzed the resiliency of the closed-loop system. Section IV contains simulation results on a benchmark adaptive cruise control problem of an autonomous vehicle under DoS attacks by leveraging the proposed algorithm. Finally, Section V concludes this article.

C. Notations

Throughout this article, $\mathbb{R}$ denotes the set of real numbers, $\mathbb{R}_+$ the set of nonnegative real numbers, and $\mathbb{Z}_+$ the set of nonnegative integers. $\mathbb{C}^-$ indicates the open left-half complex plane. $|\cdot|$ represents the Euclidean norm for vectors and the induced norm for matrices. The symbol $\otimes$ indicates the Kronecker product operator and $\text{vec}(A) = [a_1^T, a_2^T, \ldots, a_m^T]^T$, where $a_i \in \mathbb{R}^n$ are the columns of $A \in \mathbb{R}^{n\times m}$. When $n = m$, $\sigma(A)$ is the spectrum of A. For an arbitrary column vector $v \in \mathbb{R}^n$, $\text{vecv}(v) = [v_1^2, v_1v_2, \ldots, v_1v_n, v_2^2, v_2v_3, \ldots, v_{n-1}v_n, v_n^2]^T \in \mathbb{R}^{\frac{1}{2}n(n+1)}$. $\text{vecs}(P) = [p_{11}, 2p_{12}, \ldots, 2p_{1m}, p_{22}, 2p_{23}, \ldots, 2p_{m-1,m}, p_{mm}]^T \in \mathbb{R}^{\frac{1}{2}m(m+1)}$ for a symmetric matrix $P \in \mathbb{R}^{m\times m}$, and $\lambda_M(P)$ (resp., $\lambda_m(P)$) denotes the maximum (resp., minimum) eigenvalue of a real symmetric matrix P. $P \succ 0$ (resp., $P \prec 0$) represents that P is positive (resp., negative) definite.

II. Problem Formulation and Preliminaries

In this section, we will formulate a linear optimal output regulation problem and recall a model-based solution to this problem. Then, a DoS attack model is established and the concept of resilient optimal control is introduced under DoS attacks.

A. Linear Optimal Output Regulation

Consider a class of continuous-time linear systems described by

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

$$e(t) = Cx(t) + Fv(t) \tag{2}$$

where $u(t) \in \mathbb{R}$ and $x(t) \in \mathbb{R}^n$ are the control input and the state, respectively. The tracking error $e(t) \in \mathbb{R}$ is the difference between the output $Cx(t)$ and the reference $-Fv(t)$. The signal $v(t) \in \mathbb{R}^q$ is the exostate of an autonomous system usually referred to as exosystem

$$\dot{v}(t) = Sv(t). \tag{3}$$

The constant matrices A, B, C, D, F, and S are with proper dimensions. Throughout this article, the following assumptions are made on the overall system (1)–(3).

Assumption 1: The pair (A, B) is controllable.

Assumption 2: $\text{Rank}\begin{bmatrix} A-\lambda I & B \\ C & 0 \end{bmatrix} = n+1\ \forall \lambda \in \sigma(S)$.

The output regulation is achieved if the system (1)–(3) in closed-loop with a controller is exponentially stable at the origin when $v \equiv 0$, and the tracking error $e(t)$ asymptotically converges to 0 for all initial conditions $x(0) \in \mathbb{R}^n$ and $v(0) \in \mathbb{R}^q$. Assume that there is no DoS attacks, the closed-loop system with the following controller:

$$u(t) = -Kx(t) + Lv(t) \tag{4}$$

has achieved output regulation, where K is a feedback control gain such that $\sigma(A - BK) \in \mathbb{C}^-$, $L = KX + U$ is a feedforward control gain with the matrix $X \in \mathbb{R}^{n\times q}$ and the vector $U \in \mathbb{R}^{1\times q}$ solving the following regulator equations:

$$XS = AX + BU + D \tag{5}$$

$$0 = CX + F. \tag{6}$$

Based on the output regulation theory [10], $Xv(t)$ and $Uv(t)$ are where the state and the input should fall in steady-state if one hopes to solve the linear output regulation problem. For any $t \in \mathbb{R}_+$, $\tilde{x}(t) = x(t) - Xv(t)$ and $\tilde{u}(t) = u(t) - Uv(t)$ are differences between the actual state/input and the corresponding steady-state components. Since $\sigma(A - BK) \in \mathbb{C}^-$, one can show that $\lim_{t\to\infty} \tilde{x}(t) = 0$, $\lim_{t\to\infty} \tilde{u}(t) = 0$, and $\lim_{t\to\infty} e(t) = 0$, which implies that the systems (1)–(4) achieve disturbance rejection and asymptotic tracking.

Remark 1: Under the conditions in the Assumption 2, there always exists uniquely a pair (X, U) solving the regulator equations (5) and (6); see [10].

To take into account the transient performance and the convergence rate of the closed-loop system, an optimal output regulation problem is stated as follows.

Problem 1: The optimal output regulation problem is solved if one can find a feedback control policy u^* to achieve output regulation. Moreover, the following optimization problem:

$$\min_{\tilde{u}} \int_0^\infty e^{2t\lambda^-}(\tilde{x}^T(t)Q\tilde{x}(t) + \tilde{u}^2(t))dt \tag{7}$$

$$\text{s.t.}\quad \dot{\tilde{x}}(t) = A\tilde{x}(t) + B\tilde{u}(t) \tag{8}$$

is solved by $\tilde{u}^* = u^* - Uv$, where $Q \succ 0$ and $\lambda^- \in \mathbb{R}_+$.

By denoting $\tilde{x}_{\lambda^-}(t) = e^{t\lambda^-}\tilde{x}(t)$ and $\tilde{u}_{\lambda^-}(t) = e^{t\lambda^-}\tilde{u}(t)$, we have

$$\begin{aligned}\dot{\tilde{x}}_{\lambda^-}(t) &= e^{t\lambda^-}\dot{\tilde{x}}(t) + \lambda^-\tilde{x}_{\lambda^-}(t)\\ &= e^{t\lambda^-}(A\tilde{x}(t) + B\tilde{u}(t)) + \lambda^-\tilde{x}_{\lambda^-}(t)\\ &= (A + I\lambda^-)\tilde{x}_{\lambda^-}(t) + B\tilde{u}_{\lambda^-}(t).\end{aligned} \tag{9}$$

By (9) and the fact that the cost in (7) is equivalent to

$$\int_0^\infty (\tilde{x}_{\lambda^-}^T(t)Q\tilde{x}_{\lambda^-}(t) + \tilde{u}_{\lambda^-}^2(t))dt$$

one can observe that Problem 1 is convertible to a linear quadratic regulator problem. The solution to this problem is an optimal feedback control policy

$$\tilde{u}_{\lambda^-}^*(t) = -K^*\tilde{x}_{\lambda^-}(t) \tag{10}$$

which is equivalent to

$$\begin{aligned}u^*(t) &= e^{-t\lambda^-}\tilde{u}_{\lambda^-}^*(t) + Uv(t)\\ &= -K^*e^{-t\lambda^-}\tilde{x}_{\lambda^-}(t) + Uv(t)\\ &= -K^*\tilde{x}(t) + Uv(t)\\ &= -K^*x(t) + L^*v(t)\end{aligned} \tag{11}$$

where the optimal feedforward control gain L^* is

$$L^* = K^*X + U. \tag{12}$$

The optimal control gain K^* is

$$K^* = B^TP^*. \tag{13}$$

Based on the Popov–Belevitch–Hautus test, the Assumption 1, and $Q \succ 0$, one can obtain that the pair $(A + I\lambda^-, B)$ is controllable and the pair $(A + I\lambda^-, \sqrt{Q})$ is observable for any $\lambda^- \in \mathbb{R}_+$. From the linear optimal control theory, the real symmetric matrix $P^* \succ 0$ is the unique solution to the following algebraic Riccati equation (ARE):

$$(A + I\lambda^-)^TP^* + P^*(A + I\lambda^-) + Q - P^*BB^TP^* = 0. \tag{14}$$

Authorized licensed use limited to: New York University. Downloaded on August 07,2025 at 19:57:16 UTC from IEEE Xplore. Restrictions apply.

B. DoS Attack Model and Resilient Optimal Control

The DoS attack model considered in this article can be described as follows. Let $\mathcal{I}_s = [h_s, h_s + \tau_s)$ represent the sth ($s \in \mathbb{Z}_+$) DoS attack interval. h_s, $h_s + \tau_s$, and τ_s represent the start time, end time, and the length of the sth DoS attack. Based on the following Assumptions 3 and 4 that are somehow introduced in [4], [5], [6], [28], and [32], we define $\Pi_D(t_a, t_b) := (t_a, t_b) \bigcap \bigcup_{s=0}^{\infty} \mathcal{I}_s$ as the set where the system is under DoS attacks during the interval $[t_a, t_b]$, and denote $\Pi_N(t_a, t_b) := [t_a, t_b] \setminus \Pi_D(t_a, t_b)$ as the normal communication set.

Assumption 3 (DoS frequency): There exist constants $\eta > 1$ and $\tau_D > 0$ such that

$$n(t_a, t_b) \leq \eta + \frac{t_b - t_a}{\tau_D} \ \forall t_b > t_a \geq 0 \tag{15}$$

where $n(t_a, t_b)$ denotes the number of DoS OFF/ON transitions during the interval $[t_a, t_b]$.

Assumption 4 (DoS duration): There exist constants $T > 1$ and $\kappa > 0$ such that

$$|\Pi_D(t_a, t_b)| \leq \kappa + \frac{t_b - t_a}{T} \ \forall t_b > t_a \geq 0$$

where $|\Pi_D(t_a, t_b)|$ is the Lebesgue measure [26] of the set $\Pi_D(t_a, t_b)$.

Similar to robust optimal with respect to systems under dynamic uncertainties [13], we define the resilient optimality of a control policy with respect to systems under DoS attacks.

Definition 1: A control policy u_{ro}^* is said to be resilient optimal regarding to systems (1)–(3) if both following properties hold.

1) The optimal output regulation problem, Problem 1, is solved by u_{ro}^* in the absence of DoS attacks, i.e., $|\Pi_D(0, t_b)| = 0$ for any $t_b > 0$.
2) The closed-loop systems (1)–(3) with u_{ro}^* achieve output regulation in the presence of DoS attacks satisfying Assumptions 3 and 4.

III. Main Results

In this section, we will establish a learning-based resilient optimal control framework for control systems subject to DoS attacks. Based on this framework, we develop a learning algorithm to find a resilient optimal control policy given any upper bound of DoS attack duration. We further show that the closed-loop system achieves output regulation.

A. Learning-Based Controller Design Under DoS

A system with DoS attacks is essentially an autonomously switched system where we have no direct control over the switching mechanism [20]. We will switch between two linear subsystems, one is controllable (when the communication is normal), while the other one is an autonomous system but not necessarily stable (when the communication is denied). As shown in [35], a sufficient condition to ensure the exponential stability of this kind of switched systems is that the total activation time ratio between an exponentially stable linear subsystem and an unstable linear subsystem is smaller than a criterion related to the eigenvalues of each subsystem and that the average dwell time is sufficiently large.

Based on the aforementioned fact, we develop a learning-based resilient optimal control framework; see Fig. 1. The outline of this framework can be understood as follows. Through state and input data, we first learn a $\lambda^+ \in \mathbb{R}_+$ such that $\sigma(A - I\lambda^+) \in \mathbb{C}^-$. As the system is open loop when it is under DoS attacks, λ^+ can be used to quantify the divergence rate of the system under attack. Then, we determine the convergence rate parameter λ^- for the system with normal communication based on DoS duration criterion T, and choose the cost (7) based on λ^-. Finally, we develop a resilient optimal controller based on actor-critic structure and policy iteration to satisfy this sufficient condition without the exact knowledge of system matrices A, B, and D.

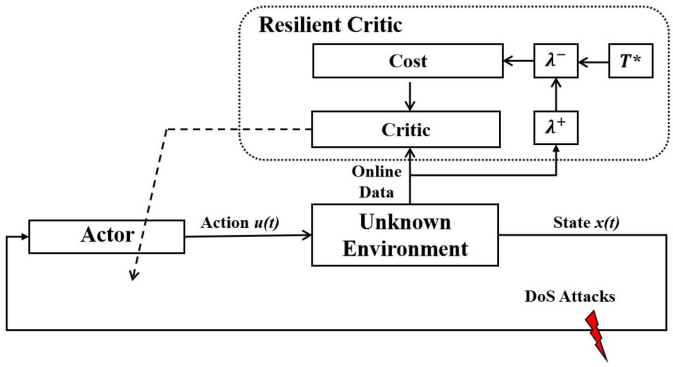

Fig. 1. Learning-based resilient optimal control framework.

To implement this framework, we begin with solving the following Lyapunov equation:

$$(A - I\lambda^+)^T P^+ + P^+(A - I\lambda^+) = -\epsilon I \tag{16}$$

using input and state data, where ϵ is a small but positive constant. Notice that one can always solve the (unique) real symmetric matrix $P^+ \succ 0$ from (16) once $\sigma(A - I\lambda^+) \in \mathbb{C}^-$. Since the state matrix A is unknown, one cannot solve (16) directly. We propose a way to calculate P^+ through input and state data. From (1), for any symmetric matrix $P \in \mathbb{R}^{n\times n}$, we have

$$\begin{aligned} &x^T(t_1)Px(t_1) - x^T(t_0)Px(t_0) \\ &= \int_{t_0}^{t_1} [x^T(A^TP + PA)x + 2x^TPBu + 2x^TPDv]d\tau. \end{aligned} \tag{17}$$

Suppose there exists a sequence $\{t_i\}_{i=0}^{\infty}$ such that the communications are allowed in all of the following intervals $[t_0, t_1], [t_2, t_3], [t_4, t_5], \ldots$. It is feasible to find this sequence through detecting attacks; see, e.g., [36]. For any two vectors a and b and a sufficiently large number $s > 0$, define

$$\begin{aligned} \delta_a = &[\text{vecv}(a(t_1)) - \text{vecv}(a(t_0)), \text{vecv}(a(t_3)) - \text{vecv}(a(t_2)), \\ &\ldots, \text{vecv}(a(t_{2s+1})) - \text{vecv}(a(t_{2s}))]^T \\ \Gamma_{a,b} = &\left[\int_{t_0}^{t_1} a \otimes b d\tau, \int_{t_2}^{t_3} a \otimes b d\tau, \ldots, \int_{t_{2s}}^{t_{2s+1}} a \otimes b d\tau\right]^T \\ \Lambda_a = &\left[\int_{t_0}^{t_1} \text{vecv}(a)d\tau, \int_{t_2}^{t_3} \text{vecv}(a)d\tau, \ldots, \int_{t_{2s}}^{t_{2s+1}} \text{vecv}(a)d\tau\right]^T. \end{aligned}$$

Equations (16) and (17) imply the following linear equation:

$$\Psi \begin{bmatrix} \text{vecs}(\mathrm{P}^+) \\ \text{vec}(B^TP^+) \\ \text{vec}(D^TP^+) \end{bmatrix} = \Phi \tag{18}$$

where $\Psi = [\delta_x - 2\lambda^+\Lambda_x, -2\Gamma_{x,u}, -2\Gamma_{x,v}], \Phi = -\epsilon\Gamma_{x,x}$.

Then, we choose the desired convergence rate parameter λ^- of the closed-loop system with normal communication and obtain the quadruple (K^*, P^*, X, U) based on online data. In order to find the pair (X, U), which is a solution to the regulator equations (5) and (6), we define a sequence of trial matrices $X_i \in \mathbb{R}^{n\times q}$ for $i = 0, 1, \ldots, (n - r)q + 1$, where $X_0 = 0, CX_1 = -F$. For $i = 2, 3, \ldots, (n - r)q + 1$, all the vectors $\text{vec}(X_i)$ form a basis for the kernel $\ker(I_q \otimes C)$. It has

Authorized licensed use limited to: New York University. Downloaded on August 07,2025 at 19:57:16 UTC from IEEE Xplore. Restrictions apply.

been shown in [7] that the solution X is always a linear combination of these basis matrices.

The idea of policy iteration is to implement both policy evaluation

$$0 = A_k^T P_k + P_k A_k + Q + K_k^T K_k + 2\lambda^- P_k \tag{19}$$

and policy improvement

$$K_{k+1} = B^T P_k \tag{20}$$

using online data by iterations, where $A_k = A - BK_k$.

Defining $x_i(t) = x(t) - X_i v(t)$, along the trajectories of (1)–(3), we have

$$\begin{aligned}\dot{x}_i &= \dot{x} - X_i S v \\ &= Ax_i + Bu + (D - \mathcal{S}(X_i))v \\ &= A_k x_i + B(u + K_k x_i) + (D - \mathcal{S}(X_i))v \end{aligned} \tag{21}$$

where $\mathcal{S}(X_i) = X_i S - AX_i$.

By (17), (19), and (20), we have

$$\begin{aligned} & x_i^T(t_1)P_k x_i(t_1) - x_i^T(t_0)P_k x_i(t_0) \\ = & \int_{t_0}^{t_1} \big[x_i^T(A_k^T P_k + P_k A_k)x_i + 2x_i^T P_k B(u + K_k x_i) \\ & + 2x_i^T P_k (D - \mathcal{S}(X_i))v\big] d\tau \\ = & \int_{t_0}^{t_1} \big[-x_i^T(Q + K_k^T R K_k + 2\lambda^- P_k)x_i \\ & + 2x_i^T K_{k+1}^T (u + K_k x_i) + 2x_i^T P_k (D - \mathcal{S}(X_i))v\big] d\tau. \end{aligned} \tag{22}$$

The aforementioned equation implies the following linear equation:

$$\Psi_{ik} \begin{bmatrix} \operatorname{vecs}(\mathrm{P_k}) \\ \operatorname{vec}(K_{k+1}) \\ \operatorname{vec}\left((D - \mathcal{S}(X_i))^T P_k\right) \end{bmatrix} = \Phi_{ik} \tag{23}$$

where

$$\begin{aligned} \Psi_{ik} &= [\delta_{x_i} + 2\lambda^- \Lambda_{x_i}, -2\Gamma_{x_i,x_i}(I \otimes K_k^T) - 2\Gamma_{x_i,u}, -2\Gamma_{x_i,v}] \\ \Phi_{ik} &= -\Gamma_{x_i,x_i} \operatorname{vec}(Q + K_k^T K_k). \end{aligned}$$

A sufficient condition to ensure (18) and (23) having a unique solution is that matrices Ψ and $\Psi_{i,k}$ are full-column rank. It is obvious that matrix $\Psi_{i,k}$ depends on the control gain K_k. In the following lemma, we will show that the uniqueness of solution to (18) and (23) can be guaranteed based on some rank condition that purely depends on the online collection.

Lemma 1: For $i = 0, 1, 2, \ldots, (n-r)q + 1$, if there exists a $s^* \in \mathbb{Z}_+$ such that for all $s > s^*$

$$\operatorname{rank}([\Gamma_{x_i,x_i}, \Gamma_{x_i,u}, \Gamma_{x_i,v}]) = \frac{(n+1)n}{2} + (1+q)n \tag{24}$$

then, the following properties hold:

1) Ψ has full-column rank for any λ^+ such that $\sigma(A - I\lambda^+) \in \mathbb{C}^-$;
2) Ψ_{ik} has column full rank for any $k \in \mathbb{Z}_+$.

Proof: To show the property 1), let us assume there exists a vector Ξ satisfying the following equation:

$$\Psi \Xi = \Phi \tag{25}$$

where $\Xi = [(\operatorname{vecs}(W))^T, (\operatorname{vec}(Y))^T, (\operatorname{vec}(Z))^T]^T$ with matrices $W = W^T \in \mathbb{R}^{n\times n}$, $Y \in \mathbb{R}^{m\times n}$, and $Z \in \mathbb{R}^{q\times n}$. Equation (25) can be converted as follows based on (17):

$$\Lambda_x \operatorname{vecs}(\Omega_1) + 2\Gamma_{x,u} \operatorname{vec}(\Omega_2) + 2\Gamma_{x,v} \operatorname{vec}(\Omega_3) = 0 \tag{26}$$

where

$$\begin{aligned} \Omega_1 &= (A - I\lambda^+)^T W + W(A - I\lambda^+) + \epsilon I \\ \Omega_2 &= B^T W - Y \\ \Omega_3 &= D^T W - Z. \end{aligned}$$

When $i = 0$, the rank condition (24) implies that the matrix $[\Lambda_x, \Gamma_{x,u}, \Gamma_{x,v}]$ is full-column rank, whereby one can show that $Y = B^T W$, $Z = D^T W$, and W solves the following Lyapunov equation:

$$(A - I\lambda^+)^T W + W(A - I\lambda^+) = -\epsilon I. \tag{27}$$

Since $\sigma(A - I\lambda^+) \in \mathbb{C}^-$, the matrix W is uniquely determined. Thus, the vector Ξ is the unique solution to (25) and Ψ is full-column rank. The proof of property 1) is thus completed. To show property 2), one can refer to [7] for mathematical proof.

Based on Lemma 1, we further show a necessary and sufficient condition related to the stability of matrix $A - I\lambda^+$ in the following lemma so that its stability can be determined without the precise knowledge of system matrices in the Algorithm 1, an off-policy learning algorithm to be developed later.

Lemma 2: Under the rank condition (24), the matrix $A - I\lambda^+$ is Hurwitz if and only if the following properties hold:

1) the matrix Ψ is full-column rank;
2) the symmetric matrix P^+ solved by (18) is positive definite.

Proof: To prove the sufficiency, suppose $A - I\lambda^+$ is Hurwitz, then the property 1) holds based on Lemma 1 and rank condition (24). We further observe that P^+ can be solved uniquely by (18) when Ψ is full-column rank. On the other hand, the matrix $P^+ \succ 0$ solving the Lyapunov equation (16) satisfies (18), which immediately shows the property 2).

To prove the necessity, suppose both properties hold. From (25)–(27), the matrix $P^+ \succ 0$ solves the Lyapunov equation (16). By the Lyapunov stability theory, $A - I\lambda^+$ must be a Hurwitz matrix. The proof is thus completed.

Now, we are ready to present an off-policy resilient policy iteration (PI) Algorithm 1, which essentially works through two phases. In the first phase (Steps 3–7), we can use Lemma 2 to ensure that one can always find a λ^+ such that $A - I\lambda^+$ is Hurwitz via online collection. Based on λ^+ and DoS duration criterion T, we set a convergence rate parameter $\lambda^- = \frac{2\lambda^+}{T-1}$ for the closed-loop system with normal communication, and approximate the resilient optimal control gain K^* and the value P^*, and the solution to regulator equations in the second phase (steps 8–19).

Remark 2: Given $\mathcal{S}(X_i)$ for $i = 0, 1, 2, \ldots, (n-r)q + 1$, one can solve (X, U) directly by tackling a convex optimization problem without relying on the knowledge of system matrices; one can refer to [7, Remark 5] for more details.

Remark 3: A control gain K_0 is said to be admissible if it is stabilizing and makes the cost (7) of the system (8) with the controller $\tilde{u}(t) = -K_0 \tilde{\xi}(t)$ to be finite. Since the system (8) is linear and the cost (7) is quadratic, one can show that any K_0 satisfying $\sigma(A + I\lambda^- - BK_0) \in \mathbb{C}^-$ is admissible. The assumption of such a stabilizing K_0 is common in all PI algorithms. By means of robust control techniques, K_0 can be obtained when we have a priori knowledge on the bounds of system matrices. Otherwise, value iteration or hybrid iteration [24] methods should be applied to learn suboptimal controllers from data.

Authorized licensed use limited to: New York University. Downloaded on August 07,2025 at 19:57:16 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: Off-Policy Resilient Policy Iteration Algorithm.

1: Select a small $c > 0$, and two constants $\lambda^+ > 0, a > 1$. Apply any locally essentially bounded input u on $[t_0, t_{2s+1}]$ such that (24) holds.
2: $i \leftarrow 0$, $P^+ \leftarrow 0$.
3: **repeat**
4: $\lambda^+ \leftarrow a\lambda^+$
5: **if** Ψ is full-column rank **then**
6: Solve P^+ from (18) **end if**
7: **until** $P^+ \succ 0$
8: $k \leftarrow -1$, $P_k \leftarrow 0$, $\lambda^- \leftarrow \frac{2\lambda^+}{T-1}$. Choose an admissible K_0.
9: **repeat**
10: $k \leftarrow k+1$
11: Solve (P_k, K_{k+1}) from (23)
12: **until** $|P_k - P_{k-1}| < c$ **and** $k > 0$
13: $k^* \leftarrow k$
14: **repeat**
15: Solve $\mathcal{S}(X_i)$ from (23)
16: $i \leftarrow i+1$
17: **until** $i = (n-r)q+1$
18: Solve (X, U) from $\mathcal{S}(X_i)$. $L_{k^*} = K_{k^*}X + U$
19: Return the resilient near-optimal control policy

$$u_{k^*}(t) = \begin{cases} -K_{k^*}x(t) + L_{k^*}v(t), & t \notin \mathcal{I}_s \\ Uv(t), & t \in \mathcal{I}_s \end{cases} \tag{28}$$

We will develop the following theorem to show that the system in closed-loop with

$$u_k(t) = \begin{cases} -K_k x(t) + L_k v(t), & t \notin \mathcal{I}_s \\ Uv(t), & t \in \mathcal{I}_s \end{cases} \tag{29}$$

is always resilient to DoS attacks if DoS frequency criterion τ_D is sufficiently large, in other words, if DoS frequency is small enough, where the feedforward control gain is $L_k = K_k X + U$. The feedback control gain K_k is obtained at the iteration $k > 0$ and the pair (X, U) is obtained at the step 18 of the Algorithm 1. Without loss of generality, the resilient near-optimal control policy (28) can be rewritten in the form of (29) by letting $k = k^*$.

Theorem 1: Under Assumptions 1–4, given any DoS duration criterion T and any $k \in \mathbb{Z}_+$, the system (1)–(3) in closed-loop with the learned controller (29) achieves output regulation if the DoS frequency criterion τ_D satisfies

$$\tau_D \geq \log\left(\frac{\lambda_M(P_k)\lambda_M(P^+)}{\lambda_m(P_k)\lambda_m(P^+)}\right)\frac{T}{(T-1)\lambda^-} := \tau_D^k. \tag{30}$$

Proof: The system (1)–(3) in closed-loop with the controller (28) can be rewritten by

$$\dot{\tilde{x}}(t) = \begin{cases} A_k\tilde{x}(t), & t \notin \mathcal{I}_s \\ A\tilde{x}(t), & t \in \mathcal{I}_s. \end{cases} \tag{31}$$

Define the following piecewise-quadratic Lyapunov candidate:

$$V(t) = \begin{cases} \tilde{x}^T(t)P_k\tilde{x}(t) := V_1, & t \notin \mathcal{I}_s \\ \mu_1\tilde{x}^T(t)P^+\tilde{x}(t) := V_0, & t \in \mathcal{I}_s \end{cases} \tag{32}$$

where $\mu_1 > 0$ is to be determined.

When the rank condition (24) holds, it can be observed from the Lemma 1 that the matrix $P_k = P_k^T \succ 0$ can be uniquely solved from (23) while satisfying (19). When the communication is normal, along the trajectories of the closed-loop system (31), we have

$$\begin{aligned} \dot{V}_1 &= \tilde{x}^T(A_k^T P_k + P_k A_k)\tilde{x} \\ &\leq -\tilde{x}^T(Q + K_k^T K_k)\tilde{x} - 2\lambda^-\tilde{x}^T P_k\tilde{x} \\ &\leq -(2\lambda^- + c_1)V_1, t \notin \mathcal{I}_s \end{aligned} \tag{33}$$

where $c_1 = \lambda_m(Q)/\lambda_M(P_k)$.

On the other hand, when the system is under DoS attacks, based on (16), we have

$$\dot{V}_0 = \mu_1\tilde{x}^T(A^T P^+ + P^+ A)\tilde{x} \leq 2\lambda^+ V_0, t \in \mathcal{I}_s. \tag{34}$$

For any $s \in \mathbb{Z}_+$, consider two intervals: $[h_s + \tau_s, h_{s+1})$ for the situation where the communications are normal and $[h_s, h_s + \tau_s)$ for the situation where the communication is denied. Then, one can observe that the Lyapunov function satisfies

$$V(t) \leq \begin{cases} e^{-(2\lambda^- + c_1)(t-h_s-\tau_s)}V(h_s+\tau_s), & t \in [h_s+\tau_s, h_{s+1}) \\ e^{2\lambda^+(t-h_s)}V(h_s), & t \in [h_s, h_s+\tau_s). \end{cases}$$

Based on the definition of Lyapunov functions V_0 and V_1, we have

$$\begin{aligned} &\mu_1\lambda_m(P^+)|\tilde{x}|^2 \leq V_0(\tilde{x}) \leq \mu_1\lambda_M(P^+)|\tilde{x}|^2 \\ &\lambda_m(P_k)|\tilde{x}|^2 \leq V_1(\tilde{x}) \leq \lambda_M(P_k)|\tilde{x}|^2 \quad \forall \tilde{x} \in \mathbb{R}^n. \end{aligned}$$

This implies that

$$\begin{aligned} &V_1(\tilde{x}) \leq \mu_2 V_0(\tilde{x}) \\ &V_0(\tilde{x}) \leq \mu_2 V_1(\tilde{x}) \quad \forall \tilde{x} \in \mathbb{R}^n \end{aligned}$$

where $\mu_1 = \sqrt{\frac{\lambda_M(P_k)\lambda_m(P_k)}{\lambda_M(P^+)\lambda_m(P^+)}}$ and $\mu_2 = \sqrt{\frac{\lambda_M(P_k)\lambda_M(P^+)}{\lambda_m(P_k)\lambda_m(P^+)}}$.

Therefore, $V(t) \leq \mu_2 V(t^-)$ holds on all the switching point t, where $t^- = \lim_{\tau \uparrow t}\tau$. For all $t \in \mathbb{R}_+$, the Lyapunov function V satisfies

$$\begin{aligned} V(t) &\leq \mu_2^{n(0,t)} e^{-(2\lambda^- + c_1)|\Pi_N(0,t)|} e^{2\lambda^+|\Pi_D(0,t)|}V(0) \\ &= e^{-2(\lambda^- + c_1)|\Pi_N(0,t)| + 2\lambda^+|\Pi_D(0,t)| + n(0,t)\log\mu_2}V(0). \end{aligned}$$

Based on the Assumptions 3 and 4, we have

$$\begin{aligned} &|n(0,t)| \leq \eta + \frac{t}{\tau_D} \\ &|\Pi_D(0,t)| \leq \kappa + \frac{t}{T} \\ &|\Pi_N(0,t)| = t - |\Pi_D(0,t)| \geq t - \kappa - \frac{t}{T} \; \forall t \in \mathbb{R}_+. \end{aligned} \tag{35}$$

Based on the fact that $\lambda^- = \frac{2\lambda^+}{T-1}$, one has

$$\begin{aligned} &-2(\lambda^- + c_1)|\Pi_N(0,t)| + 2\lambda^+|\Pi_D(0,t)| + n(0,t)\log\mu_2 \\ \leq\; &-2(\lambda^- + c_1)(t - \kappa - \frac{t}{T}) \\ &+ 2\lambda^+(\kappa + \frac{t}{T}) + (\eta + \frac{t}{\tau_D})\log\mu_2 \\ =\; &c_2 + t\left(-\frac{(\lambda^- + 2c_1)(T-1)}{T} + \frac{\log\mu_2}{\tau_D}\right) \\ \leq\; &c_2 - c_3 t \quad \forall \tau_D \geq \tau_D^k, t \in \mathbb{R}_+ \end{aligned} \tag{36}$$

where $c_2 = 2\kappa(\lambda^+ + \lambda^- + c_1) + \eta\log\mu_2$, and $c_3 = \frac{c_1(T-1)}{T}$.

Finally, along the trajectories of the closed-loop system (31), the Lyapunov function V satisfies

$$V(t) \leq e^{c_2 - c_3 t}V(0) \quad \forall \tau_D \geq \tau_D^k$$

Authorized licensed use limited to: New York University. Downloaded on August 07,2025 at 19:57:16 UTC from IEEE Xplore. Restrictions apply.

TABLE I
SYSTEM PARAMETERS

Parameter	Physical Meaning	Value
$\tau_h[s]$	Time Headway of Autonomous Vehicle	1.5
$T_L[s]$	Time Constant of Autonomous Vehicle	0.45
K_L	System Gain of Autonomous Vehicle	1
$\omega[Hz]$	Frequency of Exostates	$\pi/3$
η	DoS Frequency Regularization	1
$\tau_D[s]$	DoS Frequency Criterion	1
$\kappa[s]$	DoS Duration Regularization	0.5
$T*$	Lower Bound of DoS Duration Criterion	2

implying that

$$|\tilde{x}(t)| \leq c_4 e^{-\frac{c_3}{2}t}|\tilde{x}(0)|$$
$$|e(t)| \leq |C|c_4 e^{-\frac{c_3}{2}t}|\tilde{x}(0)| \quad \forall \tau_D \geq \tau_D^k \tag{37}$$

where

$$c_4 = \sqrt{e^{c_2}\frac{\max\{\lambda_M(P_k), \mu_1\lambda_M(P^+)\}}{\min\{\lambda_m(P_k), \mu_1\lambda_m(P^+)\}}}.$$

We can further have $\lim_{t\to\infty}(x(t) - Xv(t)) = 0$, and $\lim_{t\to\infty} e(t) = 0$, which shows that the output regulation is achieved. The proof is completed. ■

Remark 4: In order to implement the learned resilient optimal control policy derived from the Algorithm 1, we assume the exostate $v(t)$ is always measurable. If $v(t)$ is unmeasurable, one can reconstruct the exostate by the minimal polynomial of exosystem matrix S; see [7] for more details.

Remark 5: From (30) in Theorem 1, we observe that the lower bound of DoS frequency criterion τ_D^k is only related to P_k, P^+, λ^-, and T, all of which are either known or learned by Algorithm 1.

IV. EXAMPLE

In order to validate the proposed resilient control approach, we consider the adaptive cruise control problem of an autonomous vehicle. We aim to keep a desired distance between the autonomous vehicle and a preceding vehicle, i.e., a zero clearance error. According to [18], the system matrices are given as follows:

$$A = \begin{bmatrix} 0 & 1 & -\tau_h \\ 0 & 0 & -1 \\ 0 & 0 & -1/T_L \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ -K_L/T_L \end{bmatrix}$$
$$D = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad S = \begin{bmatrix} 0 & -\omega & 0 \\ \omega & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$
$$C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \quad F = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$

where the state $x \in \mathbb{R}^3$ consists of clearance error, velocity error, and acceleration of the autonomous vehicles, and the input $u \in \mathbb{R}$ is the desired acceleration. Based on the structure of exosystem matrix S, we can see that the exostate $v \in \mathbb{R}^3$ includes constant and sinusoidal signals. Dv is the acceleration of the preceding vehicle, implying that its acceleration profile is a combination of constant and sinusoidal signals; see Fig. 3(a).

The physical meanings and values of vehicle parameters and DoS-related parameters are illustrated in Table I. By the collection of online state and input data during $t \in [0, 2]s$ when an exploration noise

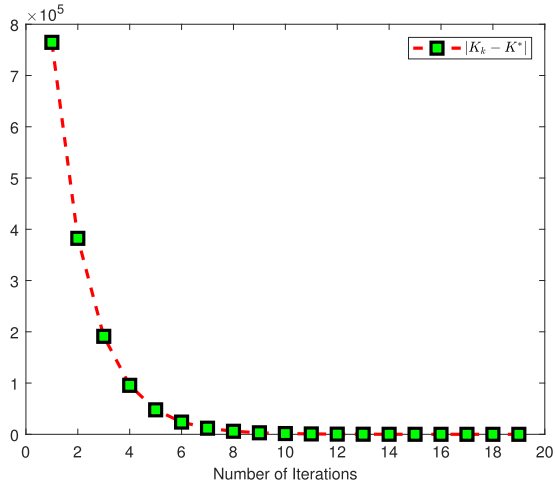

Fig. 2. Comparison between the optimal control gain K^* and the learned control gain K_k at each iteration.

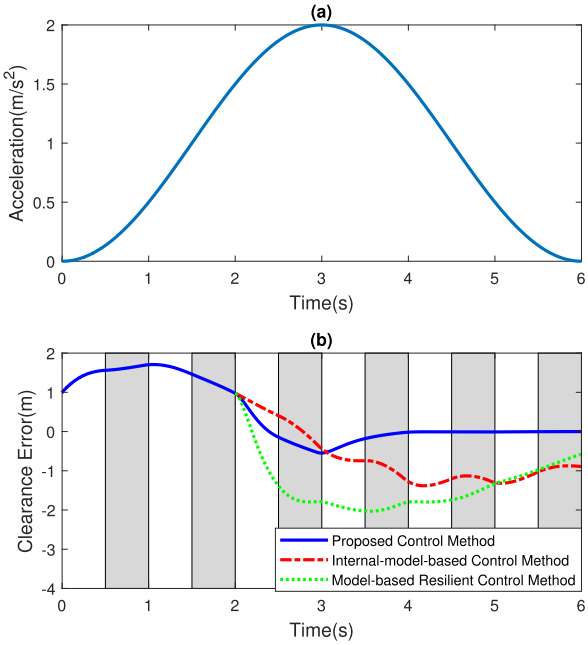

Fig. 3. Simulation results. (a) Acceleration profile of preceding vehicle. (b) Clearance error of autonomous vehicle using the proposed control method, internal-model-based control method, and model-based resilient control method.

$\xi(t) = 0.1\sin(6\pi t) + 0.1\sin(12\pi t)$ is introduced to the system input, we implement the PI Algorithm 1 to learn the resilient optimal control gain and find that the convergence is achieved as the iteration $k = 19$. The approximated feedback control gain, feedforward control gain, and value learned by PI are

$$K = \begin{bmatrix} 19.583688, -9.1462677, -4.8707726 \end{bmatrix}$$
$$P = \begin{bmatrix} 185.66245 & -182.53452 & -8.8126596 \\ -182.53452 & 219.34262 & 4.1158205 \\ -8.8126596 & 4.1158205 & 2.191847 \end{bmatrix}$$
$$L = \begin{bmatrix} -5.8656, -8.7376, -19.5893 \end{bmatrix}.$$

Authorized licensed use limited to: New York University. Downloaded on August 07,2025 at 19:57:16 UTC from IEEE Xplore. Restrictions apply.

For reference, we calculate the resilient optimal control gain, feedforward control gain, and value through solving the ARE (14) and regulator equations (5) and (6) as follows:

$$K^* = \begin{bmatrix} 19.412827, -9.165516, -4.7660936 \end{bmatrix}$$

$$P^* = \begin{bmatrix} 183.42892 & -180.67875 & -8.735772 \\ -180.67875 & 217.68209 & 4.1244822 \\ -8.735772 & 4.1244822 & 2.1447421 \end{bmatrix}$$

$$L^* = \begin{bmatrix} -5.8414, -8.7045, -19.5144 \end{bmatrix}.$$

Fig. 2 depicts the difference between the optimal control gain K^* and the learned control gain K_k at each iteration. We introduce two controllers as a comparison of performance. The first one is an internal-model-based controller [10]

$$\begin{aligned} u(t) &= -K_x x(t) - K_z z(t) \\ \dot{z}(t) &= Sz(t) + G_2 e(t) \end{aligned} \tag{38}$$

where G_2 is selected such that the pair (S, G_2) is controllable and control gains K_x, K_z are chosen such that the matrix $\begin{bmatrix} A - BK_x & -BK_z \\ G_2C & S \end{bmatrix}$ is Hurwitz.

The second one is a variant of model-based resilient controller developed in [6]

$$\begin{cases} u(t) = -K^* \hat{x}(t) \\ \dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t), & t \in \mathcal{I}_s \\ \hat{x}(t) = x(t), & t \notin \mathcal{I}_s. \end{cases} \tag{39}$$

We apply the proposed resilient optimal controller, internal-model-based controller, and model-based controller after $t = 2$ s, respectively. The simulation results are shown in Fig. 3(b) where the shaded area refers to the DoS attack intervals. One can see that the proposed control method outperforms the other two in both transient and steady-state. This is because the internal-model-based controller (38) is not aware of DoS attacks, and the state $x(t)$ and tracking error $e(t)$ are not available as feedback when the system is under DoS attacks, although the internal model principle is well known to address robust output regulation problems. Resilient controller (39) has taken the effect of DoS attacks into consideration. However, it can be only ensured that all the signals of the closed-loop system remain bounded for bounded disturbances, which is not enough to achieve asymptotic tracking.

We further determine the bound $\tau_D^{19} = 29.18$ s based on (30). However, the DoS frequency criterion $\tau_D = 1$ s can be much smaller than the bound in the simulation, which implies that the closed-loop system can defend more frequent attacks than we expect.

V. Conclusion

In this article, a learning-based resilient optimal control framework has been constructed using recent developments in reinforcement learning and ADP. Under this framework, we first estimate the divergence rate of the system under DoS attacks. Based on the estimated rate and an upper bound of DoS duration, we determine a desired convergence rate for the closed-loop system with normal communication. Then, we apply ADP methods to learn a resilient optimal control policy under the desired rate. Finally, through rigorous theoretical analysis and numerical simulations, we have shown that the control system is resilient against cyberattacks with the learned adaptive controller.

References

[1] I. Abraham and T. D. Murphey, "Active learning of dynamics for data-driven control using Koopman operators," *IEEE Trans. Robot.*, vol. 35, no. 5, pp. 1071–1083, Oct. 2019.

[2] S. Amin, A. A. Cárdenas, and S. S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *Proc. Int. Workshop Hybrid Syst., Comput. Control*, 2009, pp. 31–45.

[3] C. Chen, H. Modares, K. Xie, F. L. Lewis, Y. Wan, and S. Xie, "Reinforcement learning-based adaptive optimal exponential tracking control of linear systems with unknown dynamics," *IEEE Trans. Autom. Control*, vol. 64, no. 11, pp. 4423–4438, Nov. 2019.

[4] C. D. Persis and P. Tesi, "Input-to-state stabilizing control under denial-of-service," *IEEE Trans. Autom. Control*, vol. 60, no. 11, pp. 2930–2944, Nov. 2015.

[5] C. Deng, D. Zhang, and G. Feng, "Resilient practical cooperative output regulation for MASs with unknown switching exosystem dynamics under DoS attacks," *Automatica*, vol. 139, 2022, Art. no. 110172.

[6] S. Feng and P. Tesi, "Resilient control under denial-of-service: Robust design," *Automatica*, vol. 79, pp. 42–51, 2017.

[7] W. Gao and Z. P. Jiang, "Adaptive dynamic programming and adaptive optimal output regulation of linear systems," *IEEE Trans. Autom. Control*, vol. 61, no. 12, pp. 4164–4169, Dec. 2016.

[8] W. Gao, C. Deng, Y. Jiang, and Z. P. Jiang, "Resilient reinforcement learning and robust output regulation under denial-of-service attacks," *Automatica*, vol. 142, 2022, Art. no. 110366.

[9] W. Gao and Z. P. Jiang, "Learning-based adaptive optimal output regulation of linear and nonlinear systems: An overview," *Control Theory Technol.*, vol. 20, pp. 1–19, 2022.

[10] J. Huang, *Nonlinear Output Regulation: Theory and Applications*. Philadelphia, PA, USA: SIAM, 2004.

[11] Y. Huang, L. Huang, and Q. Zhu, "Reinforcement learning for feedback-enabled cyber resilience," *Annu. Rev. Control*, vol. 53, pp. 273–295, 2022.

[12] I. Ilahi et al., "Challenges and countermeasures for adversarial attacks on deep reinforcement learning," *IEEE Trans. Artif. Intell.*, vol. 3, no. 2, pp. 90–109, Apr. 2022.

[13] Y. Jiang and Z. P. Jiang, *Robust Adaptive Dynamic Programming*. Hoboken, NJ, USA: Wiley, 2017.

[14] Y. Jiang, B. Kiumarsi, J. Fan, T. Chai, J. Li, and F. L. Lewis, "Optimal output regulation of linear discrete-time systems with unknown dynamics using reinforcement learning," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3147–3156, Jul. 2020.

[15] Y. Jiang, L. Liu, and G. Feng, "Adaptive optimal control of networked nonlinear systems with stochastic sensor and actuator dropouts based on reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 3, pp. 3107–3120, Mar. 2024.

[16] Z. P. Jiang, T. Bian, and W. Gao, "Learning-based control: A tutorial and some recent results," *Found. Trends Syst. Control*, vol. 8, no. 3, pp. 176–284, 2020.

[17] B. Kiumarsi, K. G. Vamvoudakis, H. Modares, and F. L. Lewis, "Optimal and autonomous control using reinforcement learning: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2042–2062, Jun. 2018.

[18] S. Li, K. Li, R. Rajamani, and J. Wang, "Model predictive multi-objective vehicular adaptive cruise control," *IEEE Trans. Control Syst. Technol.*, vol. 19, no. 3, pp. 556–566, May 2011.

[19] Z. Lian, F. Guo, C. Wen, C. Deng, and P. Lin, "Distributed resilient optimal current sharing control for an islanded DC microgrid under DoS attacks," *IEEE Trans. Smart Grid*, vol. 12, no. 5, pp. 4494–4505, Sep. 2021.

[20] D. Liberzon, *Switching in Systems and Control*. Boston, MA, USA: Springer, 2003.

[21] D. Liu, S. Xue, B. Zhao, B. Luo, and Q. Wei, "Adaptive dynamic programming for control: A survey and recent advances," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 1, pp. 142–160, Jan. 2021.

[22] A. Odekunle, W. Gao, M. Davari, and Z. P. Jiang, "Reinforcement learning and non-zero-sum game output regulation for multi-player linear uncertain systems," *Automatica*, vol. 112, 2020, Art. no. 108672.

[23] W. B. Powell, *Approximate Dynamic Programming: Solving the Curse of Dimensionality*. New York, NY, USA: Wiley, 2007.

[24] O. Qasem, W. Gao, and K. G. Vamvoudakis, "Adaptive optimal control of continuous-time nonlinear affine systems via hybrid iteration," *Automatica*, vol. 157, 2023, Art. no. 111261.

[25] M. Rotulo, C. D. Persis, and P. Tesi, "Online learning of data-driven controllers for unknown switched linear systems," *Automatica*, vol. 145, 2022, Art. no. 110519.

Authorized licensed use limited to: New York University. Downloaded on August 07,2025 at 19:57:16 UTC from IEEE Xplore. Restrictions apply.

[26] W. Rudin, *Principles of Mathematical Analysis*. New York, NY, USA: McGraw-Hill, 1953.
[27] Q. Sun, K. Zhang, and Y. Shi, "Resilient model predictive control of cyber-physical systems under DoS attacks," *IEEE Trans. Ind. Inform.*, vol. 16, no. 7, pp. 4920–4927, Jul. 2020.
[28] Y. Tang, D. Zhang, P. Shi, W. Zhang, and F. Qian, "Event-based formation control for nonlinear multiagent systems under DoS attacks," *IEEE Trans. Autom. Control*, vol. 66, no. 1, pp. 452–459, Jan. 2021.
[29] K. G. Vamvoudakis and N.-M. T. Kokolakis, "Synchronous reinforcement learning-based control for cognitive autonomy," *Found. Trends Syst. Control*, vol. 8, no. 1/2, pp. 1–175, 2020.
[30] D. Vrabie, K. G. Vamvoudakis, and F. L. Lewis, *Optimal Adaptive Control and Differential Games by Reinforcement Learning Principles*. London, U.K.: IET, 2013.
[31] F. Y. Wang, H. Zhang, and D. Liu, "Adaptive dynamic programming: An introduction," *IEEE Comput. Intell. Mag.*, vol. 4, no. 2, pp. 39–47, May 2009.
[32] Y.-W. Wang, Z.-H. Zeng, X.-K. Liu, and Z.-W. Liu, "Input-to-state stability of switched linear systems with unstabilizable modes under DoS attacks," *Automatica*, vol. 146, 2022, Art. no. 110607.
[33] Y. Yang, K. G. Vamvoudakis, H. Modares, Y. Yin, and D. C. Wunsch, "Hamiltonian-driven hybrid adaptive dynamic programming," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 10, pp. 6423–6434, Oct. 2021.
[34] W. Yu, J. Tan, C. K. Liu, and G. Turk, "Preparing for the unknown: Learning a universal policy with online system identification," 2017, *arXiv:1702.02453*.
[35] G. Zhai, B. Hu, K. Yasuda, and A. N. Michel, "Stability analysis of switched systems with stable and unstable subsystems: An average dwell time approach," *Int. J. Syst. Sci.*, vol. 32, no. 8, pp. 1055–1061, 2001.
[36] 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.*, vol. 51, no. 9, pp. 4648–4660, Sep. 2021.

Authorized licensed use limited to: New York University. Downloaded on August 07,2025 at 19:57:16 UTC from IEEE Xplore. Restrictions apply.