frontiers | Frontiers in Neurorobotics

TYPE Original Research
PUBLISHED 29 January 2025
DOI 10.3389/fnbot.2025.1549414

Check for updates

OPEN ACCESS

EDITED BY
Ming-Feng Ge,
China University of Geosciences Wuhan, China

REVIEWED BY
Xiao-Wei Zhang,
Tiangong University, China
Jian Han,
Ludong University, China

*CORRESPONDENCE
Zhijian Hu
✉ zhijian.hu@laas.fr

RECEIVED 21 December 2024
ACCEPTED 10 January 2025
PUBLISHED 29 January 2025

CITATION
Zhang K, Han K, Hu Z and Tan G (2025) Privacy-preserving ADP for secure tracking control of AVRs against unreliable communication.
Front. Neurorobot. 19:1549414.
doi: 10.3389/fnbot.2025.1549414

Privacy-preserving ADP for secure tracking control of AVRs against unreliable communication

Kun Zhang[1], Kezhen Han[2], Zhijian Hu[3]* and Guoqiang Tan[4]

[1]School of Astronautics, Beihang University, Beijing, China, [2]School of Electrical Engineering, University of Jinan, Jinan, China, [3]LAAS-CNRS, University of Toulouse, CNRS, Toulouse, France, [4]Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough, United Kingdom

In this study, we developed an encrypted guaranteed-cost tracking control scheme for autonomous vehicles or robots (AVRs), by using the adaptive dynamic programming technique. To construct the tracking dynamics under unreliable communication, the AVR's motion is analyzed. To mitigate information leakage and unauthorized access in vehicular network systems, an encrypted guaranteed-cost policy iteration algorithm is developed, incorporating encryption and decryption schemes between the vehicle and the cloud based on the tracking dynamics. Building on a simplified single-network framework, the Hamilton-Jacobi-Bellman equation is approximately solved, avoiding the complexity of dual-network structures and reducing the computational costs. The input-constrained issue is successfully handled using a non-quadratic value function. Furthermore, the approximate optimal control is verified to stabilize the tracking system. A case study involving an AVR system validates the effectiveness and practicality of the proposed algorithm.

KEYWORDS

adaptive dynamic programming, encryption and decryption, tracking control, optimal control, autonomous vehicle

1 Introduction

Autonomous vehicles or robots (AVRs) have rapidly transformed from a futuristic concept to a tangible reality, driving significant advancements in automotive technology. The advancement of autonomous vehicle technology has increasingly focused on improving tracking control systems, which are crucial for effective vehicle guidance (Pan et al., 2023). However, a persistent issue is the unreliable communication between a local vehicle and a reference vehicle, leading to discrepancies in signal reception and affecting tracking precision. In addition to these developments, the emergence of connected vehicles (Li et al., 2019a; Liu et al., 2023b), which leverages cloud computing for data processing and optimization, presents both opportunities and challenges. These systems function as cyber–physical systems (He et al., 2014; Zhang et al., 2014; Mohan et al., 2020), integrating computational and physical processes to enhance real-time data exchange and improve overall traffic management (Jiang et al., 2022; Li et al., 2019b). However, during communication between the vehicle and the cloud, the network's homogeneous and civilian nature makes it, particularly, vulnerable to attacks. This vulnerability, especially in the absence of robust security protocols, exposes these systems to cyber threats, including eavesdropping.

To enhance the security of vehicular cyber-physical systems, researchers from various fields, such as communication, control systems, and information theory, have developed various strategies to address cyberattacks across different layers (Han et al., 2024; Deng and Wen, 2021; Liu et al., 2021, 2023a). Various types of attacks, including denial-of-service (DoS) attacks, false data injection (FDI) attacks, and replay attacks, have been extensively studied (Teixeira et al., 2012; Li et al., 2024; Hu et al., 2023). These types of attacks share the characteristic of being active strategies designed to disrupt system functionality or manipulate transmitted data. Although defense mechanisms have made progress in countering such threats, majority of the existing methods primarily concentrate on detecting and mitigating explicit attacks, often overlooking the fundamental challenge of ensuring communication security. In vehicular cybersecurity, one of the critical issues is the threat of eavesdropping attacks (Yang et al., 2020; Wu et al., 2022). Unlike the direct and active nature of DoS and FDI attacks, eavesdropping operates passively, enabling attackers to intercept sensitive information while remaining undetected. This makes it a significant long-term threat that can compromise communication confidentiality and can even enable more destructive attacks. Addressing this challenge requires advanced encryption and privacy-preserving techniques to ensure secure communication. Although these methods are effective, they do not ensure optimal control performance at minimal energy cost, as they do not incorporate the principles of optimal control.

Optimal tracking control has become a cornerstone of modern control theory, with adaptive dynamic programming (ADP) algorithms attracting considerable interest in recent years (Lu et al., 2020; Mu et al., 2017b). For non-linear optimal control problems, the principal challenge lies in solving the Hamilton-Jacobi-Bellman (HJB) equation—a problem that is nearly intractable through exact mathematical methods. ADP techniques have offered a promising alternative by leveraging neural networks (NNs) to approximate optimal solutions, leading to significant advancements across fields such as automatic control and artificial intelligence (Mu et al., 2017a; Guo et al., 2024). For example, El-Sousy et al. (2021) designed a three-network structure to approximate the solution of the HJB equation for permanent-magnet synchronous motor servo drives. Wang et al. (2020) proposed an dual-network to approximate local Q-functions and control policies, solving optimal consensus control for non-linear multiagent systems. Furthermore, ADP-based optimal tracking control has been widely investigated (Dong et al., 2022; Song et al., 2023), including efforts to address input-constrained systems (Yang et al., 2023; Zhang et al., 2018). However, conventional ADP approaches, particularly those employing actor-critic frameworks, are frequently hindered by significant approximation errors introduced during iterative processes and NN training, thereby restricting their practical applicability.

To address these challenges, researchers have proposed several single-network ADP methodologies designed to streamline system architectures and enhance computational efficiency in handling nonlinear systems (Xu et al., 2023; Chen et al., 2021; Zou and Zhang, 2023). Chen et al. (2021) developed an event-triggered optimal control scheme for a macro–micro stage system, using a single critic NN to solve the modified HJB equation. In Guo et al. (2024), a distributed control strategy for attitude-constrained quadrotor unmanned aerial vehicle is proposed based on a critic network. Among the core ADP algorithms, value iteration and policy iteration (PI) have been widely employed, demonstrating robust performance in numerous applications (Zhang et al., 2020; Lin et al., 2023). However, the two-stage iterative procedures inherent in these methods frequently involve information transmission, which makes them susceptible to interception by adversaries. This vulnerability necessitates additional security measures, thereby increasing computational complexity and further constraining their applicability to complex systems. Although efforts to streamline computational burdens by eliminating actor networks have yielded progress, current ADP methods still inadequately address essential issues such as input saturation and ensuring reliable system performance, leaving these critical areas as potential opportunities for future research.

Unlike the previous studies, this article proposes an encrypted guaranteed-cost tracking control scheme for input-constrained tracking system with unreliable communication, and the main contributions are summarized as follows:

1. This article introduces an encrypted guaranteed-cost tracking control scheme for AVRs under unreliable communication. Compared with existing works, this is the first attempt to integrate ADP with encryption techniques, addressing both control performance and information security challenges in vehicular networks.
2. The designed privacy-preserving control method introduces a strategy to address eavesdropping attacks in control systems. By applying consistent output masking and encryption mechanisms at both the vehicle side and the cloud side, sensitive data and critical control information are effectively protected from potential breaches. This integrated approach ensures secure data transmission while maintaining the integrity and privacy of the control system.
3. A single-network structure with enhanced computational efficiency is proposed to approximate the HJB equation. Compared to conventional dual-network designs, the single-network structure reduces computational complexity while maintaining theoretical guarantees on weight error convergence and system stability. Additionally, input saturation is explicitly addressed through the adoption of a nonlinear value function, further enhancing the robustness.

2 Preliminaries and problem formulation

Consider an AVR operating in the X-Y plane, the position and orientation of the vehicle's mass center are represented by a posture vector

$$\mathcal{Z} := \begin{bmatrix} x(t) \\ y(t) \\ \vartheta(t) \end{bmatrix},$$

where $x(t)$ and $y(t)$ denote the horizontal and vertical positions, respectively, and $\vartheta(t)$ denotes the heading direction measured counterclockwise from the X-axis. The vehicle's motion is governed by the following kinematic model:

$$\dot{\mathcal{Z}} = \mathcal{K}u(t) = \begin{bmatrix} \cos(\vartheta(t)) & \mathcal{Y}\sin(\vartheta(t)) \\ \sin(\vartheta(t)) & -\mathcal{Y}\cos(\vartheta(t)) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v(t) \\ w(t) \end{bmatrix}. \tag{1}$$

Here, $v(t)$ and $w(t)$ represent the vehicle's linear and rotational velocities, respectively, while $\mathcal{Y}$ is the distance between the vehicle's mass center and its drive axle; and $\mathcal{K}$ is the Jacobian matrix that links the control inputs to the vehicle's motion. So far, the control objectives are summarized in the following.

Control objective: For an AVR under unreliable communication, design an ADP-based robust optimal controller with secure information exchange to drive the vehicle along the target, such that the following objectives are achieved:

1) *Robust tracking control objective:* For an AVR, $\mathcal{Z}_c := [x(t); y(t); \vartheta(t)]$ to track the desired orbit $\mathcal{Z}_d := [x_d(t); y_d(t); \vartheta_d(t)]$ under malicious cyberattacks on the tracking process, as shown in Figure 1. Due to the occurrence of an attack, a small deviation arises between the received signal and the actual signal. This deviation, caused by malicious attacks, is defined as $\mathcal{Z}_a := [x_a(t); y_a(t); \vartheta_a(t)]$. We assume that $\mathcal{Z}_a$ and its derivative are bounded.

With the minor difference $\mathcal{Z}_a$ caused by unreliable communication, following the framework in Zhang et al. (2022), we derive the tracking error system as

$$\dot{\mathcal{Z}}_e = \begin{bmatrix} \cos(\vartheta_e(t))v_d(t) + y_e(t)w_d(t) - v_c(t) + \gamma_x(t) \\ \sin(\vartheta_e(t))v_d(t) - x_e(t)w_d(t) - w_c(t) + \gamma_y(t) \\ w_d(t) - w_c(t) + \gamma_\vartheta(t) \end{bmatrix}, \tag{2}$$

where $\mathcal{Z}_e := [x_e(t); y_e(t); \vartheta_e(t)]$ denotes the tracking error posture, $v_d(t)$ and $w_d(t)$ are the desired linear and rotational velocities, $v_c(t)$ and $w_c(t)$ are the control inputs of the vehicle, and $[\gamma_x(t); \gamma_y(t); \gamma_\vartheta(t)]$ captures the effect of cyberattacks on the received signals and given by

$$\begin{bmatrix} \gamma_x(t) \\ \gamma_y(t) \\ \gamma_\vartheta(t) \end{bmatrix} = \begin{bmatrix} \cos(\vartheta_c(t))\dot{x}_a + \sin(\vartheta_c(t))\dot{y}_a \\ -\sin(\vartheta_c(t))\dot{x}_a + \cos(\vartheta_c(t))\dot{y}_a \\ \dot{\vartheta}_a \end{bmatrix}.$$

This model describes the dynamic behavior of the tracking error in AVR control.

To facilitate system description and control implementation, let us consider that $z = [x_e(t); y_e(t); \vartheta_e(t)]$, $f(z) = [\cos\left(\vartheta_e(t)\right) v_d(t); \sin\left(\vartheta_e(t)\right) v_d(t); w_d(t)]$, $g(z) = [-1, y_e(t); 0, \mathcal{L} - x_e(t); 0, 1]$, and $u = [v_c(t), w_c(t)]$. The system (Equation 2) is rewritten as

$$\dot{z} = f(z) + g(z)u + \gamma, \tag{3}$$

where u is control input and satisfies the asymmetric constrained set $\mathfrak{O} = \left\{u||u| \leq \hbar\right\}$. To follow the conventional optimal tracking architecture, we can rewrite the reference trajectory as follows

$$\dot{z}_d = f_d(z_d) + g_d(z_d)u_d, \tag{4}$$

where u_d is the steady-state control input taking the following form

$$u_d = g_d^{-1}(z_d)(\dot{z}_d - f_d(z_d)), \tag{5}$$

where $g_d^{-1}(z_d)g_d(z_d) = I_n$, I_n denotes an $n \times n$ identify matrix.

Assumption 1. The unreliable communication $\gamma(t)$ is bounded by $\bar{\gamma}$, that is $\|\bar{\gamma}(t)\| \leq \bar{\gamma}$, where $\bar{\gamma}$ is positive constant.

2) *Prevent eavesdropping objective:* As shown in Figure 1, the cloud handles monitoring, scheduling, optimization, and computation tasks, while the local controller is responsible for distributing control signals, albeit with limited data storage and processing capabilities. The cyberattack considered here is eavesdropping, where unauthorized interception of data during transmission allows attackers to steal sensitive system information, such as real-time control signals and operational states. To mitigate these risks, encryption and decryption mechanisms are implemented to safeguard the confidentiality and integrity of the transmitted data, ensuring secure communication and enhancing the system's overall reliability.

3) *Optimal control objective:* Based on the optimal control strategy, the AVR can achieve a compromise between performance and cost when running along a target, such that

$$\begin{aligned} \min \quad & \mathcal{J}(z) = \int_t^\infty \gamma_1 \bar{\gamma}^2 + \mathcal{T}(z,u)\,\mathrm{d}s, \\ \text{s.t.} \quad & \dot{z} = f(z) + g(z)u, u \in \mathfrak{O}, \end{aligned} \tag{6}$$

where $\mathcal{T}(z,u) = z^T \mathcal{Q} z + \bar{\mathcal{U}}(\mu)$, which is the utility function with feedback control $\mu = u - u_d$, γ_1 is positive constant, $\mathcal{Q} = \mathcal{Q}^T > 0$, and $\bar{\mathcal{U}}(\cdot)$ is a positive definite non-quadratic integrand function.

3 Iterative algorithm design

In this section, based on the preceding analysis, the tracking problem is reformulated into a stabilization problem for the error dynamics. To address this, a cryptography-based controller is designed, which not only mitigates the impact of unreliable communication but also ensures the security of information transmission against eavesdropping.

3.1 Encryption and decryption algorithm design

To effectively counter eavesdropping attacks on data transmitted between the vehicle side and the cloud side, privacy-preserving rules are designed for both sides. The encryption and decryption formulas (Han et al., 2024) for each iteration are provided in the following.

1) *AVR to Cloud:*

Encryption process: At the vehicle side, the data z to be sent are extracted from Equation 3 and encrypted using Equation 7, resulting in the encrypted data z^r. This encrypted data are then transmitted to the cloud. The encryption formula is as follows:

$$\begin{cases} z^s = a(t)z + A\xi(t), & \text{(7a)} \\ a(t) = e^{\left(\delta_1 \sum \left\| \mathcal{V}^r(z)(t-1) \right\|_2^2\right)}, & \text{(7b)} \\ \xi(t) = \rho_1 e^{-(t \bmod \rho_2)}, & \text{(7c)} \end{cases}$$

where $a(t)$ and $\xi(t)$ are encryption operators, δ_1, ρ_1, and ρ_2 are constants, and A is the channel assignment matrices. To simplify

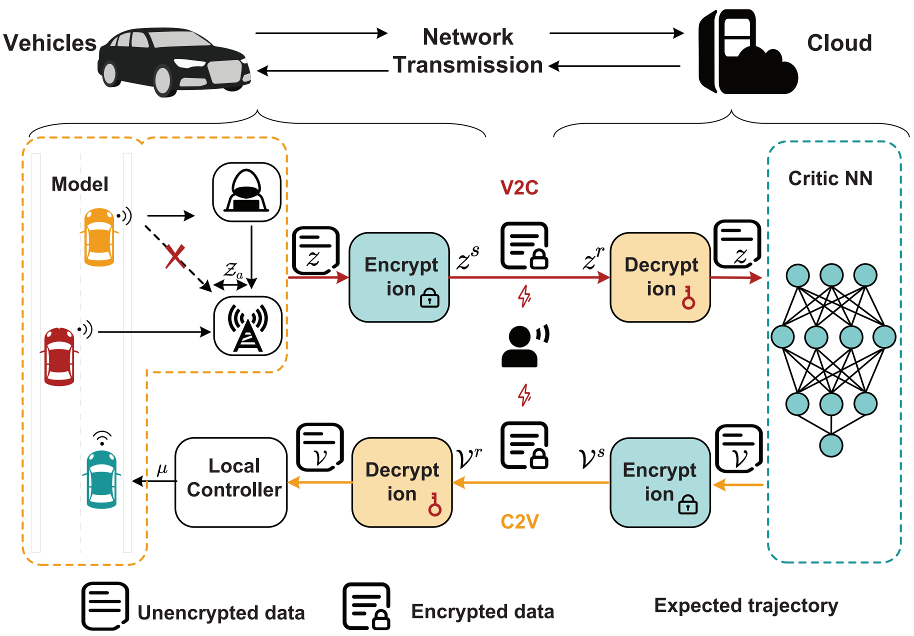

FIGURE 1
Proposed scheme for tracking process of AVRs.

the presentation of the method, it is assumed that $\mathcal{V}^s(z)(t-1)$ is already stored in the cloud. The value $\mathcal{V}(z)$ needs to be calculated on the cloud side. Its design is detailed in Section 3.2 and it serves as an essential component of the controller μ.

Decryption process: The cloud side receives the encrypted data z^r and decrypts it to recover the original data z. The decryption formula is as follows:

$$\begin{cases} z = \dfrac{z^r - A\xi(t)}{c(t)}, & \text{(8a)} \\ c(t) = e^{\left(\delta_1 \sum \left\| \mathcal{V}^s(z)(t-1) \right\|_2^2\right)}, & \text{(8b)} \end{cases}$$

where $c(t)$ is the counterpart of $a(t)$. It is observed that the design forms of the encryption operators $a(t)$ and $\xi(t)$, and encrypted expressions are shared between the vehicle side and the cloud. Furthermore, the parameters A, δ_1, ρ_1, and ρ_2 are also shared.

2) *Cloud to AVR:*

Encryption process: After policy evaluation, the computed $\mathcal{V}(z)$ is encrypted using Equation 9 and sent back to the vehicle.

$$\begin{cases} \mathcal{V}^s(z) = b(t)\mathcal{V}(z) + B\zeta(t), & \text{(9a)} \\ b(t) = e^{\left(\delta_2 \sum \|z^r\|_2^2\right)}, & \text{(9b)} \\ \zeta(t) = \varrho_1 e^{-(t \bmod \varrho_2)}, & \text{(9c)} \end{cases}$$

where $b(t)$ and $\zeta(t)$ are encryption operators, δ_2, ϱ_1, and ϱ_2 are constants, and B is the channel assignment matrices.

Decryption process: At the vehicle side, the received encrypted data $\mathcal{V}^r(z)$ is decrypted using Equation 10 to recover $\mathcal{V}(z)$ for policy improvement.

$$\begin{cases} \mathcal{V}(z) = \dfrac{\mathcal{V}^r(z) - B\zeta(t)}{d(t)}, & \text{(10a)} \\ d(t) = e^{\left(\delta_2 \sum \|z^s\|_2^2\right)}, & \text{(10b)} \end{cases}$$

where $d(t)$ is the counterpart of $b(t)$. Similarly, the design forms of the encryption operators $b(t)$ and $\zeta(t)$, and encrypted expressions are shared between the vehicle side and the cloud. Furthermore, the parameters B, δ_2, ϱ_1, and ϱ_2 are also shared. At this point, a complete iteration of privacy-preserving processing has been completed.

From the above encryption and decryption processes, it can be observed that the introduced masking signals $\xi(t)$ and $\zeta(t)$ and the encryption formula designs effectively ensure privacy during data transmission between the vehicle and the cloud. Notably, the data transmitted over the network are not the raw values z and $\mathcal{V}(z)$ but their encrypted counterparts, z^s, z^r, $\mathcal{V}^s(z)$, and $\mathcal{V}^r(z)$, which effectively prevent unauthorized entities from intercepting sensitive information.

3.2 Encrypted iterative algorithm design

The objective is to stabilize Equation 3 by constructing an encrypted iterative algorithm so that minimizing the performance index function, thereby reducing control costs and enhancing system security. Recalling Equation 6, the performance index is

$$\mathcal{V}(z) = \int_t^{\infty} (\gamma_1 \bar{\gamma}^2 + z^T \mathcal{Q} z + \bar{\mathcal{U}}(\mu))\, \mathrm{d}s, \tag{11}$$

where

$$\begin{aligned} \bar{\mathcal{U}}(\mu) &= \sum_{i=1}^{m} 2\theta_1 \int_0^{u_i - u_d} h^{-1}\left(\frac{\iota}{\theta_1}\right) r_i d\iota_i \\ &= 2\theta_1 \int_0^{u-u_d} h^{-1}\left(\frac{\iota}{\theta_1}\right) \mathcal{R} d\iota, \end{aligned} \tag{12}$$

where $\mathcal{R} = \mathrm{diag}\{[\mathrm{r}_1, ..., \mathrm{r}_m]\} > 0$, $\iota = [\iota_1, ..., \iota_m]^T$. The function $h(\cdot)$ is assumed to be a monotonic odd function satisfying $h(0) = 0$. For the purposes of this article, $h(\cdot)$ is specifically selected as $h(x) = \left(e^z - e^{-z}\right) / \left(e^z + e^{-z}\right)$.

According to the optimal control theory, Equation 11 is a Lyapunov function for the Equation 3 and the Hamiltonian function can be derived as

$$\mathcal{H}(z, \mu, \mathcal{V}(z)) = \gamma_1\bar{\gamma}^2 + z^T\mathcal{Q}z + \bar{\mathcal{U}}(\mu) + \nabla\mathcal{V}(z)\big(f(z) + g(z)u + \gamma\big), \tag{13}$$

with $\nabla\mathcal{V}(z) = \frac{\partial\mathcal{V}(z)}{\partial z}$. On defining $\mathcal{V}^*(z)$ as the minimum value of Equation 11, based on Bellman's principle of optimality, we have

$$\begin{aligned} 0 &= \mathcal{H}(z, \mu, \mathcal{V}^*(z)) \qquad (14)\\ &= \gamma_1\bar{\gamma}^2 + z^T\mathcal{Q}z + \bar{\mathcal{U}}(\mu) + \nabla\mathcal{V}^*(z)\big(f(z) + g(z)u^* + \gamma\big), \end{aligned}$$

and the optimal control u^* is obtained from $\frac{\partial\mathcal{H}(z,\mu,\mathcal{V}^*(z))}{\partial u^*} = 0$:

$$u^* = \theta_1 \tanh\left(-\frac{1}{2\theta_1}\mathcal{R}^{-1}g^T(z)\nabla\mathcal{V}^*(z)\right) + u_d. \tag{15}$$

Substituting Equation 15 into Equation 12 yields

$$\bar{\mathcal{U}}(\mu^*) = \nabla\mathcal{V}^{*T}(z)g(z)\tanh(\mathcal{D}(z)) + \theta_1^2\sum_{i=1}^{m}\ln\left(1 - \tanh^2(\mathcal{D}_i(z))\right), \tag{16}$$

where $\mathcal{D}(z) = \frac{1}{2\theta_1}\mathcal{R}^{-1}g^T(z)\nabla\mathcal{V}^*(z)$ and $\mu^* = u^* - u_d$. Then, the HJB equation can be derived as

$$\begin{aligned} \mathcal{H}(z, \mu^*, \mathcal{V}^*(z)) =& \gamma_1\bar{\gamma}^2 + z^T\mathcal{Q}z + \nabla\mathcal{V}^*(z)\big(f(z) + \gamma\big) \qquad (17)\\ &+ \theta_1^2\sum_{i=1}^{m}\ln\left(1 - \tanh^2(\mathcal{D}_i(z))\right) = 0. \end{aligned}$$

As highlighted in the preceding analysis, obtaining the optimal controller in Equation 15 necessitates solving the HJB Equation 17, a task well-known for its considerable computational and analytical challenges. To overcome this challenge, an iterative algorithm based on ADP is employed to obtain an approximate solution. The details of this iterative algorithm are presented in Algorithm 1.

```
Input: For iteration index $\varsigma = 1$, initial
       admissible control policy $u_1$, and
       computation precision $o$.
repeat
    Encryption process: State $z_\varsigma$ is encrypted into
    $z_\varsigma^s$ using Equation 7.
    Decryption process: The received data $z_\varsigma^r$ is
    decrypted into $z_\varsigma$ using Equation 8.
    Policy Evaluation: Solving the $\mathcal{V}_\varsigma(z)$ by

    $\gamma_1\bar{\gamma}^2 + z^T\mathcal{Q}z + \bar{\mathcal{U}}(\mu) + \nabla\mathcal{V}_\varsigma(z)(f(z) + g(z)u_\varsigma + \gamma) = 0.$   (18)

    Encryption process: $\mathcal{V}_\varsigma$ is encrypted into $\mathcal{V}_\varsigma^s$
    using Equation10.
    Decryption process: The received data $\mathcal{V}_\varsigma^r$ is
    decrypted into $\mathcal{V}_\varsigma$ using Equation 9.
    Policy Improvement: Updating $u_{\varsigma+1}$ as

    $u_{\varsigma+1} = -\theta_1 \tanh\left(\frac{1}{2\theta_1}\mathcal{R}^{-1}g^T(z)\nabla\mathcal{V}_\varsigma(z)\right) + u_d.$   (19)

    Set $\varsigma = \varsigma + 1$.
until $\|\mathcal{V}_{\varsigma+1}(z) - \mathcal{V}_\varsigma(z)\| \le o$;
return $u_\varsigma$, $\mathcal{V}_\varsigma(z)$.
```

Algorithm 1. Encrypted guaranteed cost policy iteration algorithm.

Lemma 1. By utilizing the encrypted PI process as described in Algorithm 1, which incorporates encryption and decryption steps for secure control of the tracking error dynamics in an AVR, the resulting control u_ς ensures the asymptotic stability of the system dynamics. Additionally, $\mathcal{V}_\varsigma(z)$ will converge to the optimal value function $\mathcal{V}^*(z)$ as $\varsigma \to \infty$, ensuring that u_ς converges to the optimal control u^*.

Proof. Initially, without iterations, the control u_1 is considered admissible. For $\forall u_\varsigma$ produced during iterations, consider the Lyapunov function $\mathcal{V}_\varsigma(z)$, which satisfies

$$\begin{aligned} \dot{\mathcal{V}}_\varsigma(z) &= \nabla\mathcal{V}_\varsigma(z)\dot{z}\\ &= \nabla\mathcal{V}_\varsigma(z)\big(f(z) + g(z)u_\varsigma + \gamma\big). \end{aligned} \tag{20}$$

According to HJB Equation 17, we can drive

$$\nabla\mathcal{V}_\varsigma(z)\big(f(z) + g(z)u_\varsigma + \gamma\big) = -\gamma_1\bar{\gamma}^2 - z^T\mathcal{Q}z - \bar{\mathcal{U}}(\mu_\varsigma), \tag{21}$$

where $\mu_\varsigma = u_\varsigma - u_d$. Then, substituting Equation 21 into Equation 22 yields

$$\dot{\mathcal{V}}_\varsigma(z) = -\gamma_1\bar{\gamma}^2 - z^T\mathcal{Q}z - \bar{\mathcal{U}}(\mu_\varsigma) \le 0. \tag{22}$$

Therefore, the iteration process ensures that the error dynamics remain asymptotically stable. Moreover, policy improvement is achieved by minimizing the associated value function, consistent with the Kleinman method, guaranteeing convergence. As the iteration count $\varsigma \to \infty$, $\mathcal{V}_\varsigma(z) \to \mathcal{V}^*(z)$, and $u_\varsigma \to u^*$ hold. This concludes the proof. □

Based on Lemma 1, the iterative process, enhanced with secure encryption and decryption, converges, leading to optimal control as the approximation errors diminish.

4 Critic neural network design

In this section, this study employs the fundamental update equations of PI to design a NN, utilizing the critic neural network (CNN) to approximate the solution of the HJB Equation 17 during each iteration step. Therefore, based on the universal approximation property of NNs, there exist ideal weights $\mathcal{W}^*$ such that the ideal value function can be approximated as

$$\mathcal{V}^*(z) = \mathcal{W}^{*T}\varphi(z) + \epsilon_1(z), \tag{23}$$

where $\varphi(z) \in \mathbb{R}^{\alpha}$ denotes activation functions and α is the number of neurons. Utilizing Equation 23, HJB Equation 17 becomes

$$\begin{aligned}\gamma_1 \bar{\gamma}^2 + z^T \mathcal{Q} z + (\mathcal{W}^{*T} \nabla \varphi(z) + \nabla \epsilon_1^T(z))\big(f(z) + \gamma\big) \\ + \theta_1^2 \sum_{i=1}^{m} \ln\left(1 - \tanh^2(\mathcal{H}_i(z))\right) = 0,\end{aligned} \tag{24}$$

where

$$\begin{aligned}\mathcal{H}_i(z) &= \mathcal{H}_{1i}(z) + \mathcal{H}_{2i}(z) \\ &= \frac{1}{2\theta}\mathcal{R}^{-1} g^T(z) \nabla \varphi^T(z) \mathcal{W}^* + \frac{1}{2\theta_1}\mathcal{R}^{-1} g^T(z) \nabla \epsilon_1^T(z),\end{aligned} \tag{25}$$

with $\nabla\varphi(z) = \frac{\partial \varphi_1}{\partial z}$ and $\nabla \epsilon_1(z) = \frac{\partial \varphi}{\partial z}$. Therefore, by defining residual error ϵ_H, Equation 24 can be rewritten as

$$\begin{aligned}\gamma_1 \bar{\gamma}^2 + z^T \mathcal{Q} z + \mathcal{W}^{*T} \nabla \varphi(z)\big(f(z) + \gamma\big) + \epsilon_H \\ + \theta_1^2 \sum_{i=1}^{m} \ln\left(1 - \tanh^2(\mathcal{H}_{1i}(z))\right) = 0,\end{aligned} \tag{26}$$

where

$$\begin{aligned}\epsilon_H \ &= \ \nabla \epsilon_1^T(z)\big(f(z) + \gamma\big) - \theta_1^2 \sum_{i=1}^{m} \frac{1}{\mathcal{O}_{1i}(z)} \tanh(\mathcal{O}_{2i}(z)) \\ &\quad \left(1 - \tanh^2(\mathcal{O}_{2i}(z))\right),\end{aligned} \tag{27}$$

with $\mathcal{O}_{1i}(z) \in [1 - \tanh^2(\mathcal{D}_i(z)), 1 - \tanh^2(\mathcal{H}_{1i}(z))]$, $\mathcal{O}_{2i}(z) \in [\mathcal{D}_i(z), \mathcal{H}_{1i}(z)]$. Note that if the number of hidden layer neurons α is sufficiently large, the residual error ϵ_H will approach zero. Based on the Lipschitz assumption of the system dynamics, this ϵ_H is bounded within a compact set, that is, $\|\epsilon_H\| \le \bar{\epsilon}_H$. Therefore, based on Equation 23 the ideal optimal control is

$$u^* = \theta_1 \tanh\left(-\frac{1}{2\theta_1}\mathcal{R}^{-1} g^T(z) \nabla \varphi^T \mathcal{W}^*\right) + u_d + \epsilon_2 \tag{28}$$

where $\epsilon_2 = -\frac{1}{2}\sum_{i=1}^{m}(1 - \tanh^2(\psi_i))\mathcal{R}^{-1} g^T(z) \nabla \epsilon_1$, $\psi_i \in [\mathcal{D}_i, \mathcal{H}_{1i}]$.

Since the ideal weight is unknown, the approximated value function is

$$\hat{\mathcal{V}}(z) = \hat{\mathcal{W}}^T \varphi(z), \tag{29}$$

where $\hat{\mathcal{W}}$ is approximated value of $\mathcal{W}^*$. Then, we can get

$$\hat{u} = -\theta_1 \tanh\left(\frac{1}{2\theta_1}\mathcal{R}^{-1} g^T(z) \nabla \varphi^T(z) \hat{\mathcal{W}}\right) + u_d. \tag{30}$$

Thus, approximated Hamiltonian function is

$$\begin{aligned}\mathcal{H}(z, \hat{\mu}, \hat{\mathcal{V}}(z)) \ &= \ \gamma_1 \bar{\gamma}^2 + z^T \mathcal{Q} z + \hat{\mathcal{W}}^T \nabla \varphi(z)\big(f(z) + \gamma\big) \\ &+ \ \theta_1^2 \sum_{i=1}^{m} \ln\left(1 - \tanh^2(\hat{\mathcal{H}}_{1i}(z))\right) \\ : &= \ \hat{\epsilon}_H,\end{aligned} \tag{31}$$

where $\hat{\epsilon}_H$ is the residual error due to NN approximation error.

Furthermore, let us consider $\mathcal{E} = \frac{1}{2}\hat{\epsilon}_H^T \hat{\epsilon}_H$, and to ensure that $\hat{\mathcal{W}}$ converge toward the optimal weights $\mathcal{W}^*$, the weight update formula (Zhang et al., 2018) is

$$\begin{aligned}\dot{\hat{\mathcal{W}}}_1 = &- \eta \frac{\tau}{\varpi^2}\hat{\epsilon}_H + \frac{\eta}{2}\kappa \nabla \varphi_1\big(g(I - \mathcal{M}(\hat{\mathcal{H}}_1))g^T\big)\nabla \mathcal{V}_a \\ &+ \eta\big(-\theta_1 \nabla \varphi g \mathcal{S}\frac{\tau^T}{\varpi}\hat{\mathcal{W}} - (\mathcal{K}_2 - \mathcal{K}_1 \tau^T)\hat{\mathcal{W}}\big),\end{aligned} \tag{32}$$

where η is learning rate, $\tau = \nabla\varphi(z)\big(f(z) + g(z)\hat{u} + \gamma\big)$, $\varpi = \tau^T \tau + 1$, and $\mathcal{K}_1$ and $\mathcal{K}_2$ are a tuning matrix. $\mathcal{M} = \mathrm{diag}\{\tanh^2(\hat{\mathcal{H}}_{1i})\}$, $\mathcal{S} = \mathrm{sgn}(\hat{\mathcal{H}}_1) - \tanh(\hat{\mathcal{H}}_1)$. Based on the Lemma 2 by Zhang et al. (2018), $\mathcal{V}_a$ denotes Lyapunov function, and if $\nabla \mathcal{V}_a\big(f(z) + g\hat{u} + \gamma\big) > 0$, then $\kappa = 0$, else $\kappa = 1$. Defining $\tilde{\mathcal{W}} = \mathcal{W}^* - \hat{\mathcal{W}}$, we obtain

$$\begin{aligned}\dot{\tilde{\mathcal{W}}} \ = \ &-\eta \tau \tau^T \tilde{\mathcal{W}} + \eta \frac{\tau}{\varpi}(\theta_1 \tilde{\mathcal{W}}^T \nabla \varphi g \mathcal{S} + \breve{\epsilon}_H) \\ &- \ \frac{\eta}{2}\kappa \nabla \varphi_1 g\big(I - \mathcal{M}(\hat{\mathcal{H}}_1)\big)g^T \nabla \mathcal{V}_a \\ &+ \ \eta \theta_1 \nabla \varphi g \mathcal{S}\frac{\tau^T}{\varpi}\hat{\mathcal{W}} + \eta(\mathcal{K}_2 - \mathcal{K}_1 \tau^T)\hat{\mathcal{W}},\end{aligned} \tag{33}$$

with $\breve{\epsilon}_H = \theta_1 \mathcal{W}^T \nabla \varphi g\big(\mathrm{sgn}(\mathcal{H}_1) - \mathrm{sgn}(\hat{\mathcal{H}}_1)\big) + 2\theta_1^2 \bar{\mathcal{H}} - \epsilon_H$, $\bar{\mathcal{H}} = \sum_{i=1}^{m} \ln \frac{1 + \exp(-2\mathcal{H}_{1i})}{1 + \exp(-2\hat{\mathcal{H}}_{1i})}$.

Theorem 1. For the optimal control policy described in Equation 30, the weight tuning law of the CNN is determined by the update formula provided in Equation 32. Under this design, the error dynamic system z and the weight errors $\tilde{\mathcal{W}}$ are uniformly ultimately bounded (UUB).

Proof. Define the Lyapunov function as $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$, where

$$\mathcal{L}_1 = \frac{1}{2}\tilde{\mathcal{W}}^T \eta^{-1} \tilde{\mathcal{W}}, \quad \mathcal{L}_2 = \mathcal{V}_a(z). \tag{34}$$

First, along Equation 33, the derivative of $\mathcal{L}_2$ is

$$\begin{aligned}\dot{\mathcal{L}}_1 =& \tilde{\mathcal{W}}^T \eta^{-1} \dot{\tilde{\mathcal{W}}}_1 \\ =& \tilde{\mathcal{W}}^T \eta^{-1}\Big\{ -\eta \tau \tau^T \tilde{\mathcal{W}} + \eta \frac{\tau}{\varpi}(\theta_1 \tilde{\mathcal{W}}^T \nabla \varphi g \mathcal{S} + \breve{\epsilon}_H) - \frac{\eta}{2}\kappa \nabla \varphi_1\big(g(I - \mathcal{M}(\hat{\mathcal{H}}_1))g^T\big)\nabla \mathcal{V}_a \\ &+ \eta \theta_1 \nabla \varphi g \mathcal{S}\frac{\tau^T}{\varpi}\hat{\mathcal{W}} + \eta(\mathcal{K}_2 - \mathcal{K}_1 \tau^T)\hat{\mathcal{W}}\Big\} \\ =& -\tilde{\mathcal{W}}^T \tau \tau^T \tilde{\mathcal{W}} + \theta_1 \tilde{\mathcal{W}}^T \nabla \varphi g \mathcal{S}\frac{\tau^T}{\varpi}\tilde{\mathcal{W}}^T + \breve{\epsilon}_H \frac{\tau^T}{\varpi}\tilde{\mathcal{W}}^T - \frac{1}{2}\kappa \nabla \mathcal{V}_a^T g\big(I - \mathcal{M}(\hat{\mathcal{H}}_1)\big)g^T \nabla \varphi^T \tilde{\mathcal{W}} \\ &+ \theta_1 \tilde{\mathcal{W}}^T \nabla \varphi g \mathcal{S}\frac{\tau^T}{\varpi}\mathcal{W}^* - \theta_1 \tilde{\mathcal{W}}^T \nabla \varphi g \mathcal{S}\frac{\tau^T}{\varpi}\tilde{\mathcal{W}} + \tilde{\mathcal{W}}^T(\mathcal{K}_2 - \mathcal{K}_1 \tau^T)\hat{\mathcal{W}} \\ =& -\tilde{\mathcal{W}}^T \tau \tau^T \tilde{\mathcal{W}} + \breve{\epsilon}_H \frac{\tau^T}{\varpi}\tilde{\mathcal{W}}^T - \frac{1}{2}\kappa \nabla \mathcal{V}_a^T g\big(I - \mathcal{M}(\hat{\mathcal{H}}_1)\big)g^T \nabla \varphi^T \tilde{\mathcal{W}} \\ &+ \theta_1 \tilde{\mathcal{W}}^T \nabla \varphi g \mathcal{S}\frac{\tau^T}{\varpi}\mathcal{W}^* + \tilde{\mathcal{W}}^T(\mathcal{K}_2 - \mathcal{K}_1 \tau^T)\hat{\mathcal{W}} \\ =& -\mathcal{P}^T \mathcal{A} \mathcal{P} + \mathcal{P}^T \mathcal{B} - \frac{1}{2}\kappa \nabla \mathcal{V}_a^T g\big(I - \mathcal{M}(\hat{\mathcal{H}}_1)\big)g^T \nabla \varphi^T \tilde{\mathcal{W}},\end{aligned} \tag{35}$$

where

$$\mathcal{P} = \begin{bmatrix} \tilde{\mathcal{W}}^T \tau \\ \tilde{\mathcal{W}}^T \end{bmatrix}, \mathcal{A} = \begin{bmatrix} I & -\frac{1}{2}\mathcal{K}_1^T \\ -\frac{1}{2}\mathcal{K}_1 & \mathcal{K}_2 \end{bmatrix},$$

$$\mathcal{B} = \begin{bmatrix} -\frac{1}{\varpi}\breve{\epsilon}_H \\ (\theta_1 \tilde{\mathcal{W}}^T \nabla \varphi g \mathcal{S}\frac{\tau^T}{\varpi} + \mathcal{K}_2 - \mathcal{K}_1 \tau^T)\mathcal{W}^* \end{bmatrix}.$$

Supposing that $\|\mathcal{W}^*\| \le \bar{\mathcal{W}}_1, \bar{\mathcal{W}}_1 > 0$, and due to $\breve{\epsilon}_H$ is bound, $\|\mathcal{B}\| \le \bar{\mathcal{B}}, \bar{\mathcal{B}} > 0$. Therefore, $\dot{\mathcal{L}}_1$ is

$$\dot{\mathcal{L}}_1 \le -\lambda_{\min}(\mathcal{A})\|\mathcal{P}\|^2 + \bar{\mathcal{B}}\|\mathcal{P}\| - \frac{1}{2}\kappa\nabla\mathcal{V}_a^T g\big(I - \mathcal{M}(\hat{\mathcal{H}}_1)\big)g^T\nabla\varphi^T\tilde{\mathcal{W}}. \tag{36}$$

Owing to κ of $\mathcal{L}_2$, $\dot{\mathcal{L}}$ is divided into two parts. For $\kappa = 0$, we have

$$\dot{\mathcal{L}} \le \nabla\mathcal{V}_a^T\dot{z} - \lambda_{\min}(\mathcal{A})\|\mathcal{P}\|^2 + \bar{\mathcal{B}}\|\mathcal{P}\|. \tag{37}$$

From a study by Rudin et al. (1964), we can know $\nabla\mathcal{V}_a^T\dot{z} < -\|\nabla\mathcal{V}_a\|z_m < 0$, $\|z\| \le z_m, z_m > 0$, thus, $\dot{\mathcal{L}}$ becomes

$$\dot{\mathcal{L}} \le -\|\nabla\mathcal{V}_a\|z_m - \lambda_{\min}(\mathcal{A})\left(\|\mathcal{P}\| - \frac{\bar{\mathcal{B}}}{2\lambda_{\min}(\mathcal{A})}\right)^2 + \frac{\bar{\mathcal{B}}^2}{4\lambda_{\min}(\mathcal{A})}.$$

Moreover, $\dot{\mathcal{L}} < 0$ if

$$\|\nabla\mathcal{V}_a\| > \frac{\bar{\mathcal{B}}^2}{4z_m\lambda_{\min}(\mathcal{A})}, \tag{38}$$

or

$$\|\mathcal{P}\| > \frac{\bar{\mathcal{B}}}{2\lambda_{\min}(\mathcal{A})}. \tag{39}$$

According to Equation 39, we can derive

$$\|\tilde{\mathcal{W}}\| > \frac{2\bar{\mathcal{B}}}{\sqrt{5}\lambda_{\min}(\mathcal{A})}. \tag{40}$$

For $\kappa = 1$, $\dot{\mathcal{L}}$ is

$$\begin{aligned}\dot{\mathcal{L}} &\le \nabla\mathcal{V}_a^T\dot{z} - \lambda_{\min}(\mathcal{A})\|\mathcal{P}\|^2 + \bar{\mathcal{B}}\|\mathcal{P}\| \\ &- \frac{1}{2}\kappa\nabla\mathcal{V}_a^T g\big(I - \mathcal{M}(\hat{\mathcal{H}}_1)\big)g^T\nabla\varphi^T\tilde{\mathcal{W}} \\ &\le \nabla\mathcal{V}_a^T(f + g\hat{u} + \gamma) - \lambda_{\min}(\mathcal{A})\|\mathcal{P}\|^2 + \bar{\mathcal{B}}\|\mathcal{P}\| \\ &- \frac{1}{2}\nabla\mathcal{V}_a^T g\big(I - \mathcal{M}(\hat{\mathcal{H}}_1)\big)g^T\nabla\varphi^T\tilde{\mathcal{W}}.\end{aligned} \tag{41}$$

Regarding $\tanh(\mathcal{H}_1) - \tanh(\hat{\mathcal{H}}_1) := \breve{\mathcal{H}}$, using the Taylor series, we know

$$\breve{\mathcal{H}} = \frac{1}{2\theta_1}\big(I - \mathcal{M}(\hat{\mathcal{H}}_1)\big)g^T\nabla\varphi^T\tilde{\mathcal{W}} + o\big((\mathcal{H}_1 - \hat{\mathcal{H}}_1)^2\big),$$

where $o\big((\mathcal{H}_1 - \hat{\mathcal{H}}_1)^2\big)$ is the higher order term and satisfies

$$\begin{aligned}&\|o\big((\mathcal{H}_1 - \hat{\mathcal{H}}_1)^2\big)\| \\ &\le \|\breve{\mathcal{H}}\| + \tfrac{1}{2\theta_1}\big(I - \mathcal{M}(\hat{\mathcal{H}}_1)\big)g^T\nabla\varphi^T\tilde{\mathcal{W}} \\ &\le \|\tanh(\mathcal{H}_1)\| + \|\tanh(\hat{\mathcal{H}}_1)\| + \tfrac{1}{2\theta_1}\big(I - \mathcal{M}(\hat{\mathcal{H}}_1)\big)g^T\nabla\varphi^T\tilde{\mathcal{W}}\end{aligned} \tag{42}$$

$$\begin{aligned}&= \big(\textstyle\sum_{i=1}^m|\tanh(\mathcal{H}_1)|^2\big)^{\frac{1}{2}} \\ &+ \big(\textstyle\sum_{i=1}^m|\tanh(\hat{\mathcal{H}}_1)|^2\big)^{\frac{1}{2}} + \tfrac{1}{2\theta_1}\big(I - \mathcal{M}(\hat{\mathcal{H}}_1)\big)g^T\nabla\varphi^T\tilde{\mathcal{W}} \\ &\le 2\sqrt{m} + \tfrac{1}{\theta_1}\bar{g}\bar{\varphi}\|\tilde{\mathcal{W}}\|,\end{aligned}$$

where $\|g\| \le \bar{g}, \bar{g} > 0$ and $\|\nabla\varphi\| \le \bar{\varphi}, \bar{\varphi} > 0$.

Recalling Equations 28–30, the term in Equation 41 with respect to $\nabla\mathcal{V}_a g$ can be written as

$$\begin{aligned}&\nabla\mathcal{V}_a^T\left(g\hat{u} - \frac{1}{2}g\big(I - \mathcal{M}(\hat{\mathcal{H}}_1)\big)g^T\nabla\varphi^T\tilde{\mathcal{W}}\right) \\ &= -\theta_1\nabla\mathcal{V}_a^T g\tanh(\mathcal{H}_1) + \theta_1\nabla\mathcal{V}_a^T go\big((\mathcal{H}_1 - \hat{\mathcal{H}}_1)^2\big) \\ &= \nabla\mathcal{V}_a^T gu^* - \nabla\mathcal{V}_a^T\epsilon_2 + \theta_1\nabla\mathcal{V}_a^T go\big((\mathcal{H}_1 - \hat{\mathcal{H}}_1)^2\big).\end{aligned} \tag{43}$$

Until now, Equation 41 can be rewritten as

$$\begin{aligned}\dot{\mathcal{L}} &\le \nabla\mathcal{V}_a^T(f + gu^* + \gamma) - \lambda_{\min}(\mathcal{A})\|\mathcal{P}\|^2 + \bar{\mathcal{B}}\|\mathcal{P}\| - \nabla\mathcal{V}_a^T\epsilon_2 + \theta_1\nabla\mathcal{V}_a^T go\big((\mathcal{H}_1 - \hat{\mathcal{H}}_1)^2\big) \\ &\le \|\nabla\mathcal{V}_a\|\|f + gu^*\| + \|\nabla\mathcal{V}_a\|\|\gamma\| - \lambda_{\min}(\mathcal{A})\|\mathcal{P}\|^2 + \bar{\mathcal{B}}\|\mathcal{P}\| - \nabla\mathcal{V}_a^T\epsilon_2 \\ &\quad + \theta_1\nabla\mathcal{V}_a^T go\big((\mathcal{H}_1 - \hat{\mathcal{H}}_1)^2\big) \\ &\le -\lambda_{\min}(\mathcal{A})\|\mathcal{P}\|^2 + \bar{\mathcal{B}}\|\mathcal{P}\| - \lambda_{\min}(\mathcal{C})\|\nabla\mathcal{V}_a\|^2 + \bar{\epsilon}_2\bar{g}\|\nabla\mathcal{V}_a\| + 2\theta_1\sqrt{m}\bar{g}\|\nabla\mathcal{V}_a\| \\ &\quad + \bar{\gamma}\|\nabla\mathcal{V}_a\| + \bar{g}^2\bar{\varphi}\|\nabla\mathcal{V}_a\|\|\tilde{\mathcal{W}}\| \\ &= -\lambda_{\min}(\mathcal{A})\|\mathcal{P}\|^2 + \bar{\mathcal{B}}\|\mathcal{P}\| - \lambda_{\min}(\mathcal{C})\|\nabla\mathcal{V}_a\|^2 + \omega\|\nabla\mathcal{V}_a\| + \bar{g}^2\bar{\varphi}\|\nabla\mathcal{V}_a\|\|\tilde{\mathcal{W}}\|,\end{aligned} \tag{44}$$

where $\|\epsilon_2\| \le \bar{\epsilon}_2, \bar{\epsilon}_2 > 0$. Let ℓ_1 and ℓ_2 satisfy $0 < \ell_1 < 1$, $0 < \ell_2 < 1$, and $\ell_1 + \ell_2 = 1$. Then, Equation 44 can be rewritten as

$$\begin{aligned}\dot{\mathcal{L}} &\le -\frac{4\ell_2\lambda_{\min}(\mathcal{C})\lambda_{\min}(\mathcal{A}) - \bar{g}^4\bar{\varphi}^2}{4\ell_2\lambda_{\min}(\mathcal{C})}\left(\|\mathcal{P}\| - \frac{2\ell_2\lambda_{\min}(\mathcal{C})\bar{\mathcal{B}}}{4\ell_2\lambda_{\min}(\mathcal{C})\lambda_{\min}(\mathcal{A}) - \bar{g}^4\bar{\varphi}^2}\right)^2 \\ &\quad - \ell_1\lambda_{\min}(\mathcal{C})\left(\|\nabla\mathcal{V}_a\| - \frac{\omega}{2\ell_1\lambda_{\min}(\mathcal{C})}\right)^2 - \ell_1\lambda_{\min}(\mathcal{C})\left(\|\nabla\mathcal{V}_a\| - \frac{\bar{g}^2\bar{\varphi}}{2\ell_1\lambda_{\min}(\mathcal{C})}\|\tilde{\mathcal{W}}\|\right)^2 \\ &\quad + \frac{\ell_2\lambda_{\min}(\mathcal{C})\bar{\mathcal{B}}^2}{4\ell_2\lambda_{\min}(\mathcal{C})\lambda_{\min}(\mathcal{A}) - \bar{g}^4\bar{\varphi}^2} + \frac{\omega^2}{4\ell_1\lambda_{\min}(\mathcal{C})} \\ &= -\frac{\omega_1}{4\ell_2\lambda_{\min}(\mathcal{C})}\left(\|\mathcal{P}\| - \frac{2\ell_2\lambda_{\min}(\mathcal{C})\bar{\mathcal{B}}}{\omega_1}\right)^2 - \ell_1\lambda_{\min}(\mathcal{C})\left(\|\nabla\mathcal{V}_a\| - \frac{\omega}{2\ell_1\lambda_{\min}(\mathcal{C})}\right)^2 \\ &\quad - \ell_1\lambda_{\min}(\mathcal{C})\left(\|\nabla\mathcal{V}_a\| - \frac{\bar{g}^2\bar{\varphi}}{2\ell_1\lambda_{\min}(\mathcal{C})}\|\tilde{\mathcal{W}}\|\right)^2 + \omega_2.\end{aligned}$$

Therefore, $\dot{\mathcal{L}} < 0$ if

$$\|\nabla\mathcal{V}_a\| > \frac{\bar{g}^2\bar{\varphi}}{2\ell_1\lambda_{\min}(\mathcal{C})} + \sqrt{\frac{\omega_1}{\ell_1\lambda_{\min}(\mathcal{C})}}, \tag{45}$$

or

$$\|\mathcal{P}\| > \frac{2\ell_2\lambda_{\min}(\mathcal{C})\bar{\mathcal{B}}}{\omega_1} + 2\sqrt{\frac{\ell_2\lambda_{\min}(\mathcal{C})\omega_2}{\omega_1}}. \tag{46}$$

Similar to Equation 40, we can derive

$$\|\tilde{\mathcal{W}}\| > \frac{4\ell_2\lambda_{\min}(\mathcal{C})\bar{\mathcal{B}}}{\sqrt{5}\omega_1} + 4\sqrt{\frac{\ell_2\lambda_{\min}(\mathcal{C})\omega_2}{5\omega_1}}. \tag{47}$$

By considering the two cases, $\kappa = 0$ and $\kappa = 1$, and based on the derived results as expressed in Equations 38–40 and Equations 45–47, we can conclude that the function $\nabla\mathcal{V}_a$ and the error weights $\tilde{\mathcal{W}}$ are UUB. Furthermore, knowing that $\mathcal{V}_a$ is in polynomial form, it follows that the error z is also UUB.

Remark 1. The algorithm designed in this article is depicted in Figure 2, where Algorithm 1 is implemented using a CNN. The CNN generates the estimated value function $\hat{\mathcal{V}}$, which is subsequently used to derive the approximated optimal control law $\hat{u}$ based on Equation 30. In contrast to the constrained optimal control designs presented in the studies by Zou and Zhang

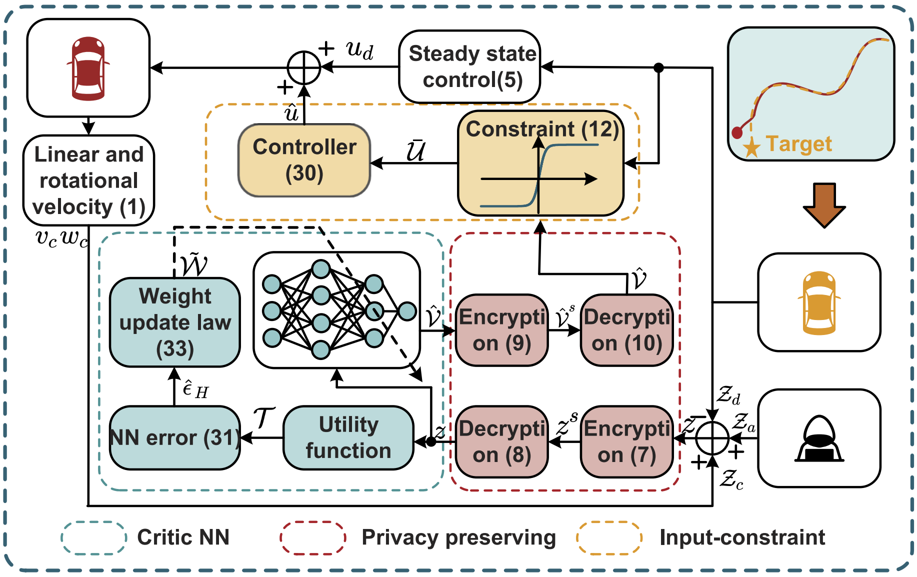

FIGURE 2
Illustration of tracking for AVRs subject to privacy protection.

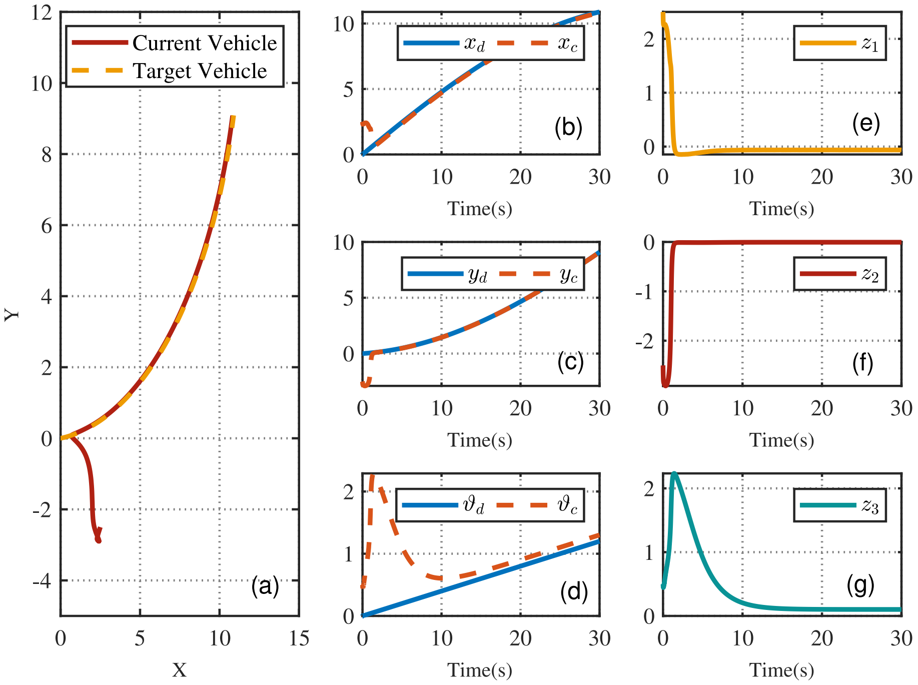

FIGURE 3
AVR driving trajectories. **(A)** The X-Y plot of tracking trajectories. **(B–D)** Tracking trajectories. **(E–G)** Tracking errors.

(2023); Chen et al. (2021), this work integrates privacy-preserving mechanisms during information transmission by leveraging encryption and decryption techniques. This incorporation not only safeguards data confidentiality but also enhances the overall security and reliability of the proposed algorithm.

5 Simulation results

To analyze the tracking performance of the AVR, we conduct simulations based on a predefined tracking error dynamic model. The tracking error dynamics $\dot{\mathcal{Z}}_e$ is modeled

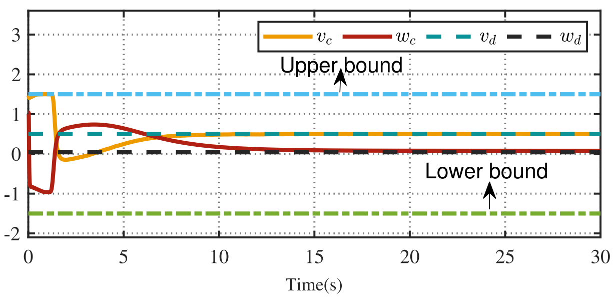

FIGURE 4
The constrained control input.

as

$$\begin{bmatrix}\dot{x}_e\\ \dot{y}_e\\ \dot{\vartheta}_e\end{bmatrix} = \begin{bmatrix}\cos(\vartheta_e)v_d\\ \sin(\vartheta_e)v_d\\ w_d\end{bmatrix} + \begin{bmatrix}-1 & y_e\\ 0 & \mathcal{Y}-x_e\\ 0 & -1\end{bmatrix}u+\gamma, \tag{48}$$

where $\mathcal{Y}$ represents the distance from the vehicle's center of mass to the rear axle, set to $\mathcal{Y} = -1.2\,\text{m}$ in this article. The desired reference trajectory is initialized with the state:

$$[x_d(0), y_d(0), \vartheta_d(0)]^T = [0, 0, 0]^T,$$

and the vehicle's trajectory is initialized with the state:

$$[x_c(0), y_c(0), \vartheta_c(0)]^T = [-2.5, 2.5, -0.5]^T.$$

Consequently, the initial value of error denotes

$$[x_e(0), y_e(0), \vartheta_e(0)]^T = [2.5, -2.5, 0.5]^T.$$

The reference trajectory's desired velocities are $v_d = 0.5$ and $w_d = 0.04$. Under the input constraints, $\hbar = 1.5$, meaning the constraint range is $[-1.5, 1.5]$. The unreliable communication γ is defined as

$$\gamma(t) = \sigma \begin{bmatrix}\sin(\sigma)x_e\\ \cos(\sigma)y_e\\ \sin(\sigma)x_e y_e\end{bmatrix},$$

where σ is a random variable uniformly distributed in $\sigma \in [-0.1, 0.1]$. For the performance evaluation, we define the cost function using the weighting matrices

$$\mathcal{Q} = \begin{bmatrix}10 & 0 & 0\\ 0 & 10 & 0\\ 0 & 0 & 10\end{bmatrix}, \quad \mathcal{R} = \begin{bmatrix}1.5 & 0\\ 0 & 0.5\end{bmatrix}.$$

The activation function vector of CNN is $\varphi(z) = \big[z_1^4, z_2^4, z_3^4, z_1^2z_2^2, z_2^2z_3^2, z_1^2z_3^2, z_1^2z_2, z_2^2z_3, z_1z_2^2, z_3^3, \sin(z_1), \sin(z_2), \sin(z_3), \cos(z_1), \cos(z_2), \cos(z_3)\big]^T$. $\rho_1 = 1.1$, $\rho_2 = 1.03$, $\varrho_1 = 3.2$, $\varrho_2 = 1.08$, $\delta_1 = 0.3 \times 10^{-5}$, $\delta_2 = 0.4 \times 10^{-2}$, $A = 1$, and $B = 1$.

Using the proposed method, Figure 3A illustrates the two-dimensional trajectory of the AVR. The vehicle quickly adjusts its direction and begins tracking the reference trajectory with good accuracy. After the initial phase, the vehicle follows the desired trajectory smoothly and closely. Figures 3B–G depict the tracking performance and error, demonstrating that the position error gradually reduces to zero, while the directional error also diminishes to zero, effectively ensuring precise position tracking throughout the process.

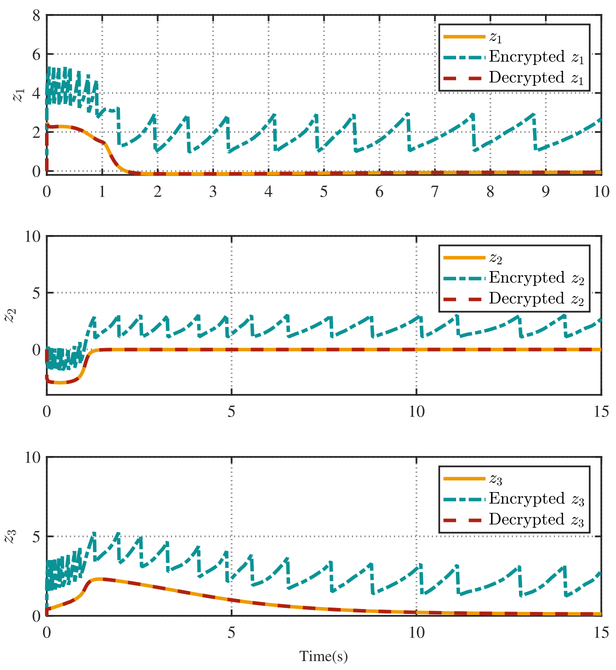

FIGURE 5
Encrypted error and decrypted error.

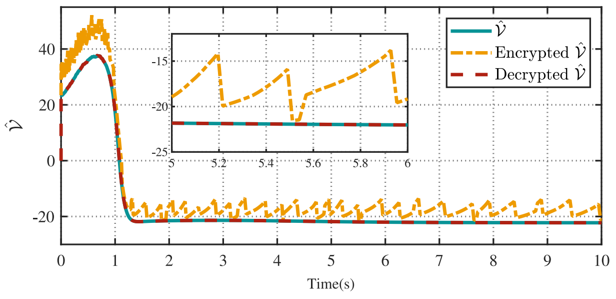

FIGURE 6
Encrypted value function and decrypted value function.

Figure 4 displays the evolution of the designed controller during the vehicle's tracking process. The dashed lines indicate the upper and lower bounds of the input constraints, which are set to $[-1.5, 1.5]$. The privacy-preserving characteristics of the proposed scheme are illustrated in Figure 5. It is evident that masking the vehicle-side output z effectively safeguards its privacy from potential attackers. Meanwhile, as shown in Figure 6, masking on the cloud side further prevents the leakage of critical information related to the designed control strategy. Therefore, these results ensure robust privacy protection during data transmission.

6 Conclusion

This study develops an encrypted guaranteed-cost tracking control scheme to address the challenges of information security and computational efficiency in AVR systems using the adaptive dynamic programming technique. By leveraging ADP and integrating encryption mechanisms between the vehicle and the cloud, the proposed method ensures stable tracking performance under unreliable communication. The input constraints are successfully managed using a nonlinear value function, while the CNN facilitates an efficient solution to the HJB equation. Simulation results from a case study confirm the stability and effectiveness of the designed algorithm, demonstrating its potential for real-world applications in AVR networks. Future work will focus on ensuring the security of cloud-based computations by processing encrypted data, further enhancing the safety and reliability of cloud operations in vehicular network systems.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

KZ: Conceptualization, Methodology, Writing – original draft. KH: Methodology, Writing – review & editing. ZH: Conceptualization, Supervision, Writing – review & editing. GT: Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work is supported by Beijing Nova Program (20240484516), the Fundamental Research Funds for the Central Universities (KG16314701), and Beihang World TOP University Cooperation Program.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Chen, X., Chen, X., Bai, W., and Guo, Z. (2021). Event-triggered optimal control for macro–micro composite stage system via single-network ADP method. *IEEE Trans. Indust. Elect.* 68, 4190–4198. doi: 10.1109/TIE.2020.2984462

Deng, C., and Wen, C. (2021). Mas-based distributed resilient control for a class of cyber-physical systems with communication delays under dos attacks. *IEEE Trans. Cybern.* 51, 2347–2358. doi: 10.1109/TCYB.2020.2972686

Dong, H., Zhao, X., and Luo, B. (2022). Optimal tracking control for uncertain nonlinear systems with prescribed performance via critic-only ADP. *IEEE Trans. Syst. Man, Cybernet.: Syst.* 52, 561–573. doi: 10.1109/TSMC.2020.3003797

El-Sousy, F. F. M., Amin, M. M., and Al-Durra, A. (2021). Adaptive optimal tracking control via actor-critic-identifier based adaptive dynamic programming for permanent-magnet synchronous motor drive system. *IEEE Trans. Ind. Appl.* 57, 6577–6591. doi: 10.1109/TIA.2021.3110936

Guo, Z., Li, H., Ma, H., and Meng, W. (2024). Distributed optimal attitude synchronization control of multiple quavs via adaptive dynamic programming. *IEEE Trans. Neural Netw. Learn. Syst.* 35:8053–8063. doi: 10.1109/TNNLS.2022.3224029

Han, K., Zhang, K., Wang, Z.-P., and Su, R. (2024). Resilient predictive load frequency control of multi-area interconnected power systems with privacy preserving and active detection against stealthy cyber attacks. *IEEE Intern. Things J.* 7, 4387–4394. doi: 10.1109/JIOT.2024.3507291

He, W., Yan, G., and Xu, L. D. (2014). Developing vehicular data cloud services in the IoT environment. *IEEE Trans. Indust. Inform.* 10, 1587–1595. doi: 10.1109/TII.2014.2299233

Hu, S., Ge, X., Chen, X., and Yue, D. (2023). Resilient load frequency control of islanded ac microgrids under concurrent false data injection and denial-of-service attacks. *IEEE Trans. Smart Grid* 14, 690–700. doi: 10.1109/TSG.2022.3190680

Jiang, M., Wu, T., Wang, Z., Gong, Y., Zhang, L., and Liu, R. P. (2022). A multi-intersection vehicular cooperative control based on end-edge-cloud computing. *IEEE Trans. Vehicular Technol.* 71, 2459–2471. doi: 10.1109/TVT.2022.3143828

Li, Y., Tang, C., Li, K., He, X., Peeta, S., and Wang, Y. (2019a). Consensus-based cooperative control for multi-platoon under the connected vehicles environment. *IEEE Trans. Intellig. Transport. Syst.* 20, 2220–2229. doi: 10.1109/TITS.2018.2865575

Li, Y., Tang, C., Peeta, S., and Wang, Y. (2019b). Nonlinear consensus-based connected vehicle platoon control incorporating car-following interactions and heterogeneous time delays. *IEEE Trans. Intellig. Transport. Syst.* 20, 2209–2219. doi: 10.1109/TITS.2018.2865546

Li, Z., Shi, Y., Xu, S., Xu, H., and Dong, L. (2024). Distributed model predictive consensus of mass against false data injection attacks and denial-of-service attacks. *IEEE Trans. Automat. Contr.* 69, 5538–5545. doi: 10.1109/TAC.2024.3371895

Lin, Z., Ma, J., Duan, J., Li, S. E., Ma, H., Cheng, B., et al. (2023). Policy iteration based approximate dynamic programming toward autonomous driving in constrained dynamic environment. *IEEE Trans. Intellig. Transp. Syst.* 24, 5003–5013. doi: 10.1109/TITS.2023.3237568

Liu, K., Zhang, H., Zhang, Y., and Sun, C. (2023a). False data-injection attack detection in cyber–physical systems with unknown parameters: a deep reinforcement learning approach. *IEEE Trans. Cybern.* 53, 7115–7125. doi: 10.1109/TCYB.2022.3225236

Liu, R., Hao, F., and Yu, H. (2021). Optimal SINR-based dos attack scheduling for remote state estimation via adaptive dynamic programming approach. *IEEE Trans. Syst. Man, Cybernet.: Syst.* 51, 7622–7632. doi: 10.1109/TSMC.2020.2981478

Liu, T., Cui, L., Pang, B., and Jiang, Z.-P. (2023b). A unified framework for data-driven optimal control of connected vehicles in mixed traffic. *IEEE Trans. Intellig. Vehicl.* 8, 4131–4145. doi: 10.1109/TIV.2023.3287131

Lu, J., Wei, Q., and Wang, F.-Y. (2020). Parallel control for optimal tracking via adaptive dynamic programming. *IEEE/CAA J. Automat. Sinica* 7, 1662–1674. doi: 10.1109/JAS.2020.1003426

Mohan, A. M., Meskin, N., and Mehrjerdi, H. (2020). A comprehensive review of the cyber-attacks and cyber-security on load frequency control of power systems. *Energies* 13:15. doi: 10.3390/en13153860

Mu, C., Ni, Z., Sun, C., and He, H. (2017a). Air-breathing hypersonic vehicle tracking control based on adaptive dynamic programming. *IEEE Trans. Neural Netw. Learn. Syst*. 28, 584–598. doi: 10.1109/TNNLS.2016.2516948

Mu, C., Ni, Z., Sun, C., and He, H. (2017b). Data-driven tracking control with adaptive dynamic programming for a class of continuous-time nonlinear systems. *IEEE Trans. Cybern*. 47, 1460–1470. doi: 10.1109/TCYB.2016.2548941

Pan, H., Zhang, C., and Sun, W. (2023). Fault-tolerant multiplayer tracking control for autonomous vehicle via model-free adaptive dynamic programming. *IEEE Trans. Reliab*. 72, 1395–1406. doi: 10.1109/TR.2022.3208467

Rudin, W. (1964). *Principles of Mathematical Analysis, Volume 3*. New York: McGraw-Hill.

Song, S., Gong, D., Zhu, M., Zhao, Y., and Huang, C. (2023). Data-driven optimal tracking control for discrete-time nonlinear systems with unknown dynamics using deterministic adp. *IEEE Trans. Neural Netw. Learn. Syst*. 36, 1184–1198. doi: 10.1109/TNNLS.2023.3323142

Teixeira, A., Pérez, D., Sandberg, H., and Johansson, K. H. (2012). “Attack models and scenarios for networked control systems,” in *Proceedings of the 1st International Conference on High Confidence Networked Systems* (New York, NY: Association for Computing Machinery), 55–64.

Wang, W., Chen, X., Fu, H., and Wu, M. (2020). Model-free distributed consensus control based on actor–critic framework for discrete-time nonlinear multiagent systems. *IEEE Trans. Syst. Man, Cybernet.: Syst*. 50, 4123–4134. doi: 10.1109/TSMC.2018.2883801

Wu, H., Li, M., Gao, Q., Wei, Z., Zhang, N., and Tao, X. (2022). Eavesdropping and anti-eavesdropping game in uav wiretap system: A differential game approach. *IEEE Trans. Wireless Commun*. 21, 9906–9920. doi: 10.1109/TWC.2022.3180395

Xu, Y., Li, T., Yang, Y., Tong, S., and Chen, C. L. P. (2023). Simplified adp for event-triggered control of multiagent systems against fdi attacks. *IEEE Trans. Syst. Man, Cybernet.: Syst*. 53, 4672–4683. doi: 10.1109/TSMC.2023.3257031

Yang, W., Zheng, Z., Chen, G., Tang, Y., and Wang, X. (2020). Security analysis of a distributed networked system under eavesdropping attacks. *IEEE Trans. Circuits Systems II: Express Briefs* 67, 1254–1258. doi: 10.1109/TCSII.2019.2928558

Yang, X., Xu, M., and Wei, Q. (2023). Approximate dynamic programming for event-driven H_∞ constrained control. *IEEE Trans. Syst. Man, Cybernet.: Syst*. 53, 5922–5932. doi: 10.1109/TSMC.2023.3277737

Zhang, H., Qu, Q., Xiao, G., and Cui, Y. (2018). Optimal guaranteed cost sliding mode control for constrained-input nonlinear systems with matched and unmatched disturbances. *IEEE Trans. Neural Netw. Learn. Syst*. 29, 2112–2126. doi: 10.1109/TNNLS.2018.2791419

Zhang, K., Liang, X., Lu, R., and Shen, X. (2014). Sybil attacks and their defenses in the internet of things. *IEEE Intern. Things J*. 1, 372–383. doi: 10.1109/JIOT.2014.2344013

Zhang, K., Zhang, H., Cai, Y., and Su, R. (2020). Parallel optimal tracking control schemes for mode-dependent control of coupled markov jump systems via integral rl method. *IEEE Trans. Autom. Sci. Eng*. 17, 1332–1342. doi: 10.1109/TASE.2019.2948431

Zhang, K., Zhang, H., Xue, W., and Zhang, R. (2022). “A robust control scheme for autonomous vehicles path tracking under unreliable communication,” in *2022 IEEE 11th Data Driven Control and Learning Systems Conference (DDCLS)* (Chengdu: IEEE), 1413–1418. doi: 10.1109/DDCLS55054.2022.9858512

Zou, H., and Zhang, G. (2023). Dynamic event-triggered-based single-network adp optimal tracking control for the unknown nonlinear system with constrained input. *Neurocomputing* 518, 294–307. doi: 10.1016/j.neucom.2022.11.015