Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 34(10): 7118–7151.
DOI: 10.3934/era.2026308
Received: 06 June 2026
Revised: 31 July 2026
Accepted: 10 August 2026
Published: 18 August 2026

Research article

Asymptotically consistent identification of binary quantized systems under Poisson-distributed replay attacks

Yadong Wang[1,2], **Wenyang Jiang**[1,2] **and Wenke Liu**[1,2,*]

[1] School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

[2] Key Laboratory of Knowledge Automation for Industrial Processes, Ministry of Education, Beijing 100083, China

* **Correspondence:** Email: winkliu_745107073@126.com.

Abstract: This paper investigated the identifiability of attack frequency and the parameter estimation problem in binary quantized finite impulse response (FIR) systems under Poisson-distributed replay attacks. First, the replay attack is modeled as a Poisson stochastic process, and the convergence properties of the attack frequency in the long-term statistical sense were analyzed, along with definitions and criteria for attack frequency identifiability. Under the assumption that the attack frequency is known, a compensated identification algorithm was established for parameter estimation, and the strong consistency and asymptotic normality of the estimators were established. Furthermore, for the case of unknown attack frequency, a joint estimation algorithm based on empirical statistics was proposed to achieve simultaneous identification of the attack frequency and system parameters. In addition, the optimal design criterion for the compensation ratio parameter was derived to reduce estimation errors. The proposed framework falls within the scope of resilient identification, where the estimator continues to recover system parameters from long-term statistics even under intermittent replay corruptions of the communication channel. Finally, numerical simulations under various attack frequencies, delay distributions, and noise levels were conducted to verify the effectiveness and robustness of the attack frequency identifiability analysis and the proposed algorithms.

Keywords: Poisson distribution; replay attack; attack frequency; identifiability; parameter estimation

1. Introduction

With the rapid development of a new wave of industrial revolution, Industry 4.0 has gradually become an important direction for modern manufacturing. The continuous integration of technologies such as smart manufacturing, industrial Internet, and artificial intelligence has driven industrial systems

toward digitalization, networking, and intelligence [1]. Meanwhile, digital transformation also provides new opportunities for the sustainable development and resource optimization of industrial systems [2]. However, under the context of Industry 4.0, the extensive interconnection of large-scale equipment and widespread network communication also exposes industrial systems to increasingly complex security challenges [3, 4].

As a crucial framework that tightly couples computation, communication, and control, cyber-physical systems (CPSs) have found extensive applications in smart grids, industrial automation, autonomous vehicles, intelligent healthcare, and other critical domains [5]. CPSs collect physical environment information in real time through sensing devices and utilize networks to transmit data and make control decisions, thereby achieving closed-loop interactions between the information space and the physical space [6]. As the scale of these systems continues to expand, data quality, communication reliability, and system security have gradually become critical factors affecting the stable operation of CPSs [7, 8].

In practical engineering, quantization techniques are widely employed in networked control systems and system identification to reduce communication bandwidth consumption and data storage pressure. In particular, binary quantization mechanisms are of significant value in resource-constrained environments due to their simplicity, low communication cost, and ease of hardware implementation [9, 10]. However, the quantization process results in partial information loss, thereby increasing the difficulty of system parameter identification and state estimation. For the identification of quantized observation systems, various identification frameworks based on probabilistic statistics and kernel function methods have been proposed [11]. In addition, under bandwidth-limited conditions, the co-design of communication mechanisms and control strategies has also attracted widespread attention [12].

With the increasing reliance of cyber-physical systems (CPSs) on network communication, network security has gradually become a research hotspot. Among various attack types, denial-of-service (DoS) attacks and replay attacks are the most typical forms [13]. Compared with DoS attacks, which directly interrupt communication, replay attacks intercept historical normal data and resend it at later times, thereby misleading system controllers or identifiers to make incorrect decisions, which makes them more covert [14]. In recent years, extensive research has been conducted on the detection of replay attacks. For example, Zhao et al. proposed a detection method based on dynamic delay estimation [15]; Mo et al. introduced physical watermark signals in control inputs to detect forged sensor data [16]; Guo et al. developed a replay attack detection mechanism based on output encoding [17]. In addition, for the problem of optimal control under replay attacks, related studies have proposed solutions based on delay estimation [18].

Beyond attack detection, a large body of work has also addressed state estimation and secure control under network attack environments. For instance, Li et al. proposed an improved Kalman filtering strategy under replay attacks with attack-detection-based compensation [19]. Ahmed et al. further studied physically implementable watermark-based detection schemes [20]. Meanwhile, replay attacks are not limited to industrial control systems; they have also received attention in voice and multimedia systems. Veesa et al. employed linear prediction residuals to detect voice replay attacks [21], while Hamadouche et al. proposed a replay-attack detection method based on perceptual image hashing [22].

Moreover, secure state estimation and resilient control under attacks have been extensively investigated in the literature. Related studies have addressed secure state estimation for CPSs under sensor attacks [23] and resilient control strategies for nonlinear systems under cyber threats [24]. These works provide important perspectives on maintaining system performance in adversarial environments and

complement the parameter identification framework developed in this paper.

In recent years, the development of artificial intelligence and data-driven methods has further promoted research in system identification and secure control [25]. Regarding the identification of quantized systems under network attacks, several important results have been achieved. Guo et al. studied parameter identification of binary-observed FIR systems under DoS attacks and proposed a consistent estimation algorithm [26]. Cui et al. further investigated identification under event-triggered communication mechanisms [27]. Additionally, for secure control of multi-agent systems under network attacks, Yu et al. developed an optimal consensus control strategy based on Stackelberg game theory [28], and Jia et al. studied multi-timescale consensus algorithms in the presence of data tampering, using only binary measurement signals [29].

Despite these advances, studies addressing parameter estimation in quantized systems affected by replay attacks are still relatively scarce. Guo et al. studied parameter identification for binary-observed FIR systems subject to replay attacks and designed a consistent defensive identification algorithm [30]. Subsequently, Liu et al. investigated joint estimation for multi-level quantized observation systems under data tampering attacks [31], and Zhang et al. proposed a compensation-based anti-replay-attack parameter-estimation method for quantized nonlinear Hammerstein systems [32].

However, several limitations remain. On one hand, most studies assume that all data are subject to attacks, whereas in practice, attackers often target only part of the data to reduce resource consumption and increase stealthiness. On the other hand, existing research typically assumes that the attack probability or model is known, whereas in real systems, the attack frequency is often difficult to obtain in advance. Moreover, when the attack frequency is unknown, whether the attack sequence itself is identifiable and how to achieve joint estimation of attack and system parameters still lack systematic theoretical investigation.

Compared with existing studies, the main differences and contributions of this work are as follows: (i) A stochastic replay-attack model based on the Poisson distribution is established, providing a principled framework for characterizing the randomness and intermittency of attack behaviors, in contrast to traditional models that assume all data are attacked. (ii) All defense scenarios are systematically categorized according to whether the attack frequency is known or unknown, and tailored algorithms are developed for each case. For the unknown-frequency case, the concept of attack frequency identifiability is introduced with rigorous theoretical conditions, bridging the gap between attack detectability and parameter estimability. (iii) A joint estimation framework with optimal data allocation is proposed, where the optimal ratio α^* is derived to minimize the asymptotic estimation variance.

This work investigates asymptotically consistent parameter estimation for binary quantized finite impulse response (FIR) models subject to Poisson-distributed replay attacks within the CPS framework. Considering that the attacker implements replay attacks randomly according to a Poisson distribution, the parameter identification problem is investigated for both known and unknown attack frequency scenarios, and the convergence and asymptotic properties of the proposed estimation algorithms are analyzed.

The main contributions of this paper are summarized as follows:

1) A stochastic replay-attack model based on the Poisson distribution is established. Compared with traditional replay-attack models in which all data are assumed to be attacked, the proposed model better captures the randomness and intermittency of attack behaviors in practical network attack scenarios.

2) For scenarios with a known attack frequency, we propose a parameter identification algorithm that incorporates an attack compensation mechanism, and we prove the asymptotic consistency of the

resulting estimator. When the attack frequency is unknown, we further examine the identifiability of the attack frequency itself. Building on this foundation, a joint estimation algorithm is subsequently constructed to enable the simultaneous estimation of system and attack parameters.

3) A theoretical analysis of the convergence and asymptotic characteristics is performed for the identification algorithms developed herein, and numerical simulations under various attack frequencies, delay distributions, and noise levels are performed to confirm their effectiveness and robustness under replay attacks.

The remainder of this paper is organized as follows. Section 2 introduces the system model and problem formulation. Section 3 investigates the parameter identification problem when the attack frequency is known. Section 4 studies the joint estimation problem when the attack frequency is unknown. Section 5 presents numerical simulation results. Finally, Section 6 concludes the paper and discusses future research directions.

2. Problem formulation

Consider a single-input single-output FIR system described by

$$\begin{aligned} y_k &= a_1u_k + a_2u_{k-1} + \cdots + a_nu_{k-n+1} + d_k \\ &= \pi_k^{\mathrm{T}}\theta + d_k, \end{aligned} \tag{2.1}$$

where d_k denotes the system noise, $\pi_k = [u_k, \ldots, u_{k-n+1}]^{\mathrm{T}}$ represents the system input vector, $\theta = [a_1, \ldots, a_n]^{\mathrm{T}}$ is the unknown system parameter vector, and y_k is the system output.

The output is measured by a binary sensor with threshold C, which can be represented by the indicator function

$$s_k = I_{\{y_k \le C\}} = \begin{cases} 1, & y_k \le C, \\ 0, & \text{otherwise}. \end{cases} \tag{2.2}$$

Figure 1 depicts that the original data s_k^0 are conveyed via a communication network to the estimation center, although they may be vulnerable to replay attacks along the way.

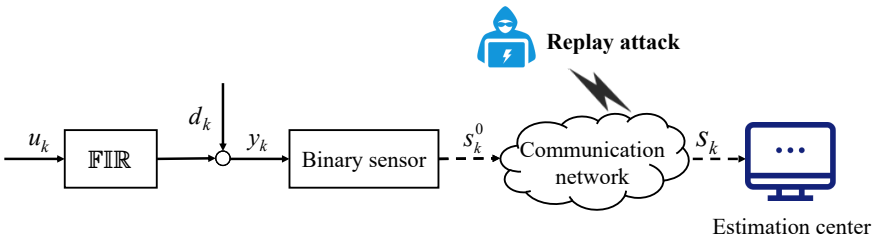


Figure 1. Schematic diagram of the system structure.

The data received by the estimation center at time instant k are denoted by s_k, which satisfy

$$s_k = \begin{cases} s_{k-\tau_k}^0, k \in \Upsilon, \\ s_k^0, \quad k \notin \Upsilon, \end{cases} \tag{2.3}$$

where Υ denotes the set of attack instants $\{\gamma_1, \gamma_2, \ldots\}$, and $s^0_{k-\tau_k}$ denotes the measurement sample obtained at instant $k - \tau_k$ and replayed at time k. The variable τ_k characterizes the replay lag associated with the attack.

The set $\Upsilon = \{\gamma_1, \gamma_2, \ldots\}$ represents the collection of time instants when replay attacks occur, while $\mathbf{p} = [p_1, \ldots, p_m]^{\mathrm{T}}$ describes the probability distribution of the replay delay intensity τ_k at each attack instant. Thus, the combination of sequential attack instants and random replay intensities reflects that the attacker, while selecting the attack time points, can randomly choose different replay delays at each chosen point to mislead the detector. Together, $(\Upsilon, \mathbf{p})$ constitutes a sequence-type replay-attack strategy with random replay intensity at given attack times.

Whenever an attack occurs, i.e., $k \in \Upsilon$, the attacker randomly selects the replay intensity $\tau_k \in \{1, \ldots, m\}$ according to the probability distribution $\mathbf{p} = [p_1, \ldots, p_m]$, namely,

$$\Pr(\tau_k = i) = p_i, \quad i = 1, \ldots, m. \tag{2.4}$$

Equations (2.3) and (2.4) characterize the data tampering mechanism caused by the attacks. To facilitate further theoretical analysis, the attack instants and attack intensities are modeled in a unified framework.

Remark 2.1. *Throughout this paper, the following information is assumed to be known to the legitimate estimator: the system input sequence $\{u_k\}$, the binary sensor threshold C, the noise distribution function $\phi(\cdot)$, the density function $\varphi(\cdot)$, the maximum replay delay m, and the replay delay distribution $\mathbf{p} = [p_1, \ldots, p_m]^{\mathrm{T}}$. The attack instant set Υ is always unknown. Two scenarios are distinguished regarding the attack frequency ϵ_i: (i) the attack frequency is known; (ii) the attack frequency is unknown and must be estimated from data. The assumption that $\mathbf{p}$ is known is motivated by practical engineering considerations: the maximum replay lag m is typically determined by the communication protocol buffer size, which is known to the defender. The case where $\mathbf{p}$ is also unknown constitutes a more challenging problem and is left as a direction for future work.*

Assume that the attack instant set $\Upsilon = \{\gamma_1, \gamma_2, \ldots\}$ is generated by a Poisson process with intensity $\lambda > 0$. Let $N(t)$ denote the number of attacks occurring within a time interval of length t. Then,

$$N(t) \sim \text{Poisson}(\lambda t),$$

that is,

$$\Pr(N(t) = r) = \frac{(\lambda t)^r}{r!} e^{-\lambda t}, \quad r = 0, 1, 2, \ldots,$$

and the numbers of attacks occurring in different intervals are mutually independent, where λ represents the average number of attacks per unit time.

To connect the continuous-time Poisson process with the discrete-time identification framework, define the discrete-time attack indicator as

$$A_k = \mathbf{1}_{\{N(k+1)-N(k)\geq 1\}}, \tag{2.5}$$

so that $\Pr(A_k = 1) = 1 - e^{-\lambda}$. The attack set in discrete time is then formally given by

$$\Upsilon_N = \{k \in \mathcal{T}_N : A_k = 1\}. \tag{2.6}$$

This explicitly relates the Poisson intensity λ to the per-sample attack probability.

Define $\mathcal{T}_N = \{1, 2, \ldots, N\}$ and $\Upsilon_N = \Upsilon \cap \mathcal{T}_N$. Let $\#(\cdot)$ denote the cardinality of a set. Then, the total number of attacks occurring within the interval $\mathcal{T}_N$ is $\#\Upsilon_N$.

For the periodic position $i = 1, \ldots, n$, define

$$\Upsilon_{N,i} = \{k \in \Upsilon_N : k \equiv i \ (\mathrm{mod}\ n)\},$$

where $\#\Upsilon_{N,i}$ denotes the frequency of attacks associated with the i-th position.

Under the above model, the attack process can be uniformly represented as $\mathcal{A} = (\Upsilon, \tau_k)$ or simply denoted by $(\Upsilon, \mathbf{p})$.

Remark 2.2. *In the considered CPS setting, the legitimate remote estimator observes the known periodic input sequence and the received (possibly corrupted) binary-valued data $\{s_k\}$, and aims to consistently identify the FIR system parameter θ despite replay-corrupted observations. The attacker, on the other hand, intercepts legitimate past sensor measurements and replays them at later time instants. The attacker's objectives may include: (i) causing systematic bias in the identified model parameters; (ii) hiding abnormal plant behavior by replaying normal historical data; (iii) evading detection mechanisms. The motivation for the estimator to continue identification under attack conditions is that the plant must remain operational and the communication channel is inherently vulnerable. Rather than attempting to perfectly detect every attacked sample, the proposed resilient identification framework leverages long-term statistical compensation to recover consistent parameter estimates from corrupted observations.*

The objective of the estimation center is to design compensation algorithms based on the system input and the received attacked data, so as to recover the true data or improve the parameter estimation accuracy.

Consider system (2.1) with a periodic input signal. Denote $\psi_k = \pi_k$. If the input period n equals the parameter dimension, then $u_k = u_{k+n}$ $(k \geq 1)$, and consequently $\psi_k = \psi_{k+n}$. The input matrix is constructed as

$$\Psi = [\psi_1, \ldots, \psi_n]^{\mathrm{T}}. \tag{2.7}$$

Remark 2.3. *When the input period equals the number of system parameters, the persistent excitation condition is guaranteed to hold, thereby ensuring the invertibility of the regression matrix and the identifiability of the system. This specific case provides an analytically tractable framework for theoretical derivation and has been widely recognized in the system identification literature. In addition, related studies on quantized inputs have shown that periodic or quantized excitation can be naturally extended to more general classes of inputs. Therefore, this assumption can be regarded both as a condition of practical relevance and as a fundamental theoretical condition.*

For a given integer N, define $L_N = \lfloor N/n \rfloor$. The algorithm employed to estimate the parameter θ is given by

$$\begin{aligned} \widehat{\theta_N} &= \Psi^{-1}[\zeta_{N,1}, \ldots, \zeta_{N,n}]^{\mathrm{T}}, \\ \zeta_{N,i} &= C - \phi^{-1}\left(\frac{1}{L_N}\sum_{L=1}^{L_N} s_{(L-1)n+i}\right), \quad i = 1, \ldots, n, \end{aligned} \tag{2.8}$$

where C is the threshold and $\widehat{\theta_N}$ denotes the parameter estimate.

Assumption 2.1. *d_k is a sequence of independent and identically distributed Gaussian random variables with mean 0 and variance δ^2, whose distribution function and density function are denoted by $\phi(\cdot)$ and $\varphi(\cdot)$, respectively.*

3. Defense algorithm design based on attack frequency

In Section 2, the attack instant set Υ is modeled as a Poisson process with intensity $\lambda > 0$. This section first establishes the relationship between the Poisson process and the attack frequency, and then designs the corresponding defense algorithms based on the attack frequency.

3.1. Definition and properties of attack frequency

Theorem 3.1. *Suppose that the attack instant set Υ is generated by a Poisson process with intensity $\lambda > 0$. Then, for any periodic position $i = 1, \ldots, n$, it holds that*

$$\frac{\#\Upsilon_{N,i}}{L_N} \to \epsilon_i, \quad w.p.1, \quad N \to \infty, \tag{3.1}$$

where $\epsilon_i \in [0, 1]$ denotes the attack frequency at periodic position i. Moreover,

$$\epsilon_i = 1 - e^{-\lambda}, \qquad i = 1, \ldots, n. \tag{3.2}$$

Proof. Since the attack instant set Υ is generated by a Poisson process with intensity $\lambda > 0$, for any unit sampling interval, we have $N(k+1) - N(k) \sim \text{Poisson}(\lambda)$.

Therefore, the probability that at least one attack occurs within a unit interval is

$$\Pr\{N(k+1) - N(k) \geq 1\} = 1 - \Pr\{N(k+1) - N(k) = 0\} = 1 - e^{-\lambda}.$$

Hence, the attack event in discrete time can be equivalently represented as a Bernoulli random variable with success probability $\epsilon_i = 1 - e^{-\lambda}$.

For a fixed periodic position i, $\#\Upsilon_{N,i}$ denotes the number of attacks occurring at that position.

Since the attack events at different sampling instants are mutually independent in discrete time, and the probability of attack occurrence at each instant is ϵ_i, $\#\Upsilon_{N,i}$ can be regarded as the sum of L_N Bernoulli random variables.

By the strong law of large numbers,

$$\frac{\#\Upsilon_{N,i}}{L_N} \to \epsilon_i, \quad w.p.1, \quad N \to \infty.$$

The proof is complete. □

According to the Poisson attack model, under discrete sampling conditions, the attack behavior of the system can be equivalently represented as a Bernoulli-type stochastic process.

According to the knowledge of the attack frequency, the problem considered in this paper is divided into the following two cases:

- The attack frequency is known;
- The attack frequency is unknown.

3.2. Known attack frequency case

When the attack frequency vector $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$ is known, although the attack instant set Υ cannot be directly observed, the expectation of the observed data can still be accurately decomposed and analytically characterized due to the stable statistical structure of Poisson replay attacks.

For the periodic position $i = 1, 2, \ldots, n$, define the statistic

$$\eta_i = \mathbb{E}(s_{(L-1)n+i}). \tag{3.3}$$

For the scenario in which the attack frequency is known, the received data s_k still follow the replay-attack model introduced in Section 2, where the attack instant set Υ is generated by a Poisson process, and the corresponding frequency parameter ϵ_i is available to the estimator. Therefore, s_k can be regarded as a randomly mixed observation sequence composed of unattacked samples s_k^0 and replayed samples $s_{k-\tau_k}^0$.

Theorem 3.2. *Under the conditions of Assumption* 2.1 *and Theorem* 3.1*, the observation statistic η_i satisfies*

$$\eta_i = (1 - \epsilon_i)\phi(C - \pi_i\theta) + \epsilon_i \sum_{j=1}^{m} p_j \phi(C - \pi_{i-j}\theta), \tag{3.4}$$

where $\phi(C - \pi_{i-j}\theta)$ denotes the value of the distribution function of the random noise d_k at the point $C - \pi_{i-j}\theta$, i.e.,

$$\phi(C - \pi_{i-j}\theta) = \Pr(d_k \le C - \pi_{i-j}\theta), \tag{3.5}$$

which characterizes the probability that the system output $y_k = \pi_{i-j}\theta + d_k$ is less than the threshold C when the replay attack delay is j.

Proof. According to the law of total probability, η_i can be decomposed as

$$\eta_i = \mathbb{E}(s_k | k \notin \Upsilon) \Pr(k \notin \Upsilon) + \mathbb{E}(s_k | k \in \Upsilon) \Pr(k \in \Upsilon). \tag{3.6}$$

First, consider the unattacked case $k \notin \Upsilon$. In this case, $s_k = s_k^0 = I_{\{y_k \le C\}}$. From the system model $y_k = \pi_i\theta + d_k$, we have

$$\mathbb{E}(s_k | k \notin \Upsilon) = \Pr(\pi_i\theta + d_k \le C) = \phi(C - \pi_i\theta).$$

Next, consider the attacked case $k \in \Upsilon$, where $s_k = s_{k-\tau_k}^0$. Taking the conditional expectation with respect to τ_k yields

$$\mathbb{E}(s_k | k \in \Upsilon) = \sum_{j=1}^{m} p_j \mathbb{E}(s_{k-j}^0).$$

Since the input satisfies the periodicity condition $\psi_k = \psi_{k+n}$, it follows that

$$\mathbb{E}(s_{k-j}^0) = \phi(C - \pi_{i-j}\theta).$$

Substituting the above result gives

$$\mathbb{E}(s_k | k \in \Upsilon) = \sum_{j=1}^{m} p_j \phi(C - \pi_{i-j}\theta).$$

Finally, combining

$$\Pr(k \in \Upsilon) = \epsilon_i, \quad \Pr(k \notin \Upsilon) = 1 - \epsilon_i,$$

and substituting them into (3.6), the conclusion follows directly. The proof is complete. □

Define the sample statistic

$$\hat{\eta}_{N,i} = \frac{1}{L_N} \sum_{L=1}^{L_N} s_{(L-1)n+i} \tag{3.7}$$

and construct the lagged mixture term

$$\hat{\nu}_{N,i} = \sum_{j=1}^{m} p_j \hat{\eta}_{N,i-j}. \tag{3.8}$$

Based on the above structure, define the compensated statistic

$$\widetilde{\eta}_{N,i} = \frac{\hat{\eta}_{N,i} - \epsilon_i \hat{\nu}_{N,i}}{1 - \epsilon_i}. \tag{3.9}$$

Theorem 3.3. *Under the condition of known attack frequency, it holds that*

$$\widetilde{\eta}_{N,i} \to \phi(C - \pi_i \theta), \quad w.p.1, \quad N \to \infty. \tag{3.10}$$

Proof. By the strong law of large numbers, we have

$$\hat{\eta}_{N,i} \xrightarrow{w.p.1} \eta_i.$$

Furthermore, according to the definition of the empirical compensation term,

$$\hat{\nu}_{N,i} = \sum_{j=1}^{m} p_j \hat{\eta}_{N,i-j} \xrightarrow{w.p.1} \bar{\nu}_i,$$

where

$$\bar{\nu}_i := \sum_{j=1}^{m} p_j \eta_{i-j}. \tag{3.11}$$

According to Theorem 3.2, the expectation of the received binary observation can be expressed as

$$\eta_i = (1 - \epsilon_i)\phi(C - \pi_i \theta) + \epsilon_i \bar{\nu}_i. \tag{3.12}$$

Therefore,

$$\begin{aligned} \widetilde{\eta}_{N,i} &\xrightarrow{w.p.1} \frac{\eta_i - \epsilon_i \bar{\nu}_i}{1 - \epsilon_i} \\ &= \frac{(1 - \epsilon_i)\phi(C - \pi_i \theta) + \epsilon_i \bar{\nu}_i - \epsilon_i \bar{\nu}_i}{1 - \epsilon_i} \\ &= \phi(C - \pi_i \theta). \end{aligned}$$

The proof is complete. □

Remark 3.1. *The quantity $\bar{\nu}_i$ represents the limiting value of the empirical compensation term $\widehat{\nu}_{N,i}$, since $\widehat{\nu}_{N,i}$ is constructed from the received data that may contain replay-corrupted samples. Therefore,*

$$\bar{\nu}_i = \sum_{j=1}^{m} p_j \eta_{i-j}.$$

It should be distinguished from the theoretical replay contribution term

$$\nu_i = \sum_{j=1}^{m} p_j \phi(C - \pi_{i-j}\theta), \tag{3.13}$$

which characterizes the contribution of replayed historical observations in the expectation model of Theorem 3.2. The two quantities have different roles in the analysis, and the compensation proof in Theorem 3.3 only requires the convergence of $\widehat{\nu}_{N,i}$ to $\bar{\nu}_i$.

3.3. Convergence analysis

Given prior knowledge of the attack frequencies ϵ_i, the remote estimator constructs the parameter estimate from the compensated statistics $\widetilde{\eta}_{N,i}$ after obtaining the observation sequence s_k. The estimation algorithm for the parameter θ is given by

$$\begin{aligned} \widehat{\theta}_N &= \Psi^{-1}[\zeta_{N,1}, \ldots, \zeta_{N,n}]^{\mathrm{T}}, \\ \zeta_{N,i} &= C - \phi^{-1}(\widetilde{\eta}_{N,i}), \quad i = 1, \ldots, n, \end{aligned} \tag{3.14}$$

where $\widetilde{\eta}_{N,i}$ is given by (3.9):

$$\widetilde{\eta}_{N,i} = \frac{\hat{\eta}_{N,i} - \epsilon_i \hat{\nu}_{N,i}}{1 - \epsilon_i}. \tag{3.15}$$

Theorem 3.4. *Under the assumption that the attack frequencies are known, the parameter estimator is strongly consistent, i.e.,*

$$\widehat{\theta}_N \xrightarrow{w.p.1} \theta, \quad N \to \infty. \tag{3.16}$$

Proof. By Theorem 3.3, it follows that

$$\widetilde{\eta}_{N,i} \xrightarrow{w.p.1} \phi(C - \pi_i\theta), \quad N \to \infty.$$

Since $\phi^{-1}(\cdot)$ is continuous on its domain, the continuous mapping theorem yields

$$\zeta_{N,i} = C - \phi^{-1}(\widetilde{\eta}_{N,i}) \xrightarrow{w.p.1} \pi_i\theta, \quad N \to \infty.$$

Therefore,

$$\widehat{\theta}_N = \Psi^{-1}\zeta_N \xrightarrow{w.p.1} \theta, \quad N \to \infty.$$

This completes the proof. □

3.4. Asymptotic normality analysis

Theorem 3.5. *Under Assumption* 2.1 *and assuming that the attack frequencies ϵ_i are known, the parameter estimator $\widehat{\theta}_N$ satisfies*

$$\sqrt{N}(\widehat{\theta}_N - \theta) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \Sigma_1), \quad N \to \infty, \tag{3.17}$$

where

$$\Sigma_1 = \Psi^{-1}\mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)\Psi^{-\mathrm{T}}, \tag{3.18}$$

with

$$\sigma_i^2 = \frac{n\,\mathrm{Var}(R_{L,i})}{\varphi_i^2}, \tag{3.19}$$

where $\varphi_i = \varphi(C - \pi_i\theta)$, $R_{L,i} = \frac{s_{(L-1)n+i} - \epsilon_i\nu_i}{1-\epsilon_i}$, and ν_i is given by Eq (3.13).

Proof. From the attack model described earlier and Theorem 3.2, at the i-th periodic position, we have

$$\mathbb{E}(s_{(L-1)n+i}) = (1 - \epsilon_i)\phi(C - \pi_i\theta) + \epsilon_i\nu_i.$$

To characterize the effects of Poisson attacks and replay delays on the observations, we define the standardized random variable

$$R_{L,i} = \frac{s_{(L-1)n+i} - \epsilon_i\nu_i}{1 - \epsilon_i},$$

whose expectation is

$$\begin{aligned}\mathbb{E}(R_{L,i}) &= \frac{\mathbb{E}(s_{(L-1)n+i}) - \epsilon_i\nu_i}{1 - \epsilon_i} \\ &= \frac{(1 - \epsilon_i)\phi(C - \pi_i\theta)}{1 - \epsilon_i} = \phi(C - \pi_i\theta).\end{aligned}$$

On the other hand, since $s_{(L-1)n+i} \in \{0, 1\}$, its second moment can be computed directly as

$$\eta_i = (1 - \epsilon_i)\phi(C - \pi_i\theta) + \epsilon_i\nu_i,$$

so that

$$\mathrm{Var}(s_{(L-1)n+i}) = \eta_i(1 - \eta_i).$$

Hence,

$$\mathrm{Var}(R_{L,i}) = \frac{\eta_i(1 - \eta_i)}{(1 - \epsilon_i)^2}.$$

Since the attack process Υ is modeled as a Poisson process, it possesses independent increments over different time windows. Moreover, the replay delays τ_k are independent and identically distributed. Therefore, the sequence $\{R_{L,i}\}_{L\geq 1}$ forms an independent and identically distributed (i.i.d.) sequence with finite second moments. Let

$$\bar{R}_{N,i} = \frac{1}{L_N}\sum_{L=1}^{L_N} R_{L,i}.$$

By the central limit theorem,

$$\sqrt{L_N}(\bar{R}_{N,i} - \phi(C - \pi_i\theta)) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \mathrm{Var}(R_{L,i})).$$

On the other hand, define the empirical statistic

$$\eta_{N,i} = \frac{1}{L_N} \sum_{L=1}^{L_N} s_{(L-1)n+i},$$

which satisfies

$$\bar{R}_{N,i} = \frac{\eta_{N,i} - \epsilon_i \nu_i}{1 - \epsilon_i}.$$

Therefore, the compensated statistic can be expressed as

$$\widetilde{\eta}_{N,i} = \frac{\eta_{N,i} - \epsilon_i \nu_i}{1 - \epsilon_i}.$$

Hence,

$$\sqrt{L_N}(\widetilde{\eta}_{N,i} - \phi(C - \pi_i\theta)) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \mathrm{Var}(R_{L,i})\right).$$

Since $L_N = \lfloor N/n \rfloor$ and $N/L_N \to n$ as $N \to \infty$, it follows that

$$\sqrt{N}(\widetilde{\eta}_{N,i} - \phi(C - \pi_i\theta)) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, n\,\mathrm{Var}(R_{L,i})\right).$$

Define the inverse-function estimator

$$\xi_{N,i} = C - \phi^{-1}(\widetilde{\eta}_{N,i}),$$

and note that

$$\phi^{-1}(\phi(C - \pi_i\theta)) = C - \pi_i\theta.$$

Since $\widetilde{\eta}_{N,i} \xrightarrow{p} \phi(C - \pi_i\theta)$ and $\phi^{-1}(\cdot)$ is differentiable at $\phi(C - \pi_i\theta)$, a first-order Taylor expansion of $\phi^{-1}(\cdot)$ around this point yields

$$\phi^{-1}(\widetilde{\eta}_{N,i}) = \phi^{-1}(\phi_i) + \frac{1}{\varphi_i}(\widetilde{\eta}_{N,i} - \phi_i) + o(|\widetilde{\eta}_{N,i} - \phi_i|).$$

Therefore,

$$\zeta_{N,i} - \pi_i\theta = C - \phi^{-1}(\widetilde{\eta}_{N,i}) - (C - \pi_i\theta) = -\frac{1}{\varphi_i}(\widetilde{\eta}_{N,i} - \phi_i) + o(|\widetilde{\eta}_{N,i} - \phi_i|).$$

Multiplying both sides by $\sqrt{N}$ gives

$$\sqrt{N}(\zeta_{N,i} - \pi_i\theta) = -\frac{1}{\varphi_i}\sqrt{N}(\widetilde{\eta}_{N,i} - \phi_i) + o_p(1).$$

Since it is known that

$$\sqrt{N}(\widetilde{\eta}_{N,i} - \phi_i) \xrightarrow{d} \mathcal{N}(0, n\,\mathrm{Var}(R_{L,i})),$$

applying the continuous mapping theorem and Slutsky's theorem yields

$$\sqrt{N}(\zeta_{N,i} - \pi_i\theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{n\,\mathrm{Var}(R_{L,i})}{\varphi_i^2}\right).$$

Moreover, from the asymptotic normality of $\widetilde{\eta}_{N,i}$ and the first-order Taylor expansion of $\phi^{-1}(\cdot)$, the asymptotic variance is determined jointly by the linearization coefficient and the variance of the original statistic, giving

$$\sigma_i^2 = \frac{n\,\mathrm{Var}(R_{L,i})}{\varphi_i^2}.$$

Finally, using the linear estimator

$$\widehat{\theta}_N = \Psi^{-1}[\zeta_{N,1}, \ldots, \zeta_{N,n}]^{\mathrm{T}},$$

we obtain

$$\sqrt{N}(\widehat{\theta}_N - \theta) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \Sigma_1),$$

where

$$\Sigma_1 = \Psi^{-1}\mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)\Psi^{-\mathrm{T}}.$$

This completes the proof. □

4. Unknown attack frequencies

When the attack frequencies are unknown, the estimation center cannot directly obtain the attack frequency parameters ϵ_i. Therefore, it is necessary to determine from long-term observation data whether the attacks exhibit a stable pattern and, furthermore, whether the corresponding frequencies can be estimated.

Let Υ denote the set of attack instants, and let $N_\Upsilon(t)$ represent the number of attacks occurring over the time interval $[0, t]$.

Intuitively, the estimability of the attack frequency depends on whether the number of attacks per unit time exhibits a stable asymptotic behavior.

Since the construction of the compensation term in the subsequent parameter estimation procedure relies on a statistical estimate of the attack frequency, it is essential to first ensure that the attack frequency can be consistently recovered from long-term observations. Motivated by this consideration, we next introduce the notion of attack frequency identifiability.

4.1. Identifiability of attack frequencies

In the field of cybersecurity, the detectability of an attack sequence is typically subject to the following conditions: (i) Regularity and persistence: the occurrence pattern should display sufficient repetitiveness and predictability. (ii) Distributional stability: the sequence should preserve consistent statistical properties across temporal or spatial dimensions. Only when the attack sequence simultaneously meets the basic requirements of regularity, persistence, and distribution consistency can its characteristic patterns be reliably extracted. These conditions together constitute the theoretical foundation for the detectability of attack sequences.

In practical systems, attack processes are often modeled as random point processes. In the classical setting, the attack arrival process can be further described by a Poisson-type stochastic process, whose statistical characteristics are mainly determined by its arrival intensity. Depending on whether the intensity function remains stable, the attack process typically exhibits the following two representative scenarios:

- One scenario is that the attack process possesses a stable arrival intensity, meaning that the average number of attacks per unit time remains constant.
- The other scenario is that the attack intensity varies significantly over time, causing attack events to exhibit pronounced clustering within certain local time intervals.

In the former case, since the attack process has a stable average arrival rate, its statistical frequency can be consistently recovered through long-term observations. In contrast, in the latter case, the attack intensity exhibits significant time-varying characteristics, and the statistical behavior may differ substantially across different time intervals, thereby reducing the stability of attack frequency estimation.

The identifiability of the attack frequency serves as a fundamental prerequisite for the subsequent parameter estimation procedure. Since both the compensated statistics and the parameter estimators are constructed based on empirical averages, it is essential that the attack frequency can be consistently recovered from long-term observations. If the attack frequency is not identifiable, the empirical statistics may fail to converge to stable limits, leading to systematic bias. The asymptotic normality of unidentifiable sequences is likewise unidentifiable.

Definition 4.1 (Identifiability of attack frequencies)**.** *If there exists a constant* $\epsilon \in (0, 1)$ *such that*

$$\frac{N_\Upsilon(t)}{t} \to \epsilon, \quad w.p.1 \text{ as } t \to \infty, \tag{4.1}$$

then the attack process is said to be globally identifiable, and ϵ *is referred to as the attack frequency.*

Furthermore, if for any deterministic time interval $[T_1, T_2]$ *with* $T_2 - T_1 \to \infty$ *it holds that*

$$\frac{N_\Upsilon(T_2) - N_\Upsilon(T_1)}{T_2 - T_1} \to \epsilon, \quad w.p.1 \text{ as } t \to \infty, \tag{4.2}$$

then the attack process is said to satisfy local uniform identifiability.

The above definition indicates that an attack frequency is not only required to have a stable limit over the entire time scale but also to maintain consistent statistical behavior over any sufficiently long local time interval. Only under this dual condition—global identifiability and local uniform identifiability—can the attack frequency be reliably estimated from long-term statistical observations.

Further discussion

The identifiability of the attack frequency serves as a fundamental prerequisite for the subsequent parameter-estimation procedure. Since both the compensated statistics and the parameter estimators are constructed based on empirical averages, it is essential that the attack frequency can be consistently recovered from long-term observations.

If the attack frequency is not identifiable, the statistical characteristics observed over different time intervals may differ significantly. As a result, the corresponding empirical statistics may fail to converge to stable limits, leading to systematic bias in estimators constructed from sample-average approximations. Therefore, identifiability essentially guarantees that time averages accurately reflect the underlying statistical properties of the attack process.

Example 1: Non-stationary Poisson attack (Non-identifiable case)

Consider a class of non-homogeneous Poisson attack processes whose arrival intensity function is given by

$$\lambda(t) = \begin{cases} 1, & t \in [k^3, k^3 + k^{1/2}], \\ 0, & \text{otherwise.} \end{cases}$$

That is, attacks occur only within local time windows, while no attacks occur at other times. Consequently, the attack process exhibits pronounced local clustering along the time axis.

Since the length of the k-th clustering interval is only $k^{1/2}$, whereas its corresponding time position is approximately k^3, the proportion of the clustering interval relative to the overall timeline approaches zero as time increases. Therefore, we have

$$\frac{N_\Upsilon(t)}{t} \to 0, \quad t \to \infty.$$

This indicates that, in terms of the global average, the long-term attack frequency tends to zero.

However, within each local clustering interval, the attack intensity per unit time satisfies

$$\frac{N_\Upsilon(k^3 + k^{1/2}) - N_\Upsilon(k^3)}{k^{1/2}} \xrightarrow{\text{w.p.1}} 1.$$

Hence, there is a clear discrepancy between the global and local statistics: the global average tends to zero, while the local time intervals still exhibit high attack intensity. As a result, it is impossible to describe the long-term statistical behavior with a single stable parameter, and the attack frequency does not satisfy the identifiability condition.

Example 2: Stationary Poisson attack (Identifiable case)

Consider another class of attack processes, where the number of attacks satisfies

$$N_\Upsilon(t) \sim \text{Poisson}(\lambda t), \quad \lambda > 0.$$

That is, the attack process follows a homogeneous Poisson process with intensity λ.

By the law of large numbers for Poisson processes, we have

$$\frac{N_\Upsilon(t)}{t} \xrightarrow{\text{w.p.1}} \lambda, \quad t \to \infty,$$

indicating that the average number of attacks per unit time converges stably to the constant λ.

Furthermore, for any deterministic time interval $[T_1, T_2]$, since a Poisson process has stationary and independent increments, its increment satisfies

$$N_\Upsilon(T_2) - N_\Upsilon(T_1) \sim \text{Poisson}(\lambda(T_2 - T_1)).$$

Thus, we have

$$\frac{N_\Upsilon(T_2) - N_\Upsilon(T_1)}{T_2 - T_1} \xrightarrow{\text{w.p.1}} \lambda, \quad T_2 - T_1 \to \infty.$$

This indicates that, regardless of which sufficiently long time interval is observed, the statistical frequency remains consistent.

Moreover, let the inter-arrival time between consecutive attacks be defined as $D_k = \gamma_{k+1} - \gamma_k$, then, for a homogeneous Poisson process,

$$D_k \sim \text{Exp}(\lambda),$$

with

$$\mathbb{E}(D_k) = \frac{1}{\lambda}.$$

By the law of large numbers, it further follows that

$$\frac{1}{m}\sum_{k=1}^{m} D_k \xrightarrow{\text{w.p.1}} \frac{1}{\lambda}.$$

Therefore, this attack process not only has a stable limit in terms of overall frequency, but its local statistical behavior and inter-attack interval structure are also consistent. Consequently, the long-term statistical behavior can be characterized using the single parameter λ, and the attack process satisfies the identifiability condition for attack frequency.

When the attack process meets the identifiability condition, the attack frequency can be regarded as having a stable statistical limit ϵ_i. Accordingly, an attack frequency estimator can be constructed using empirical statistics:

$$\widehat{\epsilon}_{N,i} = \frac{\#\Upsilon_{N,i}}{L_N},$$

where $\#\Upsilon_{N,i}$ denotes the number of times the i-th position is attacked within the first L_N periodic blocks.

If the attack process satisfies the identifiability condition, then

$$\widehat{\epsilon}_{N,i} \xrightarrow{\text{w.p.1}} \epsilon_i, \quad N \to \infty.$$

These results indicate that the true attack frequency can be stably recovered from long-term observations, which provides a theoretical foundation for the construction of compensated statistics and the analysis of parameter consistency in subsequent sections.

4.2. Design of the compensation algorithm

When the attack frequencies are unknown, the estimation center cannot directly obtain the attack frequency parameters ϵ_i, and therefore cannot directly use the previously defined compensated statistics to perform parameter identification.

To address this issue, this paper proposes an empirical-statistics-based compensation-identification algorithm. The core idea is as follows: All observation data are divided into two parts, where one part is specifically used for attack frequency estimation, and the other part is used for system parameter identification. By constructing attack-event statistics from specially selected data points, an initial estimate of the attack frequency is obtained. This estimate is then used to construct compensated statistics, which in turn allows consistent identification of the system parameters.

The compensation mechanism in the unknown attack-frequency case involves two key components. (i) Attack frequency estimation via probing: The data set $\mathcal{D}_N^{(1)}$ consists of specially constructed positions where the transmitter sends a fixed symbol "1". Since the attacker can only replace the current data with historical data, receiving $s_k \neq 1$ at these positions provides definitive evidence of an attack. By

counting such events, the attack frequency ϵ_i can be consistently estimated. (ii) Parameter estimation via compensated statistics: using the estimated attack frequency $\widehat{\epsilon}_{N,i}$, a compensated statistic is constructed that mirrors the known-frequency case. The consistency of $\widehat{\epsilon}_{N,i}$ ensures convergence to the true value.

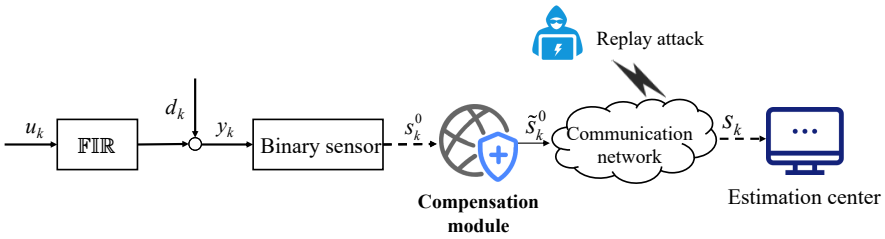


Figure 2. System structure with compensation module.

Let the maximum attack lag length be m, and define $Z = m + 1$. Furthermore, introduce a proportion parameter

$$\alpha \in (0, 1), \tag{4.3}$$

where α represents the proportion of data used for attack frequency estimation, and $1 - \alpha$ represents the proportion of data used for system parameter identification. Let the total sample length be N, and denote by M_N the number of positions used for attack frequency estimation, satisfying $\frac{2ZM_N}{N} \to \alpha, \quad N \to \infty$. Let the original quantized output be s_k^0, and denote the transmitted data after compensation as $\widetilde{s}_k^0$.

First, select M_N positions $r_1, r_2, \ldots, r_{M_N}$ from the set $\mathcal{T}_N = \{1, 2, \ldots, N\}$, satisfying

$$r_{\ell+1} - r_\ell > 2Z, \quad \ell = 1, 2, \ldots, M_N, \tag{4.4}$$

with $r_{M_N+1} = N$.

Based on the above positions, construct the set

$$\mathcal{D}_N = \bigcup_{\ell=1}^{M_N} \mathcal{B}_\ell,$$

where $\mathcal{B}_\ell = \{r_\ell, r_\ell + 1, \ldots, r_\ell + 2Z - 1\}$.

Further, define the transmitted data $\widetilde{s}_k^0$ as follows:

$$\widetilde{s}_k^0 = \begin{cases} 0, & k \in \{r_\ell, \ldots, r_\ell + Z - 2\} \text{ for some } \ell, \\ 1, & k = r_\ell + Z - 1 \text{ for some } \ell, \\ s_k^0, & k = r_\ell + Z, \ldots, r_\ell + 2Z - 1 \text{ for some } \ell, \\ s_k^0, & k \notin \mathcal{D}_N. \end{cases} \tag{4.5}$$

After the defense processing module is introduced, the attacker targets the modified sequence $\widetilde{s}_k^0$, and the received data for $k \in \Upsilon$ satisfy $s_k = \widetilde{s}_{k-\tau_k}^0$.

The above procedure is illustrated in Figure 3.

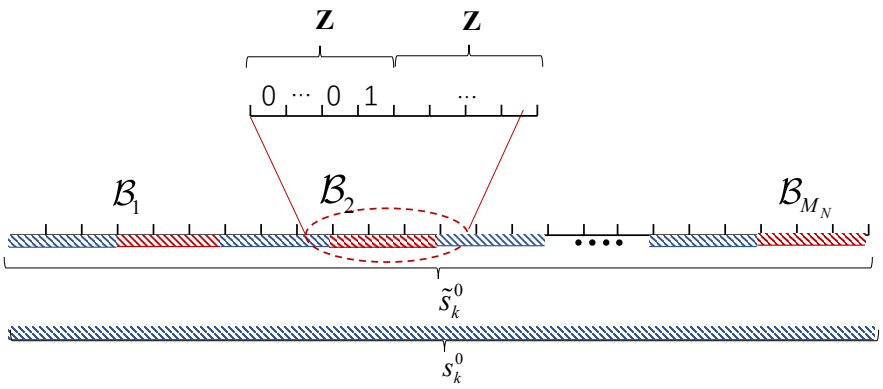


Figure 3. Schematic of the compensation procedure.

Let the number of elements in the set $\mathcal{D}_N$ be $D_N = 2M_N Z$. Since $\frac{D_N}{N} = \frac{2ZM_N}{N} \to \alpha$, asymptotically, approximately αN data points are used for attack frequency estimation, while the remaining $(1-\alpha)N$ data points are used for system parameter identification.

The core purpose of this construction is to establish a statistical relationship between the attack frequency and the received data.

Since attacks can only replace historical data, the transmitter sends a fixed "1" at $k = r_\ell + Z - 1$. If the estimation center ultimately receives $s_k \neq 1$, this indicates that an attack must have occurred at that moment. Therefore, the proportion of attacks can be statistically estimated using these special positions, allowing estimation of the attack frequency.

On the other hand, for the interval $k = r_\ell + Z, \ldots, r_\ell + 2Z - 1$, the data are not subjected to additional processing, so that the normal statistical structure is preserved as much as possible.

Based on the above construction, define the set of positions for attack frequency estimation as

$$\mathcal{D}_N^{(1)} = \{r_\ell + Z - 1\}_{\ell=1}^{M_N} \tag{4.6}$$

and the set of data used for system identification as

$$\mathcal{D}_N^{(2)} = \mathcal{T}_N \setminus \mathcal{D}_N. \tag{4.7}$$

Using the sets $\mathcal{D}_N^{(1)}$ and $\mathcal{D}_N^{(2)}$, the following parameter-estimation algorithm can be obtained:

$$\begin{aligned}
\widehat{\theta}_N &= \Psi^{-1}[\chi_{N,1}, \ldots, \chi_{N,n}]^{\mathrm{T}}, \\
\chi_{N,i} &= C - \phi^{-1}\Big(\frac{o_i - \widehat{\epsilon}_{N,i}\sum_{j=1}^{m} p_j \widehat{\eta}_{N,i}^{(j)}}{1 - \widehat{\epsilon}_{N,i}}\Big), \\
o_i &= \frac{1}{L_N}\sum_{L=1}^{L_N} I_{\{(L-1)n+i \in \mathcal{D}_N^{(2)}\}} s_{(L-1)n+i}, \\
\widehat{\epsilon}_{N,i} &= \frac{2Z\sum_{L=1}^{L_N} I_{\{(L-1)n+i \in \mathcal{D}_N^{(1)}\}}(1 - s_{(L-1)n+i})}{D_N}.
\end{aligned} \tag{4.8}$$

Here, $\widehat{\epsilon}_{N,i}$ represents the estimated attack frequency. Since the positions in the set $\mathcal{D}_N^{(1)}$ are fixed to transmit "1", the term $1 - s_{(L-1)n+i}$ precisely indicates whether an attack occurred at that position.

$\widehat{\eta}_{N,i}^{(j)}$ denotes the empirical statistic corresponding to an attack lag of j steps, and its definition coincides with that adopted for the known attack frequency scenario. Since $\widehat{\eta}_{N,i}^{(j)}$ is constructed directly from historical received data, the above compensated statistic depends solely on observable data.

The above algorithm first estimates the attack frequency using approximately αN specially selected data points, then performs system parameter identification using the remaining approximately $(1-\alpha)N$ data points. Through the compensation mechanism, the true statistical structure is restored, thereby achieving asymptotically consistent identification of the parameters of the binary quantized system when the attack frequency is unknown.

Remark 4.1. *In the present framework, the replay delay distribution* $\mathbf{p}$ *is assumed to be known. If* $\mathbf{p}$ *is also unknown, jointly estimating* $\mathbf{p}$*,* ϵ_i*, and* θ *would require additional structural assumptions or a multi-stage identification procedure. This extension is left as a promising direction for future work.*

4.3. Convergence analysis

Theorem 4.1. *Suppose that Assumption* 2.1 *holds and that the attack frequencies satisfy the identifiability condition. Then, the parameter estimator* $\widehat{\theta}_N$ *obtained from the above compensation algorithm satisfies*

$$\widehat{\theta}_N \to \theta, \quad w.p.1, \quad N \to \infty. \tag{4.9}$$

Proof. We first establish the strong consistency of the attack frequency estimator $\widehat{\epsilon}_{N,i}$.

For

$$(L-1)n + i \in \mathcal{D}_N^{(1)},$$

according to the construction of the compensation algorithm, the transmitter sends the fixed symbol "1", namely, $\widehat{s}^0_{(L-1)n+i} = 1$.

Since an attack can only replace the transmitted data with historical data, the received data $s_{(L-1)n+i}$ satisfies the following:

- If no attack occurs, then $s_{(L-1)n+i} = 1$;
- If an attack occurs, then $s_{(L-1)n+i} \neq 1$.

Therefore, by the law of total probability,

$$\begin{aligned} & \Pr\{s_{(L-1)n+i} = 1\} \\ = \; & \Pr\{s_{(L-1)n+i} = 1 | \widehat{s}^0_{(L-1)n+i} = 1\} \Pr\{\widehat{s}^0_{(L-1)n+i} = 1\} \\ & + \Pr\{s_{(L-1)n+i} = 1 | \widehat{s}^0_{(L-1)n+i} = 0\} \Pr\{\widehat{s}^0_{(L-1)n+i} = 0\}. \end{aligned}$$

Since $\widehat{s}^0_{(L-1)n+i} = 1$, it follows that

$$\Pr\{\widehat{s}^0_{(L-1)n+i} = 1\} = 1$$

and

$$\Pr\{\widehat{s}^0_{(L-1)n+i} = 0\} = 0.$$

Hence, we have

$$\begin{aligned} \Pr\{s_{(L-1)n+i} = 1\} &= \Pr\{s_{(L-1)n+i} = 1 | \widehat{s}^0_{(L-1)n+i} = 1\} \\ &= 1 - \epsilon_i. \end{aligned}$$

Therefore, $\epsilon_i = 1 - \Pr\{s_{(L-1)n+i} = 1\}$. On the other hand, according to the definition of the attack frequency estimator, we have

$$\widehat{\epsilon}_{N,i} = \frac{2Z \sum_{L=1}^{L_N} I_{\{(L-1)n+i \in \mathcal{D}_N^{(1)}\}}(1 - s_{(L-1)n+i})}{D_N}.$$

Since $1 - s_{(L-1)n+i}$ is an independent and identically distributed Bernoulli random variable, its expectation is given by

$$\mathbb{E}[1 - s_{(L-1)n+i}] = \epsilon_i,$$

thus, by the strong law of large numbers, we have

$$\frac{2Z \sum_{L=1}^{L_N} I_{\{(L-1)n+i \in \mathcal{D}_N^{(1)}\}}(1 - s_{(L-1)n+i})}{D_N} \to \epsilon_i, \quad w.p.1, \quad N \to \infty.$$

That is,

$$\widehat{\epsilon}_{N,i} \to \epsilon_i, \quad w.p.1, \quad N \to \infty.$$

Next, consider the compensated statistic $\widetilde{\eta}_{N,i}$. Define

$$\widetilde{\eta}_{N,i} = \frac{o_i - \widehat{\epsilon}_{N,i} \sum_{j=1}^{m} p_j \widehat{\eta}_{N,i}^{(j)}}{1 - \widehat{\epsilon}_{N,i}}.$$

From the consistency results under the known attack frequency case, the empirical statistics satisfy

$$o_i \to \eta_i, \quad w.p.1, \quad N \to \infty,$$

and

$$\widehat{\eta}_{N,i}^{(j)} \to \phi(C - \pi_{i,j}\theta), \quad w.p.1, \quad N \to \infty.$$

Combining this with

$$\widehat{\epsilon}_{N,i} \to \epsilon_i, \quad w.p.1, \quad N \to \infty,$$

we obtain

$$\begin{aligned} \widetilde{\eta}_{N,i} &\to \frac{\eta_i - \epsilon_i \sum_{j=1}^{m} p_j \phi(C - \pi_{i,j}\theta)}{1 - \epsilon_i} \\ &= \phi(C - \pi_i\theta), \quad w.p.1, \quad N \to \infty. \end{aligned}$$

Therefore,

$$C - \phi^{-1}(\widetilde{\eta}_{N,i}) \to \pi_i\theta, \quad w.p.1, \quad N \to \infty.$$

Finally, from the definition of the parameter estimator

$$\widehat{\theta}_N = \Psi^{-1}[\chi_{N,1}, \ldots, \chi_{N,n}]^{\mathrm{T}}$$

and by the continuous mapping theorem, we have

$$\widehat{\theta}_N \to \theta, \quad w.p.1, \quad N \to \infty.$$

This completes the proof of the theorem. □

4.4. Demonstration of asymptotic normality

Theorem 4.2. *For the compensated estimation method introduced, the estimator $\widehat{\theta}_N$ exhibits the following asymptotic normal behavior:*

$$\sqrt{N}(\widehat{\theta}_N - \theta) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \Sigma(\alpha)), \quad N \to \infty, \tag{4.10}$$

where the covariance matrix is given by

$$\Sigma(\alpha) = \Psi^{-1}\mathrm{diag}(\sigma_1^2(\alpha), \ldots, \sigma_n^2(\alpha))\Psi^{-\mathrm{T}}, \tag{4.11}$$

and for each $i = 1, 2, \ldots, n$,

$$\sigma_i^2(\alpha) = \frac{1}{\varphi_i^2(1-\epsilon_i)^2}\left[\frac{\eta_i(1-\eta_i)}{1-\alpha} + \frac{(\eta_i - \nu_i)^2\epsilon_i(1-\epsilon_i)}{\alpha}\right]. \tag{4.12}$$

Proof. For each channel $i = 1, 2, \ldots, n$, consider the following two-part sample splitting:

The number of samples used for attack frequency estimation is $D_N \sim \alpha N$, while the number of samples used for system identification is $(1-\alpha)N$.

At the specially constructed positions, $1 - s_{(L-1)n+i}$ is an independent and identically distributed Bernoulli random variable with

$$\mathbb{E}[1 - s_{(L-1)n+i}] = \epsilon_i.$$

Since the attack frequency estimator is based on D_N samples, by the central limit theorem we obtain:

$$\sqrt{D_N}(\widehat{\epsilon}_{N,i} - \epsilon_i) \xrightarrow{d} \mathcal{N}(0, \epsilon_i(1-\epsilon_i)).$$

Since $D_N/N \to \alpha$, an equivalent form is

$$\sqrt{N}(\widehat{\epsilon}_{N,i} - \epsilon_i) \xrightarrow{d} \mathcal{N}\left(0, \frac{\epsilon_i(1-\epsilon_i)}{\alpha}\right).$$

The system identification part uses $(1-\alpha)N$ samples, hence by the law of large numbers and the central limit theorem,

$$\sqrt{N}(\eta_{N,i} - \eta_i) \xrightarrow{d} \mathcal{N}\left(0, \frac{\eta_i(1-\eta_i)}{1-\alpha}\right).$$

The compensated statistic is defined as

$$\widetilde{\eta}_{N,i} = \frac{\eta_{N,i} - \widehat{\epsilon}_{N,i}\nu_i}{1 - \widehat{\epsilon}_{N,i}}, \quad \nu_i = \sum_{j=1}^{m} p_j\phi(C - \pi_{i,j}\theta).$$

By a first-order Taylor expansion, we obtain

$$\widetilde{\eta}_{N,i} - \phi(C - \pi_i\theta) = \frac{\eta_{N,i} - \eta_i}{1-\epsilon_i} + \frac{(\widehat{\epsilon}_{N,i} - \epsilon_i)(\eta_i - \nu_i)}{(1-\epsilon_i)^2} + o_p(N^{-1/2}).$$

By independence and Slutsky's theorem, the two error terms jointly determine the asymptotic distribution, and thus

$$\sqrt{N}(\widetilde{\eta}_{N,i} - \phi(C - \pi_i\theta)) \xrightarrow{d} \mathcal{N}(0, \tau_i^2),$$

where

$$\tau_i^2 = \frac{\eta_i(1-\eta_i)}{(1-\epsilon_i)^2(1-\alpha)} + \frac{(\eta_i-\nu_i)^2\epsilon_i(1-\epsilon_i)}{(1-\epsilon_i)^2\alpha}.$$

Expanding $\phi^{-1}(\cdot)$ at $\phi(C-\pi_i\theta)$ yields

$$\sqrt{N}(\chi_{N,i}-\pi_i\theta) = \frac{1}{\varphi_i}\sqrt{N}(\widetilde{\eta}_{N,i}-\phi(C-\pi_i\theta)) + o_p(1).$$

Therefore,

$$\sqrt{N}(\chi_{N,i}-\pi_i\theta) \xrightarrow{d} \mathcal{N}(0,\sigma_i^2), \quad \sigma_i^2 = \frac{\tau_i^2}{\varphi_i^2}.$$

Finally, from the linear transformation

$$\widehat{\theta}_N = \Psi^{-1}[\chi_{N,1},\ldots,\chi_{N,n}]^{\mathrm{T}},$$

and by the Cramér–Wold theorem, we obtain

$$\sqrt{N}(\widehat{\theta}_N-\theta) \xrightarrow{d} \mathcal{N}(0,\Sigma(\alpha)).$$

This completes the proof. □

4.5. Optimality of the parameter α

To characterize the data allocation between attack frequency estimation and system parameter identification, we introduce a proportion parameter $\alpha \in (0,1)$. Here, α represents the proportion of data used for attack frequency estimation, while $1-\alpha$ represents the proportion of data used for system parameter identification.

Theorem 4.3. *Under the conditions that the attack frequency is identifiable and satisfies the uniform convergence condition, the asymptotic covariance matrix of the compensated estimator $\widehat{\theta}_N$ can be expressed as*

$$\Sigma(\alpha) = \Psi^{-1}\mathrm{diag}(\sigma_1^2(\alpha),\ldots,\sigma_n^2(\alpha))\Psi^{-\mathrm{T}}, \tag{4.13}$$

where

$$\sigma_i^2(\alpha) = \frac{1}{\varphi_i^2(1-\epsilon_i)^2}\left(\frac{A_i}{\alpha}+\frac{B_i}{1-\alpha}\right). \tag{4.14}$$

Here,

$$A_i = \epsilon_i(1-\epsilon_i)\,(\nu_i-\eta_i)^2, \quad B_i = n\,\eta_i(1-\eta_i), \tag{4.15}$$

and the optimal weight α^* that minimizes the overall error $\mathrm{tr}(\Sigma(\alpha))$ *is*

$$\alpha^* = \frac{\sqrt{A}}{\sqrt{A}+\sqrt{B}}, \tag{4.16}$$

with

$$A = \sum_{i=1}^{n}\frac{\epsilon_i(1-\epsilon_i)(\nu_i-\eta_i)^2}{\varphi_i^2(1-\epsilon_i)^2}, \tag{4.17}$$

$$B = \sum_{i=1}^{n}\frac{n\,\eta_i(1-\eta_i)}{\varphi_i^2(1-\epsilon_i)^2}. \tag{4.18}$$

Proof. From the theorem conditions, the asymptotic covariance matrix of the compensated estimator is given by

$$\Sigma(\alpha) = \Psi^{-1}\mathrm{diag}(\sigma_1^2(\alpha), \ldots, \sigma_n^2(\alpha))\Psi^{-\mathrm{T}}.$$

Hence, the overall estimation error can be characterized by the trace function:

$$\mathrm{tr}(\Sigma(\alpha)) = \sum_{i=1}^{n} \sigma_i^2(\alpha)\, \nu_i^{\mathrm{T}}\nu_i,$$

where ν_i denotes the i-th column of the matrix Ψ^{-1}. Substituting $\sigma_i^2(\alpha)$ yields

$$\mathrm{tr}(\Sigma(\alpha)) = \sum_{i=1}^{n} \frac{\nu_i^{\mathrm{T}}\nu_i}{\varphi_i^2(1-\epsilon_i)^2}\left(\frac{A_i}{\alpha} + \frac{B_i}{1-\alpha}\right).$$

Define

$$A = \sum_{i=1}^{n} \frac{A_i\, \nu_i^{\mathrm{T}}\nu_i}{\varphi_i^2(1-\epsilon_i)^2}, \quad B = \sum_{i=1}^{n} \frac{B_i\, \nu_i^{\mathrm{T}}\nu_i}{\varphi_i^2(1-\epsilon_i)^2},$$

then we obtain

$$\mathrm{tr}(\Sigma(\alpha)) = \frac{A}{\alpha} + \frac{B}{1-\alpha}.$$

Taking the derivative with respect to α gives

$$\frac{d}{d\alpha}\mathrm{tr}(\Sigma(\alpha)) = -\frac{A}{\alpha^2} + \frac{B}{(1-\alpha)^2}.$$

Setting the derivative to zero yields $\frac{A}{\alpha^2} = \frac{B}{(1-\alpha)^2}$. Taking the positive square root (since $\alpha \in (0,1)$), we obtain $\frac{\sqrt{A}}{\alpha} = \frac{\sqrt{B}}{1-\alpha}$. Rearranging gives the optimal solution

$$\alpha^* = \frac{\sqrt{A}}{\sqrt{A} + \sqrt{B}}.$$

On the other hand, the function $\frac{A}{\alpha} + \frac{B}{1-\alpha}$ is strictly convex on $\alpha \in (0,1)$, hence this stationary point is the global minimizer. This completes the proof. □

5. Numerical simulations

Consider a single-input single-output second-order discrete-time system:

$$\begin{aligned} y_k &= \pi_k^{\mathrm{T}}\theta + d_k, \\ s_k &= I_{\{y_k \le C\}}, \quad k = 1, \ldots, N, \end{aligned} \tag{5.1}$$

where the vector of unknown parameters is defined as $\theta = [3, -1]^{\mathrm{T}}$.

A deterministic periodic input with period length 3 is considered, whose values within each cycle are $\psi_1 = [1, 2], \quad \psi_2 = [2, 1]$.

We set the threshold as $C = 0$, the data length as $N = 10,000$, and the noise satisfies $d_k \sim \mathcal{N}(0, 10^2)$, which is assumed to be i.i.d. as described in the previous sections.

The attack process is modeled as a Poisson point process. Let the attack event set Υ satisfy $N_\Upsilon(t) \sim \mathrm{Poisson}(\lambda t)$, where λ denotes the attack intensity parameter.

5.1. *Verification of attack frequency identifiability*

According to Definition 4.1, we construct two types of Poisson attack processes that satisfy and violate the identifiability condition, respectively.

First, consider a homogeneous Poisson attack process with constant intensity. The number of attack events satisfies $N_\Upsilon(t) \sim \text{Poisson}(\lambda t), \quad \lambda = \frac{1}{3}$. Since this process has a stable average arrival rate, it follows that $\frac{N_\Upsilon(t)}{t} \xrightarrow{\text{w.p.1}} \frac{1}{3}, \quad t \to \infty$. Furthermore, consider a non-homogeneous Poisson attack process with local clustering behavior, whose intensity function is defined as $\lambda(t) = \begin{cases} 1, & t \in [k^3, k^3 + k^{1/2}], \\ 0, & \text{otherwise.} \end{cases}$ This attack process occurs only within local time windows, while being almost absent in other intervals, leading to a clearly non-uniform clustering structure along the time axis. In numerical experiments, for the purpose of comparison with the homogeneous Poisson attack process, the same average reference attack frequency $\lambda = 1/3$ is used as a baseline.

Based on the proposed attack frequency compensation algorithm, the attack frequency estimate $\widehat{\epsilon}_{N,i}$ is identified, and the simulation results are shown in Figures 4 and 5.

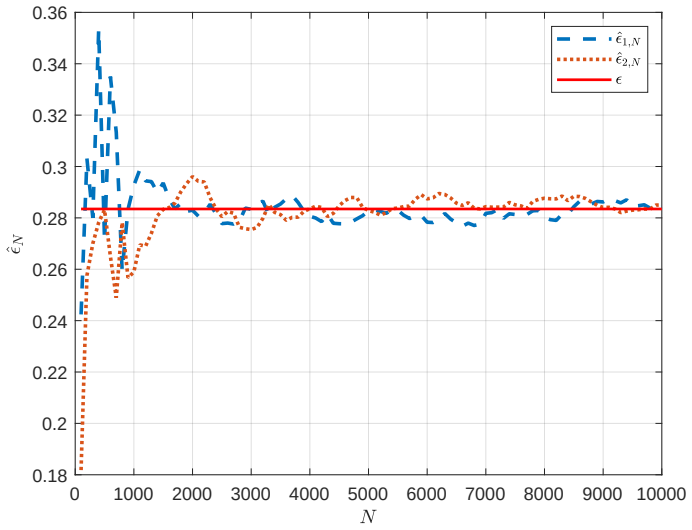


Figure 4. Attack frequency estimation results under a homogeneous Poisson attack process.

From Figure 4, it can be seen that under the homogeneous Poisson attack setting, the estimated attack frequency gradually converges to the true value $1 - e^{-\lambda} = 0.284$ as the data length increases. This indicates that the attack process satisfies a stable statistical law, and its attack frequency can be reliably recovered through long-term observations.

In contrast, Figure 5 shows that under the non-homogeneous Poisson attack setting, due to significant variations in attack intensity across different time intervals, the estimated attack frequency exhibits large fluctuations and fails to converge to a constant value. This demonstrates that the local statistical behavior of the attack process is inconsistent with its global statistical behavior, thereby verifying the important role of attack frequency identifiability in ensuring subsequent consistent parameter estimation.

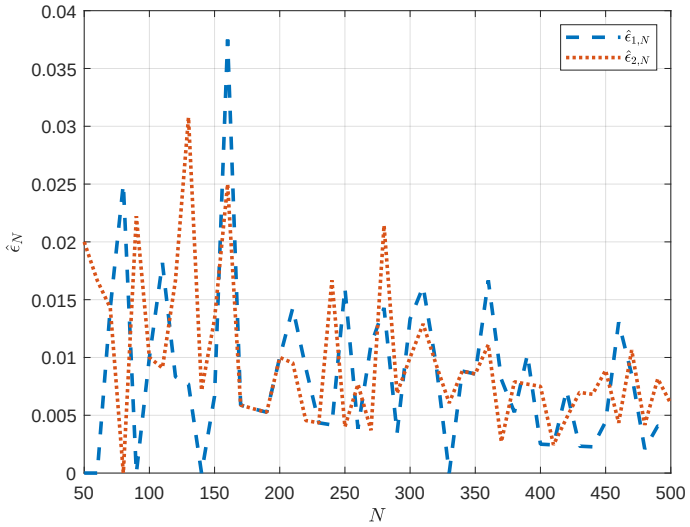


Figure 5. Attack frequency estimation results under a non-homogeneous Poisson attack process.

5.2. Known attack frequency

Consider a discrete-time system subject to replay attacks during transmission. Assume that at each sampling instant k, an attack occurs independently with known probabilities $\epsilon = [\epsilon_1, \epsilon_2]$, where $\epsilon_1 = 0.3$, $\epsilon_2 = 0.6$. We set the data length as $N = 10,000$ and run the proposed compensated identification algorithm under this setting. As the sample size increases, the estimator $\widehat{\theta}_N$ gradually stabilizes under stochastic perturbations and eventually converges to the true parameter $\theta = [3, -1]^T$.

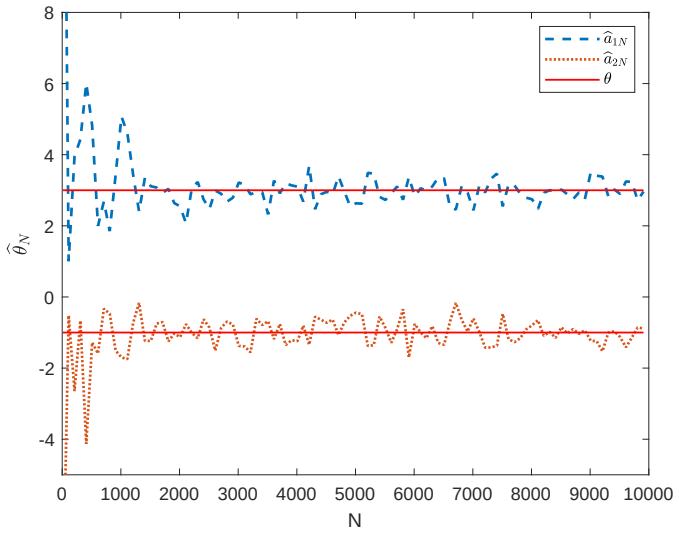
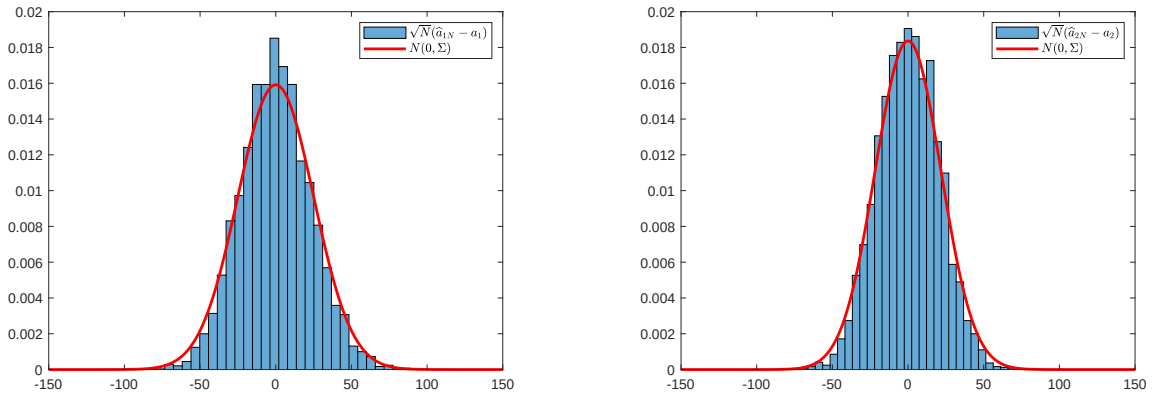


Figure 6. Convergence behavior of the algorithm.

As shown in Figure 6, the parameter estimation trajectory exhibits noticeable fluctuations at the initial stage due to the combined effect of finite-sample random attack disturbances and measurement noise. As N increases, this effect gradually diminishes, the estimation error converges significantly, and

eventually stabilizes. This demonstrates that the proposed algorithm has good convergence performance and numerical stability under discrete attack environments.

To further verify the theoretical results, a Monte Carlo simulation with 5000 runs is performed for statistical analysis to study the asymptotic distribution of the estimator $\widehat{\theta_N}$.



(a) Asymptotic normality of the $\hat{a}_1$

(b) Asymptotic normality of the $\hat{a}_2$

Figure 7. Asymptotic normality of $\hat{\theta}_N$.

As shown in Figure 7, the empirical histograms of the parameter components $\hat{a}_1$ and $\hat{a}_2$ are presented, respectively. It can be observed that, under sufficiently large sample sizes, the distributions of all components exhibit clear symmetry and gradually concentrate around zero mean.

Furthermore, the normalized estimation errors are analyzed, $\sqrt{N}(\widehat{a}_{i,N} - a_i), \quad i = 1, 2$ whose empirical distributions are highly consistent with the Gaussian distribution $\mathcal{N}(0, \Sigma_1)$, where the covariance matrix Σ_1 is given by the theoretical result. This indicates that the estimator not only converges in probability, but also its fluctuation structure satisfies asymptotic normality.

Therefore, the numerical results are consistent with the theoretical conclusion of Theorem 3.4, which verifies the asymptotic properties of the proposed estimation method from a simulation perspective.

5.3. Robustness analysis under different conditions

To further verify the robustness and applicability of the proposed algorithm, additional simulations are conducted under different attack frequencies, replay delay distributions, and noise levels.

Different attack frequencies ϵ: Simulations with three attack frequency vectors are performed: $\epsilon = [0.1,\ 0.1]^{\mathrm{T}}$ (light attacks), $\epsilon = [0.3,\ 0.6]^{\mathrm{T}}$ (moderate attacks), and $\epsilon = [0.7,\ 0.9]^{\mathrm{T}}$ (heavy attacks). The results are shown in Figure 8. The parameter estimates $\widehat{\theta_N}$ converge to the true value $\theta = [3,\ -1]^{\mathrm{T}}$ under all three attack intensities. The convergence speed decreases with increasing attack frequency as expected, yet asymptotic consistency is preserved even under heavy attacks.

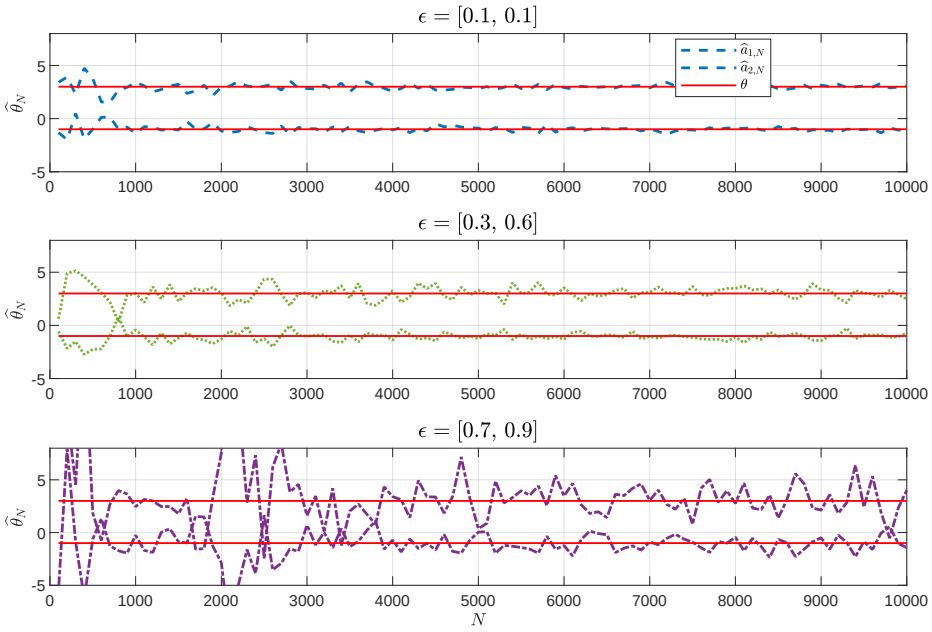


Figure 8. Convergence of $\widehat{\theta}_N$ under different attack frequencies ϵ.

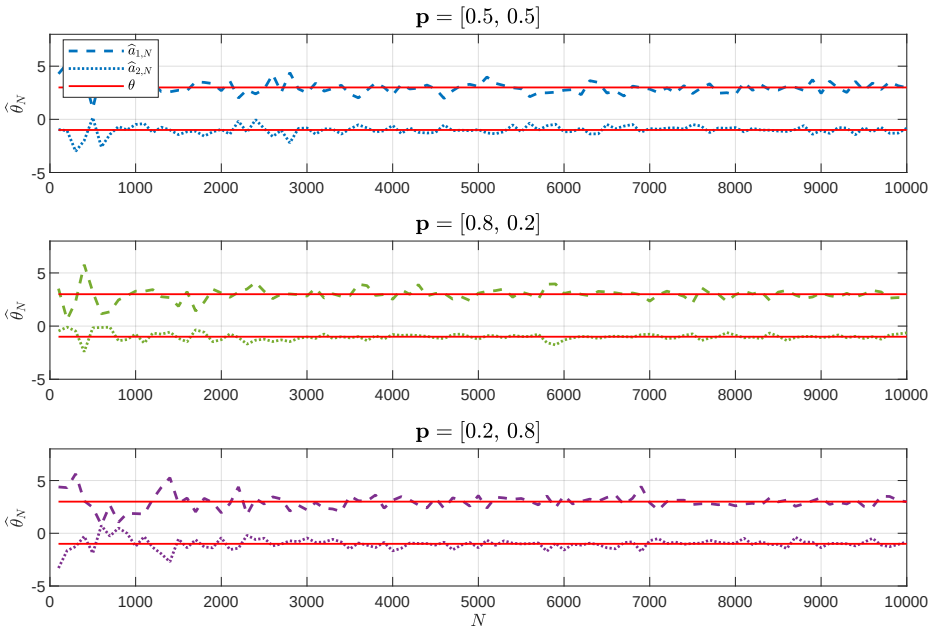


Figure 9. Convergence of $\widehat{\theta}_N$ under different replay delay distributions $\mathbf{p}$.

Different replay delay distributions p: Simulations with three delay distributions are conducted: a uniform distribution $\mathbf{p} = [0.5, 0.5]^{\mathrm{T}}$ and two skewed distributions $\mathbf{p} = [0.8, 0.2]^{\mathrm{T}}$ and $\mathbf{p} = [0.2, 0.8]^{\mathrm{T}}$. As shown in Figure 9, the algorithm maintains consistent convergence regardless of the delay distribution

shape, confirming that the compensation mechanism correctly accounts for the replay lag structure.

Different noise levels δ: Simulations with noise standard deviations $\delta = 5$, 10, and 20 are conducted. As shown in Figure 10, the estimates converge to the true parameters under all three noise levels. The asymptotic variance increases with δ, consistent with the theoretical expression in Theorem 3.5.

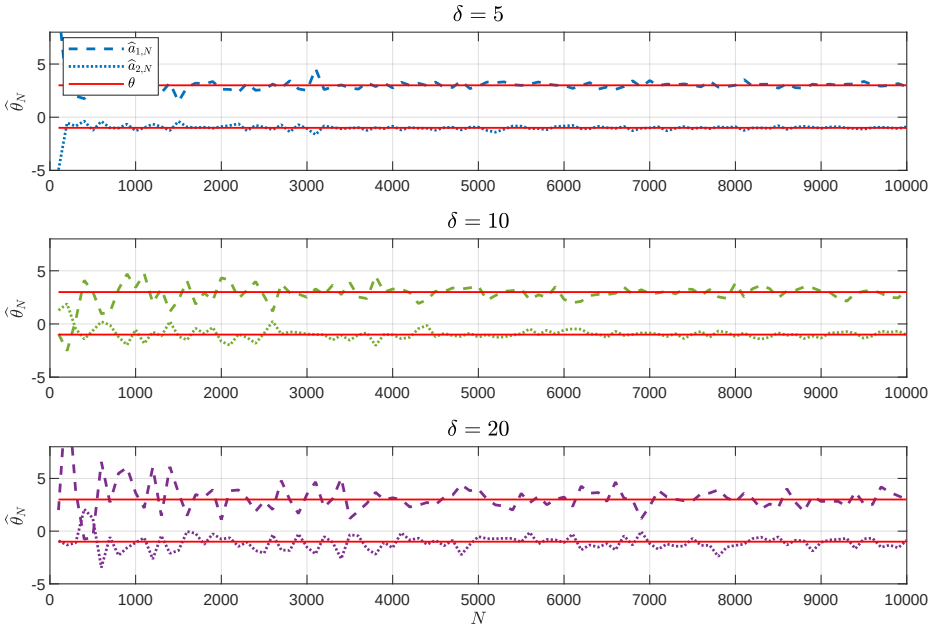


Figure 10. Convergence of $\widehat{\theta_N}$ under different noise levels δ.

These additional simulation results confirm the robustness and wide applicability of the proposed identification framework under a range of practical conditions.

5.4. Unknown attack frequency case

Consider a Poisson-type random attack process with unknown intensity in the attack sequence. Let the attack frequency vector be $\epsilon = [0.3, 0.6]^{\mathrm{T}}$, and the attack times are generated by a homogeneous Poisson process. The proposed attack frequency compensation algorithm is then applied to estimate the unknown parameters.

In the simulation, long-term observation data are first used to statistically identify the attack frequency ϵ_i, and then the parameter estimator $\widehat{\theta_N}$ is constructed based on the compensated statistics.

Figure 11 shows the convergence process of the attack frequency estimator $\widehat{\epsilon_{N,i}}$; Figure 12 presents the convergence results of the parameter estimator $\widehat{\theta_N}$.

As shown in Figure 11, under a homogeneous Poisson attack, the estimated attack frequencies $\widehat{\epsilon_{N,i}}$ gradually converge to the true attack frequencies $[0.3, 0.6]^{\mathrm{T}}$.

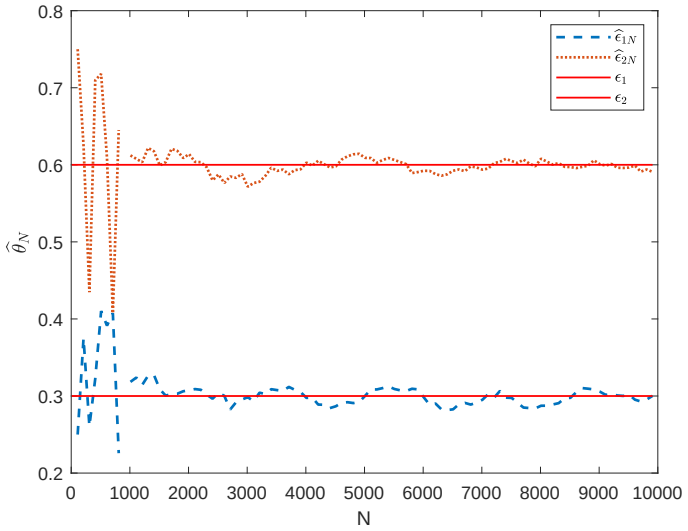


Figure 11. Convergence of attack frequency estimation under Poisson attacks.

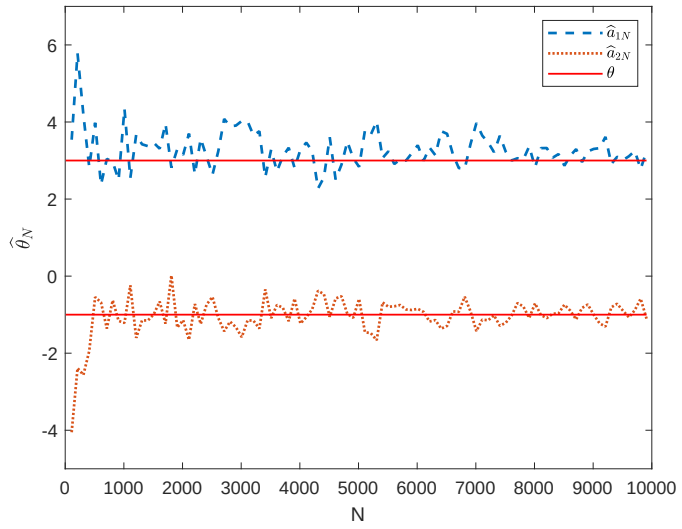
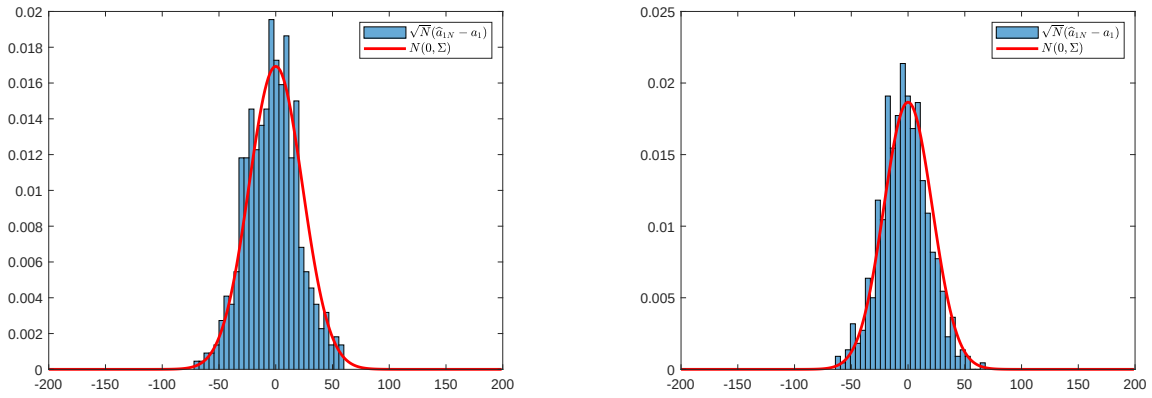


Figure 12. Convergence of parameter estimation under Poisson attacks.

Meanwhile, from Figure 12, the parameter estimates $\widehat{\theta}_N$ gradually converge to the true parameters $\theta = [3, -1]^{\mathrm{T}}$, indicating that the proposed compensation algorithm effectively eliminates the statistical bias caused by unknown attack frequencies. These results are consistent with the strong consistency conclusion stated in Theorem 4.1.

To further verify the asymptotic normality of the parameter estimator given in Theorem 4.2, 5000 Monte Carlo simulations were conducted. In Figure 13, panels (a) and (b) show the asymptotic distributions of the two parameter components $\hat{a}_1$ and $\hat{a}_2$, respectively.

As shown in Figure 13, the empirical distributions of $\sqrt{N}(\widehat{a}_{1N} - a_1)$ and $\sqrt{N}(\widehat{a}_{2N} - a_2)$ gradually approach the Gaussian distribution curves, indicating asymptotic normality of the estimation errors. The observation aligns with Theorem 4.2.



(a) Asymptotic normality of the $\hat{a}_1$

(b) Asymptotic normality of the $\hat{a}_2$

Figure 13. Asymptotic normality of the $\widehat{\theta}_N$.

Furthermore, employing the sampling ratio derived from Theorem 4.3, the optimal compensation proportion is calculated as $\alpha^* = 0.051$. To verify the effectiveness of this optimal ratio, additional simulations were conducted with $\alpha = 0.2$ and $\alpha = 0.8$ as comparison groups, using the asymptotic variance of $\sqrt{N}(\widehat{\theta}_N - \theta)$ serving as the performance criterion. The outcomes of these simulations are illustrated in Figure 14.

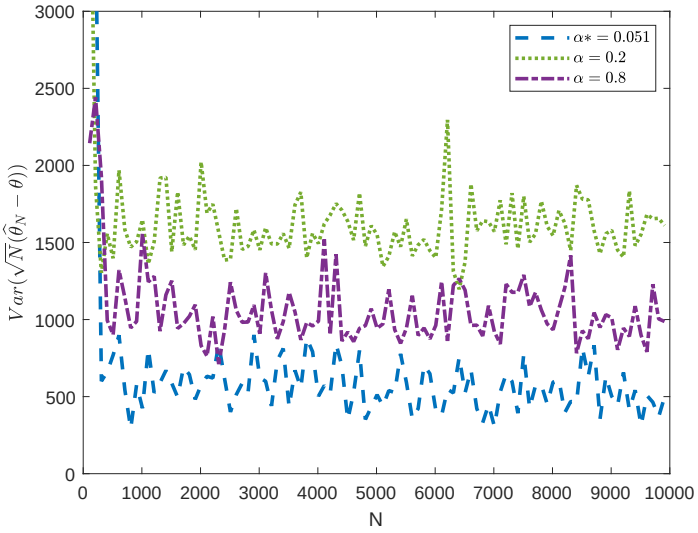


Figure 14. Comparison of asymptotic variances under different compensation ratios.

It can be observed that when the theoretical optimal ratio $\alpha^* = 0.051$ is used, the asymptotic variance of the parameter estimation error is minimized. This indicates that this ratio achieves an optimal balance between the attack frequency estimation error and the parameter statistical error, thereby validating the correctness of Theorem 4.3.

6. Conclusions

This study addresses the problem of estimating parameters in multi-parameter FIR systems with binary observations under replay attacks, framed within a system identification context. By modeling the attack strategy through an equivalent attack frequency, the proposed algorithms' convergence properties and asymptotic normality are examined for scenarios with both known and unknown attack frequencies.

When the attack frequency is known, an enhanced parameter estimation method utilizing identification error compensation is introduced, effectively correcting corrupted data and ensuring consistent estimation. For situations where the attack frequency cannot be determined, the identifiability of the replay attack sequence is thoroughly investigated, and a joint estimation method is proposed to achieve effective system parameter estimation. In addition, a design criterion for achieving optimal estimation performance under the compensation mechanism is provided. Theoretical analysis and numerical simulations under various attack frequencies, delay distributions, and noise levels demonstrate the effectiveness and robustness of the proposed methods under different attack conditions, providing a theoretical foundation for parameter estimation of quantized systems in replay-attack environments.

The FIR (finite impulse response) system considered in this study is characterized by its simple structure, ease of implementation, and high flexibility, and has therefore found extensive applications in fields such as signal processing, control engineering, and communications. Its linear nature and well-understood behavior make it a suitable benchmark for investigating system identification and anti-attack mechanisms. In practice, however, many real systems exhibit more complicated dynamics, including time-varying and nonlinear features. Furthermore, extending the proposed framework to scenarios where the replay delay distribution is also unknown represents an important direction for future investigation. Motivated by this, future work may extend the proposed framework to more general system classes, such as time-varying and nonlinear systems, so as to further improve the applicability and robustness of the developed methods in complex and realistic scenarios.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Nos. 62573044 and 62433020).

Conflict of interest

The authors declare there are no conflicts of interest.

References

1. M. Sony, S. Naik, Key ingredients for evaluating Industry 4.0 readiness for organizations: a literature review, *Benchmarking: Int. J.*, **27** (2020), 2213–2232. https://doi.org/10.1108/BIJ-09-2018-0284

2. M. Ghobakhloo, Industry 4.0, digitization, and opportunities for sustainability, *J. Cleaner Prod.*, **252** (2020), 119869. https://doi.org/10.1016/j.jclepro.2019.119869

3. J. Rymarczyk, Technologies, opportunities and challenges of the industrial revolution 4.0: theoretical considerations, *Entrepreneurial Bus. Econ. Rev.*, **8** (2020), 185–198. https://doi.org/10.15678/EBER.2020.080110

4. Q. Zhang, Y. Zhao, W. Xue, J. Guo, Parameter identification for FIR systems with simultaneously quantized inputs and outputs against replay attacks: a stochastic defense scheme, *IEEE Trans. Autom. Control*, **2026** (2026). https://doi.org/10.1109/TAC.2026.3687459

5. Z. Rahman, X. Yi, I. Khalil, Blockchain-based AI-enabled industry 4.0 CPS protection against advanced persistent threat, *IEEE Internet Things J.*, **10** (2022), 6769–6778. https://doi.org/10.1109/JIOT.2022.3147186

6. V. Lesch, M. Züfle, A. Bauer, L. Iffländer, C. Krupitzer, S. Kounev, A literature review of IoT and CPS—What they are, and what they are not, *J. Syst. Software*, **200** (2023), 111631. https://doi.org/10.1016/j.jss.2023.111631

7. A. Goknil, P. Nguyen, S. Sen, D. Politaki, H. Niavis, K. J. Pedersen, et al., A systematic review of data quality in CPS and IoT for industry 4.0, *ACM Comput. Surv.*, **55** (2023), 327. https://doi.org/10.1145/3593043

8. A. H. El-Kady, S. Halim, M. M. El-Halwagi, F. Khan, Analysis of safety and security challenges and opportunities related to cyber-physical systems, *Process Saf. Environ. Prot.*, **173** (2023), 384–413. https://doi.org/10.1016/j.psep.2023.03.012

9. X. Hou, H. Wu, J. Cao, Practical finite-time synchronization for Lur'e systems with performance constraint and actuator faults: A memory-based quantized dynamic event-triggered control strategy, *Appl. Math. Comput.*, **487** (2025), 129108. https://doi.org/10.1016/j.amc.2024.129108

10. J. Wu, W. Sun, S. F. Su, Y. Wu, Adaptive asymptotic tracking control for input-quantized nonlinear systems with multiple unknown control directions, *IEEE Trans. Cybern.*, **53** (2022), 5216–5225. https://doi.org/10.1109/TCYB.2022.3184492

11. G. Bottegal, H. Hjalmarsson, G. Pillonetto, A new kernel-based approach to system identification with quantized output data, *Automatica*, **85** (2017), 145–152. https://doi.org/10.1016/j.automatica.2017.07.053

12. N. Negi, A. Chakrabortty, Optimal co-designs of communication and control in bandwidth-constrained cyber-physical systems, *Automatica*, **142** (2022), 110288. https://doi.org/10.1016/j.automatica.2022.110288

13. C. De Persis, P. Tesi, Resilient control under denial-of-service, *IFAC Proc. Vol.*, **47** (2014), 134–139. https://doi.org/10.3182/20140824-6-ZA-1003.02184

14. G. Franze, F. Tedesco, W. Lucia, Resilient control for cyber-physical systems subject to replay attacks, *IEEE Control Syst. Lett.*, **3** (2019), 984–989. https://doi.org/10.1109/LCSYS.2019.2920507

15. D. Zhao, B. Yang, Y. Li, H. Zhang, Replay attack detection for cyber-physical control systems: A dynamical delay estimation method, *IEEE Trans. Ind. Electron.*, **72** (2024), 867–875. https://doi.org/10.1109/TIE.2024.3406859

16. Y. Mo, S. Weerakkody, B. Sinopoli, Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs, *IEEE Control Syst. Mag.*, **35** (2015), 93–109. https://doi.org/ 10.1109/MCS.2014.2364724

17. H. Guo, Z. H. Pang, J. Sun, J. Li, An output-coding-based detection scheme against replay attacks in cyber-physical systems, *IEEE Trans. Circuits Syst. II Express Briefs*, **68** (2021), 3306–3310. https://doi.org/10.1109/TCSII.2021.3063835

18. P. Yu, X. Zhang, R. Jia, Q. Dong, J. Guo, Optimal control for linear cyber physical systems under replay attacks via time delay estimation approach, *Int. J. Syst. Sci.*, **56** (2025), 183–192. https://doi.org/10.1080/00207721.2024.2389475

19. X. Li, A. Lei, L. Zhu, M. Ban, Improving Kalman filter for cyber physical systems subject to replay attacks: An attack-detection-based compensation strategy, *Appl. Math. Comput.*, **466** (2024), 128444. https://doi.org/10.1016/j.amc.2023.128444

20. C. M. Ahmed, V. R. Palleti, V. K. Mishra, A practical physical watermarking approach to detect replay attacks in a CPS, *J. Process Control*, **116** (2022), 136–146. https://doi.org/10.1016/j.jprocont.2022.06.002

21. S. Veesa, M. Singh, Implicit processing of linear prediction residual for replay attack detection, *Int. J. Speech Technol.*, **27** (2024), 781–791. https://doi.org/10.1007/s10772-024-10125-5

22. M. Hamadouche, Z. Khalil, H. Tebbi, M. Guerroumi, Y. Zafoune, A replay attack detection scheme based on perceptual image hashing, *Multimedia Tools Appl.*, **83** (2024), 8999–9031. https://doi.org/10.1007/s11042-023-15300-5

23. J. Wang, D. Wang, H. Yan, H. Shen, Composite antidisturbance $\mathcal{H}_\infty$ control for hidden Markov jump systems with multi-sensor against replay attacks, *IEEE Trans. Autom. Control*, **69** (2023), 1760–1766. https://doi.org/10.1109/TAC.2023.3326861

24. H. Shen, Y. Wang, J. Wu, Ju H. Park, J. Wang, Secure control for Markov jump cyber-physical systems subject to malicious attacks: A resilient hybrid learning scheme, *IEEE Trans. Cybern.*, **54** (2024), 7068–7079. https://doi.org/10.1109/TCYB.2024.3448407

25. P. Yu, H. Wan, B. Zhang, Q. Wu, B. Zhao, C. Xu, et al., Review on system identification, control, and optimization based on artificial intelligence, *Mathematics*, **13** (2025), 952. https://doi.org/10.3390/math13060952

26. J. Guo, R. Jia, R. Su, Y. Zhao, Y. Song, Identification of FIR systems with binary-valued observations against denial-of-service attacks, *Appl. Math. Comput.*, **450** (2023), 127989. https://doi.org/10.1016/j.amc.2023.127989

27. X. Cui, Q. Zhang, P. Yu, F. Jing, J. Guo, FIR system identification under congruential summation-triggered communication scheme with binary observations, *Nonlinear Anal. Hybrid Syst.*, **52** (2024), 101464. https://doi.org/10.1016/j.nahs.2024.101464

28. P. Yu, Y. Hu, Y. Wang, R. Jia, J. Guo, Optimal consensus control strategy for multi-agent systems under cyber attacks via a Stackelberg game approach, *IEEE Trans. Autom. Sci. Eng.*, **22** (2025), 18875–18888. https://doi.org/10.1109/TASE.2025.3591858

29. R. Jia, T. Wang, W. Xue, J. Guo, Y. Zhao, Multitime scale consensus algorithm of multiagent systems with binary-valued data under tampering attacks, *IEEE Trans. Ind. Inf.*, **21** (2025), 9377–9388. https://doi.org/10.1109/TII.2025.3598451

30. J. Guo, Q. Zhang, Y. Zhao, Identification of FIR systems with binary-valued observations under replay attacks, *Automatica*, **172** (2025), 112001. https://doi.org/10.1016/j.automatica.2024.112001

31. W. Liu, Y. Wang, J. Guo, Consistent identification for FIR systems with multilevel quantized observations subjected to data tampering attacks: A joint estimation approach, *IEEE Trans. Ind. Electron.*, **73** (2026), 7676–7687. https://doi.org/10.1109/TIE.2025.3642258

32. Q. Zhang, J. Guo, Compensation strategy-based intrusion tolerant parameter estimation for quantized nonlinear Hammerstein systems under replay attacks, *IEEE Trans. Autom. Sci. Eng.*, **23** (2025), 1343–1359. https://doi.org/10.1109/TASE.2025.3647889

AIMS Press