Reinforcement Learning-Based Tracking Control for Networked Control Systems With DoS Attacks

Jinliang Liu, Yanhui Dong, Lijuan Zha, Xiangpeng Xie, *Senior Member, IEEE*, and Engang Tian

***Abstract*— This paper is concerned with the reinforcement learning-based tracking control problem for a class of networked systems subject to denial-of-service (DoS) attacks. Taking the effects of DoS attacks into consideration, a novel value function is proposed, which considers the cost of the control input, external disturbance and tracking error. Then, using the structure of the value function, the tracking Bellman equation and Hamilton function are defined. By employing the Bellman optimality theory, the optimal control strategy and the game algebraic Riccati equation (GARE) are solved with the Hamilton function. Next, the desired tracking performance is guaranteed as the solution of the GARE is found. Furthermore, an attacks-based Q-learning algorithm is projected to find the solution to the optimal tracking problem without the system dynamics and the convergence of the Q-learning algorithm is given. Finally, the F-404 aircraft engine system is given to verify the effectiveness of the proposed control strategy.**

***Index Terms*— Discrete-time systems, cyber security, denial-of-service (DoS) attacks, tracking control.**

I. Introduction

OWING to the flexibility and reliability, networked control systems (NCSs) have been widely applied in sensor networks, smart transportation, industrial control, automatic driving, water conservancy facilities and other fields [1], [2]. Increasing research enthusiasm towards NCSs have been witnessed in the past few decades [3], [4], [5]. In NCSs, communication network facilitates the data exchange between different physical components. However, due to the intrinsic characteristic of wireless network, a series of network-induced phenomenon emerges, such as network security and constrained communication capabilities [6], [7], [8], [9], which may affect the stability of NCSs. Hence, it is important to design an efficacious security control scheme for ensuring the stability of NCSs.

Because of the inherent fragility of communication network, NCSs are facing many security threats from attackers, the negative impacts of which may do destructiveness to NCSs and can not be neglected. Fortunately, at present, numerous secure control algorithms have been proposed and some important results have been reported [10], [11], [12]. These existing works are mainly focused on denial-of-service (DoS) attacks and deception attacks [13]. DoS attacks try to block the interconnection of communication modules, which will cause data transmission failure or even let the system instability [14], [15]. Deception attacks confuse the systems by injecting false data into the transmission channel [16]. For NCSs, the real time and accuracy of data are crucial to the performance of state estimation and control. Therefore, many researchers have begun to pay attention to the security of NCSs. For example, the authors in [17] discussed the control problem for stochastic NCSs under aperiodic DoS jamming attacks. In [18], a resilient event-driven fuzzy controller was put forward for nonlinear NCSs against non-periodic DoS attacks. Moreover, Zhao et al. [19] provided an active security control approach for switched systems subject to asynchronous DoS attacks. In [20], the event-based security control issue was addressed for discrete-time NCSs under the influences of deception attacks and DoS attacks. Different from the above results, this paper applies a game-theoretic approach to design a secure tracking control scheme, which is capable of simultaneously resisting the influence of DoS attacks and external disturbance [21], [22]. However, as far as we know, limited works can be found on the topic of secure tracking control problem using reinforcement learning (RL) technique. In addition, NCSs are indeed impacted by unreliable communication and limited perception range in practical environments. It is virtually unattainable to acquire the exact parameters of NCSs. Therefore, we are committed to devising a security control strategy without the precise system dynamics, which is the impetus for our research.

As a potent machine learning technique, RL provides a valid approach to address the optimization control problem for complex NCSs [23], [24], [25], [26]. At the present research phase, the RL technique has been adopted to tackle the challenges in intelligent control scenarios characterized by unknown system dynamics across various control domains,

Manuscript received 9 October 2023; revised 30 January 2024 and 26 February 2024; accepted 4 March 2024. Date of publication 11 March 2024; date of current version 6 May 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62373252, Grant 62273174, and Grant 61903182; in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20230063 and Grant BK20211290; and in part by the Qing Lan Project. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Valeria Loscri. *(Corresponding author: Lijuan Zha.)*

Jinliang Liu is with the School of Computer Science, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: liujinliang@vip.163.com).

Yanhui Dong is with the College of Information Engineering, Nanjing University of Finance and Economics, Nanjing 210023, China (e-mail: dongyanhui98@outlook.com).

Lijuan Zha is with the School of Science, Nanjing Forestry University, Nanjing 210037, China (e-mail: zhalijuan@vip.163.com).

Xiangpeng Xie is with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China (e-mail: xiexiangpeng1953@163.com).

Engang Tian is with the School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China (e-mail: tianengang@163.com).

Digital Object Identifier 10.1109/TIFS.2024.3376250

such as design robotic motion control [27], threshold attack detection scheme [28], energy-efficient control [29] and tracking control of multiagent systems [30]. For instance, Xu et al. investigated the security tracking control of heterogeneous multi-agent systems under distributed DoS attacks by using data-driven collaborative learning technique [31]. In [32], the point-to-point consensus tracking control problem was studied for multi-agent systems with completely unknown dynamics. The authors in [33] developed a parallel tracking control optimization algorithm for fuzzy interconnected systems. In [34], a model-free optimal control policy was proposed for multi-rate sampled system via Q-Learning (QL) framework. However, few results about optimal security control problem for NCSs have been published considering exogenous interference and DoS attacks simultaneously by RL technique, which is still a challenging issue.

This work presents an optimal tracking control mechanism for discrete-time NCSs (DTNCSs) with complex external disturbances and DoS attacks using RL method. The primary contributions of the paper can be summarized below:

(1) Considering the complex external disturbances and DoS attacks, an augmented system model is established, a new Bellman equation and value function are proposed based on the augmented model.
(2) Different from some previous research results, the influence of DoS attacks is considered in the GARE. The existence and uniqueness of the solution to the GARE are obtained while achieving the expected tracking performance.
(3) An attacks-based QL algorithm is proposed to derive the target controller gain and disturbance gain, which is without the system dynamics. In addition, the convergence of the algorithm is discussed.

II. Problem Formulation

A. System Descriptions

Consider the DTNCSs as follows:

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + D\omega_k \\ y_k = Cx_k \end{cases} \tag{1}$$

where $x_k \in \mathbb{R}^n$, $y_k \in \mathbb{R}^p$ and $u_k \in \mathbb{R}^q$ are the system state, measurement output, and control input, respectively. $\omega_k \in \mathbb{R}^s$ indicates the external disturbance; A, B, C and D are unknown system matrices with suitable dimensions.

In Fig. 1, the diagram of DTNCSs subject to DoS attacks is depicted. The communication channel between the controller and actuator is connected via an unprotected network, which is vulnerable to potential disruption from DoS attacks. Meanwhile, the opponent can eavesdrop on the transmission signal and launch attacks at any time.

Assumption 1: (A, B) is controllable and (A, C) is observable.

Assumption 2: [35] The transmission signal will be completely lost if the communication channel is under DoS attacks.

The primary aim of this work is to establish an optimal tracking control strategy for DTNCSs (1) against DoS attacks such that a desired reference signal can be tracked by system output.

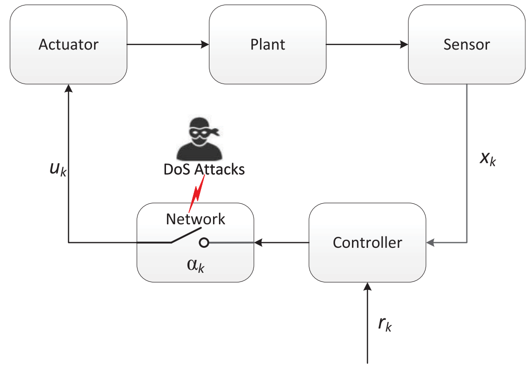


Fig. 1. Diagram of DTNCSs under DoS attacks.

The reference model is inferred as follows:

$$r_{k+1} = Fr_k \tag{2}$$

where $r_k \in \mathbb{R}^p$ is the reference signal.

Remark 1: It is worth noting that F in (2) does not need to be a Hurwitz matrix, which can cover various kinds of reference trajectories. Based on this consideration, the discount factor can be introduced in the value function for the purpose of guaranteeing the boundedness of the value function [36].

When the data transmission channel is normal, the controller is designed as follows:

$$u_k^d = K\bar{x}_k \tag{3}$$

where $\bar{x}_k = \left[x_k^T \; r_k^T\right]^T$, K stands for the controller gain to be devised.

Due to the openness characteristic of communication network, the transmitted signal may be corrupted by DoS attacks, which aims at blocking the communication channels. In this paper, the occurrence of malicious DoS attacks is described by the following stochastic variable:

$$\alpha_k = \begin{cases} 0, & \text{if DoS attacks are active} \\ 1, & \text{otherwise.} \end{cases} \tag{4}$$

In the presence of DoS attacks, the practical control input u_k is expressed as

$$u_k = \alpha_k u_k^d \tag{5}$$

where $\alpha_k \in \{0, 1\}$ obeys the Bernoulli distribution with $Pr\{\alpha_k = 1\} = \bar{\alpha_d}$, $Pr\{\alpha_k = 0\} = 1 - \bar{\alpha_d}$ in which $\bar{\alpha_d} \in (0, 1]$.

From DTNCSs (1), the reference trajectory (2) and control input (5), one can derive the augmented system as

$$\begin{cases} \bar{x}_{k+1} = \bar{A}\bar{x}_k + \bar{B}\alpha_k u_k^d + \bar{D}\omega_k \\ \bar{y}_k = \bar{C}\bar{x}_k \end{cases} \tag{6}$$

where

$$\bar{A} = \begin{bmatrix} A & 0 \\ 0 & F \end{bmatrix}, \bar{B} = \begin{bmatrix} B \\ 0 \end{bmatrix}, \bar{C}^T = \begin{bmatrix} C^T \\ 0 \end{bmatrix}, \bar{D} = \begin{bmatrix} D \\ 0 \end{bmatrix}.$$

Similar to the form of control signal u_k^d, the disturbance ω_k is defined as

$$\omega_k = L\bar{x}_k \tag{7}$$

where L is the disturbance gain, which will be designed in the following.

Remark 2: The real control input shown in (5) is resultant from the influence of the DoS attacks. Specifically, $\alpha_k = 1$ means the actuator successfully receives information from the controller; Otherwise, the malignant DoS attacks occur and the actual control input becomes zero.

Remark 3: In [36], a tracking control policy was presented for DTNCSs with unknown dynamics and network-induced dropout, where the impacts of the DoS attacks are not considered. Since potential destructions of the DoS attacks posed by malicious adversary may results in the control signal inaccessible to the actuator, we make some improvements based on [36]. In this paper, we aim to provide a new tracking control method for DTNCSs subject to DoS attacks and exogenous interference by RL method.

B. Problem Formulation

In this paper, the RL-based secure tracking control strategy is put forward for DTNCSs (1) against DoS attacks. Prior to proceeding, the following design objective can be given:

1. A control scheme will be designed to make the output of the system (1) track the reference model (2), that is, the tracking error $e_k = y_k - r_k$ is bounded under the designed control scheme.

2. The following condition is satisfied:

$$E\left\{\sum_{i=k}^{\infty}\varphi^{i-k}[e_i^T Q e_i + u_i^T R u_i]\right\} \le \gamma^2 E\left\{\sum_{i=k}^{\infty}\varphi^{i-k}\omega_i^T\omega_i\right\} \tag{8}$$

for all $\omega_k \in L_2[0,\infty)$. Besides, $\varphi \in (0,\ 1)$ is the discount factor, $\gamma > 0$ denotes H_∞ tracking performance level and $Q \ge 0$, $R > 0$ are known weighting matrices.

To achieve the goal, value function is defined as

$$V(\bar{x}_k) = E\left\{\sum_{i=k}^{\infty}\varphi^{i-k}J(e_i, u_i^d, \omega_i; \gamma)\right\} \tag{9}$$

where $J(\cdot)$ serves as the utility function with the following form:

$$J(e_i, u_i^d, \omega_i; \gamma) = e_i^T Q e_i + \alpha_i u_i^{d^T} R u_i^d - \gamma^2\omega_i^T\omega_i. \tag{10}$$

According to [37], regarding the control signal u_k^d and disturbance ω_k as two players, the H_∞ tracking control problem in the article can be seen as zero-sum game problem. Then, we are committed to solving a minmax problem under DoS attacks as

$$\begin{aligned} V^*(\bar{x}_k) &= \min_{u^d_{i(i\ge k)}}\max_{\omega_{i(i\ge k)}} V(\bar{x}_k) \\ &= \min_{u^d_{i(i\ge k)}}\max_{\omega_{i(i\ge k)}} E\left\{\sum_{i=k}^{\infty}\varphi^{i-k}J(e_i, u_i^d, \omega_i; \gamma)\right\} \\ &= \min_{u^d_{i(i\ge k)}}\max_{\omega_{i(i\ge k)}} E\Bigg\{\sum_{i=k}^{\infty}\varphi^{i-k}[\bar{x}_i^T\bar{Q}\bar{x}_i \\ &\quad + \alpha_i u_i^{d^T} R u_i^d - \gamma^2\omega_i^T\omega_i]\Bigg\} \end{aligned} \tag{11}$$

where $V^*(\bar{x}_k)$ represents the optimal value function and

$$\bar{Q} = \begin{bmatrix} C^TQC & -C^TQ \\ -QC & Q \end{bmatrix}.$$

Remark 4: The value function in (11) is dependent on the control action player and the disturbance. The former aims to minimize the value function, whereas the latter tries to increase the value function. The control strategy is to find an optimal sequence $(u_k^{d^*}, \omega_k^*)$ which makes the value function minimization.

Using the expression of $\bar{Q}$, the value function in (9) is rewritten as

$$\begin{aligned} V(\bar{x}_k) &= \bar{x}_k^T\bar{Q}\bar{x}_k + \bar{\alpha}_d u_k^{d^T} R u_k^d - \gamma^2\omega_k^T\omega_k \\ &\quad + E\left\{\sum_{i=k+1}^{\infty}\varphi^{i-k}[\bar{x}_i^T\bar{Q}\bar{x}_i + \alpha_i u_i^{d^T} R u_i^d - \gamma^2\omega_i^T\omega_i]\right\} \end{aligned} \tag{12}$$

then the Bellman equation is defined as

$$V(\bar{x}_k) = E\left\{\bar{x}_k^T\bar{Q}\bar{x}_k + \alpha_k u_k^{d^T} R u_k^d - \gamma^2\omega_k^T\omega_k\right\} + \varphi V(\bar{x}_{k+1}). \tag{13}$$

Similar to [38], the value function (12) has the quadratic form

$$V(\bar{x}_k) = E\left\{\bar{x}_k^T P\bar{x}_k\right\} \tag{14}$$

with $P = P^T > 0$ being some matrix.

Combining the formulas (13) and (14), we can get

$$E\left\{\bar{x}_k^T P\bar{x}_k\right\} = E\left\{\bar{x}_k^T\bar{Q}\bar{x}_k + \alpha_k u_k^{d^T} R u_k^d - \gamma^2\omega_k^T\omega_k + \varphi V(\bar{x}_{k+1})\right\}. \tag{15}$$

Remark 5: Note that many existing available tracking control approaches have been provided based on QL algorithm for different systems on the premise of reliable communication channel, which is unrealistic in some cases. In case of the DoS attacks, the above methods are not applicable. In light of this, we try to address the security tracking controller design for DTNCSs via the QL algorithm, which is still challenging at the current research stage.

III. Main Results

In this section, the optimal tracking control strategy will be derived, which is rely on the solution to GARE. Besides, the existence of the solution to GARE is discussed. Subsequently, the boundedness of tracking error is analyzed.

A. Optimal Tracking Controller Design

In what follows, Theorem 1 is utilized to derive the optimal controller gain K and disturbance gain L.

Theorem 1: Considering the value function in (14), the optimal controller gain K as well as worst disturbance gain L are designed as

$$K = (M_{11} - M_{12}(M_{22})^{-1}M_{21})^{-1}$$

$$\times (M_{12}(M_{22})^{-1}\varphi \bar{D}^T P\bar{A} - \varphi \bar{B}^T P\bar{A})$$
$$L = (M_{22} - M_{21}(M_{11})^{-1}M_{12})^{-1}$$
$$\times (M_{21}(M_{11})^{-1}\varphi \bar{B}^T P\bar{A} - \varphi \bar{D}^T P\bar{A}) \quad (16)$$

and optimal kernel matrix P is a solution of the following GARE:

$$P = \varphi \bar{A}^T P\bar{A} + \bar{Q} - \varphi^2 \left[\bar{A}^T P\bar{B} \;\; \bar{A}^T P\bar{D}\right] \times \begin{bmatrix}\beta M_{11} & M_{12}\\ \beta M_{21} & M_{22}\end{bmatrix}^{-1} \begin{bmatrix}\bar{B}^T P\bar{A}\\ \bar{D}^T P\bar{A}\end{bmatrix} \quad (17)$$

where $M_{11} = R + \varphi \bar{B}^T P\bar{B}$, $M_{12} = \varphi \bar{B}^T P\bar{D}$, $M_{21} = \bar{\alpha_d}\varphi \bar{D}^T P\bar{B}$, $M_{22} = \varphi \bar{D}^T P\bar{D} - \gamma^2 I$, $\beta = \frac{1}{\bar{\alpha_d}}$.

Proof: On the basis of Bellman optimality theory, the minmax problem expressed in formula (11) can be turned into dealing with the following Hamilton-Jacobi-Isaacs equation:

$$\min_{u_k^d \in \mathscr{A}_1(\Omega_{1,k})} \max_{\omega_k \in \mathscr{A}_2(\Omega_{2,k})} H(\bar{x}_k, u_k^d, \omega_k) = 0 \quad (18)$$

where $\mathscr{A}_1(\Omega_{1,k})$ and $\mathscr{A}_2(\Omega_{2,k})$ denote the sets of all acceptable control mechanisms and disturbance policies over $\Omega_{1,k}$ and $\Omega_{2,k}$, respectively. Besides, the Hamilton function $H(\bar{x}_k, u_k^d, \omega_k)$ is defined as

$$\begin{aligned} H(\bar{x}_k, u_k^d, \omega_k) &= E\{J(\bar{x}_k, u_k, \omega_k; \gamma)\} + \varphi V(\bar{x}_{k+1}) - V(\bar{x}_k)\\ &= \bar{x}_k^T \bar{Q}\bar{x}_k + \bar{\alpha_d} {u_k^d}^T R u_k^d - \gamma^2 \omega_k^T \omega_k\\ &\quad + E\left\{\varphi \bar{x}_{k+1}^T P\bar{x}_{k+1}\right\} - E\left\{\bar{x}_k^T P\bar{x}_k\right\}. \end{aligned} \quad (19)$$

From the method adopted in [37], according to $\partial H(\bar{x}_k, u_k, \omega_k)/\partial u_k^d = 0$ and $\partial H(\bar{x}_k, u_k, \omega_k)/\partial \omega_k = 0$, it is not difficult to derive

$$\begin{aligned} &(R + \varphi \bar{B}^T P\bar{B})u_k^d + \varphi \bar{B}^T P\bar{A}\bar{x}_k + \varphi \bar{B}^T P\bar{D}\omega_k = 0\\ &\bar{\alpha_d}\varphi \bar{D}^T P\bar{B}u_k^d + \varphi \bar{D}^T P\bar{A}\bar{x}_k + (\varphi \bar{D}^T P\bar{D} - \gamma^2 I)\omega_k = 0. \end{aligned} \quad (20)$$

Besides, $\partial^2 H(\bar{x}_k, u_k, \omega_k)/\partial^2 u_k^d > 0$ and $\partial^2 H(\bar{x}_k, u_k, \omega_k)/\partial^2 \omega_k < 0$ should be hold, that is, the following inequalities are satisfied

$$\begin{aligned} &R + \bar{B}^T P\bar{B} > 0\\ &-\gamma^2 I + \bar{D}^T P\bar{D} < 0 \end{aligned} \quad (21)$$

Then, (20) can be rewritten as

$$\begin{aligned} M_{11}u_k^d + M_{12}\omega_k &= -\varphi \bar{B}^T P\bar{A}\bar{x}_k\\ M_{21}u_k^d + M_{22}\omega_k &= -\varphi \bar{D}^T P\bar{A}\bar{x}_k \end{aligned} \quad (22)$$

Utilizing the condition (22), the following optimal control and worst disturbance policies are acquired:

$$\begin{aligned} u_k^d &= (M_{11} - M_{12}(M_{22})^{-1}M_{21})^{-1}\\ &\quad \times (M_{12}{M_{22}}^{-1}\varphi \bar{D}^T P\bar{A} - \varphi \bar{B}^T P\bar{A})\bar{x}_k\\ \omega_k &= (M_{22} - M_{21}(M_{11})^{-1}M_{12})^{-1}\\ &\quad \times (M_{21}{M_{11}}^{-1}\varphi \bar{B}^T P\bar{A} - \varphi \bar{D}^T P\bar{A})\bar{x}_k. \end{aligned} \quad (23)$$

According to (3) and (7), the gain matrices K and L in (16) are readily derived. By substituting (6) into (15), it can be deduced that

$$\begin{aligned} \bar{x}_k^T P\bar{x}_k &= \bar{x}_k^T \bar{Q}\bar{x}_k + \bar{\alpha_d}{u_k^d}^T R u_k^d - \gamma^2 \omega_k^T \omega_k\\ &\quad + \varphi E\{\bar{x}_{k+1}^T P\bar{x}_{k+1}\}\\ &= \bar{x}_k^T(\bar{Q} + \varphi \bar{A}^T P\bar{A})\bar{x}_k + \bar{\alpha_d}\varphi \bar{x}_k^T \bar{A}^T P\bar{B}u_k^d\\ &\quad + \bar{\alpha_d}{u_k^d}^T(M_{11}u_k^d + M_{12}\omega_k + \varphi \bar{B}^T P\bar{A}\bar{x}_k)\\ &\quad + \omega_k^T(M_{21}u_k^d + M_{22}\omega_k + \varphi \bar{D}^T P\bar{A}\bar{x}_k)\\ &\quad + \varphi \bar{x}_k^T \bar{A}^T P\bar{D}\omega_k. \end{aligned} \quad (24)$$

On the ground of (22), the formula (24) is developed as

$$\bar{x}_k^T P\bar{x}_k = \bar{x}_k^T(\bar{Q} + \varphi \bar{A}^T P\bar{A})\bar{x}_k + \varphi \bar{x}_k^T \left[\bar{A}^T P\bar{B} \;\; \bar{A}^T P\bar{D}\right]\begin{bmatrix}\bar{\alpha_d}u_k^d\\ \omega_k\end{bmatrix}. \quad (25)$$

In light of matrix theory, (22) can be rewritten as follows:

$$\begin{bmatrix}\bar{\alpha_d}u_k^d\\ \omega_k\end{bmatrix} = -\varphi \begin{bmatrix}\beta M_{11} & M_{12}\\ \beta M_{21} & M_{22}\end{bmatrix}^{-1}\begin{bmatrix}\bar{B}^T P\bar{A}\\ \bar{D}^T P\bar{A}\end{bmatrix}\bar{x}_k. \quad (26)$$

Under the combination of (25) and (26), the following result is elicited:

$$\bar{x}_k^T P\bar{x}_k = \bar{x}_k^T\left(\varphi \bar{A}^T P\bar{A} + \bar{Q} - \varphi^2\left[\bar{A}^T P\bar{B} \quad \bar{A}^T P\bar{D}\right] \times \begin{bmatrix}\beta M_{11} & M_{12}\\ \beta M_{21} & M_{22}\end{bmatrix}^{-1}\begin{bmatrix}\bar{B}^T P\bar{A}\\ \bar{D}^T P\bar{A}\end{bmatrix}\right)\bar{x}_k. \quad (27)$$

Obviously, the GARE (17) can be obtained. The proof of Theorem 1 is completed. ■

It has been summarized from (23) that the optimal controller gain will be obtained if solution P exists. In addition, the DoS attacks will affect the existence of P, so it is of significance to analyze the existence of solution.

B. Existence of the Solution to Riccati Equation

In this subsection, by using the general iterative method, the existence of the solution for GARE (17) is discussed. It is worth noting that the direct analysis of the solution for GARE (17) is not feasible. Therefore, we firstly define a function as follows:

$$\begin{aligned} F(X) &= \varphi \bar{A}^T X\bar{A} + \bar{Q} - \varphi^2\left[\bar{A}^T X\bar{B} \quad \bar{A}^T X\bar{D}\right]\\ &\quad \times \begin{bmatrix}\beta \bar{M}_{11} & \bar{M}_{12}\\ \beta \bar{M}_{21} & \bar{M}_{22}\end{bmatrix}^{-1}\begin{bmatrix}\bar{B}^T X\bar{A}\\ \bar{D}^T X\bar{A}\end{bmatrix}\\ F^k(X) &= F(F^{k-1}(X)) \end{aligned} \quad (28)$$

where $\bar{M}_{11} = R + \varphi \bar{B}^T X\bar{B}$, $\bar{M}_{12} = \varphi \bar{B}^T X\bar{D}$, $\bar{M}_{21} = \bar{\alpha_d}\varphi \bar{D}^T X\bar{B}$ and $\bar{M}_{22} = \varphi \bar{D}^T X\bar{D} - \gamma^2 I$. $F^k(X)$ denotes the kth iteration of $F(X)$.

It is noteworthy that $F(X)$ is defined based on the right side of (17). If we are capable to prove that $F(X)$ converges to X, the existence of the unique solution to GARE (17) will be guaranteed. Before analyzing the convergence of $F(X)$, the following lemma [37] is necessary.

Lemma 1: Define

$$\begin{aligned} \tilde{F}(X, K, L) &= \bar{Q} + \bar{\alpha_d}K^T RK + \varphi(\bar{A} + \bar{\alpha_d}\bar{B}K + \bar{D}L)^T\\ &\quad \times X(\bar{A} + \bar{\alpha_d}\bar{B}K + \bar{D}L) - \gamma^2 L^T L\\ &\quad + \varphi \bar{\alpha_d}(1 - \bar{\alpha_d})K^T \bar{B}^T X\bar{B}K\\ \tilde{F}^k(X, K, L) &= \tilde{F}(\tilde{F}^{k-1}(X), K, L). \end{aligned} \quad (29)$$

If there is a matrix $\vec{X} \geq 0$ that satisfies $\vec{X} \geq \tilde{F}(\vec{X}, K, L)$, then $\lim_{k\to\infty}\tilde{F}^k(X_0, K, L) = X^*$ for $\forall X_0 \geq 0$.

Proof: Note that $\tilde{F}(X, K, L)$ is equivalent to $F(X)$. According to the expression (29), the following two cases are conducted to prove that $\tilde{F}(X, K, L)$ is a monotonically increasing function with regard to the variable X and has an upper bound.

Case 1. Initial value $X_0 = 0$. Then, $X_k = \tilde{F}^k(0, K, L)$.

According to the increment of function $\tilde{F}^k(X, K, L)$, the monotonically increasing sequence X_k is derived, that is

$$X_1 \leq X_2 \leq \cdots \leq X_k.$$

Then, combining the condition $\vec{X} \geq \tilde{F}(\vec{X}, K, L)$ and the property of function $\tilde{F}^k(X, K, L)$, the following sequence can be acquired:

$$\tilde{F}^k(\vec{X}, K, L) \leq \cdots \leq \tilde{F}(\vec{X}, K, L) \leq \vec{X}.$$

In addition, one can derive

$$X_k \leq \tilde{F}^k(\vec{X}, K, L) \leq \vec{X}$$

which indicates that X_k converges to $\vec{X}$.

Case 2. When the initial condition $X_0 \geq 0$, there exists $\delta > 0$ satisfying $X^* \geq \delta X_0 \geq 0$, which implies the following inequality holds:

$$\tilde{F}^k(0, K, L) \leq \tilde{F}^k(\delta X_0, K, L) \leq \tilde{F}^k(X^*, K, L) \leq X^*$$

thus, $\lim_{k\to\infty} \tilde{F}^k(\delta X_0, K, L) = X^*$ can be obtained.

Based on the structure of (29), we can further get

$$\begin{aligned}&\tilde{F}^k(\delta X_0, K, L) - \tilde{F}^k(0, K, L)\\ &\quad = \delta(\tilde{F}^k(X_0, K, L) - \tilde{F}^k(0, K, L))\end{aligned}$$

which implies

$$\lim_{k\to\infty} \tilde{F}^k(X_0, K, L) = X^*.$$

The proof of Lemma 1 is completed now. ■

By applying Lemma 1, we can derive the following theorem which shows there exists only one solution to (17).

Theorem 2: If there exists a matrix $\vec{X} \geq 0$ such that $\vec{X} \geq F(\vec{X})$ holds, the following two conditions are satisfied:

- $F(X) = X$ only has one solution X^*,
- $\lim_{k\to\infty} F^k(X_0) = X^*$ for $\forall X_0 \geq 0$.

Proof: Since $F(\vec{X}_1) = \tilde{F}(\vec{X}_1, K, L) \leq \tilde{F}(\vec{X}_2, K, L) = F(\vec{X}_2)$ holds for any $\vec{X}_1 \leq \vec{X}_2$, we can get that $F(X)$ increases monotonically with variable X. From Lemma 1, it can be got that $\lim_{k\to\infty} F^k(X_0) = X^*$ for $\forall X_0 \geq 0$. Note that $F(X)$ is equivalent to the right of (17), thus, there exists only one matrix $X > 0$ such that $F(X) = X$. The proof is completed. ■

C. Stability Analysis

Theorem 3 is provided to analyze the boundedness of tracking error e_k.

Theorem 3: Consider the augmented system (6) without external disturbance. For known scalars $\bar{\alpha_d}$, the proposed control strategy u_k^d can stochastically stabilize $\bar{e}_k = \varphi^{k/2}e_k$ under the impact of DoS attacks. Then, the boundedness of tracking error e_k is further obtained.

Proof: Define an auxiliary state vector $\hat{x}_k = \varphi^{k/2}\bar{x}_k$. Owing to $e_k = [C \;\; -I]\bar{x}_k$ and $[C \;\; -I] \neq 0$, it is easy to get that $\bar{e}_k$ goes to zero if $\hat{x}_k$ goes to zero. In the following, it is illustrated that $\hat{x}_k$ and $\bar{e}_k$ can converge to zero when k goes to infinity.

Choose the Lyapunov function as

$$V(\hat{x}_k) = E\left\{\hat{x}_k^T P \hat{x}_k\right\} \tag{30}$$

where P is the solution of GARE (17).

When $\omega_k = 0$, P and u_k^d can be calculated as [38]

$$\begin{aligned}P &= \bar{Q} + \varphi\bar{A}^T P\bar{A} - \bar{\alpha_d}\varphi^2\bar{A}^T P\bar{B}(R + \varphi\bar{B}^T P\bar{B})^{-1}\bar{B}^T P\bar{A}\\ u_k^d &= -\varphi(R + \varphi\bar{B}^T P\bar{B})^{-1}\bar{B}^T P\bar{A}\bar{x}_k.\end{aligned} \tag{31}$$

Let $\Delta V(k) = V(\hat{x}_{k+1}) - V(\hat{x}_k)$, according to (30), $\Delta V(k)$ is represented by

$$\Delta V(k) = E\left\{\hat{x}_{k+1}^T P\hat{x}_{k+1}\right\} - E\left\{\hat{x}_k^T P\hat{x}_k\right\} \tag{32}$$

with

$$\hat{x}_{k+1} = \varphi^{1/2}(\bar{A} + \alpha_k\bar{B}K)\hat{x}_k. \tag{33}$$

Using (33), the equation (32) is rewritten as

$$\begin{aligned}&\Delta V(k)\\ &= E\left\{\varphi\hat{x}_k^T(\bar{A} + \alpha_k\bar{B}K)^T P(\bar{A} + \alpha_k\bar{B}K)\hat{x}_k - \hat{x}_k^T P\hat{x}_k\right\}\\ &= \hat{x}_k^T[\varphi\bar{A}^T P\bar{A} + \bar{\alpha_d}\varphi\bar{A}^T P\bar{B}K + \bar{\alpha_d}\varphi K^T\bar{B}^T P\bar{A}\\ &\quad + \bar{\alpha_d}\varphi K^T\bar{B}^T P\bar{B}K - P]\hat{x}_k.\end{aligned} \tag{34}$$

According to the formulas (3) and (31), one has

$$\begin{aligned}K &= -\varphi(R + \varphi\bar{B}^T P\bar{B})^{-1}\bar{B}^T P\bar{A}\\ P &= \bar{Q} + \varphi\bar{A}^T P\bar{A} + \bar{\alpha_d}\varphi\bar{A}^T P\bar{B}K.\end{aligned} \tag{35}$$

By substituting (35) into (34), the following result can be acquired:

$$\begin{aligned}\Delta V(k) &= \hat{x}_k^T[\bar{\alpha_d}\varphi(K^T\bar{B}^T P\bar{A} + K^T\bar{B}^T P\bar{B}K) - \bar{Q}]\hat{x}_k\\ &= \hat{x}_k^T[-\bar{\alpha_d}K^T(R + \varphi\bar{B}^T P\bar{B})K\\ &\quad + \bar{\alpha_d}\varphi K^T\bar{B}^T P\bar{B}K - \bar{Q}]\hat{x}_k\\ &= \hat{x}_k^T(-\bar{Q} - \bar{\alpha_d}K^T RK)\hat{x}_k \;<\; 0\end{aligned} \tag{36}$$

Thus, the stochastica stability of $\bar{e}_k$ is guaranteed when $\omega_k = 0$. Furthermore, the tracking error e_k is bounded under the proposed control strategy. ■

IV. RL-Boosted Model-Free Control Strategy

The above results are mainly developed with completely known knowledge of system dynamics. Next, the QL method is introduced to derive a model-free control strategy, which no longer requires known system dynamics.

Recalling the equation (15), the H_∞ tracking Q-function is given as

$$\begin{aligned}Q(\bar{x}_k, u_k^d, \omega_k) &= E\left\{\bar{x}_k^T\bar{Q}\bar{x}_k + \alpha_k {u_k^d}^T Ru_k^d - \gamma^2\omega_k^T\omega_k\right\}\\ &\quad + E\{\varphi\bar{x}_{k+1}^T P\bar{x}_{k+1}\}.\end{aligned} \tag{37}$$

Combining (6) and (37), the expression (37) can be developed as follows:

$$Q(\bar{x}_k, u_k^d, \omega_k) = \begin{bmatrix} \bar{x}_k \\ u_k^d \\ \omega_k \end{bmatrix}^T \begin{bmatrix} N_{11} & \bar{\alpha_d} N_{12} & N_{13} \\ \bar{\alpha_d} N_{21} & \bar{\alpha_d} N_{22} & \bar{\alpha_d} N_{23} \\ N_{31} & N_{32} & N_{33} \end{bmatrix} \begin{bmatrix} \bar{x}_k \\ u_k^d \\ \omega_k \end{bmatrix} = \begin{bmatrix} \bar{x}_k \\ u_k^d \\ \omega_k \end{bmatrix}^T N \begin{bmatrix} \bar{x}_k \\ u_k^d \\ \omega_k \end{bmatrix} \quad (38)$$

where

$$\begin{aligned} N_{11} &= \bar{Q} + \varphi \bar{A}^T P \bar{A} \in \mathbb{R}^{n\times n} \\ N_{12} &= \varphi \bar{A}^T P \bar{B} \in \mathbb{R}^{n\times q} \\ N_{13} &= \varphi \bar{A}^T P \bar{D} \in \mathbb{R}^{n\times s} \\ N_{21} &= \varphi \bar{B}^T P \bar{A} \in \mathbb{R}^{q\times n} \\ N_{22} &= R + \varphi \bar{B}^T P \bar{B} \in \mathbb{R}^{q\times q} \\ N_{23} &= \varphi \bar{B}^T P \bar{D} \in \mathbb{R}^{q\times s} \\ N_{31} &= \varphi \bar{D}^T P \bar{A} \in \mathbb{R}^{s\times n} \\ N_{32} &= \bar{\alpha_d} \varphi \bar{D}^T P \bar{B} \in \mathbb{R}^{s\times q} \\ N_{33} &= \varphi \bar{D}^T P \bar{D} - \gamma^2 I \in \mathbb{R}^{s\times s}. \end{aligned}$$

Based on the Q-function (37), we transform the tracking control problem into solving the following minmax problem:

$$Q^*(\bar{x}_k, u_k^d, \omega_k) = \min_{u_k^d \in \mathscr{A}_1(\Omega_{1,k})} \max_{\omega_k \in \mathscr{A}_2(\Omega_{2,k})} Q(\bar{x}_k, u_k^d, \omega_k). \quad (39)$$

In virtue of $\frac{\partial Q(\bar{x}_k, u_k^d, \omega_k)}{\partial u_k^d} = 0$ and $\frac{\partial Q(\bar{x}_k, u_k^d, \omega_k)}{\partial \omega_k} = 0$, the following gains K^* and L^* can be derived:

$$\begin{aligned} K &= (N_{22} - N_{23}(N_{33})^{-1} N_{32})^{-1} \\ &\quad \times (N_{23}(N_{33})^{-1} N_{31} - N_{21}) \\ L &= (N_{33} - N_{32}(N_{22})^{-1} N_{23})^{-1} \\ &\quad \times (N_{32}(N_{22})^{-1} N_{21} - N_{31}). \end{aligned} \quad (40)$$

For convenience, define $\Xi_k = \begin{bmatrix} \bar{x}_k^T & u_k^{dT} & \omega_k^T \end{bmatrix}^T$. Then, the condition (38) is rewritten as

$$Q(\bar{x}_k, u_k^d, \omega_k) = \Xi_k^T N \Xi_k. \quad (41)$$

Referring to the equation (13), the Q-function Bellman equation is given by

$$\Xi_k^T N \Xi_k = E\left\{ \bar{x}_k^T \bar{Q} \bar{x}_k + \alpha_k {u_k^d}^T R u_k^d - \gamma^2 \omega_k^T \omega_k \right\} + \varphi \Xi_{k+1}^T N \Xi_{k+1}. \quad (42)$$

Define $\bar{\Xi}_k = \Xi_k^T \otimes \Xi_k$, (42) is presented as

$$\bar{\Xi}_k Vec(N) = E\left\{ \bar{x}_k^T \bar{Q} \bar{x}_k + \alpha_k {u_k^d}^T R u_k^d - \gamma^2 \omega_k^T \omega_k \right\} + \varphi \bar{\Xi}_{k+1} Vec(N) \quad (43)$$

where $Vec(N) = [n_{11}, 2n_{12}, \ldots, 2n_{1l}, n_{22}, 2n_{23}, \ldots, 2n_{2l}, \ldots, n_{ll}]^T$, n_{ij} corresponds to the element in ith row and jth column, $i, j = 1, \ldots, l$, $l = n + p + q + s$.

The least-squares (LS) approach is used to obtain matrix N [39]. Because $N \in \mathbb{R}^{l\times l}$ is a symmetric matrix, at least $l \times l/2$ data are needed to calculate (43).

Algorithm 1

---

procedure System Initialization:

Start with $N^0 \geq 0$, u_k^0, ω_k^0, a small positive δ and $i = 0$.

procedure Repeat:

1. Policy Evaluation: Employ LS to solve N^{i+1}

$$\Xi_k^T N^{i+1} \Xi_k = E\left\{ \bar{x}_k^T \bar{Q} \bar{x}_k + \alpha_k {(u_k^d)^i}^T R (u_k^d)^i - \gamma^2 {\omega_k^i}^T \omega_k^i \right\} + \varphi \Xi_{k+1}^T N^i \Xi_{k+1}. \quad (44)$$

2. Policy Improvement: Update the gains K and L using N^{i+1}

$$\begin{aligned} K^{i+1} &= (N_{22}^{i+1} - N_{23}^{i+1} {N_{33}^{i+1}}^{-1} N_{32}^{i+1})^{-1} \\ &\quad \times (N_{23}^{i+1} {N_{33}^{i+1}}^{-1} N_{31}^{i+1} - N_{21}^{i+1}) \\ L^{i+1} &= (N_{33}^{i+1} - N_{32}^{i+1} {N_{22}^{i+1}}^{-1} N_{23}^{i+1})^{-1} \\ &\quad \times (N_{32}^{i+1} {N_{22}^{i+1}}^{-1} N_{21}^{i+1} - N_{31}^{i+1}). \end{aligned} \quad (45)$$

3. Stop

if $|K^{i+1} - K^i| \leq \delta$ *and* $|L^{i+1} - L^i| \leq \delta$ **then**

Output the N^{i+1}, control policy u_k^d and disturbance ω_k.

else

$i = i + 1$ and return step 1.

Next, Algorithm 1 is proposed to learn matrix N online and obtain the optimal tracking controller. The convergence analysis will be given in next theorem.

It is notable that u_k and ω_k depends on $\bar{x}_k$ only, which will result in LS approach to be infeasible. Then, the probing noises p_k and q_k need to be added in u_k^d and ω_k to ensure Algorithm 1 is solvable separately, that is, $u_k^d = K\bar{x}_k + p_k$ and $\omega_k = L\bar{x}_k + q_k$. And the invertibility of $\Xi_k^T \Xi_k$ can be proved [40].

Remark 6: The probing noises p_k and q_k introduced in control input and external disturbance are inspired by the recent work in [40], which can assure the condition of policy evaluation. In fact, the choice of probing noises is not a key issue in the research. The sinusoidal function and exponential attenuation function are often used as the probing noises in many literatures. In addition, the probing noise will not have any impact on the estimation of Q-function.

In order to prove the Theorem 4, Lemma 2 is needed.

Lemma 2: [41] In Algorithm 1, iterating N^i and iterating P^i are equivalent.

Theorem 4: If the GARE (17) is solvable, then N^i in Algorithm 1 converges to optimal value N^* within error δ. Moreover, target gains K^* and L^* are designed.

Proof: Inspired by [41], P and N are expressed as

$$P = \begin{bmatrix} I & K^T & L^T \end{bmatrix} N \begin{bmatrix} I \\ K \\ L \end{bmatrix} \quad (46)$$

which further means

$$N = \begin{bmatrix} \bar{Q} & & \\ & \bar{\alpha_d} R & \\ & & -\gamma^2 I \end{bmatrix}$$

$$+\varphi E\left\{\begin{bmatrix}\bar{A}^T\\ \alpha_k\bar{B}^T\\ \bar{D}^T\end{bmatrix}P\begin{bmatrix}\bar{A} & \alpha_k\bar{B} & \bar{D}\end{bmatrix}\right\} \tag{47}$$

Recalling (40), u_k^d and ω_k can be written as

$$u_k^d = (N_{22} - N_{23}N_{33}^{-1}N_{32})^{-1} \times (N_{23}N_{33}^{-1}N_{31} - N_{21})\bar{x}_k$$
$$\omega_k = (N_{33}-N_{32}N_{22}^{-1}N_{23})^{-1}\times(N_{32}N_{22}^{-1}N_{21}-N_{31})\bar{x}_k. \tag{48}$$

Therefore, if the matrix N can be calculated through Algorithm 1, the desired control strategy will be obtained without the system dynamics.

Lemma 2 means that whether N^i converges or not is decided by the convergence of P^i. Besides, the existence of the optimal solution is proved in Theorem 2. Thus, the convergence of N^i can be guaranteed and the gains K, L will be solved. So far, all proofs have been completed. ■

Remark 7: In this paper, the optimal tracking control problem is addressed for linear system (1) under DoS attacks by RL-learning method. Noted that the nonlinear systems with input constraints can also be tackled by RL-learning method, some related works have been published, see [26] for example.

Remark 8: It should be clarified that Q-learning is a classic reinforcement learning algorithm, which can be applied without requirements of the knowledge of system dynamics. Q-learning has been widely used to solve the optimal control problems owing to its advantages in model-free learning and off-policy learning [25], [26].

Remark 9: Note that the parameter γ in Eq. (8) is given a priori, which is often specified in real-world applications. For known scalars $\bar{\alpha_d}$ and γ and given δ, the gains K in (3) and L in (7) can be obtained according to Algorithm 1. The optimal performance can be obtained by solving the following optimization problem:

$$\min_{\bar{\alpha_d},\delta}\gamma \tag{49}$$

subject to (44), Then K and L can be derived.

V. Simulation Example

In this section, an F-404 aircraft engine system (FACS) is given to prove the correctness of the control strategy. Referring to [42] and [43], the parameter of FACS is presented as

$$A=\begin{bmatrix}1.4600 & 0 & 2.4280\\ 0.1643 & -0.4000 & -0.3788\\ 0.3107 & 0 & -2.2300\end{bmatrix}.$$

Then, the nominal system (1) is discretized as follows:

$$x_{k+1}=\begin{bmatrix}1.1612 & 0 & 0.2353\\ 0.0168 & 0.9608 & -0.0314\\ 0.0307 & 0 & 0.8036\end{bmatrix}x_k + \begin{bmatrix}0.0356\\ -0.0050\\ 0.2693\end{bmatrix}u_k+\begin{bmatrix}0\\ 0.0980\\ 0\end{bmatrix}\omega_k$$
$$y_k=\begin{bmatrix}1 & 1 & 0\end{bmatrix}x_k.$$

In the experiment, the parameters are given as $Q = 100$, $R = 0.1$, $\varphi = e^{-0.1}$, $\bar{\alpha_d} = 0.8$, and the H_∞ tracking performance index $\gamma = 8$.

The solution P of the GARE (17) is obtained via the method in Theorem 1:

$$P=\begin{bmatrix}177.4124 & 165.1484 & 12.0807 & -167.5431\\ 165.1484 & 155.6609 & 9.5620 & -157.5961\\ 12.0807 & 9.5620 & 2.3596 & -10.0188\\ -167.5431 & -157.5961 & -10.0188 & 159.6466\end{bmatrix}.$$

Based on the above mentioned solution, the corresponding controller gain K^* and disturbance gain L^* are deduced as

$$K^*=\begin{bmatrix}-15.8434 & -11.3470 & -4.0336 & 12.1123\end{bmatrix}$$
$$L^*=\begin{bmatrix}0.1379 & 0.1131 & 0.0239 & -0.1178\end{bmatrix}.$$

Then, when the system dynamics are unknown, Algorithm 1 is used to design the optimal controller. Consider the following external disturbance:

$$\omega_k = 7\cos(0.4k)rand()$$

where $rand()$ is used to generate a random number in (0, 1). To highlight the proposed algorithm, the initial system state, initial controller gain and disturbance gain are selected as $\bar{x}_0 = \begin{bmatrix}1 & 0 & 0\end{bmatrix}^T$, $K_0 = \begin{bmatrix}-3 & 0 & 0 & 3\end{bmatrix}$, $L_0 = \begin{bmatrix}-3 & 0 & 0 & 3\end{bmatrix}$, respectively. And initial matrix N_0 is an identity matrix with appropriate dimension. Upon termination of the Algorithm 1, the matrix N ultimately converges to $\begin{bmatrix}\bar{N}_1 & \bar{N}_2\end{bmatrix}$, where N_1 and N_2 are represented by

$$\bar{N}_1=\begin{bmatrix}323.0708 & 269.4763 & 49.3364\\ 269.4763 & 231.0573 & 36.2204\\ 49.3364 & 36.2204 & 11.8816\\ -278.5858 & -236.8240 & -38.4471\\ 9.3246 & 6.6742 & 2.3806\\ 17.2911 & 13.3714 & 3.6955\end{bmatrix}$$
$$\bar{N}_2=\begin{bmatrix}-278.5858 & 9.3246 & 17.2911\\ -236.8240 & 6.6742 & 13.3714\\ -38.4471 & 2.3806 & 3.6955\\ 245.0381 & -7.1189 & -13.9597\\ -7.1189 & 0.5942 & 0.6809\\ -13.9597 & 0.6809 & -62.6358\end{bmatrix}$$

which can be applied to obtain the controller gain K and disturbance gain L as follows:

$$K=\begin{bmatrix}-15.8070 & -11.3305 & -4.0238 & 12.0803\end{bmatrix}$$
$$L=\begin{bmatrix}0.1042 & 0.0903 & 0.0153 & -0.0915\end{bmatrix}.$$

In the following, two cases are employed to demonstrate tracking performance and the correctness of the designed results.

Case 1. In this case, tracking trajectory is taken as

$$r_k = 3\sin(0.03k).$$

Under the utilization of QL algorithm, the relative results are depicted in Figs. 2 and 3. In Fig. 2, the curves of system output y_k and reference signal r_k are displayed in the presence of DoS attacks. Moreover, Fig. 3 plots the response of tracking error e_k under the devised security control scheme. It is not difficult to observe that the system output can effectively track the reference signal even though the malicious DoS attacks are randomly activated during the whole simulation process.

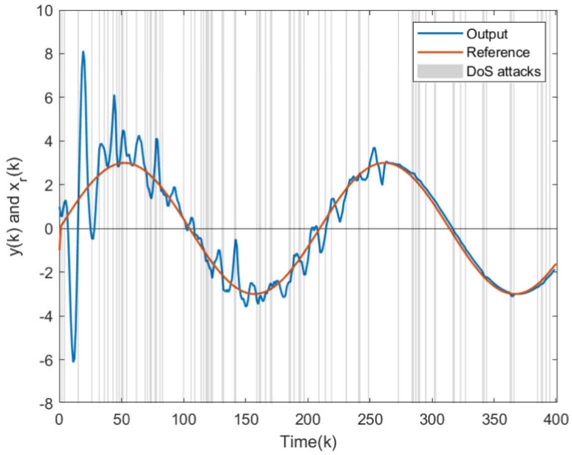


Fig. 2. The responses of y_k and r_k in Case 1.

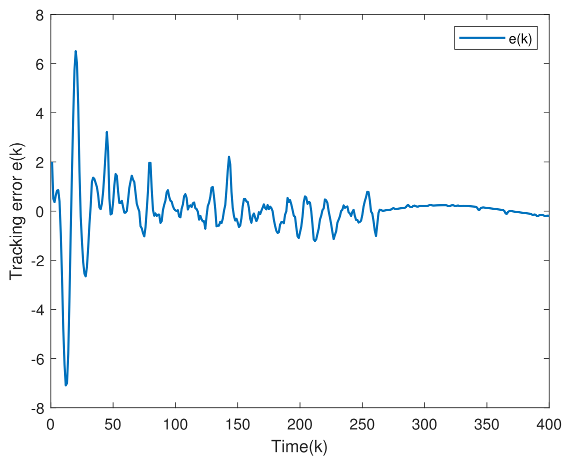


Fig. 3. Tracking error e_k in Case 1.

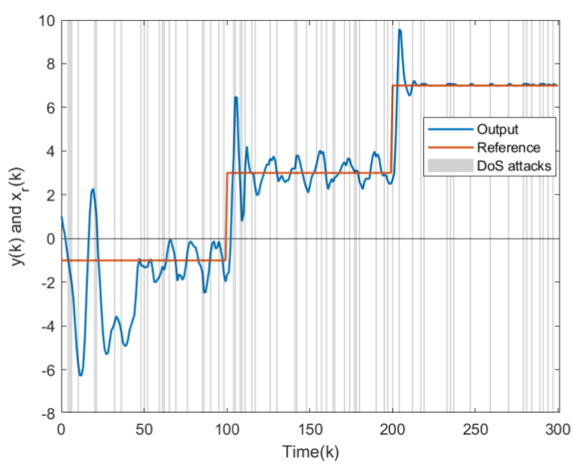


Fig. 4. The responses of y_k and r_k in Case 2.

Hence, the feasibility of the proposed tracking control strategy is validated according to the aforesaid contents.

Case 2. In this case, the tracking trajectory is selected as

$$r_k = \begin{cases} -1, & 0 \le k < 100 \\ 3, & 100 \le k < 200 \\ 7, & k \ge 200. \end{cases}$$

In Case 1, the reference trajectory of sine function trend has been verified. In the following, we discuss the case where

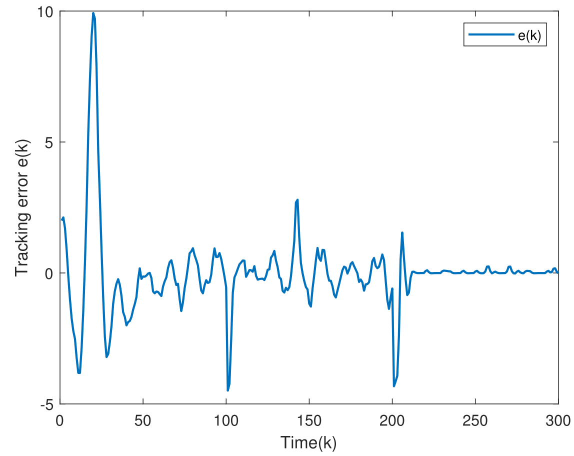


Fig. 5. Tracking error e_k in Case 2.

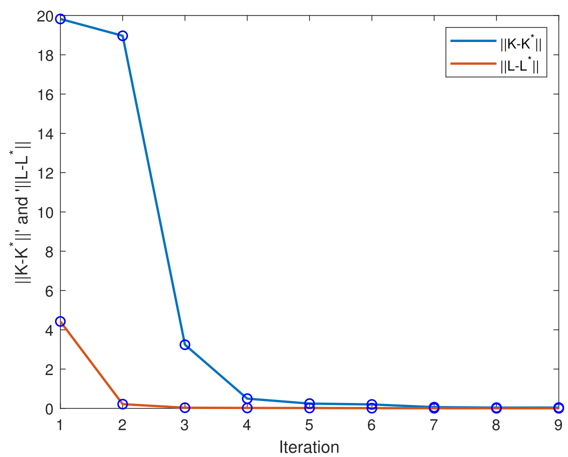


Fig. 6. The errors of K and L between optimal and calculated values.

the reference trajectory is a linear piecewise function. It is apparently witnessed from Figs. 4 and 5 that the reference model with piecewise form can be excellently tracked by output signal y_k. Fig. 6 shows the error between optimal controller gain K^* and computed gain K as well as the error between worst disturbance gain L^* and computed L matrices, respectively. Obviously, the optimal controller gain can be derived after 9 iterations. On basis of these consequences, the validity of the presented QL algorithm can also be verified.

VI. Conclusion

In this paper, the RL-assisted security tracking control scheme is put forward for DTNCSs against DoS attacks. Considering the impact of DoS attacks, a particular value function is proposed to comprehensively reflect the control input, external disturbance and tracking error. Based on augmented system, the tracking Bellman equation and the attacks-based GARE have been obtained. In addition, the existence of optimal solution under DoS attacks can be ensured. Besides, a QL algorithm has been provided to deal with the discrete-time zero-sum game problem without any system dynamics, and the convergence of the QL algorithm is proved. In the end, the simulation results of FACS demonstrate the practicability of the devised security tracking control strategy. Future work will focus on enhancing the security of NCSs, enabling them to withstand not only DoS attacks but also other potential network attacks such as false data injection attacks and deception attacks. Additionally, we will further investigate the RL-boosted security control problem for nonlinear NCSs with input constraints.

REFERENCES

[1] S. Hu, X. Ge, Y. Li, X. Chen, X. Xie, and D. Yue, "Resilient load frequency control of multi-area power systems under DoS attacks," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 936–947, 2023.

[2] Z. Cheng, D. Yue, S. Shen, S. Hu, and L. Chen, "Secure frequency control of hybrid power system under DoS attacks via lie algebra," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 1172–1184, 2022.

[3] C. Peng and H. Sun, "Switching-like event-triggered control for networked control systems under malicious denial of service attacks," *IEEE Trans. Autom. Control*, vol. 65, no. 9, pp. 3943–3949, Sep. 2020.

[4] L. Zha, Y. Guo, J. Liu, X. Xie, and E. Tian, "Protocol-based distributed security fusion estimation for time-varying uncertain systems over sensor networks: Tackling DoS attacks," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 10, pp. 119–130, 2024.

[5] X. Bu, W. Yu, Y. Yin, and Z. Hou, "Event-triggered data-driven control for nonlinear systems under frequency-duration-constrained DoS attacks," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 1449–1460, 2023.

[6] L. Zha, R. Liao, J. Liu, X. Xie, E. Tian, and J. Cao, "Dynamic event-triggered output feedback control for networked systems subject to multiple cyber attacks," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 13800–13808, Dec. 2022.

[7] Z. Cao, B. Niu, G. Zong, and N. Xu, "Small-gain technique-based adaptive output constrained control design of switched networked nonlinear systems via event-triggered communications," *Nonlinear Anal., Hybrid Syst.*, vol. 47, Feb. 2023, Art. no. 101299.

[8] Y. Yu, W. Yang, W. Ding, and J. Zhou, "Reinforcement learning solution for cyber-physical systems security against replay attacks," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 2583–2595, 2023.

[9] J. Yang, Q. Zhong, K. Shi, and S. Zhong, "Co-design of observer-based fault detection filter and dynamic event-triggered controller for wind power system under dual alterable DoS attacks," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 1270–1284, 2022.

[10] J. Liu, E. Gong, L. Zha, E. Tian, and X. Xie, "Interval type-2 fuzzy-model-based filtering for nonlinear systems with event-triggering weighted try-once-discard protocol and cyberattacks," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 3, pp. 721–732, Mar. 2024, doi: 10.1109/TFUZZ.2023.3305088.

[11] J. Liu, E. Gong, L. Zha, X. Xie, and E. Tian, "Outlier-resistant recursive security filtering for multirate networked systems under fading measurements and round-robin protocol," *IEEE Trans. Control Netw. Syst.*, vol. 10, no. 4, pp. 1962–1974, Dec. 2023.

[12] L. Zha, R. Liao, J. Liu, X. Xie, E. Tian, and J. Cao, "Outlier-resistant distributed filtering over sensor networks under dynamic event-triggered schemes and DoS attacks," *IEEE Trans. Autom. Sci. Eng.*, early access, Feb. 5, 2024, doi: 10.1109/TASE.2024.3360718.

[13] Y. Wan, C. Long, R. Deng, G. Wen, X. Yu, and T. Huang, "Distributed event-based control for thermostatically controlled loads under hybrid cyber attacks," *IEEE Trans. Cybern.*, vol. 51, no. 11, pp. 5314–5327, Nov. 2021.

[14] Y. Zhang, L. Wei, W. Fu, X. Chen, and S. Hu, "Secondary frequency control strategy considering DoS attacks for MTDC system," *Electric Power Syst. Res.*, vol. 214, Jan. 2023, Art. no. 108888.

[15] X. Chen, C. Hu, E. Tian, and C. Peng, "Event-based fuzzy resilient control of nonlinear DC microgrids under denial-of-service attacks," *ISA Trans.*, vol. 127, pp. 206–215, Aug. 2022.

[16] Z. Cheng, S. Hu, D. Yue, C. Dou, and S. Shen, "Resilient distributed coordination control of multiarea power systems under hybrid attacks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 1, pp. 7–18, Jan. 2022.

[17] N. Zhao, P. Shi, W. Xing, and J. Chambers, "Observer-based event-triggered approach for stochastic networked control systems under denial of service attacks," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 1, pp. 158–167, Mar. 2021.

[18] Y. Pan, Y. Wu, and H. Lam, "Security-based fuzzy control for nonlinear networked control systems with DoS attacks via a resilient event-triggered scheme," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 10, pp. 4359–4368, Oct. 2022.

[19] R. Zhao, Z. Zuo, Y. Wang, and W. Zhang, "Active control strategy for switched systems against asynchronous DoS attacks," *Automatica*, vol. 148, Feb. 2023, Art. no. 110765.

[20] H. Yan, J. Wang, H. Zhang, H. Shen, and X. Zhan, "Event-based security control for stochastic networked systems subject to attacks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 11, pp. 4643–4654, Nov. 2020.

[21] D. Ding, H. Liu, H. Dong, and H. Liu, "Resilient filtering of nonlinear complex dynamical networks under randomly occurring faults and hybrid cyber-attacks," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 4, pp. 2341–2352, Jul./Aug. 2022.

[22] J. Wang, D. Wang, L. Su, J. H. Park, and H. Shen, "Dynamic event-triggered h8 load frequency control for multi-area power systems subject to hybrid cyber attacks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 12, pp. 7787–7798, Dec. 2022.

[23] J. Liu, N. Zhang, Y. Li, X. Xie, E. Tian, and J. Cao, "Learning-based event-triggered tracking control for nonlinear networked control systems with unmatched disturbance," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 53, no. 5, pp. 3230–3240, May 2023.

[24] J. Liu, N. Zhang, L. Zha, X. Xie, and E. Tian, "Reinforcement learning-based decentralized control for networked interconnected systems with communication and control constraints," *IEEE Trans. Autom. Sci. Eng.*, early access, Nov. 2, 2024, doi: 10.1109/TASE.2023.3300917.

[25] L. Zhang et al., "Data-driven H_∞ optimal output feedback control for linear discrete-time systems based on off-policy Q-learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 7, pp. 3553–3567, Jul. 2023.

[26] S. Song, M. Zhu, X. Dai, and D. Gong, "Model-free optimal tracking control of nonlinear input-affine discrete-time systems via an iterative deterministic Q-learning algorithm," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 1, pp. 999–1012, Jan. 2024.

[27] C. Yang, C. Chen, W. He, R. Cui, and Z. Li, "Robot learning system based on adaptive neural control and dynamic movement primitives," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 777–787, Mar. 2019.

[28] P. Zhu, S. Jin, X. Bu, Z. Hou, and C. Yin, "Model-free adaptive control for a class of MIMO nonlinear cyberphysical systems under false data injection attacks," *IEEE Trans. Control Netw. Syst.*, vol. 10, no. 1, pp. 467–478, Mar. 2023.

[29] L. Yu, Z. Xu, T. Zhang, X. Guan, and D. Yue, "Energy-efficient personalized thermal comfort control in office buildings based on multi-agent deep reinforcement learning," *Building Environ.*, vol. 223, Sep. 2022, Art. no. 109458.

[30] Z. Peng, R. Luo, J. Hu, K. Shi, S. K. Nguang, and B. K. Ghosh, "Optimal tracking control of nonlinear multiagent systems using internal reinforce Q-learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 4043–4055, Aug. 2022.

[31] Y. Xu and Z.-G. Wu, "Data-based collaborative learning for multiagent systems under distributed denial-of-service attacks," *IEEE Trans. Cognit. Develop. Syst.*, vol. 16, no. 1, pp. 75–85, Feb. 2024.

[32] Y. Yin, X. Bu, P. Zhu, and W. Qian, "Point-to-point consensus tracking control for unknown nonlinear multi-agent systems using data-driven iterative learning," *Neurocomputing*, vol. 488, pp. 78–87, Jun. 2022.

[33] K. Zhang, H. Zhang, Y. Mu, and C. Liu, "Decentralized tracking optimization control for partially unknown fuzzy interconnected systems via reinforcement learning method," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 4, pp. 917–926, Apr. 2021.

[34] Z. Xia, M.-J. Hu, W. Dai, H. Yan, and X. Ma, "Q-learning-based multi-rate optimal control for process industries," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 70, no. 6, pp. 2006–2010, Jun. 2023.

[35] E. Mousavinejad, X. Ge, Q.-L. Han, F. Yang, and L. Vlacic, "Resilient tracking control of networked control systems under cyber attacks," *IEEE Trans. Cybern.*, vol. 51, no. 4, pp. 2107–2119, Apr. 2021.

[36] Y. Jiang, J. Fan, T. Chai, F. L. Lewis, and J. Li, "Tracking control for linear discrete-time networked control systems with unknown dynamics and dropout," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4607–4620, Oct. 2018.

[37] C. Wu, X. Li, W. Pan, J. Liu, and L. Wu, "Zero-sum game-based optimal secure control under actuator attacks," *IEEE Trans. Autom. Control*, vol. 66, no. 8, pp. 3773–3780, Aug. 2021.

[38] B. Kiumarsi, F. L. Lewis, H. Modares, A. Karimpour, and M. B. Naghibi-Sistani, "Reinforcement Q-learning for optimal tracking control of linear discrete-time systems with unknown dynamics," *Automatica*, vol. 50, no. 4, pp. 1167–1175, Apr. 2014.

[39] S. Emara, F. Wang, B. Li, and T. Zeyl, "Pareto: Fair congestion control with online reinforcement learning," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 5, pp. 3731–3748, Sep. 2022.

[40] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Model-free Q-learning designs for linear discrete-time zero-sum games with application to H-infinity control," *Automatica*, vol. 43, no. 3, pp. 473–481, Mar. 2007.

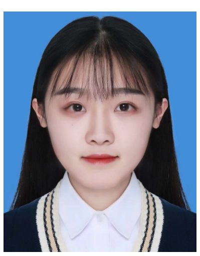
[41] C. Mu, Q. Zhao, C. Sun, and Z. Gao, "An ADDHP-based Q-learning algorithm for optimal tracking control of linear discrete-time systems with unknown dynamics," *Appl. Soft Comput.*, vol. 82, Sep. 2019, Art. no. 105593.

[42] X. Zhao, C. Liu, and E. Tian, "Finite-horizon tracking control for a class of stochastic systems subject to input constraints and hybrid cyber attacks," *ISA Trans.*, vol. 104, pp. 93–100, Sep. 2020.

[43] T. Li, X. Tang, J. Ge, and S. Fei, "Event-based fault-tolerant control for networked control systems applied to aircraft engine system," *Inf. Sci.*, vol. 512, pp. 1063–1077, Feb. 2020.

Jinliang Liu received the Ph.D. degree in control theory and control engineering from the School of Information Science and Technology, Donghua University, Shanghai, China, in 2011.

He was a Post-Doctoral Research Associate with the School of Automation, Southeast University, Nanjing, China, from 2013 to 2016. He was a Visiting Researcher/Scholar with the Department of Mechanical Engineering, The University of Hong Kong, Hong Kong, from 2016 to 2017. He was a Visiting Scholar with the Department of Electrical Engineering, Yeungnam University, Gyeongsan, South Korea, from 2017 to 2018. From 2011 to 2023, he was an Associate Professor and then a Professor with Nanjing University of Finance and Economics. In 2023, he joined Nanjing University of Information Science and Technology, Nanjing, where he is currently a Professor with the School of Computer Science. His research interests include networked control systems, complex dynamical networks, and time-delay systems.

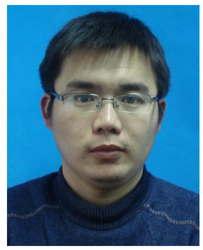


Yanhui Dong received the B.S. degree in computer science and technology from Nanjing University of Finance and Economics, Nanjing, China, in 2020, where she is currently pursuing the M.S. degree with the College of Information Engineering.

Her research interests include security control, reinforcement learning algorithms, and networked control systems.

Lijuan Zha received the Ph.D. degree in control science and engineering from Donghua University, Shanghai, China, in 2018.

From 2017 to 2024, she was a Lecturer and then an Associate Professor with the College of Information Engineering, Nanjing University of Finance and Economics. From 2018 to 2023, she was a Post-Doctoral Research Associate with the School of Mathematics, Southeast University. She is currently an Associate Professor with the School of Science, Nanjing Forestry University, Nanjing, China. Her current research interests include networked control systems, neural networks, and complex dynamical systems.



Xiangpeng Xie (Senior Member, IEEE) received the B.S. and Ph.D. degrees in engineering from Northeastern University, Shenyang, China, in 2004 and 2010, respectively.

From 2010 to 2014, he was a Senior Engineer with Metallurgical Corporation of China Ltd., Beijing, China. He is currently a Professor with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China. His research interests include fuzzy modeling and control synthesis, state estimations, optimization in process industries, and intelligent optimization algorithms.

Prof. Xie serves as an Associate Editor for IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, *International Journal of Fuzzy Systems*, and *International Journal of Control, Automation, and Systems*.

Engang Tian received the B.S. degree in mathematics from Shandong Normal University, Jinan, China, in 2002, the M.Sc. degree in operations research and cybernetics from Nanjing Normal University, Nanjing, China, in 2005, and the Ph.D. degree in control theory and control engineering from Donghua University, Shanghai, China, in 2008.

From 2011 to 2012, he was a Post-Doctoral Research Fellow with The Hong Kong Polytechnic University, Hong Kong. From 2015 to 2016, he was a Visiting Scholar with the Department of Information Systems and Computing, Brunel University London, Uxbridge, U.K. From 2008 to 2018, he was an Associate Professor and then a Professor with the School of Electrical and Automation Engineering, Nanjing Normal University. In 2018, he was appointed as an Eastern Scholar by the Municipal Commission of Education, Shanghai, and joined the University of Shanghai for Science and Technology, Shanghai, where he is currently a Professor with the School of Optical-Electrical and Computer Engineering. He has published more than 100 articles in refereed international journals. His research interests include networked control systems, cyber-attacks, and nonlinear stochastic control and filtering.