Consistent parameter identification for quantized Wiener systems under replay attacks: a data-flag fusion mechanism ⋆

Qingxiang Zhang [a], Jin Guo [a,b], Yanlong Zhao [c,d], Ji-Feng Zhang [c,e]

[a] *School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing* 100083*, P.R. China*

[b] *Key Laboratory of Knowledge Automation for Industrial Processes, Ministry of Education, Beijing* 100083*, P.R. China*

[c] *State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing* 100190*, P.R. China*

[d] *School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing* 100049*, P.R. China*

[e] *School of Automation and Electrical Engineering, Zhongyuan University of Technology, Zhengzhou* 450007*, P.R. China*

Abstract

This paper investigates the problem of parameter identification for quantized Wiener systems subject to replay attacks. Such attacks cause input-output timing misalignment, and the coupling between quantized measurements and system nonlinearities further complicates the design of identification algorithms. We propose a data-flag fusion mechanism based on binary stochastic flag that compensates for identification errors induced by timing misalignment by leveraging the statistical properties of the flag. Moreover, by exploiting the structural characteristics of the system and the statistical properties of the noise, the quantized observations and system nonlinearity are jointly formulated as a unified nonlinear equation set, whose solution enables the joint estimation of the attack strategy and system parameters. Theoretical analysis is conducted to establish the consistency and asymptotic normality of the estimators, and the optimal configuration of the flag parameter is formulated under the minimum variance criterion. The generation of stochastic flag that satisfies the required statistical properties is investigated, the proposed framework is extended to multi-threshold observations, and a robust adjustment scheme is introduced to handle extreme attack scenarios. All theorems and conclusions are validated through numerical simulations.

Key words: Wiener systems identification; Quantized input and observation; Data-flag fusion mechanism; Replay attacks; Cyber-physical systems.

1 Introduction

As a core enabling technology of the new wave of industrial transformation, Cyber-Physical Systems (CPSs) integrate information technology with the physical world through computation, communication, and control, forming an intelligent closed-loop system characterized by real-time perception, dynamic decision-making, and precise control (Pivoto et al., 2021). Driven by the global demand for digital transformation, CPSs are increasingly applied across both industrial and civilian domains, including smart manufacturing (Mustapha, 2025), smart cities (Rakha, 2024), and wise medical (Qu et al., 2024).

Motivation. Due to their strong reliance on communication networks, CPSs face unprecedented security risks when exposed to network unreliability and external attacks (Li and Ye, 2025). Among others, replay attacks, characterized by high stealth, low implementation cost, and significant disruptive potential, have been widely recognized in both academic research and real-world cases as a serious threat to the secure operation of CPSs (Liu et al., 2023; Mo and Sinopoli, 2009; Porter et al., 2021). In CPSs security research, accurate modeling and identification of system dynamics are fundamental for attack detection, anomaly diagnosis,

⋆ This research was supported by the National Natural Science Foundation of China (62173030, 62433020, 62573044, and T2293770). This paper was not presented at any conference. Corresponding author: Jin Guo.

Email addresses: zmaster1001@163.com (Qingxiang Zhang), guojin@ustb.edu.cn (Jin Guo), ylzhao@amss.ac.cn (Yanlong Zhao), jif@iss.ac.cn (Ji-Feng Zhang).

and performance recovery. Nonlinear Wiener systems are commonly adopted due to their strong modeling capacity and practical interpretability. However, most existing identification methods are developed under ideal communication assumptions and fail to account for the impact of communication insecurity. Moreover, due to the limited bandwidth and computational resources in CPSs, signal quantization is unavoidable in practice. Quantization reduces data precision and introduces randomness, which further complicates the identification problem (Guo et al., 2017; Guo et al., 2012). Replay attacks amplify this challenge by repeatedly injecting outdated data to mislead the decision makers, significantly increasing the risk of identification failure under quantized conditions. Therefore, investigating the identification of quantized Wiener systems under replay attacks poses not only theoretical challenges but also holds substantial engineering relevance. Motivated by these challenges, this paper investigates the identification problem of quantized Wiener systems under replay attacks. The proposed approach aims to enhance the security and controllability of CPSs under non-ideal communication conditions by providing both theoretical insight and practical solutions.

Related work. As a representative block-structured nonlinear system, Wiener systems have long been a key focus in system identification. When the structures are known, various parameter estimation methods have been developed, including least squares approaches (Nadi and Arefi, 2023), recursive algorithms (Naseri et al., 2022; Ozbot et al., 2023), and mixed time-frequency techniques (Shakib et al., 2022). For the unknown nonlinear part, modeling techniques such as Volterra series, kernel-based methods, orthogonal basis functions, and Gaussian process hyperparameters have been employed to enhance model expressiveness and identification accuracy (Bai, 2008; Kang et al., 2014; Nejib et al., 2016; Risuleo et al., 2019). With the advancement of artificial intelligence, nonlinear dynamic models based on neural network, such as the recurrent equilibrium network, subspace encoders, and deep neural networks, have been applied to Wiener system identification (Beintema et al., 2023; Pillonetto et al., 2025; Revay et al., 2024). However, most existing studies assume that output signals can be precisely obtained, overlooking the impact of communication constraints and network security.

Current research on quantized nonlinear system identification remains primarily theoretical. Zong et al. (2023) proposed a hybrid particle swarm gradient algorithm based on an auxiliary model for parameter estimation in dual-rate Hammerstein systems. Li et al. (2023) designed a parameter estimator for quantized Hammerstein systems using a constant filter and augmented parameter error data. Guo et al. (2017) employed the empirical measure method under persistency of excitation conditions to estimate parameters in quantized Wiener systems. Additionally, Cao et al. (2024) and Li et al. (2025) developed adaptive error self-learning estimator, respectively, for identifing quantized Wiener-Hammerstein systems. The practical applicability and robustness in complex scenarios are not considered.

In recent years, research on replay attacks has primarily focused on detection and defense mechanisms. The first category involves the use of additional information, such as timestamps (Farha et al., 2022; Jia et al., 2025; Liu et al., 2024) and random numbers (Huang et al., 2020), which aim to reveal the essence of replay attacks by exploiting temporal or numerical discrepancies. However, these methods typically require significant communication bandwidth. The second category introduces watermarking into control signals. This widely adopted defense mechanism can mitigate the impact of replay attacks but often degrades system performance (Fang et al., 2020; Fritz and Zhang, 2023; Liu et al., 2023,?; Mo and Sinopoli, 2009; Porter et al., 2021; Zhu and Martínez, 2014). The third category focuses on communication data design, aiming to enhance attack detection accuracy without compromising system performance. This includes encoding strategies (Song and Ye, 2023; Ye et al., 2019), data reconstruction (Ferrari and Teixeira, 2021; Li et al., 2023), or sending preset data (Guo et al., 2025), with unified attack inference at the receiver side. Other approaches include leveraging cryptographic techniques (Rasheed et al., 2024; Yu et al., 2025), and using delay-based communication strategies (Zhao et al., 2025). In contrast, research on system identification under network attacks remains relatively limited. As system identification forms the foundation for state estimation and controller design, it is imperative to further investigate the security of system identification under replay attacks.

Compared with denial-of-service attacks and false data injection attacks, replay attacks are not well investigated in terms of both estimation and control. In-depth research on replay attacks during system identification remains notably insufficient. Existing methods (Guo et al., 2025), which follow the anomaly detection designed for data tampering, fail to effectively address the core challenge of the temporal misalignment and often suffer from slow convergence. Furthermore, the replay attack strategy model, a high-dimensional probabilistic distribution vector, becomes extremely complex when coupled with system nonlinearities and quantization effects, creating an urgent need for a dedicated theoretical and algorithmic framework. An in-depth exploration precisely targeting the aforementioned gap and challenges is conducted in this paper.

Contributions. This paper focuses on the problem of parameter identification for quantized Wiener systems under replay attacks. To address the challenges arising from time misalignment caused by such attacks, as well as the coupling between quantized observations and system nonlinearities, a data-flag fusion transmission mech-

anism based on binary stochastic flag is proposed. By exploiting the statistical properties of the flag, the mechanism compensates for the estimation errors induced by time misalignment. Furthermore, by incorporating the structural characteristics of the system and the distribution of the noise, the quantized observations and system nonlinearities are formulated into a nonlinear equation set, whose solution enables the joint estimation of system parameters and the attack strategy. Moreover, the strong consistency and asymptotic normality of the estimators are theoretically analyzed, and an optimization problem for the design of flag parameters is formulated. A flag generation is developed to ensure the required statistical properties, and the proposed mechanism is extended to multi-threshold observation scenarios to enhance the adaptability. In addition, a robust countermeasure is designed to improve the algorithm's stability under extreme attack scenarios.

An algorithm framework based on system structure and intermediate variable estimation has been constructed to address the complex nonlinear coupling challenges arising from the transition from periodic input-linear systems to quantized input-nonlinear Wiener systems. The conditions for system identifiability under input excitation and parameter solvability are presented. Convergence of the algorithm against replay attacks is maintained. The proposed fusion mechanism ensures that all transmitted data simultaneously carries both system information and security information, thereby fully preserving the original input excitation characteristics of the design. The main innovations and contributions of this paper are summarized as follows.

- This paper addresses the problem of parameter identification for quantized nonlinear Wiener systems under replay attacks. In contrast to existing studies that focus on i) identification without attacks (Cao et al., 2024; Guo et al., 2017; Li et al., 2023, 2025; Zong et al., 2023), ii) non-quantized information studies (Fang et al., 2020; Fritz and Zhang, 2023; Liu et al., 2023; Zhu and Martínez, 2014), and iii) linear system identification (Guo et al., 2025), this work provides a systematic investigation into the identification of nonlinear systems under the combined effects of quantization and replay attacks.
- Compared with preset data schemes (Guo et al., 2025) and conventional timestamp and random number-based methods (Farha et al., 2022; Huang et al., 2020; Liu et al., 2024), this paper proposes a data-flag fusion mechanism based on binary stochastic flag with real-time and stochastic characteristics. This effectively overcomes the predictability and communication overhead issues of existing methods, and preserves both identification accuracy and the advantage of binary communication, while ensuring the robustness against replay attacks.
- The proposed mechanism and identification algorithm enable joint consistent estimation of both attack probabilities and system parameters. Theoretical properties of the estimators are analyzed, and an optimal configuration for the flag parameters is developed. A flag generation satisfying the required statistical properties is constructed, the mechanism is extended to multi-threshold quantization scenarios, and a robust adjustment scheme is proposed to handle extreme attack conditions.

Organization. The remainder of this paper is organized as follows. Section 2 introduces the identification framework for quantized Wiener systems under replay attacks. Section 3 analyzes the performance of the original and improved identification algorithm under replay attacks. Section 4 develops the defense mechanism and algorithm and analyzes their performance. Section 5 discusses several relevant technical issues. Section 6 provides numerical simulation results. Section 7 concludes this paper and outlines future research directions.

2 Problem formulation

In this work, we consider a Single-Input Single-Output (SISO) discrete-time Wiener system, where the linear dynamic component is a FIR system of order n_1, and the static nonlinear component consists of n_2 nonlinear basis functions. The system is described as follows.

$$\begin{cases} y_k &= \sum_{i=0}^{n_2} \eta_i f_i(x_k) + w_k, \\ x_k &= \theta_1 u_k + \theta_2 u_{k-1} + \cdots + \theta_{n_1} u_{k-n_1+1}, \end{cases} \tag{1}$$

where $\eta_0, f_0(\cdot) \equiv 1$, eliminating scale ambiguity caused by all free parameters ensures that the mapping from input-output data to system parameter combinations is unique; w_k denotes the system noise satisfying Assumption 2.1 below; u_k is the quantized input; x_k serves as an intermediate variable; y_k is the system output. The parameters to be identified are defined as $\theta = [\theta_1, \ldots, \theta_{n_1}]^T$ and $\eta = [\eta_1, \ldots, \eta_{n_2}]^T$ for the linear and nonlinear component, respectively. The superscript ^{T} indicates vector or matrix transposition. The binary-valued measurement output s_k^0 is generated through the indicator function.

$$s_k^0 = I_{\{y_k \le C\}} = \begin{cases} 1, \ y_k \le C; \\ 0, \ \text{others}, \end{cases} \tag{2}$$

where C denotes the threshold of the binary sensor. As illustrated in Fig. 1, s_k^0 is transmitted over an unsecured communication network to a remote data center, where the received data at time k is denoted as s_k.

Assumption 2.1 *The noise $\{w_k\}$ is composed of independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and variance σ^2. Its cumulative distribution function is denoted by $\Phi(\cdot)$.*

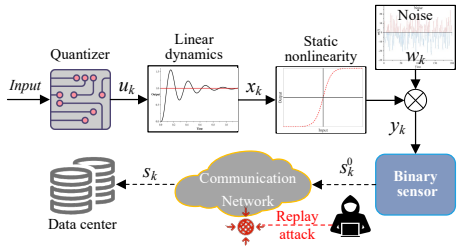

Fig. 1. System architecture diagram

Remark 2.1 *The noise can be relaxed to a ϕ-mixing process. For the unknown variance, it can be estimated by treating it as an unknown parameter (Wang et al., 2010).*

Building upon the fundamental replay attack model $s_k = s^0_{k-\delta}$ (Li et al., 2023; Mo and Sinopoli, 2009; Zhao et al., 2025), the replay delay or intensity δ is often bounded by a realistic upper limit. Given the established data association between sender and receiver in the communication network, the relationship between s_k and s^0_k at time k is formulated as follows.

$$s_k = s^0_{k-\delta_k}, \tag{3}$$
$$\Pr\left\{s_k = s^0_{k-\delta_k}\right\} = \lambda_{\delta_k}, \tag{4}$$
$$\delta_k \in \mathbb{U} = \{0, 1, \ldots, \mu\}, \tag{5}$$

where δ_k is a discrete integer-valued random variable representing the replay intensity at time k. $\delta_k \neq 0$ implies that a replay attack has been launched. μ defines the upper bound on δ_k. Denote $\Lambda = [\lambda_0, \lambda_1, \ldots, \lambda_\mu]^T$ as the probability vector of δ_k, satisfying $\mathbb{1}\Lambda = [1, 1, \ldots, 1]\Lambda = 1$. The above random replay attack strategy can thus be compactly characterized by the tuple (μ, Λ), determined by its intensity bound and probability vector.

To counter replay attacks, defenders must design effective countermeasures and develop consistent identification algorithms based on the available information, including u_k, the threshold, the noise distribution, and the received data s_k. The consistent goal is to ensure that, as the sample size tends to infinity, the estimates of the system parameters η and θ strongly converge to their true values. In what follows, we first analyze the performance of the identification algorithm under replay attacks. Then, we propose a defense mechanism to mitigate the impact of such attacks. Finally, we discuss extensions of the proposed mechanism.

3 Preliminaries

3.1 Original identification algorithms

Assume that the quantized input u_k can take a distinct values, i.e., $u_k \in \{r_1, r_2, \cdots, r_a\}$. Define the input regression pattern as $\pi_k = [u_k, u_{k-1}, \ldots, u_{k-n_1+1}]$, which has $h = a^{n_1}$ possible values, denoted by

$$\begin{cases} \tau_1 = [r_1, r_1, \cdots, r_1]_{1\times n_1}, \\ \tau_2 = [r_1, r_1, \cdots, r_2]_{1\times n_1}, \\ \qquad\vdots \\ \tau_h = [r_a, r_a, \cdots, r_a]_{1\times n_1}. \end{cases} \tag{6}$$

Assumption 3.1 *Persistent excitation condition. The input regression pattern π_k satisfies the persistent excitation requirement if there exists a strictly positive probability measure such that $p_l \triangleq \lim\limits_{N\to\infty} \frac{\sum_{k=1}^{N} I_{\{\pi_k=\tau_l\}}}{N}$, $l = 1, 2, \ldots, h$. Without loss of generality, assume that the set of persistently exciting patterns corresponds to indices $\iota \in \mathcal{H} = \{1, 2, \ldots, h_0\}$, where $n_1 + n_2 \leq h_0 \leq h$, ensuring the identifiability of the system.*

Remark 3.1 *The persistent excitation condition extends the classical full-rank requirement to quantized inpus by ensuring all regression patterns occur with positive probability. This probabilistic formulation directly guarantees the non-singularity of the information matrix through the full column rank of Ω_{h_0}.*

Let $\Gamma = [\Gamma_1, \ldots, \Gamma_{n_1}]^T$ denote a full-rank matrix composed of n_1 input regression patterns, where each $\Gamma_i \in \{\tau_j\}$ for $i = 1, \ldots, n_1$ and $j \in \mathcal{H}$. The set $\{\Gamma_1, \ldots, \Gamma_{n_1}\}$ is referred to as the basic persistently exciting pattern set. Define $\Upsilon = \Gamma\theta$. Set $X_1, X_2, \ldots, X_n$ be $n = n_1 + n_2$ unknown variables. Denote $\mathcal{X}_1 = [X_1, \ldots, X_{n_2}]^T$, $\mathcal{X}_2 = [X_{n_2+1}, \ldots, X_n]^T$. Consider the following equation set with $X_0 \equiv 1$.

$$\bar{\xi} = \begin{bmatrix} \bar{\xi}_1 \\ \bar{\xi}_2 \\ \vdots \\ \bar{\xi}_{h_0} \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n_2} X_i f_i(\tau_1\Gamma^{-1}\mathcal{X}_2) \\ \sum_{i=0}^{n_2} X_i f_i(\tau_2\Gamma^{-1}\mathcal{X}_2) \\ \vdots \\ \sum_{i=0}^{n_2} X_i f_i(\tau_{h_0}\Gamma^{-1}\mathcal{X}_2) \end{bmatrix}. \tag{7}$$

Assumption 3.2 *There exists a compact set $\Theta \subseteq \mathbb{R}^{h_0}$, such that ξ is an interior point of Θ, where*

$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_{h_0} \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n_2} \eta_i f_i(\tau_1\Gamma^{-1}\Upsilon) \\ \sum_{i=0}^{n_2} \eta_i f_i(\tau_2\Gamma^{-1}\Upsilon) \\ \vdots \\ \sum_{i=0}^{n_2} \eta_i f_i(\tau_{h_0}\Gamma^{-1}\Upsilon) \end{bmatrix}. \tag{8}$$

For $\forall\bar{\xi} \in \Theta$, (7) admits a unique solution, denoted by $[\mathcal{X}_1^T, \mathcal{X}_2^T]^T = \mathcal{L}(\bar{\xi})$, and $\mathcal{L}(\bar{\xi})$ is bounded and continuous at $\bar{\xi}$.

For convenience, we denote the solution in Assumption 3.2 as $\mathcal{X}_1 = \mathcal{L}_1(\bar{\xi})$, $\mathcal{X}_2 = \mathcal{L}_2(\bar{\xi})$. Consider the system (1)

with binary measurements (2). In the absence of replay attacks, under Assumptions 2.1, 3.1, and 3.2, the identification algorithm defined by (9)-(11) yields consistent estimates of the system parameters η and θ.

$$\nu_{N,\iota} = C - \mathcal{F}(\frac{1}{N_\iota}\sum_{k=1}^{N} s_k I_{\{\pi_k=\tau_\iota\}}), \tag{9}$$

$$\left[\widehat{\eta}_N^T, \widehat{\Upsilon}_N^T\right]^T = \pounds(\nu_N), \tag{10}$$

$$\widehat{\theta}_N = \Gamma^{-1}\widehat{\Upsilon}_N, \tag{11}$$

where $\nu_N = [\nu_{N,1}, \ldots, \nu_{N,h_0}]^T$; $\mathcal{F}(\cdot)$ denotes the inverse function of $\Phi(\cdot)$; $N_\iota = \sum_{k=1}^{N} I_{\{\pi_k=\tau_\iota\}}$ with $\sum_{\iota=1}^{h_0} N_\iota = N$, $\iota \in \mathcal{H}$; $\widehat{\eta}_N$ is the estimate of the nonlinear parameter η, and $\widehat{\Upsilon}_N$ is the prediction for Υ; $\widehat{\theta}_N$ denotes the estimate of the linear parameter θ.

A brief proof is provided for the design and analysis of subsequent algorithms. From (1), there is $y_k = \sum_{i=0}^{n_2}\eta_i f_i(\tau_k\theta) + w_k = \sum_{i=0}^{n_2}\eta_i f_i(\tau_k\Gamma^{-1}\Upsilon) + w_k$. The probability of event $s_k = 1$ in the absence of an attack is

$$\begin{aligned}\mathbb{E}\{s_k\} &= \Pr\{s_k = 1\}\\ &= \Pr\left\{w_k \le C - \sum_{i=0}^{n_2}\eta_i f_i(\tau_k\Gamma^{-1}\Upsilon)\right\}\\ &= \Phi(C - \sum_{i=0}^{n_2}\eta_i f_i(\tau_k\Gamma^{-1}\Upsilon))\\ &\triangleq \Phi_k,\end{aligned} \tag{12}$$

where $\mathbb{E}\{\cdot\}$ represents mathematic expectation. Due to $\pi_k \in \{\tau_1, \ldots, \tau_{h_0}\}$, we have

$$\mathbb{E}\{s_k\} \in \{\mathbb{E}\{s_k I_{\{\pi_k=\tau_1\}}\}, \ldots, \mathbb{E}\{s_k I_{\{\pi_k=\tau_{h_0}\}}\}\}. \tag{13}$$

From Law of Large Numbers, for each pattern τ_ι,

$$\frac{1}{N_\iota}\sum_{k=1}^{N} s_k I_{\{\pi_k=\tau_\iota\}} \to \Phi_\iota, \text{ w.p.1, as } N \to \infty. \tag{14}$$

Combining with (9), we obtain

$$\nu_{N,\iota} \to \sum_{i=0}^{n_2}\eta_i f_i(\tau_\iota\Gamma^{-1}\Upsilon) = \xi_\iota, \text{w.p.1, as } N \to \infty. \tag{15}$$

By (7), (8) and (10), $\widehat{\eta}_N = \pounds_1(\nu_N) \to \eta = \pounds_1(\xi)$ and $\widehat{\Upsilon}_N = \pounds_2(\nu_N) \to \Upsilon = \Gamma\theta = \pounds_2(\xi)$, w.p.1, as $N \to \infty$, which yields that $\widehat{\theta}_N = \Gamma^{-1}\widehat{\Upsilon}_N \to \Gamma^{-1}\Upsilon = \theta$.

Remark 3.2 *Assumption 2.1 is the standard condition for ensuring the asymptotic normality of the estimator. Assumption 3.1 ensures that the input signal can continuously and sufficiently excite all dynamic modes of the system. Assumption 3.2 is the core model-related condition guaranteeing global structural identifiability.*

Remark 3.3 *The algorithm* (10) *and* (11) *transforms the complex nonlinear and quantized coupling identification problem into a clear and theoretically verifiable two-stage process by introducing an intermediate variable.*

3.2 Algorithm analysis under replay attack

Lemma 3.1 *(Hall and Heyde, 1980) Consider a martingale difference sequence* $\{X_k, \mathscr{F}_k, k \ge 1\}$. *If* $\mathbb{E}^2\{\sum_{k=1}^{N} X_k\} < \infty$ *and* $\sum_{k=1}^{N}\frac{\mathbb{E}\{X_k^2\}}{k^2} < \infty$, *then* $\frac{1}{N}\sum_{k=1}^{N} X_k \to 0$, w.p.1, *as* $N \to \infty$.

Assumption 3.3 *Let* $\mathscr{F}$ *be a* σ*-algebra. The quantized input sequence* $\{u_k\}$ *is assumed to be an i.i.d. stochastic Process. Furthermore,* u_k *is measurable with respect to* $\mathscr{F}_{k-1} = \sigma\{w_i, \delta_i, i \le k-1\}$, *and satisfies* $|\mathbb{E}\{u_k\}| < \infty$.

Remark 3.4 *Assumption 3.3 differs from most studies on deterministic inputs. This ensures the convergence of identification algorithms under replay attacks, with martingale difference theory serving as proof tools.*

Due to the replay nature of the attack, the conditional probability relationships between regression patterns play a key role in analyzing the algorithm's performance. Define the matrix $\Psi^\delta \in \mathbb{R}^{h_0\times h_0}$, whose element $\psi_{i,j}^\delta = \Pr\{\pi_{k-\delta} = \omega_i | \pi_k = \omega_j\}$, $i, j \in \mathcal{H}$ and $\delta \in \mathbb{U}$. This matrix characterizes the influence of the replayed input pattern (delayed by δ steps) on the current pattern. It satisfies the following properties. (i) Each column of Ψ^δ sums to 1. $\sum_{i=1}^{h_0}\psi_{i,j}^\delta = 1$; (ii) For $\delta = 0$, the matrix reduces to the identity $\Psi^0 = \mathrm{I}_{h_0}$, where I_{h_0} denotes the h_0-dimensional identity matrix; (iii) Specifically, $\psi_{i,j}^0$ is the Kronecker delta, i.e., $\psi_{i,i}^0 = 1$, $\psi_{i,j}^0 = 0$ for $i \ne j$.

Theorem 3.1 *Consider system* (1) *and the binary measurement* (2). *Suppose the system is subjected to a random replay attack* (μ, Λ). *Under Assumptions 2.1, 3.1, 3.2, and 3.3, the parameter estimates generated by the identification algorithm* (9)-(11) *are convergent. However, the estimates do not converge to the true parameters* η *and* θ. *Specifically, we have*

$$\widehat{\eta}_N \to \pounds_1(\overline{\zeta}), \tag{16}$$

$$\widehat{\theta}_N \to \Gamma^{-1}\pounds_2(\overline{\zeta}), \tag{17}$$

$$\overline{\zeta} = [C - \mathcal{F}(\zeta_1), \ldots, C - \mathcal{F}(\zeta_{h_0})]^T, \tag{18}$$

$$\zeta_i = \sum_{\delta=0}^{\mu}\lambda_\delta\sum_{j=1}^{h_0}\psi_{j,i}^\delta\Phi_j, \quad i \in \mathcal{H}. \tag{19}$$

Proof. In the presence of the attack, from Assumption 3.3, u_k is $\mathscr{F}_{k-1}$-measurable. Considering the regression

property of the pattern π_k, and using (3), (4), (5), and (12), we obtain

$$\begin{aligned}
&\mathbb{E}\{s_k|\mathscr{F}_{k-1}\}\\
&=\Pr\{s_k=1\}\\
&=\sum_{\delta=0}^{\mu}\Pr\left\{s_k=s_{k-\delta}^0\right\}\Pr\left\{s_{k-\delta}^0=1\right\}\\
&=\sum_{\delta=0}^{\mu}\Pr\left\{s_k=s_{k-\delta}^0\right\}\sum_{j=1}^{h_0}\Pr\{\pi_{k-\delta}=\tau_j\}\Phi_j\\
&=\sum_{\delta=0}^{\mu}\Pr\left\{s_k=s_{k-\delta}^0\right\}\sum_{j=1}^{h_0}\Pr\{\pi_{k-\delta}=\tau_j|\pi_k=\tau_k\}\Phi_j\\
&=\sum_{\delta=0}^{\mu}\lambda_\delta\sum_{j=1}^{h_0}\psi_{j,k}^{\delta}\Phi_j. \qquad (20)
\end{aligned}$$

From (13), $\mathbb{E}\{s_k|\mathscr{F}_{k-1}\}\in\{\mathbb{E}\{s_kI_{\{\pi_k=\tau_1\}}|\mathscr{F}_{k-1}\},\ldots,\mathbb{E}\{s_kI_{\{\pi_k=\tau_{h_0}\}}|\mathscr{F}_{k-1}\}\}$. Let $\upsilon_k=(s_k-\zeta_k)I_{\{\pi_k=\tau_i\}}$. Then, by $\mathbb{E}\{s_k|\mathscr{F}_{k-1}\}=\zeta_k$, we have $\mathbb{E}\{\upsilon_k|\mathscr{F}_{k-1}\}=0$, implying that $\{\upsilon_k\}$ is a martingale difference sequence. Since $\upsilon_k\in(-1,1)$, $\mathbb{E}^2\{\sum_{k=1}^N\upsilon_k\}<\infty$ and $\sum_{k=1}^{\infty}\frac{\mathbb{E}\{\upsilon_k^2\}}{k^2}\le\sum_{k=1}^{\infty}\frac{1}{k^2}<\infty$. Then, by Lemma 3.1, $\frac{1}{N_i}\sum_{k=1}^N(s_k-\zeta_k)I_{\{\pi_k=\tau_i\}}\to 0$, as $N\to\infty$. This implies that as $N\to\infty$,

$$\begin{aligned}
&\frac{1}{N_i}\sum_{k=1}^{N}s_kI_{\{\pi_k=\tau_i\}}\\
&=\frac{1}{N_i}\sum_{k=1}^{N}(s_k-\zeta_k)I_{\{\pi_k=\tau_i\}}+\zeta_i\to\zeta_i. \qquad (21)
\end{aligned}$$

Combining with (9)-(11), the theorem is proved. □

Remark 3.5 *In this paper, replay attacks target the measurement data transmission process without affecting the input signal. Therefore, the persistent excitation condition remains valid.*

3.3 Improved identification algorithm with known attack strategy

Algorithms in (16)-(17) shows that although the original algorithm retains convergence under attack, it is inherently biased and cannot recover the true parameters without additional improving or attack detection mechanisms. We assume that the attack strategy is known, and propose an improved identification algorithm by constructing an attack matrix to achieve consistent parameter estimation despite the presence of attacks.

Let $\psi_{N,i,j}^{\delta}=\frac{\sum_{k=1}^N I_{\{\pi_{k-\delta}=\tau_i\}}I_{\{\pi_k=\tau_j\}}}{\sum_{k=1}^N I_{\{\pi_k=\tau_j\}}}$ denote the empirical frequency that, given $\pi_k=\tau_j$, the delayed input pattern $\pi_{k-\delta}=\tau_i$ occurs, based on N data samples. The matrix Ψ_N^{δ}, composed of elements $\psi_{N,i,j}^{\delta}$, corresponds to the theoretical matrix Ψ^{δ}. Using this, we construct the $h_0\times h_0$-dimension empirical attack strategy matrix $\Omega_N(\mu,\Lambda)=\sum_{\delta=0}^{\mu}\lambda_\delta(\Psi_N^{\delta})^T$. Let $\beta_N(\mu,\Lambda)=\Omega_N^{-1}(\mu,\Lambda)$ denote the inverse of this matrix, with elements $\beta_{N,i,j}(\mu,\Lambda)$, $i,j\in\mathcal{H}$. Then, by incorporating $\beta_{N,i,j}(\mu,\Lambda)$, we improve the original algorithm (9)-(11) and propose the compensation-based identification algorithm as follows, which is capable of achieving consistency.

$$\widehat{\eta}_N=\pounds_1(\mathscr{C}_N), \qquad (22)$$

$$\widehat{\theta}_N=\Gamma^{-1}\pounds_2(\mathscr{C}_N), \qquad (23)$$

$$\mathscr{C}_N=\left[C-\mathcal{F}(\varsigma_{N,1}),\ldots,C-\mathcal{F}(\varsigma_{N,h_0})\right]^T, \qquad (24)$$

$$\varsigma_{N,i}=\sum_{i=1}^{h_0}\beta_{N,i,j}(\mu,\Lambda)\frac{1}{N_j}\sum_{k=1}^{N}s_kI_{\{\pi_k=\tau_j\}},\ i\in\mathcal{H}. \qquad (25)$$

Theorem 3.2 *Under the condition of Theorem 3.1, for a known attack strategy (μ,Λ), if the inverse matrix $\beta(\mu,\Lambda)$ of $\Omega(\mu,\Lambda)$ exists, then the parameter estimate provided by the identification algorithm (22)-(25) is consistent, i.e., $\widehat{\eta}_N\to\eta$ and $\widehat{\theta}_N\to\theta$, w.p.1, as $N\to\infty$.*

Proof. According to the statistical properties, $\mathbb{E}\{\psi_{N,i,j}^{\delta}\}=\psi_{i,j}^{\delta}$. From Law of Large Numbers, $\Psi_N^{\delta}\to\Psi^{\delta}$, as $N\to\infty$. Likewise,

$$\begin{aligned}
&\Omega_N(\mu,\Lambda)\to\sum_{\delta=0}^{\mu}\lambda_\delta(\Psi^{\delta})^T\triangleq\Omega(\mu,\Lambda),\\
&\beta_N(\mu,\Lambda)\to\Omega^{-1}(\mu,\Lambda)\triangleq\beta(\mu,\Lambda). \qquad (26)
\end{aligned}$$

Under replay attacks, from (19) and (21), it follows that as $N\to\infty$,

$$\begin{bmatrix}\frac{1}{N_1}\sum_{k=1}^N s_kI_{\{\pi_k=\tau_1\}}\\ \vdots\\ \frac{1}{N_{h_0}}\sum_{k=1}^N s_kI_{\{\pi_k=\tau_{h_0}\}}\end{bmatrix}\to\begin{bmatrix}\zeta_1\\ \vdots\\ \zeta_{h_0}\end{bmatrix}=\Omega(\mu,\Lambda)\begin{bmatrix}\Phi_1\\ \vdots\\ \Phi_{h_0}\end{bmatrix}.$$

By the above, (25), and (26), as $N\to\infty$, we have

$$\begin{aligned}
\begin{bmatrix}\varsigma_{N,1}\\ \vdots\\ \varsigma_{N,h_0}\end{bmatrix}&=\beta_N(\mu,\Lambda)\begin{bmatrix}\frac{1}{N_1}\sum_{k=1}^N s_kI_{\{\pi_k=\tau_1\}}\\ \vdots\\ \frac{1}{N_{h_0}}\sum_{k=1}^N s_kI_{\{\pi_k=\tau_{h_0}\}}\end{bmatrix}\\
&\to\beta(\mu,\Lambda)\Omega(\mu,\Lambda)\begin{bmatrix}\Phi_1\\ \vdots\\ \Phi_{h_0}\end{bmatrix}=\begin{bmatrix}\Phi_1\\ \vdots\\ \Phi_{h_0}\end{bmatrix}. \qquad (27)
\end{aligned}$$

Finally, from (10), (11), (14), and (15), the consistency of the estimator follows, completing the proof. □

The execution of the compensation-based identification algorithm depends heavily on the non-singularity of $\Omega_N(\mu,\Lambda)$. This problem is also equivalent to the non-singularity of $\Omega(\mu,\Lambda)$. We give the following theorem.

Theorem 3.3 *Assuming that the regression pattern is connected and irreducible, the matrix* $\Omega(\mu,\Lambda)$ *is singular if and only if* $\lambda_0 = 0$.

Proof. Given that the regression pattern is connected and irreducible, for any $i,j \in \mathcal{H}$ and $\delta \geq 1$, $\psi_{i,j}^{\delta+1} = \sum_{l\in\mathcal{H}} \psi_{i,l}^{\delta}\psi_{l,j}^{1}$, which implies $\Psi^{\delta+1} = \Psi^{\delta}\Psi^{1} = (\Psi^{1})^{\delta+1}$.

At $\delta = n_1$, we have $\psi_{i,j}^{n_1} = \Pr\{\pi_{k-n_1} = \omega_i | \pi_k = \omega_j\} = \frac{\Pr\{\pi_{k-n_1}=\omega_i, \pi_k=\omega_j\}}{\Pr\{\pi_k=\omega_j\}}$. This can be decomposed as a product of input probabilities $\Pr\{u_{k-n_1} = \pi_{k-n_1,1}\}\cdots\Pr\{u_{k-2n_1+1} = \pi_{k-n_1,n_1}\}$. Similarly, at $\delta = n_1+1$,

$$\begin{aligned}\psi_{i,j}^{n_1+1} &= \Pr\{\pi_{k-n_1-1} = \omega_i | \pi_k = \omega_j\} \\ &= \Pr\{u_{k-n_1-1} = \pi_{k-n_1-1,1}\} \\ &\cdots \Pr\{u_{k-2n_1} = \pi_{k-n_1-1,n_1}\} I_{\{u_{k-n_1}\in\{r_1,\cdots,r_a\}\}}.\end{aligned}$$

Let ℓ_i denote an eigenvalue of Ψ^1. Then, $\ell_i^{n_1}$ and $\ell_i^{n_1+1}$ are eigenvalues of Ψ^{n_1} and Ψ^{n_1+1}, respectively. Since $\Psi^{n_1} = \Psi^{n_1+1}$, we have $\ell_i^{n_1} = \ell_i^{n_1+1}$, implying ℓ_i is 0 or 1. Given that the matrix order $h_0 > n_1$, there exists exactly one eigenvalue equal to 1 and the remaining $h_0 - 1$ eigenvalues are 0. From $\Omega^T(\mu,\Lambda) = \sum_{\delta=0}^{\mu}\lambda_\delta\Psi^\delta = \lambda_0 I_{h_0} + \lambda_1\Psi^1 + \cdots + \lambda_\mu\Psi^\mu$, it follows that $|\Omega(\mu,\Lambda)| = |\Omega^T(\mu,\Lambda)| = 1\cdot\underbrace{\lambda_0\cdots\lambda_0}_{h_0-1} = (\lambda_0)^{h_0-1}$. Therefore, $\Omega(\mu,\Lambda)$ is singular if and only if $\lambda_0 = 0$. □

In the case of extreme conditions $\lambda_0 = 0$, a transmission side adjustment scheme has been developed to ensure the feasibility of the identification algorithm. See Subsection 5.3 for details.

4 Defense mechanism and algorithm design

To achieve consistent parameter estimation, this section introduces a defense mechanism against replay attacks by a binary stochastic flag generator and preprocessing the transmitted data at the transmission-side.

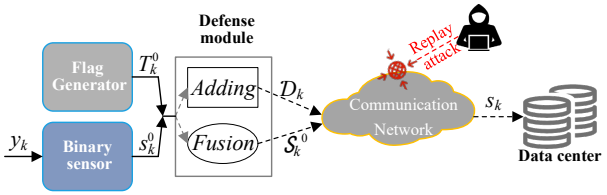

Fig. 2. Defense module diagram

As illustrated in Fig. 2, the proposed defense mechanism inserts a defense module before data transmission to process the original observation s_k^0. Two distinct processing strategies are adopted. The adding-based approach, which generates $\mathcal{D}_k$, is detailed in Subsection 4.1; The fusion-based approach, which generates $\mathcal{S}_k^0$, is presented in Subsection 4.2. Then, the processed data are transmitted to the data center for subsequent identification.

4.1 Sending mechanism based on adding flag

Under the adding-based approach, the attacker targets $\mathcal{D}_k$, such that

$$\begin{cases} s_k = \mathcal{D}_{k-\delta_k}, \\ \Pr\{s_k = \mathcal{D}_{k-\delta_k}\} = \lambda_{\delta_k}, \end{cases} \tag{28}$$

where $\delta_k \in \mathbb{U}$. $\mathcal{D}_k$ is constructed by concatenating a binary flag T_k^0 with the raw sensor measurement s_k^0, i.e., $\mathcal{D}_k = T_k^0 \oplus s_k^0 = T_k^0 | s_k^0$. At the data center, a separation operation is performed to recover the attacked flag T_k and sensor data z_k, denoted by $\ominus_0(s_k) = T_k$ and $\ominus_1(s_k) = z_k$, respectively. Accordingly, $T_k = T_{k-\delta_k}^0$ and $z_k = s_{k-\delta_k}^0$.

To estimate attack strategies, the amount of information carried by the defense design must be no less than the maximum offset μ plus one that an actual attack could potentially cause. This is sufficient to uniquely distinguish every possible timing misalignment pattern resulting from an attack (i.e., from no attack to the maximum delay of μ steps, totaling $\mu+1$ states). To simplify the analysis of the problem, we define $\overline{\mu} = \mu + 1$. In what follows, our goal is to estimate the attack probability Λ. In practice, even when μ is completely unknown, we can select sufficiently large $\overline{\mu}$ such that the designed mechanism still achieves consistent estimation, as demonstrated in the simulation results of Figs. 11 and 12.

A binary flag sequence is defined as a stochastic sequence with specific statistical properties that is actively generated by the sender and injected into measurement data. Its core function is to serve as a covert timing carrier, enabling the receiver to detect and compensate for timing inaccuracies caused by replay attacks within the data. The design of the flag sequence $\{T_k^0\}$ satisfies the following statistical properties. (i) Periodicity. The sequence follows a periodic stochastic structure with period $\overline{\mu}$; (ii) Binary stochastic feature. Within each period, the flag at position j is generated independently as a Bernoulli trial with parameter $G_j \in [0,1]$, i.e., $\Pr\{T_k^0 = 1\} = G_j$; (iii) Independence. Flag generation is statistically independent across positions within each period, and across periods for the same position. It is also independent for s_k^0.

These pre-designed statistical properties are the core enabling the defense mechanism. The designed identification algorithm, by comparing and utilizing changes in these statistical properties, constructs the estimation equations, thereby enabling the simultaneous resolution of both the attack strategy and the system parameters. According to the nature of the flag and replay attack, we establish the following linear equation set with linear constraints $0 \leq \lambda_0, \lambda_1, \ldots, \lambda_\mu \leq 1$ and $\mathbb{1}\Lambda = 1$.

$$\begin{cases} \sum_{\delta=0}^{\mu} \lambda_\delta G_{1-\delta} &= \rho_1(\mu, \Lambda) \\ \sum_{\delta=0}^{\mu} \lambda_\delta G_{2-\delta} &= \rho_2(\mu, \Lambda) \\ \qquad\qquad \vdots & \\ \sum_{\delta=0}^{\mu} \lambda_\delta G_{\overline{\mu}-\delta} &= \rho_{\overline{\mu}}(\mu, \Lambda) \end{cases}. \tag{29}$$

We can rewrite (29) as

$$\Pi \cdot \Lambda = [\rho_1, \rho_2, \ldots, \rho_{\overline{\mu}}]^T \triangleq \rho. \tag{30}$$

Π is a right circulant matrix with the first row element $G_1, G_{\overline{\mu}}, \ldots, G_2$, entirely determined by preset values. Guo et al. (2025) provided a sufficient but not necessary condition for the invertibility of Π, that is, $\overline{\mu}$ is a prime number and $G_i \neq G_j$ with $i \neq j$.

Remark 4.1 *Tagging/Labeling typically refers to metadata used to classify, identify, or authenticate data content or sources. Indicators are metrics reflecting faults or attacks. Watermarks are employed for attack detection purposes. The flag draws inspiration from the active watermark injection to restore temporal information.*

Based on the sending mechanism based on adding flag, and the "principle of necessary equivalence", the identification algorithm is designed as follows.

$$\widehat{\Lambda}_N = \Pi^{-1}\mathcal{L}_N, \tag{31}$$

$$\mathcal{L}_{N,j} = \frac{1}{L_{N,j}} \sum_{k=1}^{N} T_k I_{\{\mathrm{mod}(k-1,\overline{\mu})+1=j\}},\ j \in \mathcal{U}, \tag{32}$$

$$\widehat{\eta}_N = [\mathscr{Q}_{N,1}, \ldots, \mathscr{Q}_{N,n_2}]^T, \tag{33}$$

$$\widehat{\theta}_N = \Gamma^{-1}[\mathscr{Q}_{N,n_2+1}, \ldots, \mathscr{Q}_{N,n}]^T, \tag{34}$$

$$\mathscr{Q}_N = \mathscr{L}(\xi_N), \tag{35}$$

$$\xi_N = [C - \mathcal{F}(\varrho_{N,1}), \ldots, C - \mathcal{F}(\varrho_{N,h_0})]^T, \tag{36}$$

$$\varrho_{N,i} = \sum_{\iota=1}^{h_0} \beta_{N,i,\iota}(\widehat{\Lambda}_N) \frac{1}{N_\iota} \sum_{k=1}^{N} z_k I_{\{\pi_k=\tau_\iota\}},\ i \in \mathcal{H}, \tag{37}$$

where $\widehat{\Lambda}_N$ is the estimate for Λ; $\mathcal{L}_N = [\mathcal{L}_{N,1}, \ldots, \mathcal{L}_{N,\overline{\mu}}]^T$; $L_{N,j} = \sum_{k=1}^{N} I_{\{\mathrm{mod}(k-1,\overline{\mu})+1=j\}}$; $\mathrm{mod}(a_1, a_2)$ is remainder function, representing the remainder of a_1 divided by a_2; $\mathcal{U} = \{1, \ldots, \overline{\mu}\}$.

Remark 4.2 *The sender and receiver must reach consensus on the period and generation parameter G_j (not the specific flag data), as well as the fusion rules, before initiating communication.*

Theorem 4.1 *Under the condition of Theorem 3.1, for the unknown attack probability Λ, based on the sending mechanism of adding flag, the identification algorithm defined by* (31)-(37) *is consistent and yields $\widehat{\Lambda}_N \to \Lambda$, $\widehat{\eta}_N \to \eta$, and $\widehat{\theta}_N \to \theta$,* w.p.1, *as $N \to \infty$.*

Proof. According to (28), we have

$$\begin{aligned} \mathbb{E}\{T_k\} &= \Pr\{T_k = 1\} \\ &= \Pr\left\{T_k = T^0_{k-\delta_k}, T^0_{k-\delta_k} = 1\right\} \\ &= \sum_{\delta=0}^{\mu} \Pr\{\delta_k = \delta\} \Pr\left\{T^0_{k-\delta} = 1\right\} \\ &= \sum_{\delta=0}^{\mu} \lambda_\delta G_{k-\delta} \\ &\triangleq \rho_k. \end{aligned} \tag{38}$$

Due to the periodic nature of $\{T^0_k\}$, it follows that $\mathbb{E}\{T_k I_{\{\mathrm{mod}(k-1,\overline{\mu})+1=j\}}\} = \mathbb{E}\{T_j\} = \rho_j$. By Law of Large Numbers, $\mathcal{L}_{N,j} \to \mathbb{E}\{T_k I_{\{\mathrm{mod}(k-1,\overline{\mu})+1=j\}}\} = \rho_j$, w.p.1, as $N \to \infty$. Hence, there is $\widehat{\Lambda}_N = \Pi^{-1}\mathcal{L}_N \to \Pi^{-1}\rho = \Lambda$, w.p.1, as $N \to \infty$. The theorem is proved from the proof of Theorem 3.2. □

Traditional timestamps or random numbers-based mechanisms contain rich information (specific time values or numerical values), requiring high communication bandwidth and resolution for transmission and verification. The adding flag mechanism occupies only 1 bit of bandwidth and does not require complex data parsing. Although this mechanism has seen improvements in bandwidth aspects, it still incurs a certain communication cost.

4.2 Sending mechanism of data-flag fusion

To address bandwidth burden, we design a data-flag fusion mechanism. At time k, the sensor data and the flag to be transmitted are fused into a new data $\mathcal{S}^0_k$, which occupies only 1 bit. This design ensures compatibility with the data format of binary communication networks, thereby preserving communication efficiency and the benefits of binary signaling. As illustrated in Fig. 2, the defense module generates the fused data $\mathcal{S}^0_k$ through the fusion-based approach, defined by the rule

$$\mathcal{S}^0_k = I_{\{s^0_k \cdot T^0_k = 1\}} = \begin{cases} 1, & s^0_k = T^0_k = 1, \\ 0, & \text{others.} \end{cases} \tag{39}$$

Under random replay attacks, the data received by the data center at time k becomes $s_k = \mathcal{S}^0_{k-\delta_k}$, similar to

the form in (28). When $s_k = 1$, both $T^0_{k-\delta_k} = 1$ and $s^0_{k-\delta_k} = 1$ must hold.

$$\begin{cases} \sum_{\delta=0}^{\mu} \lambda_\delta G_{1-\delta} \sum_{\iota=1}^{h_0} \psi^\delta_{\iota,1} \Phi_\iota &= z_{1,1} \\ \sum_{\delta=0}^{\mu} \lambda_\delta G_{2-\delta} \sum_{\iota=1}^{h_0} \psi^\delta_{\iota,1} \Phi_\iota &= z_{1,2} \\ &\vdots \\ \sum_{\delta=0}^{\mu} \lambda_\delta G_{\overline{\mu}-\delta} \sum_{\iota=1}^{h_0} \psi^\delta_{\iota,1} \Phi_\iota &= z_{1,\overline{\mu}} \\ &\vdots \\ \sum_{\delta=0}^{\mu} \lambda_\delta G_{j-\delta} \sum_{\iota=1}^{h_0} \psi^\delta_{\iota,i} \Phi_\iota &= z_{i,j} \\ &\vdots \\ \sum_{\delta=0}^{\mu} \lambda_\delta G_{\overline{\mu}-\delta} \sum_{\iota=1}^{h_0} \psi^\delta_{\iota,h_0} \Phi_\iota &= z_{h_0,\overline{\mu}} \end{cases}. \tag{40}$$

For $h_0 + \overline{\mu}$ unknown variables Φ_i and λ_δ, $\delta \in \mathbb{U}$, $i \in \mathcal{H}$, $j \in \mathcal{U}$, we establish the equation set in (40). Under the condition that the equation has a solution, the procedure for computing the solution is given as follows.

Step 1. Expand the equation set with respect to subscript j, resulting in a vector equation for each j. $\sum_{\delta=0}^{\mu} \lambda_\delta G_{j-\delta} (\Psi^\delta)^T \Phi = z_j$, where $z_j = [z_{1,j}, \ldots, z_{h_0,j}]^T$; $\Phi = [\Phi_1, \ldots, \Phi_{h_0}]^T$.

Step 2. Left-multiply both sides of the above equation by $((\Psi^1)^T)^{n_1}$, yielding $\sum_{\delta=0}^{\mu} \lambda_\delta G_{j-\delta} ((\Psi^1)^T)^{n_1} (\Psi^\delta)^T \Phi = ((\Psi^1)^T)^{n_1} z_j$. By the result of Theorem 3.3, this can be simplified to $\sum_{\delta=0}^{\mu} \lambda_\delta G_{j-\delta} ((\Psi^1)^T)^{n_1} \Phi = ((\Psi^1)^T)^{n_1} z_j$.

Step 3. Let $\sum_{\delta=0}^{\mu} \lambda_\delta G_{1-\delta} = \kappa_1 = 1$. Then, $\sum_{\delta=0}^{\mu} \lambda_\delta G_{2-\delta} = ((\Psi^1)^T)^{n_1} z_2 / ((\Psi^1)^T)^{n_1} z_1 \triangleq \kappa_2$, …, $\sum_{\delta=0}^{\mu} \lambda_\delta G_{\overline{\mu}-\delta} = ((\Psi^1)^T)^{n_1} z_{\overline{\mu}} / ((\Psi^1)^T)^{n_1} z_1 \triangleq \kappa_{\overline{\mu}}$, where X_2/X_1 denotes the norm ratio of two vectors, i.e., $\frac{||X_2||}{||X_1||}$. According to (30), we obtain $[1, \kappa_2, \ldots, \kappa_{\overline{\mu}}]^T = \kappa\rho$.

Step 4. Output the solution.

$$\Lambda = \frac{\Pi^{-1}[\kappa_1, \ldots, \kappa_{\overline{\mu}}]^T}{\mathbb{1}\Pi^{-1}[\kappa_1, \ldots, \kappa_{\overline{\mu}}]^T}, \tag{41}$$

$$\Phi = \frac{1}{\overline{\mu}} \sum_{j=1}^{\overline{\mu}} (\sum_{\delta=0}^{\mu} \lambda_\delta G_{j-\delta} (\Psi^\delta)^T)^{-1} z_j. \tag{42}$$

Denote the solution to the equation set (40) obtained via Steps 1-4 as $(\Phi, \Lambda) = \Xi([z_{i,j}])$, where $[z_{i,j}]$ represents the column vector obtained by arranging $(i,j) = (1,1), (1,2), \ldots, (h_0, \overline{\mu})$.

The identification algorithm (43)-(48) comprises three components. First, both communication parties must pre-share the flag bit generation rules and fusion rules. Second, the receiver calculates empirical frequencies at different positions based on the received data, utilizing the flag bit index and input regressions. Third, based on (40), the attack strategy and system parameter estimation are computed according to Steps 1-4.

$$\varphi_{N,i,j} = \frac{1}{\acute{N}_{i,j}} \sum_{k=1}^{N} s_k I_{\{\pi_k = \tau_i\}} I_{\{\mathrm{mod}(k-1,\overline{\mu})+1=j\}}, \tag{43}$$

$$(\varkappa_N, \aleph_N) = \Xi([\varphi_{N,i,j}]), \tag{44}$$

$$\widehat{\Lambda}_N = \aleph_N, \tag{45}$$

$$\widehat{\eta}_N = \pounds_1(\varpi_N), \tag{46}$$

$$\widehat{\theta}_N = \Gamma^{-1} \pounds_2(\varpi_N), \tag{47}$$

$$\varpi_N = [C - \mathcal{F}(\varkappa_{N,1}), \ldots, C - \mathcal{F}(\varkappa_{N,h_0}]^T, \tag{48}$$

where $\acute{N}_{i,j} = \sum_{k=1}^{N} I_{\{\pi_k = \tau_i\}} I_{\{\mathrm{mod}(k-1,\overline{\mu})+1=j\}}$; $i \in \mathcal{H}$; $j \in \mathcal{U}$.

Theorem 4.2 *With the condition of Theorem 3.1, for the unknown Λ, using the data-flag fusion mechanism* (39) *and the identification algorithm* (43)-(48)*, the estimates are consistent.*

Proof. By (20), (38), and (39), it follows that

$$\begin{aligned} \mathbb{E}\{s_k | \mathscr{F}_{k-1}\} &= \Pr\{s_k = 1\} \\ &= \sum_{\delta=0}^{\mu} \Pr\left\{s_k = \mathcal{S}^0_{k-\delta}\right\} \Pr\left\{\mathcal{S}^0_{k-\delta} = 1\right\} \\ &= \sum_{\delta=0}^{\mu} \lambda_\delta \Pr\left\{T^0_{k-\delta} = 1\right\} \Pr\left\{s^0_{k-\delta} = 1\right\} \\ &= \sum_{\delta=0}^{\mu} \lambda_\delta G_{k-\delta} \sum_{\iota=1}^{h_0} \psi^\delta_{\iota,k} \Phi_\iota \\ &\triangleq \hbar_{k,k}. \end{aligned} \tag{49}$$

$\hbar_{i,j} = \sum_{\delta=0}^{\mu} \lambda_\delta G_{j-\delta} \sum_{\iota=1}^{h_0} \psi^\delta_{\iota,i} \Phi_\iota$. Since $\pi_k \in \{\tau_1, \ldots, \tau_{h_0}\}$, and the flag sequence is periodically generated, it follows that the process $(s_k - \hbar_{k,k}) I_{\{\pi_k = \tau_i\}} I_{\{\mathrm{mod}(k-1,\overline{\mu})+1=j\}}$ constitutes a martingale difference sequence. Hence,

$$\varphi_{N,i,j} \to \hbar_{i,j}, \text{ w.p.1, as } N \to \infty. \tag{50}$$

According to (40), replace $z_{i,j}$ with $\hbar_{i,j}$. Since Π is invertible, combining (30), we have $\Pi^{-1}[\kappa_1, \ldots, \kappa_{\overline{\mu}}]^T = \kappa\Pi^{-1}\rho = \kappa\Lambda$. As $\mathbb{1}\Lambda \equiv 1$, it holds that $\Lambda = \frac{\Pi^{-1}[\kappa_1, \ldots, \kappa_{\overline{\mu}}]^T}{\mathbb{1}\Pi^{-1}[\kappa_1, \ldots, \kappa_{\overline{\mu}}]^T}$. Therefore, we obtain $(\Phi, \Lambda) = \Xi([\hbar_{i,j}])$. From (50), we get

$$\begin{aligned} &(\varkappa_N, \aleph_N) = \Xi([\varphi_{N,i,j}]) \to (\Phi, \Lambda) = \Xi([\hbar_{i,j}]), \\ &\text{w.p.1, as } N \to \infty. \end{aligned} \tag{51}$$

Furthermore, from (9), (15), and (48), we have $\varpi_N \to \xi$, w.p.1, as $N \to \infty$. Finally, by (10) and (11), the theorem is proved. □

Remark 4.3 *The fusion mechanism can be applied almost directly to the Hammerstein system with a similar block structure. The primary modification lies in the unified nonlinear equation system* (8).

4.3 Algorithm performance analysis

Lemma 4.1 *(Kay, 1993) Let $X_k \sim \mathcal{N}(\mu_X, \Sigma_X)$ be a k-dimensional Gaussian random vector. Then, for any matrix $A \in \mathbb{R}^{m\times k}$ and $B \in \mathbb{R}^m$, the affine transformation $AX_k + B \sim \mathcal{N}(A\mu_X + B, A\Sigma_X A^T)$.*

Lemma 4.2 *(Chow and Teicher, 1997) Let $\{X_k\}$, $\{Y_k\}$, $\{Z_k\}$, and $\{W_k\}$ be sequences of random variables. Suppose $X_k \xrightarrow{d} X$, $Y_k \xrightarrow{d} Y$, $Z_k \xrightarrow{p} m$, and $W_k \xrightarrow{p} n$, where m and n are finite constants. Then, it follows that $W_kX_k + Y_k + Z_k \xrightarrow{d} nX + Y + m$, $k \to \infty$, where $\xrightarrow{d}$ denotes convergence in distribution and $\xrightarrow{p}$ denotes convergence in probability.*

Let $\mathrm{diag}(X_i)$ be the diagonal matrix obtained by arranging the elements X_i on the main diagonal in order of subscript i. Denote $\zeta = [\zeta_1, \ldots, \zeta_{h_0}]^T$, $\mathscr{C} = \beta(\Lambda)\zeta$,

$$\mathscr{D}(\xi) = [\pounds_1^d(\xi), \ldots, \pounds_{h_0}^d(\xi)]^T, \tag{52}$$

where $\pounds_i^d(\xi) = \frac{\partial \pounds_i(\xi)}{\partial \xi} = [\frac{\partial \pounds_i(\xi)}{\partial \xi_1}, \ldots, \frac{\partial \pounds_i(\xi)}{\partial \xi_{h_0}}]^T$, $i \in \mathcal{H}$. Similarly,

$$\mathscr{X}(\Lambda) = [\mathscr{C}_1^d(\Lambda), \ldots, \mathscr{C}_{h_0}^d(\Lambda)]^T, \tag{53}$$
$$\mathcal{J}(\hbar) = [\Xi_1^d(\hbar), \ldots, \Xi_{h_0+\overline{\mu}}^d(\hbar)]^T. \tag{54}$$

The asymptotic normality of the algorithm (31)-(37) is as follows.

Theorem 4.3 *Under the condition of Theorem 4.1, if the partial derivatives of $\pounds$ with respect to ξ exist, then as $N \to \infty$, the estimate $\mathscr{Q}_N$ given by* (31)-(37) *is asymptotically normal.*

$$\sqrt{N}(\mathscr{Q}_N - \mathscr{Q}) \xrightarrow{d} (0, \Sigma_0), \tag{55}$$

where $\mathscr{Q} = [\eta^T, \Upsilon^T]^T = \pounds(\xi)$; ξ is given by (8); $\Sigma_0 = \mathcal{R}\mathscr{X}(\Lambda)\Pi^{-1}\mathrm{diag}(\overline{\mu}(\rho_j - \rho_j^2))(\mathcal{R}\mathscr{X}(\Lambda)\Pi^{-1})^T + \mathcal{R}\beta\mathrm{diag}(\frac{\zeta_i - \zeta_i^2}{p_i})(\mathcal{R}\beta)^T$; $\mathcal{R} = \mathscr{D}(\xi)\mathrm{diag}(\mathscr{H}(\Phi_i))$; *$\mathscr{D}(\xi)$ and $\mathscr{X}(\Lambda)$ are given by* (52) *and* (53)*, respectively; ζ_i and ρ_j are defined in* (19) *and* (38)*, respectively; $\mathscr{H}(\cdot) = \frac{1}{\dot{\Phi}(x)}$ denotes the reciprocal of the derivative of $\Phi(\cdot)$, that is, the reciprocal of the noise probability density function; $p_i = \lim_{N\to\infty}\frac{N_i}{N}$; $i \in \mathcal{H}$, $j \in \mathcal{U}$.*

Proof. See Appendix A for details.

Theorem 4.4 *Under the condition of Theorem 4.2, if the partial derivative $\frac{\partial \Xi}{\partial \hbar}$ exists, then as $N \to \infty$, the estimate $\chi_N = [\varkappa_N^T, \aleph_N^T]^T$ obtained from* (43)-(44) *satisfies the following asymptotic normality.*

$$\sqrt{N}(\chi_N - \chi) \xrightarrow{d} (0, \Sigma_1), \tag{56}$$

where $\chi = [\Phi^T, \Lambda^T]^T$; $\Sigma_1 = \mathcal{J}(\hbar)\Sigma_2\mathcal{J}^T(\hbar)$; $\Sigma_2 = \mathrm{diag}(\frac{\hbar_\iota - \hbar_\iota^2}{\acute{p}_\iota})$; $\mathcal{J}(\hbar)$ is given in (54)*; $\acute{p}_{i,j} = \lim_{N\to\infty}\frac{\acute{N}_{i,j}}{N}$; $\hbar = [\hbar_{1,1}, \hbar_{1,2}, \ldots, \hbar_{i,j}, \ldots, \hbar_{h_0,\overline{\mu}}]^T$; $\iota = (i,j) = (1,1), (1,2), \ldots, (h_0, \overline{\mu})$.*

Proof. From (49) and (50), we get $\mathbb{E}\{(s_k - \mathbb{E}\{s_k\})^2\} = \hbar_{i,j} - \hbar_{i,j}^2$. Again, by (43), Central Limit Theorem, the independence of the noise and flag, and $\lim_{N\to\infty}\frac{\acute{N}_{i,j}}{N} = \acute{p}_{i,j}$, as $N \to \infty$, we have

$$\begin{bmatrix} \sqrt{\frac{N}{\acute{N}_{1,1}}}\sqrt{\acute{N}_{1,1}}(\varphi_{N,1,1} - \hbar_{1,1}) \\ \vdots \\ \sqrt{\frac{N}{\acute{N}_{h_0,\overline{\mu}}}}\sqrt{\acute{N}_{h_0,\overline{\mu}}}(\varphi_{N,h_0,\overline{\mu}} - \hbar_{h_0,\overline{\mu}}) \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma_2). \tag{57}$$

By (51) and Mean Value Theorem, there exists $\overline{\varphi}_{i,j}$ between $\hbar_{i,j}$ and $\varphi_{N,i,j}$ such that

$$\chi_N - \chi = \mathcal{J}(\overline{\varphi})(\varphi_N - \hbar), \tag{58}$$

where $\varphi_N = [\varphi_{N,1,1}, \varphi_{N,1,2}, \ldots, \varphi_{N,i,j}, \ldots, \varphi_{N,h_0,\overline{\mu}}]^T$; $\overline{\varphi}$ corresponds to φ_N and $\hbar$. With $N \to \infty$, there are $\overline{\varphi}_{i,j} \to \hbar_{i,j}$ such that according to (57), (58) and Lemma 4.1, (56) holds. □

According to (56), the covariance matrix

$$\Sigma_1(G) = \mathcal{J}(G)\mathrm{diag}(\frac{\hbar_{i,j}(G) - \hbar_{i,j}(G)^2}{\acute{p}_{i,j}})\mathcal{J}^T(G) \tag{59}$$

is highly dependent on the identification parameters $G = [G_1, G_2, \ldots, G_{\overline{\mu}}]^T$. Therefore, by adjusting G, the estimation error can be minimized. Furthermore, from (30) and (41), it can be observed that when estimating Λ via $\Pi(G)^{-1}\rho$, a large condition number of $\Pi(G)$ (when it is nearly singular) leads to numerical instability. In such cases, small perturbations in ρ may be significantly amplified during inversion, resulting in substantial estimation bias for Λ. Therefore, it is necessary to explicitly constrain the condition number of $\Pi(G)$, defined as $\mathrm{cond}(\Pi(G)) = \frac{\sigma_{\max}(\Pi(G))}{\sigma_{\min}(\Pi(G))}$, where $\sigma_{\max}$ and $\sigma_{\min}$ denote the largest and smallest singular values of $\Pi(G)$, respectively. Based on the above, we formulate the following constrained optimization problem for the estimates, aiming to minimize the trace of the covariance matrix

$\mathrm{TR}(\Sigma_1(G))$.

$$\min_{G=[G_1,G_2,\ldots,G_{\overline{\mu}}]^T\in[0,1]^{\overline{\mu}}} \mathrm{TR}(\Sigma_1(G)) \tag{60}$$

$$\text{s.t. } \mathrm{cond}(\Pi(G)) \le \mathrm{cond}_0, \tag{61}$$

where cond_0 is a predefined threshold, typically set to a small value (e.g., $\mathrm{cond}_0 = 10$) to ensure the numerical stability.

Due to the non-convexity of both the objective function and the constraint, obtaining the global optimum in closed form is intractable. Therefore, numerical methods are required to solve the problem. Two common strategies are as follows. (I). Grid search. Since $G_j \in [0,1]$, the interval can be uniformly discretized with step size $\Delta_j \in (0,1)$. The resulting Cartesian product forms a finite search space over which the optimal solution satisfying the constraint can be sought via exhaustive search (Liu et al., 2025). (II). Convex relaxation. The nonconvex problem can be relaxed into a semidefinite programming formulation, allowing for approximate global optimization using convex optimization techniques (Liu et al., 2025). A similar approach can be applied to Theorem 4.3; for brevity, it is omitted here.

Remark 4.4 *In addition to $\{G_j\}$, the system input configuration, quantization threshold selection, and flag period can all serve as system-level optimization metrics.*

5 Several technical issues

5.1 Extension of multi-threshold measurement

Compared with the limited information in binary measurements, multi-threshold measurements offer greater flexibility in handling complex environments, dynamic changes, and multi-objective requirements. Therefore, in this subsection, we extend the data-flag fusion mechanism to this scenario to enhance the adaptability.

The system structure in (1) remains unchanged. The binary measurement is generalized to a multi-threshold setting, where the output y_k is measured through a sensor with a finite number (m) of thresholds $-\infty = C_0 < C_1 < \cdots < C_m < C_{m+1} = +\infty$. The quantized output s_k^0 is then given by

$$s_k^0 = \sum_{q=1}^{m+1} \varepsilon_q I_{\{C_{q-1}<y_k\le C_q\}}, \tag{62}$$

where $s_k^0 \in \mathcal{E} = \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{m+1}\}$, and $\varepsilon_i < \varepsilon_j$ for $i < j$. The design rule for the flag remains the same as described in Section 4. The data-flag fusion mechanism is extended as follows.

$$\mathcal{S}_k^0 = \begin{cases} s_k^0, & T_k^0 = 1, \\ \varepsilon_{m+1}, & \text{others.} \end{cases} \tag{63}$$

The random replay attack strategy remains unchanged, and the data center receives $s_k = \mathcal{S}_{k-\delta_k}^0$ at time k. Upon receiving s_k, the data center performs preprocessing to obtain

$$\S(s_k) = [s_k^1, s_k^2, \ldots, s_k^m],\ s_k^i = I_{\{s_k\le\varepsilon_i\}}, \tag{64}$$

with $i = 1, 2, \ldots, m$. Based on the mechanism (63) and the preprocessing step (64), the following algorithm can be designed.

$$\widehat{\eta}_N = \pounds_1(\mathscr{L}_N), \tag{65}$$

$$\widehat{\theta}_N = \Gamma^{-1}\pounds_2(\mathscr{L}_N), \tag{66}$$

$$\mathscr{L}_N = [\mathcal{W}_N^1, \ldots, \mathcal{W}_N^m]_{h_0\times m}\mathscr{T}, \tag{67}$$

$$\mathcal{W}_N^\iota = \left[C_\iota - \mathcal{F}(\mathscr{A}_{N,1}^\iota), \ldots, C_\iota - \mathcal{F}(\mathscr{A}_{N,h_0}^\iota)\right]^T, \tag{68}$$

$$\widehat{\Lambda}_N = [\mathscr{B}_N^1, \ldots, \mathscr{B}_N^m]_{(\mu+1)\times m}\mathscr{T}, \tag{69}$$

$$(\mathscr{A}_N^\iota, \mathscr{B}_N^\iota) = \Xi([\gamma_{N,i,j}^\iota]), \tag{70}$$

$$\gamma_{N,i,j}^\iota = \frac{1}{\acute{N}_{i,j}} \sum_{k=1}^{N} s_k^\iota I_{\{\pi_k=\tau_i\}} I_{\{\mathrm{mod}(k-1,\overline{\mu})+1=j\}}, \tag{71}$$

where $\acute{N}_{i,j}$ is defined in (43); $i \in \mathcal{H}$, $j \in \mathcal{U}$, and $\iota \in \{1, 2, \ldots, m\}$; $\mathscr{T} = [\mathscr{T}_1, \ldots, \mathscr{T}_m]^T$ satisfies $\mathbb{1}\mathscr{T} = 1$, and each $\mathscr{T}_\iota \in [0,1]$. $\mathcal{W}_N^\iota$ and $\mathscr{B}_N^\iota$ represent the ι-th unbiased estimate of ξ and Λ, respectively. By choosing an appropriate weighting vector $\mathscr{T}$, a minimum-variance (or most efficient) estimate can be obtained. Therefore, $\mathscr{L}_N$ and $\widehat{\Lambda}_N$ are referred to as Quasi-convex combination estimators (Wang et al., 2010).

Theorem 5.1 *Consider system (1) under the multi-threshold measurement (62), subject to replay attacks with probability Λ. Under Assumptions 2.1, 3.1, 3.2, and 3.3, the proposed algorithm (65)-(71) enables the consistent estimation.*

Proof. $\Phi(C_\iota - \sum_{i=0}^{n_2}\eta_i f_i(\tau_k\Gamma^{-1}\Upsilon)) \triangleq \Phi_k^\iota$. From Assumption 3.3, (62), (63), and (64), we obtain

$$\begin{aligned}
&\mathbb{E}\{s_k^\iota|\mathscr{F}_{k-1}\} \\
&= \Pr\{s_k^\iota = 1\} \\
&= \sum_{\delta=0}^{\mu} \Pr\{s_k = \mathcal{S}_{k-\delta}^0\} \Pr\{T_{k-\delta}^0 = 1\} \Pr\{s_k^0 \le \varepsilon_i\} \\
&= \sum_{\delta=0}^{\mu} \lambda_\delta G_{k-\delta} \sum_{l=1}^{h_0} \psi_{l,k}^\delta \Phi(C_\iota - \sum_{i=0}^{n_2}\eta_i f_i(\tau_l\Gamma^{-1}\Upsilon)) \\
&= \sum_{\delta=0}^{\mu} \lambda_\delta G_{k-\delta} \sum_{l=1}^{h_0} \psi_{l,k}^\delta \Phi_l^\iota
\end{aligned}$$

$\triangleq \mathscr{P}^{\iota}_{k,k}$. (72)

$\mathscr{P}^{\iota}_{i,j} = \sum_{\delta=0}^{\mu} \lambda_\delta G_{j-\delta} \sum_{l=1}^{h_0} \psi^{\delta}_{l,i} \Phi^{\iota}_l$. It is evident that $(s^{\iota}_k - \mathscr{P}^{\iota}_{k,k}) I_{\{\pi_k=\tau_i\}} I_{\{\mathrm{mod}(k-1,\overline{\mu})+1=j\}}$ forms a martingale difference sequence, implying $\gamma^{\iota}_{N,i,j} \to \mathscr{P}^{\iota}_{i,j}$, as $N \to \infty$, according to Lemma 3.1. In the equation set (40), replacing Φ with $\Phi^{\iota} = [\Phi^{\iota}_1, \ldots, \Phi^{\iota}_{h_0}]^T$ and using (72), we obtain $(\mathscr{A}^{\iota}_N, \mathscr{B}^{\iota}_N) = \Xi([\gamma^{\iota}_{N,i,j}]) \to (\Phi^{\iota}, \Lambda) = \Xi([\mathscr{P}^{\iota}_{i,j}])$, w.p.1, as $N \to \infty$. From (7) and (68), it follows that $\mathcal{W}^{\iota}_N$ is an unbiased estimator of the ι-th component of ξ. Moreover, by (67), $\mathscr{L}_N \to \xi$, w.p.1, as $N \to \infty$. Finally, by (10) and (11), the theorem is proved. □

Remark 5.1 *If the measurement data does not take the $m+1$-th value, then the specific data value is less than the maximum threshold. That is, for each threshold, the binary indicators after comparing the measurement data are both 1. This is equivalent to constructing m parallel binary measurement channels, each corresponding to a specific quantization interval. The remaining content is identical to the binary case.*

Remark 5.2 *Under binary measurements, each sample requires a finite number of index checks and accumulations, with a time complexity of $O(N)$. By pre-computing the eigenvalue decomposition of Ψ, each matrix inversion only updates coefficients, resulting in an overall complexity of $O(1)$ for this stage. Under multi-threshold measurement, the complexity for m statistics increases to $O(mN)$, while the computationally intensive matrix polynomial inversion can be shared across m thresholds. Thus, the total online time complexity for multi-threshold measurement is $O(mN) + O(1)$, where the first term represents data traversal overhead and the second $O(1)$ term encapsulates all fixed costs independent of N but dependent on the system dimensions h_0 and m.*

5.2 Practical source of flag

The flag T_k originates from a binary comparison output of a gain module driven by a periodic input with period $\overline{\mu}$. Specifically, we define

$$t_k = B_k A + e_k, \tag{73}$$

where $e_k \sim \mathcal{N}(0, \sigma_e^2)$ is an i.i.d. Gaussian variable independent of the system noise w_k; B_k is a periodic input with period $\overline{\mu}$, taking values in $\{b_1, \ldots, b_{\overline{\mu}}\}$ and satisfying $B_k = B_{k-\overline{\mu}}$; A is the gain coefficient; T^0_k is obtained by comparing t_k with a threshold C_T, expressed as

$$T^0_k = I_{\{t_k \le C_T\}} = \begin{cases} 1, \ t_k \le C_T; \\ 0, \ \text{others}. \end{cases} \tag{74}$$

Given the periodic sequence $\{B_k\}$, gain coefficient A, distribution of e_k, and threshold C_T, the flag sequence generated by (73) and (74) exhibits the desired statistical properties (Wang et al., 2010). The parameter $G_j = \Pr\{T^0_k = 1\} = F_e(C_T - b_j A)$, where $F_e(\cdot)$ is the cumulative distribution function of e_k. This variable ensures the stochastic feature of the flag, and its effects are already reflected in the design parameters $\{G_j\}$. It does not adversely affect the convergence of the main identification algorithm.

5.3 Handling extreme attack scenarios

As indicated by Theorem 3.3, the identification algorithm (22)-(24) becomes invalid when $\lambda_0 = 0$. Furthermore, from (42), we see that when $\lambda_0 = 0$, the matrix $\sum_{\delta=0}^{\mu} \lambda_\delta G_{j-\delta} (\Psi^{\delta})^T$ has a zero eigenvalue, making it non-invertible and preventing valid estimation of Φ. To address this extreme situation, we propose a communication adjustment scheme to ensure algorithmic feasibility and stability. Specifically, we define

$$\mathcal{Z}_k = \mathcal{S}^0_{k+\epsilon} = I_{\{s^0_{k+\epsilon}=T^0_{k+\epsilon}=1\}}, \tag{75}$$

where $\mathcal{Z}_k$ is the data transmitted from the sender to the data center. $s^0_{k+\epsilon}$ and $T^0_{k+\epsilon}$ denote the sensor measurement and flag, respectively. ϵ is chosen based on Λ such that $\epsilon = \min\{i \in \mathbb{U} | \lambda_i \neq 0\}$. In this case, we have

$$\begin{aligned} &\Pr\{s_k = 1\} \\ &= \sum_{\delta=0}^{\mu} \Pr\left\{s_k = \mathcal{S}^0_{k+\epsilon-\delta}\right\} \Pr\left\{\mathcal{S}^0_{k+\epsilon-\delta} = 1\right\} \\ &= \sum_{\delta=0}^{\mu} \Pr\left\{s_k = \mathcal{S}^0_{k-(\delta-\epsilon)}\right\} \Pr\left\{\mathcal{S}^0_{k-(\delta-\epsilon)} = 1\right\} \\ &= \sum_{\upsilon=0}^{\mu-\epsilon} \lambda_\upsilon G_{k-\upsilon} \sum_{\iota=1}^{h_0} \psi^{\upsilon}_{\iota,k} \Phi_\iota. \end{aligned}$$

Thus, the effective impact of the attack probability becomes $[\lambda_0, \lambda_1, \ldots, \lambda_{\mu-\epsilon}, 0, \ldots, 0]^T$. This transformation circumvents the case $\lambda_0 = 0$ and ensures the robustness of the algorithm under extreme conditions.

Remark 5.3 *This strategy serves as a "disaster recovery mechanism" with full-time replay attacks rendering communication channels completely unreliable. While this introduces additional time delays, it offers certain potential benefits.*

6 Numerical simulation

Consider a SISO discrete-time Wiener system.

$$\begin{cases} y_k = 1 + \eta_1 \frac{x_k^2}{20} + \eta_2 \frac{2^{x_k}}{20} + w_k, \\ x_k = \theta_1 u_k + \theta_2 u_{k-1}, \end{cases}$$

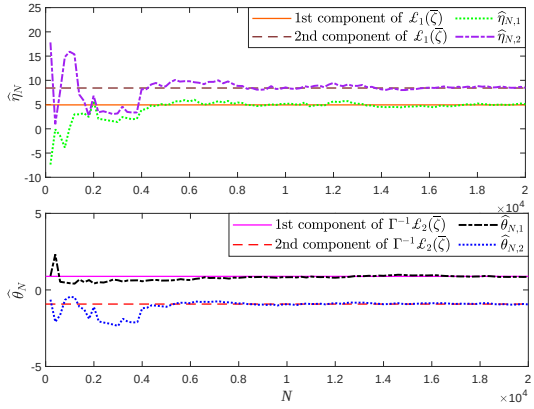

Fig. 3. Convergence of algorithms (9)-(11) under replay attacks

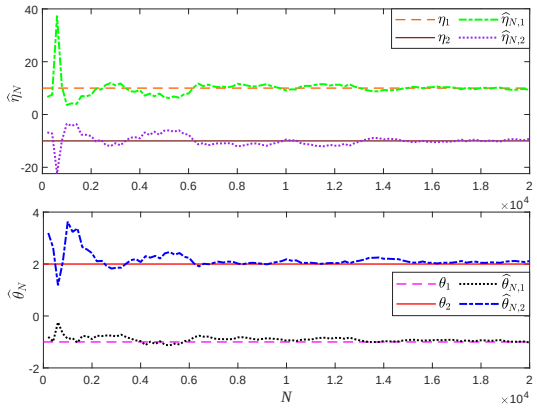

Fig. 4. Consistency of algorithms (22)-(25) with known attacks

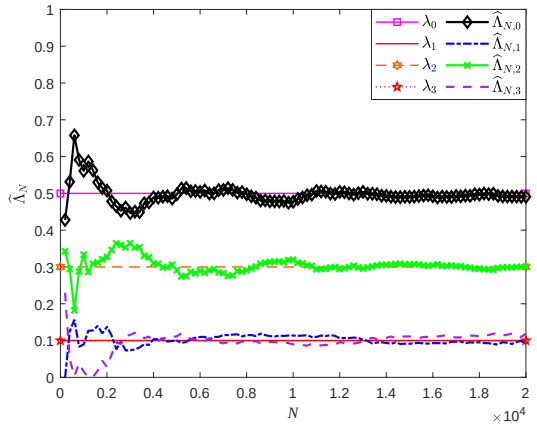

Fig. 5. Identification effect of algorithms (31)-(37) for Λ

where order $n_1 = n_2 = 2$; Unknown parameters $\theta = [\theta_1, \theta_2]^T = [-1, 2]^T$ and $\eta = [\eta_1, \eta_2]^T = [10, -10]^T$; The quantized input $u_k \in \{-1, 1\}$, and the regression patterns $\tau_1 = [-1, -1]$, $\tau_2 = [-1, 1]$, $\tau_3 = [1, -1]$, $\tau_4 = [1, 1]$ satisfy the persistent excitation condition in Assumption 3.3. The full-rank matrix is defined as $\Gamma = [\tau_3, \tau_4]^T$. The binary sensor threshold $C = 0$, and the noise $w_k \sim \mathcal{N}(0, 5^2)$, satisfying Assumption 2.1. The measured data s_k^0 is subject to random replay attacks characterized by (μ, Λ) during transmission to the data center. A data sample $N = 20000$.

Under a replay attack strategy $(\mu, [\lambda_0, \lambda_1, \lambda_2, \lambda_3]^T) = (3, [0.5, 0.1, 0.3, 0.1]^T)$, the convergence of the algorithms is shown in Fig. 3, where the estimates $\widehat{\eta}_N$ and $\widehat{\theta}_N$ converge to $\mathcal{L}_1(\bar{\zeta})$ and $\Gamma^{-1}\mathcal{L}_2(\bar{\zeta})$, respectively, verifying Theorem 3.1. For the known attack strategy, parameters η and θ are estimated using algorithms (22)-(25), with results shown in Fig. 4. These results confirm consistent estimation, in accordance with Theorem 3.2.

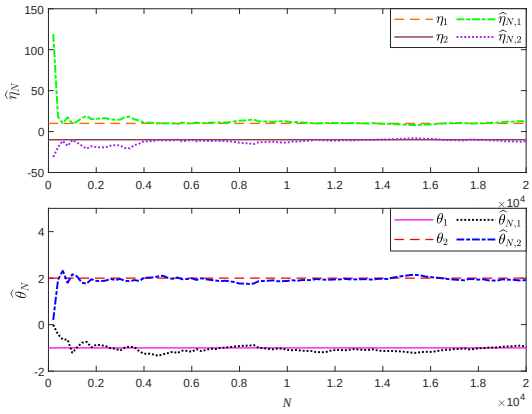

Fig. 6. Identification effect of algorithms (31)-(37) for η and θ

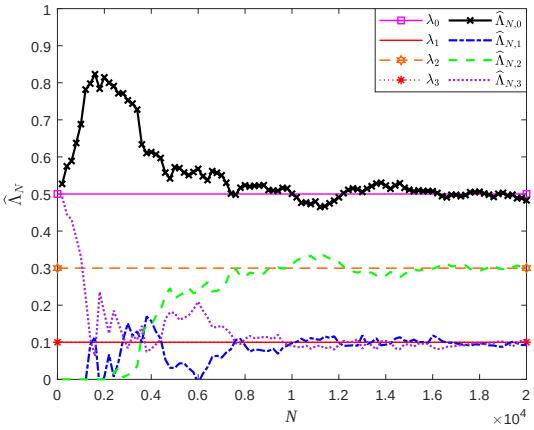

Fig. 7. Performance of algorithms (43)-(48) for Λ

For the unknown attack strategy case, set $A = 1$, the input b_k cycles through $[0.4, 2, 1, 4]$, and $e_k \sim \mathcal{N}(0, 2^2)$ is an i.i.d. Gaussian variable independent of w_k. Set the threshold $C_T = 3$ to generate the flag T_k^0. $\overline{\mu} = \mu + 1 = 4$. (I). Using the adding-based mechanism, compute $\mathcal{D}_k = T_k^0 \oplus s_k^0$ and apply algorithms (31)-(37) for joint estimation of Λ and parameters η, θ, as shown in Figs. 5 and 6. The results confirm consistent estimation, validating Theorem 4.1. (II). The data-flag fusion mechanism yields new data $\mathcal{S}_k^0 = I_{\{s_k^0 \cdot T_k^0 = 1\}}$. Algorithms (43)-(48) are applied with results shown in Figs. 7 and 8,

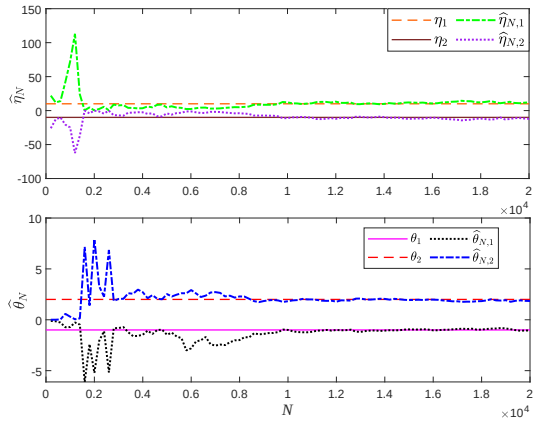

Fig. 8. Performance of algorithms (43)-(48) for η and θ

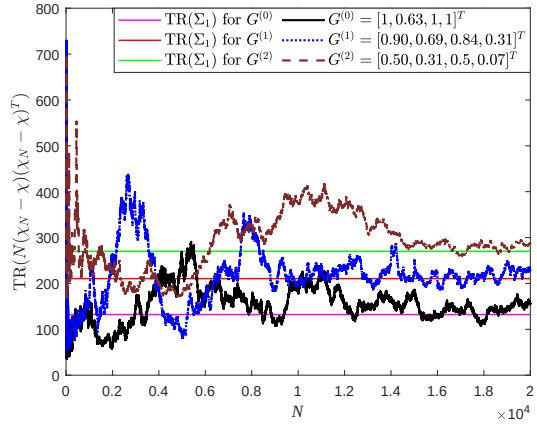

Fig. 9. Asymptotic performance of algorithms (43)-(48) for estimating χ in (56)

verifying Theorem 4.2. (III). To evaluate performance, we approximate the true error TR(Σ_1) by averaging TR$(N(\chi_N - \chi)^2)$ under 150 trajectories. For comparison, use optimal parameters $G^{(0)} = [1, 0.63, 1, 1]^T$ and suboptimal parameters $G^{(1)} = [0.90, 0.69, 0.84, 0.31]^T$, $G^{(2)} = [0.50, 0.31, 0.5, 0.07]^T$, and plot the results in Fig. 9. All cases show asymptotic convergence, with the optimal setting yielding the lowest error. Similarly, for Theorem 4.3 concerning asymptotic normality and parameter optimization, use $\acute{G}^{(0)} = [0.18, 1, 1, 1]^T$ and suboptimal $\acute{G}^{(1)} = [0.16, 0.69, 0.84, 0.67]^T$, $\acute{G}^{(2)} = [0.067, 0.16, 0.50, 0.70]^T$ to obtain Fig. 10, which matches theoretical predictions. The additional $\widehat{\mu} = \overline{\mu}+2 = \mu+3$ is selected. According to the simulation results in Figs. 11 and 12, in the attack strategy estimation $\widehat{\Lambda}_N$, the components corresponding to $\delta > \mu$ converge to values close to zero, while the other components and system parameters are accurately estimated, consistent with the theoretical analysis.

For the multi-threshold sensor case with $m = 3$ thresholds $C_1, C_2, C_3 = -2, 0, 5$, and measurement data $\mathcal{E} = \{1, 2, 3, 4\}$, generate new data $\mathcal{S}_k^0$ using (63). To test robustness, consider an attack strategy $(\mu, [\lambda_0, \lambda_1, \lambda_2, \lambda_3]^T) = (3, [0.4, 0.1, 0, 0.5]^T)$. Upon receiving s_k, the data center processes it using (64) and applies algorithms (65)-(71). The results in Figs. 13 and 14 validate Theorem 5.1. Under the extreme condition $(\mu, [\lambda_0, \lambda_1, \lambda_2, \lambda_3]^T) = (3, [0, 0.3, 0.5, 0.2]^T)$ with $\lambda_0 = 0$, leading to $\epsilon = 1$, the adjustment scheme (75) is used to transmit $\mathcal{Z}_k = \mathcal{S}_{k+1}^0$. The data center applies algorithms (43)-(48), with results shown in Figs. 15 and 16. The estimate $\widehat{\Lambda}_N$ converges to $[0.3, 0.5, 0.2, 0]^T$, ensuring the correctness of the output (42), thus validating the effectiveness of the adjustment scheme.

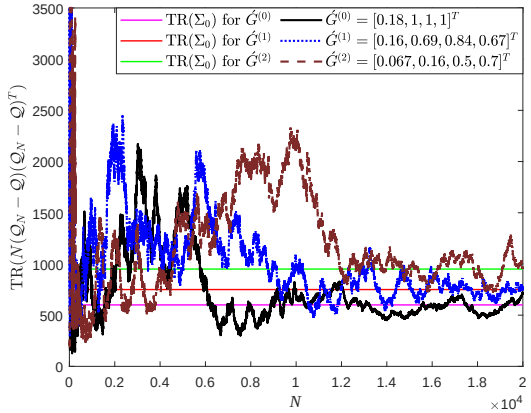

Fig. 10. Asymptotic performance of algorithms (31)-(37) for estimating $\mathscr{Q}$ in (55)

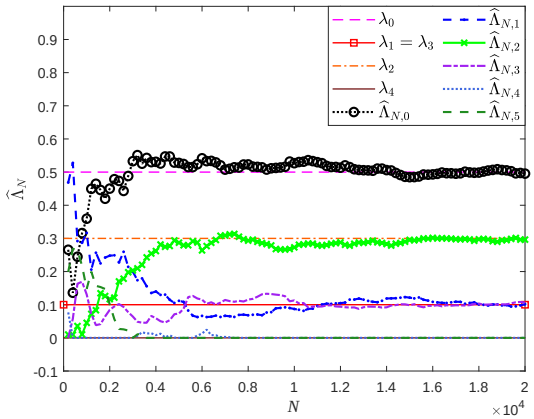

Fig. 11. Λ estimation using algorithms (43)-(48) with $\widehat{\mu} > \mu + 1$

Finally, there is a comparison and analysis of methods. The adding-based mechanism, the two stage method (Guo et al., 2025), and the fusion mechanism are used for performance comparison. Set $(\mu, [\lambda_0, \lambda_1, \lambda_2, \lambda_3]^T) = (3, [0.5, 0.1, 0.3, 0.1]^T)$ and $\overline{\mu} = 6$. System settings remain unchanged. (i) Communication overhead. Neither the two stage method nor the fusion mechanism introduces additional bit overhead, preserving the channel's data transmission efficiency. The adding-based mechanism, however, requires an extra bit of overhead. (ii) Estimation effect. The curves depicting the estimation er-

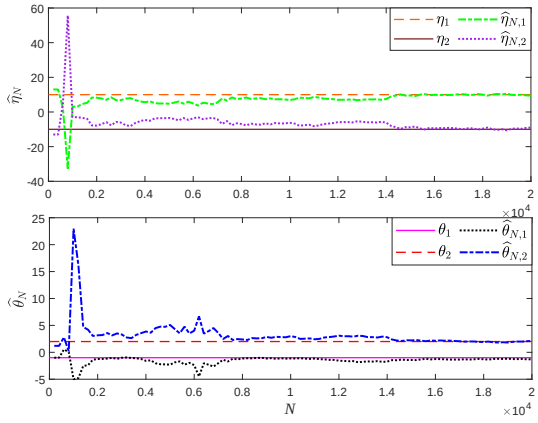

Fig. 12. Parameter estimations using algorithms (43)-(48) with $\widehat{\mu} > \mu + 1$

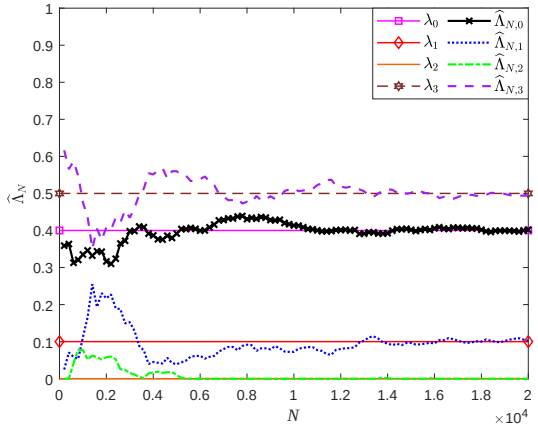

Fig. 13. Consistency of algorithms (65)-(71) for Λ with multi-threshold measurement

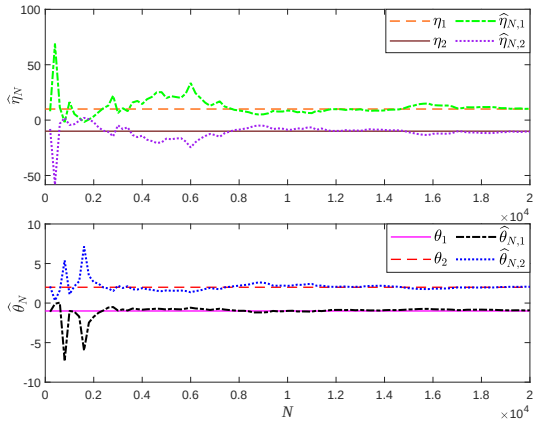

Fig. 14. Consistency of algorithms (65)-(71) for η and θ with multi-threshold measurement

ror of attack strategies, η, and θ (measured by the L_2-norm $||\cdot||$ between estimated and true values) as increasing N are shown in Fig. 17. The results indicate that compared with the two stage method, the fusion mechanism exhibits lower overall error and faster convergence

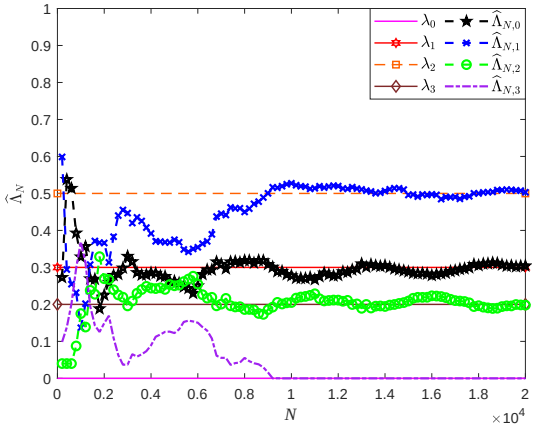

Fig. 15. Estimate effect of algorithms (43)-(48) for Λ using adjustment scheme

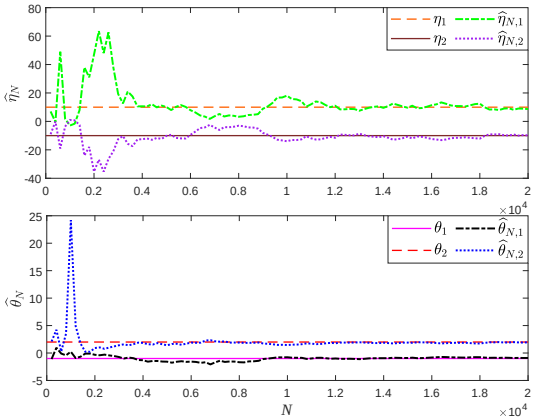

Fig. 16. Estimate effect of algorithms (43)-(48) for η and θ using adjustment scheme

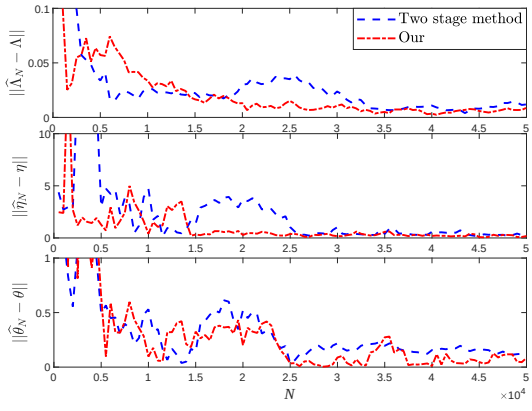

Fig. 17. Comparison results on estimation errors

speed. Previous results in Figs. 5-6 and Figs. 7-8 indicate that the adding-based mechanism exhibits faster convergence, which is attributable to increased communication overhead. (iii) To assess predictability, a window length $W_l = 50$ is selected, and the information entropy of this window is computed as $-\log_2 \sum_{i=1}^{W_l} s(i)/W_l -$

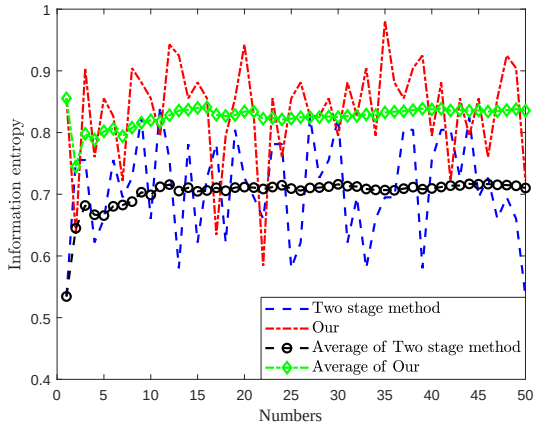

Fig. 18. Comparison results on security (information entropy)

$\log_2 \sum_{i=1}^{W_l}(1-s(i))/W_l$. As the number of windows increases, the curve of information entropy changes as shown in Fig. 18. The information entropy has converged, while the proposed fusion mechanism exhibits higher information entropy, greater uncertainty, and enhanced security.

7 Concluding remarks

This paper tackles the problem of parameter identification for quantized Wiener systems in CPSs under replay attacks. A data-flag fusion mechanism based on binary stochastic flag is proposed, which retains the advantages of binary communication while significantly enhancing real-time resilience against attacks. A parameter identification method based on adding flag is developed for scenarios with unknown attack probabilities. Furthermore, the limitation of 1-bit communication is overcome by designing a data-flag fusion mechanism and corresponding joint identification algorithm for system parameters and attack probabilities. The asymptotic properties of the estimators and optimal flag configuration strategies are rigorously analyzed. The proposed mechanism is also extended to multi-threshold measurement scenarios, improving adaptability and generality, while an adjustment scheme enhances robustness under extreme conditions.

Future research directions include i) extending the proposed identification framework to more complex system models with dynamic nonlinearities or feedback structures, ii) designing joint defense strategies against multiple types of attacks such as denial-of-service and tampering, and iii) exploring online identification mechanisms under resource-constrained and distributed settings to enhance practical deployment and performance.

Appendix A. Proof of Theorem 4.3

Since $z_k = \ominus_1(s_k)$ only takes 0 or 1, $\mathbb{E}\{(z_k - \mathbb{E}\{z_k\})^2\} = \mathbb{E}\{z_k^2\} - (\mathbb{E}\{z_k\})^2 = \mathbb{E}\{z_k\} - (\mathbb{E}\{z_k\})^2$. Due to the i.i.d. property of the noise, according to Central Limit Theorem, as $N \to \infty$,

$$\begin{bmatrix} \sqrt{N_1}(\overline{z}_{N,1} - \zeta_1) \\ \vdots \\ \sqrt{N_{h_0}}(\overline{z}_{N,h_0} - \zeta_{h_0}) \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, \mathrm{diag}(\zeta_i - \zeta_i^2)), \tag{76}$$

where $\overline{z}_{N,i} = \frac{1}{N_i}\sum_{k=1}^{N} z_k I_{\{\pi_k=\tau_i\}}$; N_i is given by (9). Likewise, by the periodicity and independence of the flag, there is $\lim_{N\to\infty} \frac{N}{L_{N,j}} = \overline{\mu}$, so

$$\begin{bmatrix} \sqrt{N}(\mathcal{L}_{N,1} - \rho_1) \\ \vdots \\ \sqrt{N}(\mathcal{L}_{N,\overline{\mu}} - \rho_{\overline{\mu}}) \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, \mathrm{diag}(\overline{\mu}(\rho_j - \rho_j^2))),\ N \to \infty.$$

From Lemma 4.1, (30) and (31), we have

$$\begin{aligned} &\Pi^{-1} \begin{bmatrix} \sqrt{N}(\mathcal{L}_{N,1} - \rho_1) \\ \vdots \\ \sqrt{N}(\mathcal{L}_{N,\overline{\mu}} - \rho_{\overline{\mu}}) \end{bmatrix} \\ &= \sqrt{N}(\widehat{\Lambda}_N - \Lambda) \\ &\xrightarrow{d} \mathcal{N}(0, \Pi^{-1}\mathrm{diag}(\overline{\mu}(\rho_j - \rho_j^2))\Pi^{-T}),\ N \to \infty. \end{aligned} \tag{77}$$

By (27) and (36), combined with Mean Value Theorem, there exist intermediate values ξ_i^* between $\xi_{N,i}$ and ξ_i, and $\overline{\varrho}_i$ between $\varrho_{N,i}$ and Φ_i, such that

$$\begin{aligned} &\sqrt{N}(\mathscr{Q}_N - \mathscr{Q}) \\ &= \sqrt{N}(\mathcal{L}(\xi_N) - \mathcal{L}(\xi)) \\ &= \sqrt{N}\mathscr{D}(\xi^*)(\xi_N - \xi) \\ &= \sqrt{N}\mathscr{D}(\xi^*) \begin{bmatrix} C - \mathcal{F}(\varrho_{N,1}) - (C - \mathcal{F}(\Phi_1)) \\ \vdots \\ C - \mathcal{F}(\varrho_{N,h_0}) - (C - \mathcal{F}(\Phi_{h_0})) \end{bmatrix} \\ &= \sqrt{N}\mathscr{D}(\xi^*)\mathrm{diag}(\mathscr{H}(\overline{\varrho}_i))(\beta(\Lambda) \begin{bmatrix} \zeta_1 \\ \vdots \\ \zeta_{h_0} \end{bmatrix} - \beta_N(\widehat{\Lambda}_N) \begin{bmatrix} \overline{z}_{N,1} \\ \vdots \\ \overline{z}_{N,h_0} \end{bmatrix}) \\ &= \sqrt{N}\mathscr{D}(\xi^*)\mathrm{diag}(\mathscr{H}(\overline{\varrho}_i))\Big\{\mathscr{C}(\Lambda) - \mathscr{C}_N(\widehat{\Lambda}_N) \\ &+ \beta_N(\widehat{\Lambda}_N) \begin{bmatrix} \zeta_1 - \overline{z}_{N,1} \\ \vdots \\ \zeta_{h_0} - \overline{z}_{N,h_0} \end{bmatrix}\Big\}, \end{aligned} \tag{78}$$

where $\mathscr{C}_N = \beta_N(\widehat{\Lambda}_N)\zeta$. $\mathscr{C}(\Lambda) - \mathscr{C}_N(\widehat{\Lambda}_N) = \mathscr{C}(\Lambda) - \mathscr{C}(\widehat{\Lambda}_N) + \mathscr{C}(\widehat{\Lambda}_N) - \mathscr{C}_N(\widehat{\Lambda}_N)$. Similarly, Mean Value Theorem is applied to the derivatives of $\mathscr{C}(\Lambda)$ and $\mathscr{C}(\widehat{\Lambda}_N)$ with respect to Λ. As $N \to \infty$, we have $\xi_i^* \to \xi_i$, $\overline{\varrho}_i \to \Phi_i$, $\mathscr{C}_N \to \mathscr{C}$, and $\widehat{\Lambda}_N \to \Lambda$. Combining Lemmas 4.1 and 4.2, together with (76), (77), and (78), it follows that $\sqrt{N}(\mathscr{Q}_N - \mathscr{Q}) \xrightarrow{d} (0, \Sigma_0)$. The theorem is proved. □

References

Bai, E.W. (2008). Non-Parametric Nonlinear System Identification: A Data-Driven Orthogonal Basis Function Approach. *IEEE Transactions on Automatic Control*, 53(11), 2615–2626.

Beintema, G.I., Schoukens, M., & Tóth, R. (2023). Deep subspace encoders for nonlinear system identification. *Automatica*, 156, 111210.

Cao, H., Li, L., Feng, Y., & Li, L. (2024). Self-error learning framework-based algorithm for parameter recovery of extended Wiener–Hammerstein systems subject to quantised measurements. *ISA Transactions*, 150, 374–387.

Chow, Y.S., & Teicher, H. (1997). *Probability theory: Independence, interchangeability,martingales* (3rd ed.). New York: Springer–Verlag.

Fang, C., Qi, Y., Cheng, P., & Zheng, W.X. (2020). Optimal periodic watermarking schedule for replay attack detection in cyber–physical systems. *Automatica*, 112, 108698.

Farha, F., Ning, H., Yang, S., Xu, J., Zhang, W., & Choo, K.K.R. (2022). Timestamp Scheme to Mitigate Replay Attacks in Secure ZigBee Networks. *IEEE Transactions on Mobile Computing*, 21(1), 342–351.

Ferrari, R.M.G., & Teixeira, A.M.H. (2021). A Switching Multiplicative Watermarking Scheme for Detection of Stealthy Cyber-Attacks. *IEEE Transactions on Automatic Control*, 66(6), 2558–2573.

Fritz, R., & Zhang, P. (2023). Detection and Localization of Stealthy Cyberattacks in Cyber-Physical Discrete Event Systems. *IEEE Transactions on Automatic Control*, 68(12), 7895–7902.

Guo, J., Wang, L.Y., Yin, G., Zhao, Y., & Zhang, J.-F. (2017). Identification of Wiener systems with quantized inputs and binary-valued output observations. *Automatica*, 78, 280–286.

Guo, J., Zhang, J.-F., & Zhao, Y. (2012). Adaptive tracking of a class of first-order systems with binary-valued observations and fixed thresholds. *Journal of Systems Science and Complexity*, 5(6), 1041–1051.

Guo, J., Zhang, Q., & Zhao, Y. (2025). Identification of FIR Systems with binary-valued observations under replay attacks. *Automatica*, 172, 112001.

Hall, P., & Heyde, C.C. (1980). *Martingale Limit Theory and Its Application*. New York: Academic Press.

Huang, J., Ho, D.W.C., Li, F., Yang, W., & Tang, Y. (2020). Secure remote state estimation against linear man-in-the-middle attacks using watermarking. *Automatica*, 121, 109182.

Jia, R., Wang, T., & Xue, W. (2025). Multitime Scale Consensus Algorithm of Multiagent Systems With Binary-Valued Data Under Attacks. *IEEE Transactions on Industrial Informatics*, 21(12), 9377–9388.

Kang, Y., Zhai, D. H., Liu, G. P., Zhao, Y. B., & Zhao, P. (2014). Stability analysis of a class of hybrid stochastic retarded systems under asynchronous switching. *IEEE Transactions on Automatic Control*, 59(6), 1511–1523.

Kay, S.M. (1993). *Fundamentals of statistical signal processing: Estimation theory*. Upper Saddle River, NJ: Prentice Hall.

Li, L., Wang, F., Zhang, J., & Liu, X. (2023). Hammerstein system identification using robust estimator based on quantized observation. In *12th IEEE Data Driven Control and Learning Systems Conference* (pp. 95–99).

Li, L., Wang, Y., Wang, X., Zhang, H., & Ren, X. (2025). An Error Learning Scenario-Based Scheme to Quantized Identification in Wiener-Hammerstein Systems Subject to Deadzone Nonlinearity. *IEEE Transactions on Automation Science and Engineering*, 22, 5375-5387.

Li, P., & Ye, D. (2025). Stochastic encryption against stealthy attacks in CPSs: A zero-sum game approach. *Automatica*, 172, 112012.

Li, T., Wang, Z., Zou, L., Chen, B., & Yu, L. (2023). A dynamic encryption-decryption scheme for replay attack detection in cyber–physical systems. *Automatica*, 151, 110926.

Liu, F., Rapakoulias, G., & Tsiotras, P. (2025). Optimal Covariance Steering for Discrete-Time Linear Stochastic Systems. *IEEE Transactions on Automatic Control*, 70(4), 2289–2304.

Liu, H., Li, Y., Han, Q.L., & Raïssi, T. (2023). Watermark-Based Proactive Defense Strategy Design for Cyber-Physical Systems With Unknown-but-Bounded Noises. *IEEE Transactions on Automatic Control*, 68(6), 3300–3315.

Liu, H., Zhang, Y., Li, Y., & Niu, B. (2023). Proactive attack detection scheme based on watermarking and moving target defense. *Automatica*, 155, 111163.

Liu, W., Wang, Y., & Guo, J. (2025). Consistent Identification for FIR Systems with Multi-Level Quantized Observations Subjected to Data Tampering Attacks: A Joint Estimation Approach. *IEEE Transactions on Industrial Electronics*, doi: 10.1109/TIE.2025.3642258.

Liu, Z.Q., Ge, X., Xie, H., Han, Q.L., Zheng, J., & Wang, Y.L. (2024). Secure Leader–Follower Formation Control of Networked Mobile Robots Under Replay Attacks. *IEEE Transactions on Industrial Informatics*, 20(3), 4149–4159.

Mo, Y., & Sinopoli, B. (2009). Secure control against replay attacks. In *47th Annual Allerton Conference on Communication, Control, and Computing* (pp. 911–918).

Mustapha, K.B. (2025). A survey of emerging applications of large language models for problems in mechanics, product design, and manufacturing. *Advanced Engineering Informatics*, 64, 103066.

Nadi, M., & Arefi, M.M. (2023). Hierarchical iterative identification of output nonlinear Box-Jenkins Wiener model with ARMA noise. *ISA Transactions*, 143, 321–333.

Naseri, F., Schaltz, E., Stroe, D.I., Gismero, A., & Farjah, E. (2022). An Enhanced Equivalent Circuit Model with Real-Time Parameter Identification for Battery State-of-Charge Estimation. *IEEE Transactions on Industrial Electronics*, 69(4), 3743–3751.

Nejib, H., Taouali, O., & Bouguila, N. (2016). Identification of nonlinear systems with kernel methods. In *IEEE International Conference on Systems, Man, and Cybernetics* (pp. 000577–000581).

Ozbot, M., Lughofer, E., & Skrjanc, I. (2023). Evolving Neuro-Fuzzy Systems-Based Design of Experiments in Process Identification. *IEEE Transactions on Fuzzy Systems*, 31(6), 1995–2005.

Pillonetto, G., Aravkin, A., Gedon, D., & Ljung, L. (2025). A.H. Ribeiro, T.B. Schön, Deep networks for system identification: A survey. *Automatica*, 171, 111907.

Pivoto, D.G.S., de Almeida, L.F.F., da Rosa Righi, R., Rodrigues, J.J.P.C., Lugli, A.B., & Alberti, A.M. (2021). Cyber-physical systems architectures for industrial internet of things applications in Industry 4.0: A literature review. *Journal of Manufacturing Systems*, 58, 176–192.

Porter, M., Hespanhol, P., Aswani, A., Johnson-Roberson, M., & Vasudevan, R. (2021). Detecting Generalized Replay Attacks via Time-Varying Dynamic Watermarking. *IEEE Transactions on Automatic Control*, 66(8), 3502–3517.

Qu, Z., Shi, W., Liu, B., Gupta, D., & Tiwari, P. (2024). IoMT-Based Smart Healthcare Detection System Driven by Quantum Blockchain and Quantum Neural Network. *IEEE Journal of Biomedical and Health Informatics*, 28(6), 3317–3328.

Rakha, H.A. (2024). Development, Modeling and Assessment of Connected Automated Vehicle Applications. *IEEE Transactions on Intelligent Transportation Systems*, 25(4), 306–331.

Rasheed, A., Baza, M., Badr, M.M., Alshahrani, H., & Choo, K.K.R. (2024). Efficient Crypto Engine for Authenticated Encryption, Data Traceability, and Replay Attack Detection Over CAN Bus Network. *IEEE Transactions on Network Science and Engineering*, 11(1), 1008–1025.

Revay, M., Wang, R., & Manchester, I.R. (2024). Recurrent Equilibrium Networks: Flexible Dynamic Models With Guaranteed Stability and Robustness. *IEEE Transactions on Automatic Control*, 69(5), 2855–2870.

Risuleo, R.S., Lindsten, F., & Hjalmarsson, H. (2019). Bayesian nonparametric identification of Wiener systems. *Automatica*, 108, 108480.

Shakib, M.F., Pogromsky, A.Y., Pavlov, A., & van de Wouw, N. (2022). Computationally efficient identification of continuous-time Lur'e-type systems with stability guarantees. *Automatica*, 136, 110012.

Song, Y., & Ye, D. (2023). Replay attack detection and mitigation for cyber–physical systems via RADIR algorithm with encryption scheduling. *Neurocomputing*, 558, 126698.

Wang, L., Yin, G.G., Zhang, J.-F., & Zhao, Y. (2010). *System identification with quantized observations*. Basel: Birkhäuser.

Ye, D., Zhang, T.Y., & Guo, G. (2019). Stochastic coding detection scheme in cyber-physical systems against replay attack. *Information Sciences*, 481, 432–444.

Yu, P., Hu, Y., Wang, Y., Jia, R., & Guo, J. (2025). Optimal Consensus Control Strategy for Multi-Agent Systems Under Cyber Attacks via a Stackelberg Game Approach. *IEEE Transactions on Automation Science and Engineering*, 22, 18875–18888.

Zhao, D., Yang, B., Li, Y., & Zhang, H. (2025). Replay Attack Detection for Cyber-Physical Control Systems: A Dynamical Delay Estimation Method. *IEEE Transactions on Industrial Electronics*, 72(1), 867–875.

Zhu, M., & Martínez, S. (2014). On the Performance Analysis of Resilient Networked Control Systems Under Replay Attacks. *IEEE Transactions on Automatic Control*, 59(3), 804–808.

Zong, T., Li, J., & Lu, G. (2023). Parameter identification of dual-rate Hammerstein-Volterra nonlinear systems by the hybrid particle swarm-gradient algorithm based on the auxiliary model. *Engineering Applications of Artificial Intelligence*, 117, 105526.