Electronic Research Archive

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

ERA, 34(7): 4461–4485.
DOI: 10.3934/era.2026197
Received: 11 March 2026
Revised: 22 April 2026
Accepted: 08 May 2026
Published: 25 May 2026

Theory article

Optimal attack strategy for binary measurements-based Hammerstein system identification subject to data tampering attacks

Zimeng Zhou[1], **Qingxiang Zhang**[1], **Fengwei Jing**[2,*] **and Jin Guo**[1]

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

[2] National Engineering Research Center for Advanced Rolling and Intelligent Manufacturing, University of Science and Technology Beijing, Beijing 100083, China

* **Correspondence:** Email: jingfengwei@nercar.ustb.edu.cn.

Abstract: This paper studies the formulation for an optimal attack strategy, which aims to maximize the identification error of Hammerstein systems with binary measurements subject to data tampering attacks. First, the convergence of the parameter estimates under attack is analyzed, and the absolute error between the estimated value and the true value is used as the objective function. Second, an optimization model with constraints on the maximum data tampering rate and the average data tampering rate is established to maximize the objective function. Then, the sine-cosine optimization algorithm is used to search for the optimal solution that meets the constraints, and its performance is compared with other algorithms. Finally, the effectiveness of the proposed method is verified by a numerical simulation.

Keywords: Hammerstein system; binary measurements; data tampering attack; optimal attack; optimization algorithm

1. Introduction

With the continuous integration of information technology and physical systems, the advancement of embedded systems, the internet of things (IoT), and control theory has offered new solutions to increasingly complex challenges in industries such as manufacturing, transportation, energy, and healthcare. In this context, the concept of cyber-physical systems (CPS) has emerged [1–3]. CPS integrates communication, computation, and remote coordinated control, thereby demonstrating significant application value in fields such as healthcare services, intelligent transportation, and smart grids.

However, the open and interconnected nature of CPS makes it vulnerable to severe cybersecurity

threats [4]. Due to the deep integration between cyber and physical components, cyberattacks can extend from the virtual space into the physical domain, thus directly impacting operational safety. Attackers may compromise system availability, integrity, and confidentiality through false data injection attacks [5], data tampering attacks [6], replay attacks [7], man-in-the-middle attack [8], distributed denial of service attack [9], and more, thus potentially causing physical equipment failures or safety incidents. Pasqualetti et al. [10] proposed a systematic mathematical modeling framework for CPS widely used in critical infrastructures, analyzed the limitations of monitoring under failures and attacks, and designed both centralized and distributed methods for attack detection and identification. Nevertheless, the heterogeneity, real-time requirements, and resource constraints of CPS make traditional cyber security mechanisms difficult to directly apply, thus highlighting the urgent need for novel security technologies tailored for CPS. In recent years, researchers have proposed various model-based security strategies, such as dynamic defense mechanisms [11] and secure control synthesis strategies [12], which provided new directions to enhance CPS security. One of the motivations of this paper is to enable defenders to better understand the behavior of unknown attackers, thereby designing appropriate defensive strategies to protect the system more effectively. For a comprehensive overview of privacy-preserving approaches in industrial CPS, Ding et al. [13] presented the most up-to-date results for privacy-preserving filtering, control, and optimization, thus systematically summarizing mainstream strategies and evaluation metrics.

In the field of CPS security research, malicious attacks on systems have attracted widespread attention due to their potentially destructive consequences. Attackers can mislead the system into making incorrect decisions based on falsified information, which results in abnormal operations or even safety hazards. To better understand attack mechanisms and develop effective defense strategies, researchers have conducted systematic studies from multiple perspectives, including attack modeling, optimization methods, and implementation strategies. The research on the attack strategy design has progressively evolved from single performance metrics to multi-objective optimization, where the integration of system knowledge and advanced optimization techniques is leveraged to enhance both the attack efficacy and stealthiness. For instance, to improve the attack stealth, Mao and Yang [14] proposed an optimal strategy based on corrupted innovations, which fuses intercepted data with side information to reduce computational complexity while enhancing attack performance. From a methodological perspective, Meira-Góes et al. [15] developed algorithms under a Markov decision process framework to synthesize optimal attack strategies that maximize the success probability. In the context of direct current microgrids, Asghari et al. [16] formulated highly stealthy false data injection attacks within an optimal control framework aimed at maximizing control errors. By combining Markov decision processes with expert learning, Truong et al. [17] analyzed optimal malicious expert strategies that disrupt weighted average prediction algorithms under various loss functions. In the field of remote state estimation, Wang et al. [18] applied convex optimization to design a stealthy false data injection attack that maximizes the estimation error covariance while remaining undetected. Chen et al. [19] structured the attacker's objective using a linear quadratic optimization framework, thereby reducing the optimal attack strategy to a linear feedback form under stealth constraints. For a distributed estimation, Zayyani et al. [20] introduced an optimal measurement and channel attack design from a maximum-disturbance perspective, thereby employing Lagrange multipliers to derive the suboptimal attack strategies. Adopting a game-theoretic approach, Zhang et al. [21] developed optimal distributed denial of service (DDoS) attack strategies for

cyber-physical systems by optimizing the channel selection and power allocation in multi-attacker scenarios, with the effectiveness validated through simulations. Zhang et al. [22] modeled limited attack instants to design constrained optimal stealthy attack strategies against state estimators, thereby optimizing the estimation error covariance via Lagrange multipliers and verifying the performance through simulation cases. For master-slave neural networks where both channels are subject to random deception attacks, Kazemy et al. [23] used a dual-event triggering mechanism and static output feedback, deriving linear matrix inequality (LMI)-form synchronization sufficient conditions based on the Lyapunov-Krasovskii functional method, thereby achieving master-slave synchronization while saving communication resources. For discrete-time networked systems with both network-induced delays and malicious packet losses, Zhang et al. [24] introduced a packet handler and a novel Lyapunov functional, and based on the Finsler lemma, established necessary and sufficient conditions for the validity of quartic polynomial inequalities. They derived bounded real conditions depending on the upper bound of delays and the maximum number of consecutive packet losses, achieving a joint analysis of the H_∞ performance. For multi-agent systems with uncertain power and DoS attacks, Yang et al. [25] designed a predetermined-time cooperative control protocol based on first-order filter distributed observers, neural networks, and self-triggered mechanisms, thus achieving synchronization errors within a preset range within a user-definable predetermined time.

The Hammerstein model describes common nonlinearities in life [26], such as actuator saturation, sensor dead zones, and dose-response relationships. Binary measurements appear in threshold sensors [27], network data transmissions, medical diagnostic indicators, and 1-bit quantized outputs of low-cost IoT devices. Their combination naturally appears in smart grid load monitoring, industrial safety monitoring, autonomous driving sensor fusion, and wireless structural health monitoring. Data tampering attacks, as a specific form of integrity attack, differ from availability-disrupting DoS attacks in that they cleverly corrupt data, thus making them particularly dangerous for learning-based systems that rely on data consistency. Studies on finite impulse response (FIR) systems have shown that even limited tampering can lead parameter estimates to converge to biased values.Therefore, understanding how data tampering attacks undermine the identification of such models is crucial to assess security risks and design robust learning algorithms.

It is worth noting that binary measurements can be mathematically interpreted as an extreme form of censored or truncated data, where the observation is only available as a thresholded binary outcome. In the context of Hammerstein system identification under non-ideal observation conditions, several recent works have addressed robust adaptive filtering problems. For instance, a robust adaptive Hammerstein filtering algorithm has been developed for censored regressions under contaminated Gaussian noise [28], where a maximum likelihood-based cost function is constructed to recover censored data and compensate for computational biases caused by data truncation. This approach effectively integrates the issues of contaminated Gaussian noise and censored data, thereby providing theoretical convergence and a mean-square performance analysis. While these works focus on robust adaptive filtering under non-Gaussian and censored observation conditions, the present paper addresses a different but complementary problem: optimal attack strategies for Hammerstein system identification under data tampering attacks. Nevertheless, the mathematical connections between censored regression and binary measurements suggest that techniques from a robust censored regression could potentially be adapted for a defense mechanism design, which we leave as an important direction for future research.

This paper focuses on binary measurement-based nonlinear Hammerstein systems under data tampering attacks. First, a consistent identification algorithm is proposed for the case without data tampering, and its numerical implementation is realized using a hybrid particle swarm optimization-broyden fletcher goldfarb shanno (PSO-BFGS) algorithm [29]. Then, a comparative study is conducted to systematically analyze the performance differences among the PSO algorithm, the BFGS algorithm, and their hybrid version in terms of parameter estimation. Finally, to address the data tampering attack optimization problem under energy constraints, the sine cosine algorithm (SCA) is adopted to search for the optimal attack strategy, which is further compared with other algorithms.

The challenges faced by this article are as follows:

- The Hammerstein system consists of a static nonlinear module followed by a dynamic linear module in cascade, where the parameters η and θ are highly coupled. Under binary measurements, the output contains only 1 bit of information, thus making traditional identification methods for continuous measurements inapplicable. Instead, parameters must be inferred from the probability domain, which leads to solving a system of nonlinear equations. The existence, uniqueness, and continuity of the solution in this nonlinear setting require rigorous verification.
- Unlike linear FIR systems, the impact of data tampering attacks on Hammerstein system identification propagates through nonlinear mappings $G^{-1}(\cdot)$, $\mathcal{L}_1(\cdot)$, and $\mathcal{L}_2(\cdot)$, thus making the analysis and optimization of attack effects more complex. Quantitatively characterizing the analytical relationship between the attack probabilities (m_1, m_2) and the resulting identification error $J(m_1, m_2)$ is a key challenge.
- The attacker is subject to both maximum and average data tampering rate constraints, which is more realistic than a single constraint but also makes the feasible region of the optimization problem more complex. Searching for the optimal attack strategy under dual constraints is a non-convex optimization problem that admits no closed-form solution.

The main contributions and innovations of this paper are as follows:

- Unlike linear systems, this paper systematically studies the system identification problem of binary measurement nonlinear Hammerstein systems under data tampering attacks, thus filling the gap in existing research on the combination of nonlinear systems and binary measurements.
- This paper transforms the parameter identification problem of Hammerstein systems under binary measurements into an optimization problem, provides the initialization strategy and convergence criteria for the algorithm, and offers a numerical implementation based on the PSO-BFGS algorithm.
- Under the energy constraint of data tampering, this paper proposes an optimization problem to maximize the identification error, gives a numerical solution based on the SCA algorithm, and compares it with the sparrow search algorithm (SSA) and grey wolf optimization (GWO) algorithm.

The remainder of the paper is organized as follows: Section 2 introduces the system identification problem for Hammerstein systems with binary measurements; Section 3 presents the identification algorithm and its implementation; Section 4 addresses the modeling of the optimal attack strategy; Section 5 focuses on the optimization algorithm and implementation for the optimal attack strategy; Section 6 provides numerical simulations; Section 7 discusses the shortcomings of this article and future research directions; Section 8 concludes the paper.

Remark 1.1. *The parameter vector of the nonlinear part is denoted by* $\eta = [b_1, \ldots, b_{n_1}]^T$. $\theta = [a_1, \ldots, a_{n_2}]^T$ *is the parameter vector of the linear part.* $W = \Omega\theta$ *is the output vector of the linear part, where* Ω *is the full-rank matrix constructed from the inputs.* ξ *and* $\tilde{\xi}$ *are variables of the equation system.* $\mathcal{L}(\cdot)$ *is the mapping function for parameter estimation.* P, K_{pso}, *and* K_{bfgs} *are parameters of the PSO-BFGS algorithm.* S *and* Z *are the population size and the maximum number of iterations of the SCA, respectively. pBest and gBest are the personal best and the global best in the PSO, respectively.*

2. Problem description

A single-input single-output discrete-time Hammerstein system, where the static nonlinear part is a weighted sum of n_1 nonlinear functions, and the dynamic linear part is an n_2-order FIR system, is described as follows:

$$\begin{cases} y_k = \sum_{i=1}^{n_2} a_i x_{k-i+1} + v_k, \\ x_k = b_0 + \sum_{j=1}^{n_1} b_j h_j(u_k), \ b_0 = 1, \end{cases} \tag{2.1}$$

where $\boldsymbol{\eta} = [b_1, b_2, \ldots, b_{n_1}]^T \in \mathbb{R}^{n_1}$, and $\boldsymbol{\theta} = [a_1, a_2, \ldots, a_{n_2}]^T \in \mathbb{R}^{n_2}$, and $u_k \in \mathbb{R}$, $x_k \in \mathbb{R}$, and $v_k \in \mathbb{R}$ denote the system input, intermediate variable, and system noise, respectively. The function $h_j(\cdot)$ maps from $\mathbb{R}$ to $\mathbb{R}$, where $j \in J = \{1, 2, \ldots, n_1\}$. The integers n_1 and n_2 are known. A threshold-based detector with parameter M converts the continuous-valued output y_k into a binary sequence s_k^0, which is mathematically captured by the following:

$$s_k^0 = I_{\{y_k \le M\}} = \begin{cases} 1, \ y_k \le M, \\ 0, \ \text{others}. \end{cases} \tag{2.2}$$

Remark 2.1. *Unlike linear measurements, the binary sensor* $y_q(k) = mathbbIy(k)geqM$ *only retains information about threshold crossings. Any tampering that does not flip the binary bits cannot be detected, thus creating a blind spot for stealthy attacks. This information loss, combined with the resulting nonlinear estimation problem, fundamentally affects the system's observability.*

As can be seen from Figure 1, s_k^0 is transmitted to a remote data center through an unsecured communication network and may be subject to data tampering attacks during transmission. The observed time series at the computational hub is represented as s_k for discrete time indices k, and it is related to s_k^0 as follows:

$$\begin{cases} \mathrm{P}\left\{s_k = 0 \mid s_k^0 = 1\right\} = m_1, \\ \mathrm{P}\left\{s_k = 1 \mid s_k^0 = 0\right\} = m_2. \end{cases} \tag{2.3}$$

The above equations essentially define a data tampering attack(dta) strategy, denoted concisely as (m_1, m_2).

The aim of this paper is to examine, from the perspective of an attacker, the optimal strategy that maximizes the identification error under energy constraints.

Assumption 2.2. *The system noise* $\{v_k\}$ *is an independent and identically distributed sequence of Gaussian random variables with zero mean and variance* σ^2. *Its cumulative distribution function and probability density function are denoted by* $G(\cdot)$ *and* $f(\cdot)$, *respectively.*

Assumption 2.3. *The data tampering attack is subject to an energy constraint, meaning that the data tampering rate* $\kappa = \Pr\{s_k \neq s_k^0\} \leq \bar{\kappa} < 1$.

Remark 2.4. *The Gaussian distribution is the most common noise model in system identification, and it can approximate many real physical noises using the central limit theorem. A zero mean can be achieved through data preprocessing. The Gaussian assumption guarantees the smoothness and strict monotonicity of the cumulative distribution function* $G(\cdot)$*, thus ensuring the existence of* $G^{-1}(\cdot)$.

Remark 2.5. *Attackers are typically constrained by an energy budget, thus rendering them unable to tamper with data indefinitely; the constraint* $\bar{\kappa} < 1$ *ensures that the data received by the estimation center retains a portion of the true information, thereby averting a scenario of complete unidentifiability.*

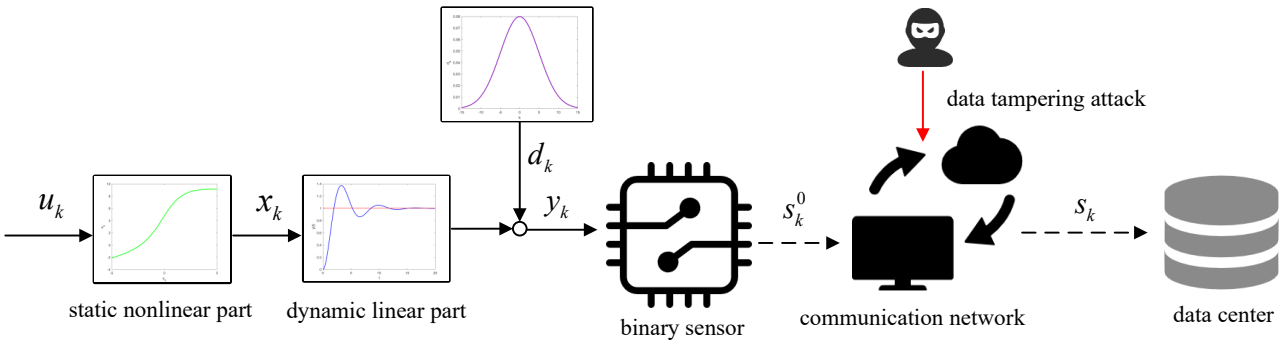

Figure 1. system block diagram.

3. Preliminary work

3.1. Identification algorithm without attack

Assume the system input u_k is a periodic sequence with period $n = n_1 + n_2$, to ensure system identifiability, that is, for all $k > 1$, we have $u_{k+n} = u_k$. Let $\pi_k = [u_k, u_{k-1}, \ldots, u_{k-n_2+1}]$ denote the vector constructed from the system input at time k. For a data sequence of length N, assume that it contains D_N complete periods, i.e.,

$$D_N = \left\lfloor \frac{N}{n} \right\rfloor, \tag{3.1}$$

where $\lfloor \cdot \rfloor$ denotes the floor function.

Remark 3.1. *The periodic input assumption is adopted because of the following: (i) it enables consistent estimation under binary measurements via the Law of Large Numbers; and (ii) it guarantees the persistent excitation and full rank of* Ω*, thus ensuring identifiability of Hammerstein system parameters. The period is set to* $n = n_1 + n_2$ *to ensure a unique solution.*

The full-rank matrix composed of n_2 vectors from the set $\{\pi_k\}$ is given by the following:

$$\boldsymbol{\Omega} = [\omega_1, \ldots, \omega_{n_2}]^T \in \mathbb{R}^{n_2 \times n_1}, \boldsymbol{W} = \boldsymbol{\Omega\theta} \in \mathbb{R}^{n_2}, \tag{3.2}$$

where $\boldsymbol{\omega_i} \in \{\pi_1^T, \pi_2^T \ldots, \pi_n^T\}$, $\omega_i \in \mathbb{R}^{n_2}$.

Denote the $(1+n_1)\times n_2$ dimensional matrix

$$\boldsymbol{\alpha_i} = \begin{bmatrix} h_0(\pi_{i,1}) & h_0(\pi_{i,2}) & \cdots & h_0(\pi_{i,n_2}) \\ h_1(\pi_{i,1}) & h_1(\pi_{i,2}) & \cdots & h_1(\pi_{i,n_2}) \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_1}(\pi_{i,1}) & h_{n_1}(\pi_{i,2}) & \cdots & h_{n_1}(\pi_{i,n_2}) \end{bmatrix} \boldsymbol{\Omega^{-1}},\ i = 1, 2, \ldots, n. \tag{3.3}$$

Let $x_1, x_2, \ldots, x_{n_1+n_2}$ be $n_1 + n_2$ unknown variables, denote $\boldsymbol{X_1} = [x_1, x_2, \ldots, x_{n_1}]^T$, $\boldsymbol{X_2} = [x_{n_1+1}, x_{n_1+2}, \ldots, x_{n_1+n_2}]^T$, and consider the following system of equations:

$$\boldsymbol{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix} = \begin{bmatrix} [\eta_0,\ \boldsymbol{X_1^T}]\alpha_1 \boldsymbol{X_2} \\ [\eta_0,\ \boldsymbol{X_1^T}]\alpha_2 \boldsymbol{X_2} \\ \vdots \\ [\eta_0,\ \boldsymbol{X_1^T}]\alpha_n \boldsymbol{X_2} \end{bmatrix}. \tag{3.4}$$

Assumption 3.2. *There exists a compact set $\Xi \subseteq \mathbb{R}^{n_1+n_2}$, such that $\tilde{\boldsymbol{\xi}}$ is an interior point of this set, where*

$$\tilde{\boldsymbol{\xi}} = \begin{bmatrix} [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_1}\boldsymbol{\Omega\theta} \\ [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_2}\boldsymbol{\Omega\theta} \\ \vdots \\ [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_n}\boldsymbol{\Omega\theta} \end{bmatrix}. \tag{3.5}$$

For any $\boldsymbol{\xi} \in \Xi$, the system of (3.4) *has a unique solution, which is denoted by the following:*

$$[\boldsymbol{X_1^T}, \boldsymbol{X_2^T}]^T = \pounds(\boldsymbol{\xi}). \tag{3.6}$$

Moreover, $\pounds(\boldsymbol{\xi})$ is bounded and continuous at the point $\boldsymbol{\xi}$.

Remark 3.3. *This assumption constitutes a generalization of the Implicit Function Theorem; its boundedness and continuity guarantee the consistency in the probability of the estimators.*

For ease of presentation, we denote $\pounds_1(\boldsymbol{\xi})$ and $\pounds_2(\boldsymbol{\xi})$ as $\boldsymbol{X_1^T}$ and $\boldsymbol{X_2^T}$, respectively. In the absence of data tampering attacks, for System (2.1) and binary measurements (2.2), under Assumption 3.2, with $N \to \infty$, the parameter estimates given by here comes wgen Eqs (3.7)–(3.9) converge to the true values η and θ as follows:

$$\nu_{i,N} = M - G^{-1}\left(\frac{1}{D_N}\sum_{q=1}^{D_N} s_{(q-1)n+i}\right), \tag{3.7}$$

$$[\hat{\boldsymbol{\eta}}_{\boldsymbol{N}}^{\boldsymbol{T}}, \hat{\boldsymbol{W}}_{\boldsymbol{N}}^{\boldsymbol{T}}]^{\boldsymbol{T}} = \pounds(\boldsymbol{\nu_N}), \tag{3.8}$$

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{N}} = \boldsymbol{\Omega^{-1}}\hat{\boldsymbol{W}}_{\boldsymbol{N}}, \tag{3.9}$$

where $F^{-1}(\cdot)$ denotes the inverse function of $F(\cdot)$, $\boldsymbol{\nu_N} = [\nu_{1,N}, \ldots, \nu_{n,N}]^T \in \mathbb{R}^n$, $\hat{\boldsymbol{\eta}}_{\boldsymbol{N}}$ represents the estimate of $\boldsymbol{\eta}$, $\hat{\boldsymbol{W}}_N$ is the estimate of $\boldsymbol{W}$, and $\hat{\boldsymbol{\theta}}_{\boldsymbol{N}}$ denotes the estimate of the system parameter $\boldsymbol{\theta}$.

The following part provides a brief proof to support the subsequent algorithm design and analysis. According to System (2.1), we have the following:

$$y_k = [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_k}\boldsymbol{W} + \nu_k. \tag{3.10}$$

In the absence of attacks, the probability of the event $s_k = 1$ is given by the following:

$$\begin{aligned} \mathbb{E}(s_k) &= \mathrm{P}\{s_k = 1\} \\ &= \mathrm{P}\{v_k \le M - [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_k W}\} \\ &= G(M - [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_k W}). \end{aligned} \tag{3.11}$$

where $\mathbb{E}(\cdot)$ represents the expectation of $\cdot$.

When the amount of data approaches infinity, by the law of frequency tending to probability, we have the following:

$$\frac{1}{D_N}\sum_{q=1}^{D_N} s_{(q-1)n+i} \to \mathbb{E}(s_i),\ \nu_{i,N} \to [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_i W} = \tilde{\xi}_i. \tag{3.12}$$

According to Eqs (3.4) and (3.5), we have the following:

$$[\hat{\boldsymbol{\eta}}_N^{\boldsymbol{T}}, \hat{\boldsymbol{W}}_N^{\boldsymbol{T}}]^T = £(\boldsymbol{\nu}_N) \to [\boldsymbol{\eta}^{\boldsymbol{T}}, \boldsymbol{W}^{\boldsymbol{T}}]^T = £(\tilde{\boldsymbol{\xi}}). \tag{3.13}$$

It follows from Eqs (3.2) and (3.9) that

$$\hat{\boldsymbol{\theta}}_N = \boldsymbol{\Omega^{-1}}\hat{\boldsymbol{W}}_N \to \boldsymbol{\theta} = \boldsymbol{\Omega^{-1}W}. \tag{3.14}$$

3.2. Numerical Implementation of the PSO-BFGS Hybrid Algorithm

To solve the complex nonlinear System (3.4), this paper introduces an equivalent optimization approach. Specifically, the original system is transformed into an objective function expressed as the sum of squared residuals (3.15), and the system is indirectly solved by minimizing this function to obtain the following optimal solution:

$$\mathcal{L}(\boldsymbol{\eta}, \boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{n} \left\| \nu_{i,N} - [\eta_0,\ \boldsymbol{\eta}^{\boldsymbol{T}}]\boldsymbol{\alpha_i\theta} \right\|^2. \tag{3.15}$$

This optimization problem is essentially a nonlinear least squares problem. To effectively solve the nonlinear least squares problem, this paper employs a hybrid optimization strategy that combines the PSO and BFGS algorithms.

PSO is a population-based global optimization algorithm inspired by swarm intelligence. It features a simple implementation, strong robustness, and is well-suited to handle non-convex problems, thus making it effective to roughly search the global optimal region of the objective function. However, due to its relatively slow convergence speed and limited precision near the optimal solution, PSO alone often cannot meet requirements for high-accuracy solutions.

A particle's current position represents its solution in the search space and is typically expressed as a vector. This position is updated iteratively based on the particle's velocity and current position, usually following the position update formula

$$x_i^{t+1} = x_i^t + v_i^{t+1}, \tag{3.16}$$

where x_i^{t+1} denotes the position of particle i at iteration $t + 1$, and v_i^{t+1} denotes the velocity of particle i at iteration $t + 1$.

A particle's velocity determines the direction and step size of its movement in the search space. Each particle has a velocity vector that represents its movement along each dimension. The velocity is updated by taking the current velocity, the particle's own best-known position (personal best), and the swarm's best-known position (global best) into account, which leads to Eq (3.17):

$$v_i^{t+1} = \omega \cdot v_i^t + c_1 \cdot \mathrm{e}_1 \cdot (\mathrm{pBest_i} - x_i^t) + c_2 \cdot \mathrm{e}_2 \cdot (\mathrm{gBest_i} - x_i^t), \tag{3.17}$$

where ω is the inertia weight. c_1 and c_2 are the cognitive (individual learning) and social (swarm learning) coefficients, respectively, which control the particle's tendency to move toward its personal best position $\mathrm{pBest_i}$ and the global best position gBest. e_1 and e_2 are random numbers uniformly distributed in the interval $[0, 1]$.

In contrast, BFGS is a classical gradient-based local optimization algorithm that features superlinear (quasi-second-order) convergence speed. It is well-suited to rapidly approach an extremum when the objective function is sufficiently smooth. The algorithm constructs an approximate Hessian matrix update to avoid directly computing second-order derivatives, thus balancing both efficiency and accuracy.

Each iteration consists of the following steps:

Step 1: Compute the search direction:

$$p^{(k)} = -(\boldsymbol{H}^{(\boldsymbol{k})})^{-1} g^{(k)}. \tag{3.18}$$

Step 2: Perform a line search to determine the step size $\lambda^{(k)}$, such that

$$x^{(k+1)} = x^{(k)} + \lambda^{(k)} p^{(k)}.$$

Step 3: Update the variables:

$$z^{(k+1)} = z^{(k)} + \lambda^{(k)} p^{(k)}. \tag{3.19}$$

Step 4: Update the Hessian approximation matrix:

$$\boldsymbol{H}^{(\boldsymbol{k+1})} = \boldsymbol{H}^{(\boldsymbol{k})} + \frac{y^{(k)}(y^{(k)})^T}{(y^{(k)})^T s^{(k)}} - \frac{\boldsymbol{H}^{(\boldsymbol{k})} s^{(k)} (s^{(k)})^T \boldsymbol{H}^{(\boldsymbol{k})}}{(s^{(k)})^T \boldsymbol{H}^{(\boldsymbol{k})} s^{(k)}}, \tag{3.20}$$

where $s^{(k)} = z^{(k+1)} - z^{(k)}$, $y^{(k)} = g^{(k+1)} - g^{(k)}$.

To combine the advantages of both methods, this paper first uses the PSO algorithm to search for an approximate global optimum in a large search space. Then, taking this approximate solution as the initial point, the BFGS algorithm is applied for local fine-tuning, thereby improving the accuracy and stability of the final solution.The pseudo code of the hybrid optimization process is shown in Algorithm 1.

The proposed algorithm consists of three main parts, whose computational complexities are summarized in Table 1.

Table 1. Computational complexity analysis of the proposed algorithm.

Step	Operation	Time Complexity
Computation of $v_{i,N}$	Summation and inverse evaluation	$O(D_N \cdot n)$
PSO global search	Evaluating $\mathcal{L}(\eta, \theta)$ P times per iteration	$O(K_{\mathrm{pso}} \cdot P \cdot n \cdot (n_1 + n_2))$
BFGS local refinement	Gradient computation, line search, Hessian update	$O(K_{\mathrm{bfgs}} \cdot (n_1 + n_2)^2)$

Algorithm 1 Implementation of the Identification Algorithm

Require: Parameters: M, μ, σ, N, n, u_k, η_0, number of particles P, maximum PSO iterations K_{pso}, maximum BFGS iterations K_{bfgs}, tolerance ϵ

Ensure: Estimated system parameters $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\theta}}$

Compute $\nu_{i,N}$ and $Z_{i,N}$ according to Eq (3.7)

Stage 1: Global Optimization Using PSO Algorithm

Initialize particle positions $z_p^{(0)} = \text{vec}(\eta^{(p,0)}, \theta^{(p,0)})$ and velocities $v_p^{(0)}$

Init $z_p^{(0)}, v_p^{(0)}$, $\text{pbest}_p = z_p^{(0)}$, $\text{gbest} = \arg\min \mathcal{L}(\text{pbest}_p)$

for $k = 0$ to K_{pso} **do**

    for each particle p **do**

        Update position and velocity using Eqs (3.16) and (3.17)

        Update pbest_p if $\mathcal{L}(z_p^{(k+1)}) < \mathcal{L}(\text{pbest}_p)$

        Update gbest if $\mathcal{L}(\text{pbest}_p) < \mathcal{L}(\text{gbest})$

    end for

end for

Stage 2: Local Optimization Using the BFGS Algorithm

Set initial point $z^{(0)} = \text{gbest}$ and $H^{(0)} = I$

for $k = 0$ to K_{bfgs} **do**

    Compute gradient $g^{(k)} = \nabla_z \mathcal{L}(z^{(k)})$

    if $\|g^{(k)}\| < \epsilon$ **then**

        break

    end if

    Compute search direction using Eq (3.18)

    Perform line search to get step size $\lambda^{(k)}$ minimizing $\mathcal{L}(z^{(k)} + \lambda^{(k)} p^{(k)})$

    Update variables using Eq (3.19)

    Update Hessian approximation using Eq (3.20)

end for

Convert final $z^{(k+1)}$ to $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\theta}}$

The overall time complexity is given by the following:

$$O\left(D_N \cdot n + K_{\text{pso}} \cdot P \cdot n \cdot (n_1 + n_2) + K_{\text{bfgs}} \cdot (n_1 + n_2)^2\right). \tag{3.21}$$

For the simulation setup in Section 6 ($N = 10,000, n = 3, n_1 = 1, n_2 = 2, P = 20, K_{\text{pso}} = 200, K_{\text{bfgs}} = 100$), the total computational cost is approximately $O(10^4)$ operations, which is manageable for typical offline identification tasks.

The PSO-BFGS hybrid algorithm does not have a provable guarantee of global convergence for arbitrary non-convex problems. However, its effectiveness in practice is supported by numerical comparisons with single-algorithm baselines. According to Li et al. [29], the effectiveness of the PSO-BFGS hybrid algorithm lies in the synergy of its two phases: the PSO stage reduces sensitivity to initial values through population-based random initialization and dynamically triggers a BFGS local search using the proposed local diversity index (LDI), while the BFGS stage starts from near-optimal initial points provided by the PSO algorithm, thereby leveraging its superlinear

convergence speed advantage. This combination significantly improves the convergence accuracy and speed while maintaining the global search capability. For medium-dimensional offline identification tasks, the overall computational cost remains acceptable.

4. Modeling of the optimal attack strategy problem

4.1. Construction of attack effectiveness metrics

To evaluate the impact of a data tampering attack on system identification, it is crucial to first understand how the attack influences the asymptotic behavior of the parameter estimates. The following theorem establishes the converged values of the estimates under a given attack strategy (m_1, m_2), which will serve as the foundation to define the attack effectiveness metric.

Theorem 4.1. *For System* (2.1) *and binary measurements* (2.2)*, under the data tampering attack strategy* (m_1, m_2)*, the convergence values of* η *and* θ *given by Eqs* (3.7)–(3.9) *of Algorithm 1 are as follows:*

$$[\hat{\boldsymbol{\eta}}_N^{\boldsymbol{T}}, \hat{\boldsymbol{\theta}}_N^{\boldsymbol{T}}]^T \rightarrow [\pounds_1(\tilde{\boldsymbol{\beta}}), \boldsymbol{\Omega^{-1}}\pounds_2(\tilde{\boldsymbol{\beta}})]^T, \tag{4.1}$$

$$\tilde{\boldsymbol{\beta}} = [M - G^{-1}(\beta_1), \ldots, M - G^{-1}(\beta_n)]^T, N \rightarrow \infty, \tag{4.2}$$

where $\tilde{\beta} \in \mathbb{R}^{n\times 1}$, $\beta_i = m_2 + (1 - m_1 - m_2)G(M - [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_i W})$, *and* $i = 1, \ldots, n$.

Proof. Under attack conditions, the probability of the event $s_k = 1$ is given by the following:

$$\begin{aligned}
\mathbb{E}(s_k) &= \mathrm{P}\{s_k = 1\} \\
&= \mathrm{P}\left\{s_k = 1 \mid s_k^0 = 1\right\}\mathrm{P}\left\{s_k^0 = 1\right\} + \mathrm{P}\left\{s_k = 1 \mid s_k^0 = 0\right\}\mathrm{P}\left\{s_k^0 = 0\right\} \\
&= (1 - m_1)G(M - [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_k W}) + m_2[1 - G(M - [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_k W})] \\
&= m_2 + (1 - m_1 - m_2)G(M - [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_k W}) \\
&= \beta_k.
\end{aligned} \tag{4.3}$$

Since u_k is periodic, $\mathbb{E}[s_k] \in \{\mathbb{E}[s_1], \mathbb{E}[s_2], \ldots, \mathbb{E}[s_n]\}$. Therefore, it implies the following:

$$\mathbb{E}(s_{(q-1)n+i}) = \mathbb{E}(s_i) = m_2 + (1 - m_1 - m_2)G(M - [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_i W}) = \beta_i. \tag{4.4}$$

According to the law of large numbers, we obtain the following:

$$\frac{1}{D_N}\sum_{q=1}^{D_N} s_{(q-1)n+i} \rightarrow \mathbb{E}(s_i) = \beta_i, \text{ w.p.1 as } N \rightarrow \infty. \tag{4.5}$$

Combining Eqs (3.11) and (3.12), Theorem 4.1 is proven. □

Theorem 4.1 reveals that under attack, the parameter estimates converge to biased values that explicitly depend on the attack probabilities m_1 and m_2 and the system's noise distribution. Compared to the attack-free case where $\hat{\boldsymbol{\eta}}_N \rightarrow \boldsymbol{\eta}$ and $\hat{\boldsymbol{\theta}}_N \rightarrow \boldsymbol{\theta}$, the attack shifts the convergence point to $[\mathcal{L}_1(\tilde{\boldsymbol{\beta}}), \boldsymbol{\Omega^{-1}}\mathcal{L}_2(\tilde{\boldsymbol{\beta}})]$, thereby introducing a systematic identification error. This result quantitatively characterizes how data tampering distorts the learning process, thereby providing the necessary basis to formulate an optimization problem that maximizes this distortion under resource constraints.

4.2. Establishment of the attack strategy optimization model

Building upon the convergence analysis in Theorem 4.1, we now proceed to construct a scalar measure of attack effectiveness and formulate the corresponding optimization problem. The attack effectiveness is defined as the identification error induced by the attack, while the attack resource consumption is captured by constraints on the data tampering rate.

Assumption 4.2. *The dta is subject to both maximum energy constraints and average energy constraints, so the energy constraint can be expressed as the data tampering rate(dtr).*

- *maximum energy limitation:*

$$\kappa_{\sup} = \sup_{k\geq 1}\{\kappa_k\} \leq \overline{\kappa}_{\sup} < 1, \tag{4.6}$$

- *average energy limitation:*

$$kappa_{\text{ave}} = \lim_{o\to\infty}\sup \frac{1}{o}\sum_{k=1}^{o}\kappa_k \leq \overline{\kappa}_{\text{ave}} < 1, \tag{4.7}$$

where $\sup\{\cdot\}$ *denotes the supremum (least upper bound), and* $\kappa_k = \mathrm{P}\{s_k \neq s_k^0\}$ *represents the instantaneous data tampering probability at time step k.*

Remark 4.3. *The maximum tampering rate constraint is designed to prevent an attacker from expending excessive energy at any single instant, thus avoiding detection; conversely, the average tampering rate constraint limits long-term energy consumption and reflects the overall energy budget. Compared to a single constraint, this dual-constraint framework aligns more closely with realistic attack scenarios.*

For the convenience of subsequent descriptions, we define the following Remark:

$$G_{\min} = G(M - \max_{1\leq i\leq n}\{[\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_i}\boldsymbol{W}\}), \tag{4.8}$$

$$G_{\max} = G(M - \min_{1\leq i\leq n}\{[\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_i}\boldsymbol{W}\}), \tag{4.9}$$

$$G_{\text{ave}} = \frac{1}{n}G(M - [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_i}\boldsymbol{W}). \tag{4.10}$$

Theorem 4.4. *For System* (2.1) *and binary measurements* (2.2)*, under dtas, if Assumptions 2.2 and 2.3 hold, then the dtr at time k under the attack strategy* (m_1, m_2) *is given by the following:*

$$\kappa_k = m_1 G(M - [\eta_0, \eta]\alpha_k W) + m_2[1 - G(M - [\eta_0, \eta]\alpha_k W)]. \tag{4.11}$$

Moreover, the maximum energy limitation is given by the following:

$$\kappa_{\sup} = m_2 + (1 - m_1 - m_2)(I_{\{m_1\geq m_2\}}G_{\max} + I_{\{m_1<m_2\}}G_{\min}). \tag{4.12}$$

The average energy limitation is given by the following:

$$\kappa_{\text{ave}} = m_2 + (1 - m_1 - m_2)G_{\text{ave}}, \tag{4.13}$$

where $G_{\min}$*,* $G_{\max}$*, and* G_{ave} *are defined in Eqs* (4.8)–(4.10)*.*

Proof. Based on System (2.1) and the attack strategy (2.3), under Assumptions 2.2 and 2.3, and using the law of total probability, conditional probability, and Eq (3.11), the dtr at time k under the dta strategy (m_1, m_2) is given by the following:

$$\begin{aligned}
\kappa_k &= \mathrm{P}\left\{s_k \neq s_k^0\right\} \\
&= \mathrm{P}\left\{s_k = 0, s_k^0 = 1\right\} + \mathrm{P}\left\{s_k = 1, s_k^0 = 0\right\} \\
&= \mathrm{P}\left\{s_k = 0 \mid s_k^0 = 1\right\}\mathrm{P}\left\{s_k^0 = 1\right\} + \mathrm{P}\left\{s_k = 1 \mid s_k^0 = 0\right\}\mathrm{P}\left\{s_k^0 = 0\right\} \\
&= m_1\,\mathrm{P}\left\{s_k^0 = 1\right\} + m_2(1 - \mathrm{P}\left\{s_k^0 = 1\right\}) \\
&= m_1 G(M - [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_i}\boldsymbol{W}) + m_2(1 - G(M - [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_i}\boldsymbol{W})) \\
&= m_2 + (1 - m_1 - m_2)G(M - [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_i}\boldsymbol{W}),
\end{aligned} \tag{4.14}$$

which implies that Eq (4.14) holds.

According to Eq (4.14), the maximum dtr is given by the following:

$$\begin{aligned}
\kappa_{\sup} &= \sup_{k\geq 1}\{\kappa_k\} \\
&= \sup_{k\geq 1}\{m_1 G(M - [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_i}\boldsymbol{W}) + m_2(1 - G(M - [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_i}\boldsymbol{W}))\} \\
&= \begin{cases} m_2 + (1 - m_1 - m_2)G_{\max}, m_1 \geq m_2, \\ m_2 + (1 - m_1 - m_2)G_{\min}, m_1 < m_2. \end{cases}
\end{aligned} \tag{4.15}$$

The average dtr is given by the following:

$$\begin{aligned}
\kappa_{\text{ave}} &= \lim_{o\to\infty}\sup\frac{1}{o}\sum_{k=1}^{o} m_1 G(M - [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_i}\boldsymbol{W}) + m_2(1 - G(M - [\eta_0, \boldsymbol{\eta}]\boldsymbol{\alpha_i}\boldsymbol{W})) \\
&= m_2 + (1 - m_1 - m_2)G_{\text{ave}}.
\end{aligned} \tag{4.16}$$

Theorem 4.4 is proven. □

Under the attack strategy (m_1, m_2), the estimation error of Algorithm 1 given by steps (3.7)–(3.9) as follows:

$$J(m_1, m_2) = \|[\pounds_1(\tilde{\boldsymbol{\beta}}), \boldsymbol{\Omega^{-1}}\pounds_2(\tilde{\boldsymbol{\beta}})]^T - [\boldsymbol{\eta^T}, \boldsymbol{\theta^T}]^T\|_\infty, \tag{4.17}$$

where $\tilde{\beta}$ is given by Eq (4.2), and $\|\cdot\|_\infty$ denotes the infinity norm of a vector, which is the maximum absolute value among all elements of the vector.

Combining Eqs (4.15)–(4.17), this problem can be formulated as the following optimization problem:

$$\min_{\kappa_{\text{ave}}}\min_{\kappa_{\sup}}\max_{(m_1,m_2)} J(m_1, m_2) \tag{4.18}$$

$$\begin{aligned}
\text{s.t. } & \kappa_{\sup} = m_2 + (m_1 - m_2)(I_{\{m_1\geq m_2\}}G_{\max} + (1 - I_{\{m_1\geq m_2\}})G_{\min}) \leq \bar{\kappa}_{\sup}, \\
& \kappa_{\text{ave}} = m_2 + (m_1 - m_2)G_{\text{ave}} \leq \bar{\kappa}_{\text{ave}}, \\
& 0 \leq m_1, m_2 \leq 1.
\end{aligned} \tag{4.19}$$

The solution to the above optimization problem is denoted as (m_1^*, m_2^*), which is referred to as the optimal attack strategy.

5. Optimal attack strategy design and solution

According to Eq (4.17), the problem has no explicit solution and is instead addressed by selecting a numerical solution using optimization algorithms.

The SCA is a population-based metaheuristic optimization algorithm proposed by Mirjalili [30] in 2016. It updates the positions of candidate solutions using sine and cosine functions to perform both global and local searches within the search space. The core idea is to leverage the oscillatory properties of sine and cosine functions to periodically adjust the solution positions, thus enabling a "swinging" exploration capability. Each solution's position update considers the distance to a target (either the global best or the current best individual) and incorporates sine and cosine wave factors to enhance the search diversity, which is designed to enhance the probability of finding a global optimum through its oscillatory search mechanism, though convergence to a global optimum is not guaranteed in all cases.

The SCA structures its iteration strategy into two threads: global exploration and local exploitation. In the global exploration thread, significant random oscillations are applied to solutions to explore unknown areas of the search space; in the local exploitation thread, minor random perturbations are applied to solutions to thoroughly search the neighborhood of current solutions. The specific update formulas are as follows:

$$X = lb + \text{rand}(ub - lb), \tag{5.1}$$

$$X_i^{t+1} = \begin{cases} X_i^t + r_1 \cdot \sin(r_2) \cdot |r_3 \cdot P_i^t - X_i^t|, & \text{if } k < 0.5, \\ X_i^t + r_1 \cdot \cos(r_2) \cdot |r_3 \cdot P_i^t - X_i^t|, & \text{if } k \geq 0.5, \end{cases} \tag{5.2}$$

$$r_1 = \alpha\left(1 - \frac{t}{T}\right), \tag{5.3}$$

where r_1 controls the step size of the search and gradually decreases with iterations, thus reflecting the transition from exploration to exploitation, r_2 controls the phase of the sine and cosine functions, thus determining the search "direction", r_3 controls the attraction strength of the target solution toward the current position, and k is a random variable between 0 and 1 used to switch between sine and cosine functions, thus increasing the uncertainty and diversity of the search.

The convergence behavior of thew SCA depends on parameters such as (r_1, r_2) and the population size. Under the appropriate parameter tuning and sufficient iterations, the SCA tends to perform well in escaping the local optima, but global convergence is not theoretically guaranteed for non-convex problems.

To systematically obtain the optimal attack strategy, Algorithm 2 outlines the complete pseudocode of the SCA-based solver. The algorithm begins by initializing a population of candidate attack strategies (m_1, m_2) within the feasible domian. In each iteration, it evaluates the attack effectiveness $J(m_1, m_2)$ under the given energy constraints $\bar{\kappa}_{\text{sup}}$ and $\bar{\kappa}_{\text{ave}}$. Then, the positions of candidate solutions are updated using the sine and cosine oscillation rules given in Eqs (5.2) and (5.3), which balance the global exploration and the local exploitation. The process repeats until the maximum iteration count is reached, thus outputting the strategy that maximizes the identification error while satisfying all constraints.

Algorithm 2 Optimization problem solving based on the sine cosine algorithm under given constraints

Require: Parameters: M, μ, σ, N, n, u_k, η_0, $h(\cdot)$, m_1, m_2, $\bar{\kappa}_{\text{sup}}$, $\bar{\kappa}_{\text{ave}}$, population size S, maximum number of iterations Z

Ensure: Optimal attack strategy (m_1^*, m_2^*) that satisfies the constraints and maximizes $\|J(m_1, m_2)\|_\infty$

Initialize population with several (m_1, m_2) values
best_val $\Leftarrow 0$
$(m_1^*, m_2^*) \Leftarrow \emptyset$
for $t = 1$ to Z **do**
    for $l = 1$ to S **do**
        Compute β_i using Eq (4.3)
        Compute $\tilde{\beta}$ using Eq (4.2)
        Compute x_i using Eq (3.2) and Algorithm 1
        Compute $F_{\min}$, $F_{\max}$, F_{ave} using Eqs (4.8)–(4.10)
        Compute λ_{sup}, λ_{ave} using Eq (4.15) and Eq (4.16)
        if $\lambda_{\text{sup}}(m_{1i}, m_{2i}) \le \bar{\lambda}_{\text{sup}}$ and $\lambda_{\text{ave}}(\tau_{0i}, \tau_{1i}) \le \bar{\lambda}_{\text{ave}}$ and $\|x_i\|_\infty >$ best_val **then**
            best_val $\Leftarrow \|x_i\|_\infty$
            $(m_1^*, m_2^*) \Leftarrow (\tau_{0i}, \tau_{1i})$
        end if
    end for
    for $l = 1$ to S **do**
        for each dimension $z \in \{m_1, m_2\}$ **do**
            Randomly generate $r_1, r_2, r_3, k \in [0, 1]$
            Update $z_i^{(t+1)}$ using Eq (5.3) based on the value of k
            Clamp $z_i^{(t+1)}$ to $[0, 1]$
        end for
    end for
end for
return (m_1^*, m_2^*)

6. Numerical simulation

Consider the following system:

$$\begin{cases} y_k = \sum_{i=1}^{n_1} a_i x_{k-i+1} + n_k, \\ x_k = b_0 + \sum_{j=1}^{n_2} b_j h_j(u_k), b_0 = 1, \\ s_k^0 = I_{\{y_k \le M\}}. \end{cases}$$

The simulation parameters are selected following the setup in [26]. Specifically, the sample length is set to $N = 10{,}000$, the threshold is set to $M = 0$, and the system noise $\{n_k\}$ is i.i.d. Gaussian with a zero mean and a variance of 5. The static nonlinearity order is $n_1 = 1$ and the dynamic linear order is $n_2 = 2$, with system parameters $\eta = b_1 = 20$ and $\boldsymbol{\theta} = [a_1, a_2]^T = [3, -2]^T$. The periodic input $\{u_k\}$ has a period of 3 which takes values $[2, -1, 3]$, and the nonlinear function is $h_1(u_k) = (0.25u_k^3 - u_k^2 - u_k - 2)/40$.

Remark 6.1. *The assumption $M = 0$ entails no loss of generality: (i) a nonzero threshold can be absorbed by shifting the noise mean; (ii) the chosen parameters guarantee $0 < \mathbb{E}[s_k] < 1$, thus ensuring*

persistent excitation; and (iii) in attack optimization, M only appears through $G(\cdot)$, and setting $M = 0$ simplifies expressions without changing the problem.

Under the scenario without data tampering attacks, as can be seen from Figure 2, when the data length N is very large, the frequency approaches the probability. Simulations of Algorithm 1 and the PSO and BFGS algorithms are performed to verify the consistency, thus producing Figures 3–5 and Tables 2–4. In Figures 3–5, the subscripts of the parameters distinguish the convergence curves obtained by different algorithms: PSO, BFGS, and the proposed PSO-BFGS hybrid algorithm. Figures 3–5 show that as the data volume increases, the frequency eventually converges to the probability. Under different algorithms, the parameters $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\theta}}$ converge to the true values $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$, and the proposed PSO-BFGS algorithm achieves smaller convergence errors.

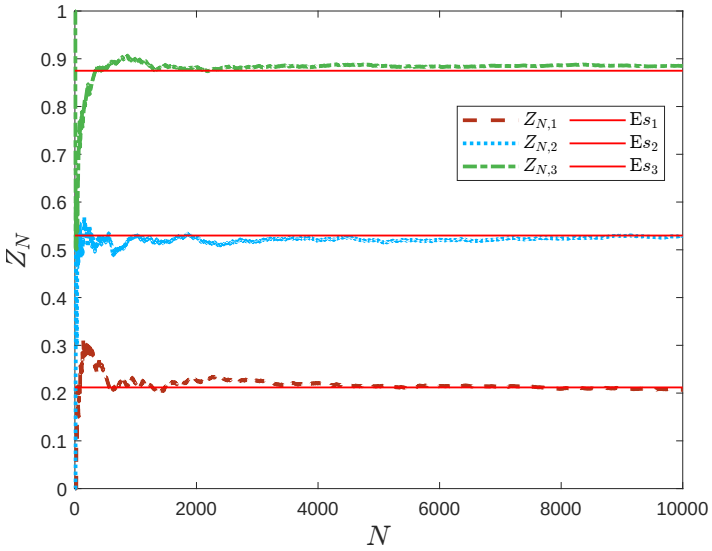

Figure 2. Convergence of the frequency under non-attacks.

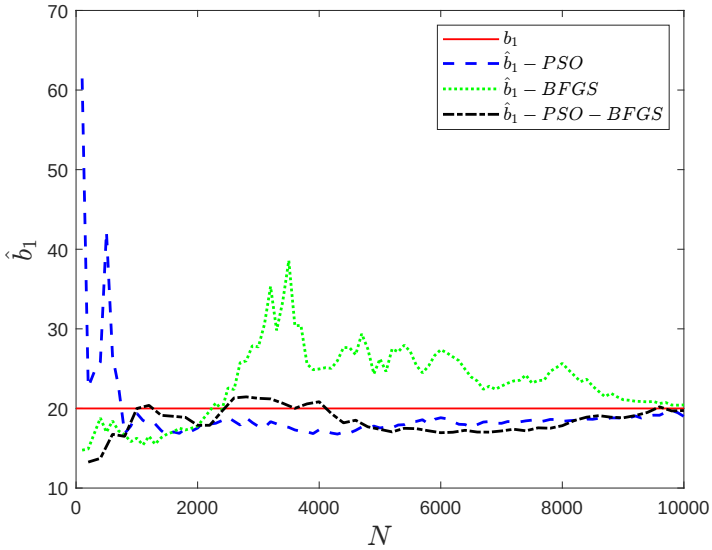

Figure 3. Comparison of $\boldsymbol{\eta}$ convergence under different algorithms without attacks.

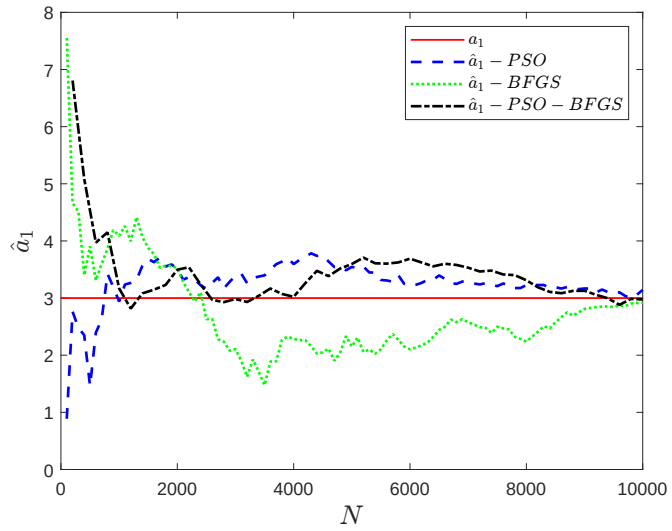

Figure 4. Comparison of a_1 convergence under different algorithms without attacks.

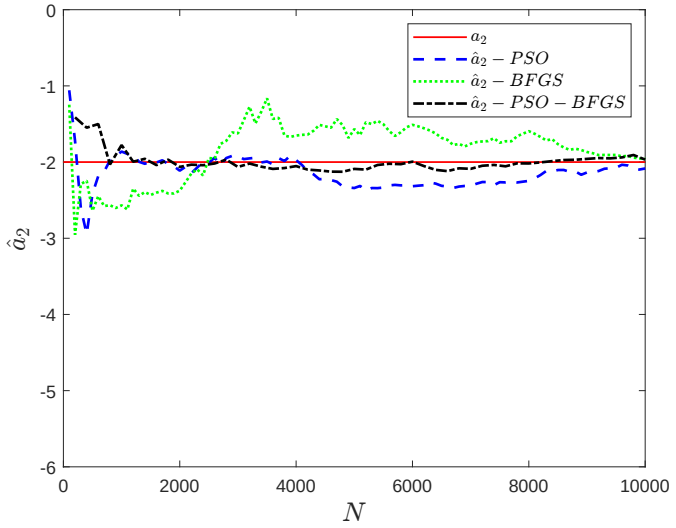

Figure 5. Comparison of a_2 convergence under different algorithms without attacks.

Table 2. Convergence and error of Parameter $\boldsymbol{\eta}$ under different identification algorithms.

	PSO	BFGS	PSO-BFGS
The value of N at the first convergence	750	2200	1200
Convergence error	−1.0027	0.4411	−0.2580

Table 3. Convergence and error of parameter a_1 under different identification algorithms.

	PSO	BFGS	PSO-BFGS
The value of N at the first convergence	750	2450	1100
Convergence error	0.1422	−0.0872	−0.0244

Table 4. Convergence and error of parameter a_2 under different identification algorithms.

	PSO	BFGS	PSO-BFGS
The value of N at the first convergence	800	2400	800
Convergence error	−0.0842	−0.0427	0.0324

When the system is subjected to a dta, the maximum dtr is set to $\bar{\kappa}_{\text{sup}} = 0.6$, and the average dtr rate is set to $\bar{\kappa}_{\text{ave}} = 0.6$. The population size S of the SCA parameter is 20, and the number of iterations K_{sca} is 100. The simulation results based on Algorithm 2 are shown in Figures 6 and 7, and the optimal attack strategy is $(m_1^*, m_2^*) = (0, 0.76097)$. Figure 6 shows that under the optimal attack strategy, the estimation error, (i.e., $J(m_1, m_2)$), reaches 15,299.7048. Additionally, as seen from Figure 8, the error value obtained under the optimal attack strategy is the largest, thereby achieving the maximum attack effect. In addition to the optimal strategy (m_1^*, m_2^*), two other attack strategies, $(m_1, m_2) = (0.1, 0.1)$ and $(m_1, m_2) = (0.3, 0.3)$, were simulated. According to Formulas (4.15) and (4.16), their maximum and average dtrs converged, as shown in Figures 9 and 10, respectively.

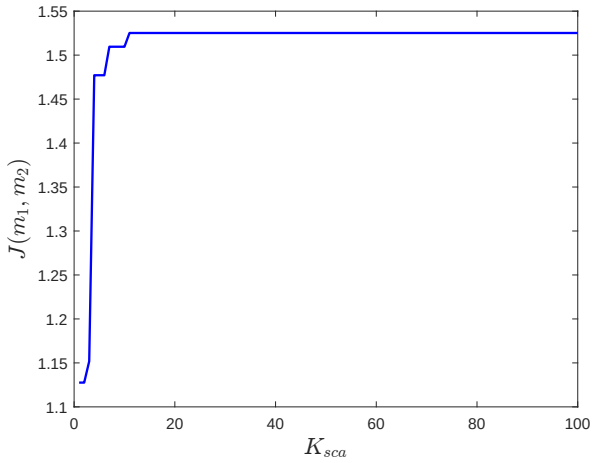

Figure 6. Convergence curve of fitness values in the sine cosine algorithm search.

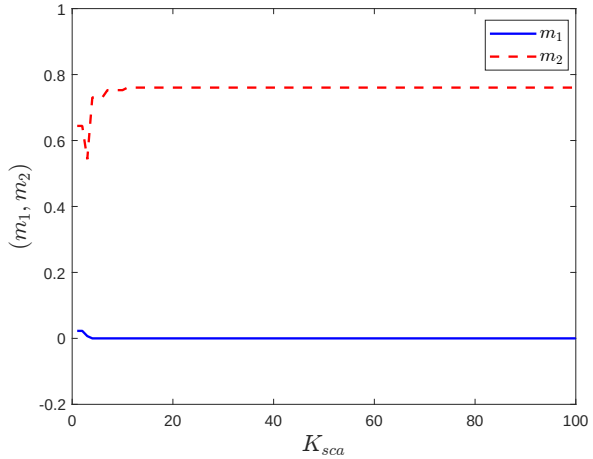

Figure 7. Convergence curve of (m_1, m_2) values in the sine cosine algorithm search.

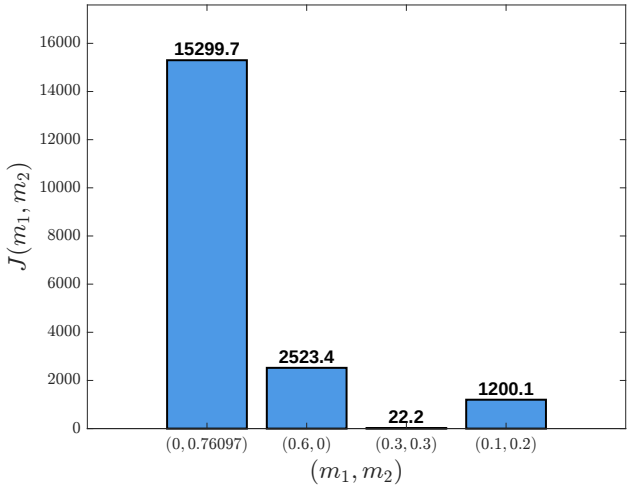

Figure 8. Bar chart comparing error values under different attack strategies.

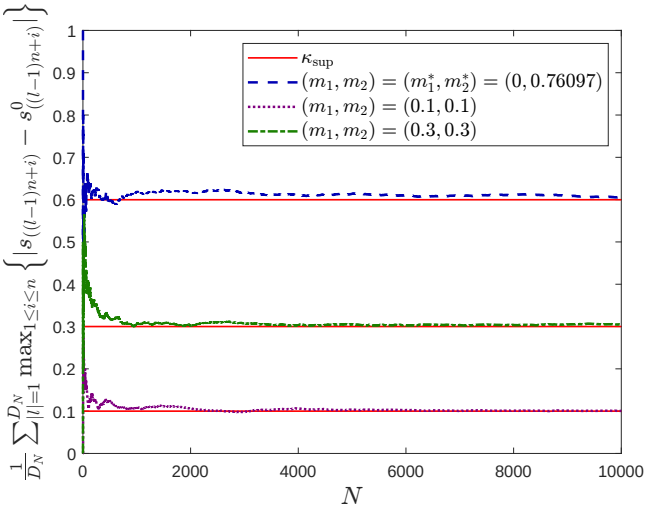

Figure 9. Convergence of the maximum data tampering rate.

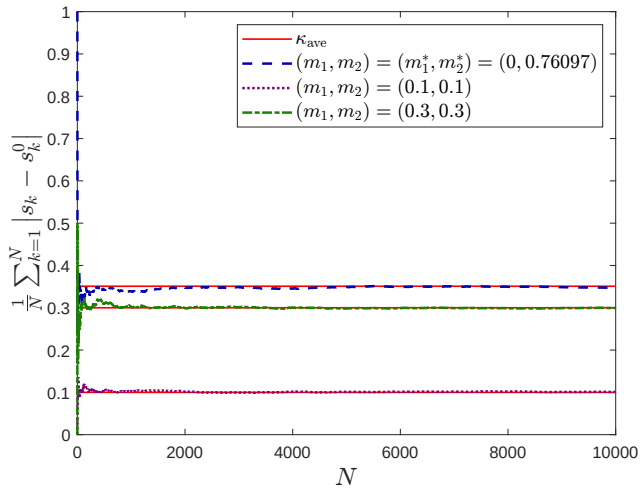

Figure 10. Convergence of the average data tampering rate.

In addition to the SCA, the SSA and GWO are employed for comparison. Set the number of iterations $K_{comparsion}$ to 100 and the population size S to 20, with the same constraints $\bar{\kappa}_{\text{sup}} = 0.6$ and $\bar{\kappa}_{\text{ave}} = 0.6$. The fitness convergence curves during the search process are shown in Figure 11, and the convergence curves of m_1 and m_2 values are shown in Figures 12 and 13, respectively. The results demonstrate that all three algorithms converge to similar final values; however, under the same number of iterations, the SCA algorithm can find a larger fitness value, thus indicating a better capability in searching for the global optimum.

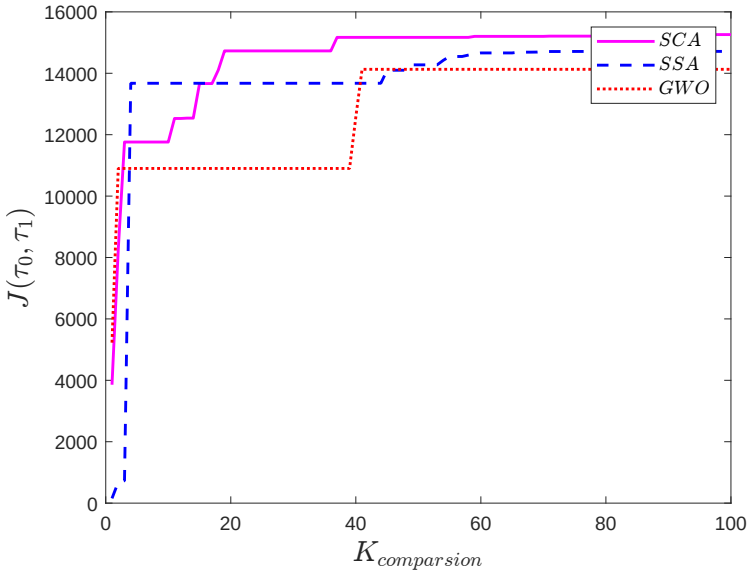

Figure 11. Comparison of fitness value convergence curves for different optimization algorithms.

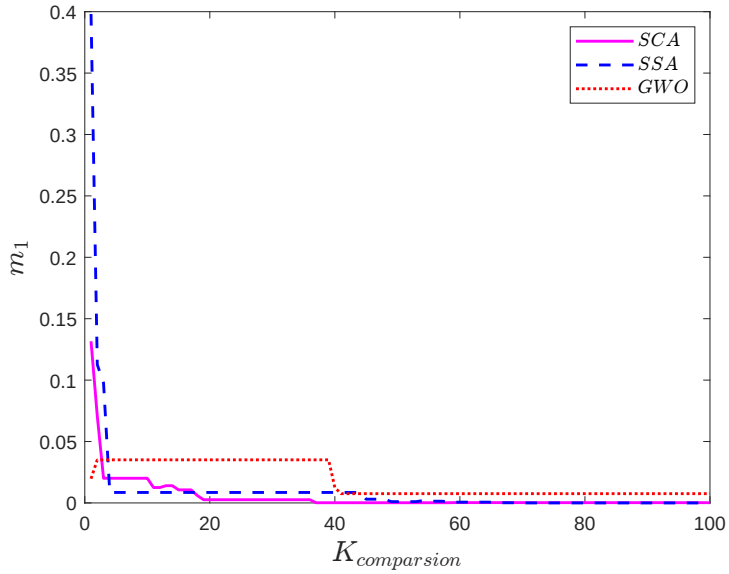

Figure 12. Comparison of m_1 value convergence curves for different optimization algorithms.

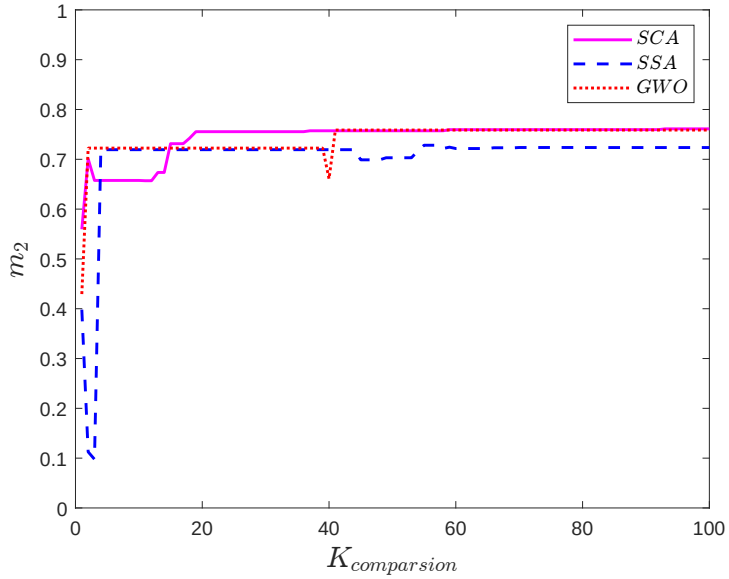

Figure 13. Comparison of m_2 value convergence curves for different optimization algorithms.

7. Discussion

7.1. Limitations of the study

The SCA exhibits a superior global search capability compared to other algorithms (SSA, GWO), as shown in Figures 11–13, thus indicating that its oscillatory exploration mechanism offers advantages in handling constrained non-convex optimization problems. However, the SCA is sensitive to parameters such as r_1 and r_2. More importantly, similar to most metaheuristic optimization methods, it does not provide theoretical global convergence guarantees for non-convex constrained problems. Therefore, the optimal attack strategy (m_1^*, m_2^*) obtained via the SCA is a numerical solution whose optimality is empirically validated through simulation rather than being theoretically proven. Additionally, this study only considers binary sensors and single-threshold detection scenarios, and assumes Gaussian system noise. In real industrial systems, sensors may have multi-level outputs, and noise distributions may be non-Gaussian, both of which could affect the effectiveness of attack strategies and the design of defense mechanisms.

7.2. Future research directions

Future work may proceed in the following directions: investigate the system identification and attack strategies under multi-level quantized outputs; and consider combined attack modes such as data tampering with false data injection or replay attacks. From a theoretical perspective, future work may explore the convexity properties of the objective function $J(m_1, m_2)$ with respect to (m_1, m_2) under specific noise distributions and system structures, thus potentially leading to provable convergence guarantees or performance bounds for the optimal attack strategy.

7.3. Attack strategy analysis from a maximum-disturbance perspective

From the perspective of control and signal processing, the optimal attack strategy studied in this paper can be viewed as a type of maximum-disturbance optimization problem. The attacker's goal is to maximize the disturbance to the system identification process under energy constraints, (i.e., to maximize the identification error $J(m_1, m_2)$). This aligns with the classic concept of "worst-case disturbance" in control theory, which typically seeks to find the disturbance signal that most severely degrades the system performance.

In this work, by tampering with binary measurement data, the attacker essentially injects structured disturbance into the observation channel of the system. This disturbance not only affects the observation at a single time instant, but, through the iterative nature of the system identification algorithm, it also leads to a systematic bias in the parameter estimation. Maximizing such disturbance is not simply a matter of increasing the tampering rate; rather, it requires an optimal allocation between the attack probabilities m_1 and m_2 to achieve the greatest disruptive effect under given energy constraints.

The simulation results show that the optimal attack strategy is $(m_1^*, m_2^*) = (0, 0.76076)$, indicating that, under the current system setup, unidirectional tampering (only tampering with $0 \to 1$ transitions) yields a greater disruptive effect than bidirectional uniform tampering. This is consistent with the principle in the maximum-disturbance theory that disturbance should target the most vulnerable part of the system: in this system, concentrating the attack energy on one direction of tampering more effectively exploits the system dynamics and nonlinearity, thereby amplifying the identification error.

In future research, by building on the maximum-disturbance framework, more complex disturbance structures, such as time-varying tampering strategies or correlated attack sequences, could be explored. Additionally, insights from disturbance rejection methods in the robust control and filtering theory may offer new directions to design defense strategies.

8. Conclusions

This paper studied the optimal attack strategy configuration problem that maximizes the identification error for a binary measurement Hammerstein system under data tampering attacks. First, in the absence of data tampering attacks, an identification algorithm based on a system of equations was proposed, along with an implementation of an identification algorithm based on the PSO-BFGS method. Moreover, from the attacker's perspective, to achieve maximum attack effectiveness under energy constraints, the attack problem was formulated as a constrained optimization problem. The sine cosine algorithm was employed to derive the optimal attack strategy from the attacker's viewpoint, and a comparative analysis with other algorithms was provided. Based on this work, the following future research directions can be explored: (i) extending the results to multi-level sensors and multi-threshold scenarios to study the system identification and attack strategies under the multi-level quantization output; (ii) considering joint attack modes, such as data tampering combined with fake data injection or replay attacks; and (iii) relaxing the periodic input assumption to more general input conditions (e.g., random or pseudo-random sequences) to enhance the practical applicability.

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 (62433020 and 62573044).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. R. Alguliyev, R. Aliguliyev, L. Sukhostat, An approach for assessing the functional vulnerabilities criticality of CPS components, *Cyber Secur. Appl.*, **3** (2025), 100058. https://doi.org/10.1016/j.csa.2024.100058
2. H. Fawzi, P. Tabuada, S. Diggavi, Secure estimation and control for cyber-physical systems under adversarial attacks, *IEEE Trans. Autom. Control*, **59** (2014), 1454–1467. https://doi.org/10.1109/TAC.2014.2303233
3. Z. Yu, H. Gao, X. Cong, N. Wu, H. H. Song, A survey on cyber-physical systems security, *IEEE Internet Things J.*, **10** (2023), 21670–21686. https://doi.org/10.1109/JIOT.2023.3289625
4. 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
5. Z. Jin, Y. Liu, J. Diao, Z. Wang, C. Sun, Z. Liu, Stealthy false data injection attacks on remote state estimation of cyber-physical systems, *Acta Autom. Sinica*, **51** (2025), 356–365. http://doi.org/10.16383/j.aas.c240527
6. J. Guo, R. Jia, R. Su, Y. Zhao, Identification of FIR systems with binary-valued observations against data tampering attacks, *IEEE Trans. Syst. Man Cybern. Syst.*, **53** (2023), 5861–5873. https://doi.org/10.1109/TSMC.2023.3276352
7. 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
8. M. Conti, N. Dragoni, V. Lesyk, A survey of man in the middle attacks, *IEEE Commun. Surv. Tutorials*, **18** (2016), 2027–2051. https://doi.org/10.1109/COMST.2016.2548426
9. L. Ming, Y. Leau, Y. Xie, Distributed denial of service attack in HTTP/2: Review on security issues and future challenges, *IEEE Access*, **12** (2024), 33296–33308. https://doi.org/10.1109/ACCESS.2024.3371013
10. F. Pasqualetti, F. Dörfler, F. Bullo, Attack detection and identification in cyber-physical systems, *IEEE Trans. Autom. Control*, **58** (2013), 2715–2729. https://doi.org/10.1109/TAC.2013.2266831

11. J. Tan, H. Jin, H. Hu, R. Hu, H. Zhang, H. Zhang, WF-MTD: Evolutionary decision method for moving target defense based on wright-fisher process, *IEEE Trans. Dependable Secure Comput.*, **20** (2022), 4719–4732. https://doi.org/10.1109/TDSC.2022.3232537
12. K. Long, V. Dhiman, M. Leok, J. Cortés, N. Atanasov, Safe control synthesis with uncertain dynamics and constraints, *IEEE Robot. Autom. Lett.*, **7** (2022), 7295–7302. https://doi.org/10.1109/LRA.2022.3182544
13. D. Ding, Q. Han, X. Ge, X. Zhang, J. Wang Privacy-preserving filtering, control and optimization for industrial cyber-physical systems, *Sci. China Inf. Sci.*, **68** (2025), 141201. https://doi.org/10.1007/s11432-024-4328-1
14. L. Mao, G. Yang, Optimal stealthy attack with side information against remote state estimation: A corrupted innovation-based strategy, *IEEE Trans. Cybern.*, **55** (2025), 897–904. https://doi.org/10.1109/TCYB.2024.3502790
15. R. Meira-Góes, R. H. Kwong, S. Lafortune, Synthesis of optimal multiobjective attack strategies for controlled systems modeled by probabilistic automata, *IEEE Trans. Autom. Control*, **67** (2022), 2873–2888. https://doi.org/10.1109/TAC.2021.3094737
16. M. Asghari, A. Ameli, M. Ghafouri, A. Kirakosyan, An optimal cyber-physical attack strategy on DC microgrids, *Int. J. Electr. Power Energy Syst.*, **157** (2024), 109900. https://doi.org/10.1016/j.ijepes.2024.109900
17. A. Truong, S. R. Etesami, J. Etesami, N. Kiyavash, Optimal attack strategies against predictors-learning from expert advice, *IEEE Trans. Inf. Forensics Secur.*, **13** (2018), 6–19. https://doi.org/10.1109/TIFS.2017.2718488
18. H. Wang, Y. Ren, Y. Fang, An optimal attack strategy based on adversary-modified historical innovations, *IEEE Internet Things J.*, **11** (2024), 33337–33345. https://doi.org/10.1109/JIOT.2024.3425894
19. Y. Chen, S. Kar, J. M. F. Moura, Optimal attack strategies subject to detection constraints against cyber-physical systems, *IEEE Trans. Control Network Syst.*, **5** (2018), 1157–1168. https://doi.org/10.1109/TCNS.2017.2690399
20. H. Zayyani, M. Salman, H. A. Hilal, Joint measurement and channel design of a malicious sensor in distributed estimation based on maximum disturbance in a sensor network, *IEEE Sensors Lett.*, **9** (2025), 7000104. https://doi.org/10.1109/LSENS.2024.3507579
21. H. Zhang, J. Yao, Z. Wang, S. Gao, H. Yan, Optimal DDoS attack strategy for cyber-physical systems: A multiattacker-defender game, *IEEE Syst. J.*, **18** (2024), 929–940. https://doi.org/10.1109/JSYST.2024.3381304
22. Y. Zhang, Z. Peng, G. Wen, J. Wang, T. Huang, Optimal stealthy linear man-in-the-middle attacks with resource constraints on remote state estimation, *IEEE Trans. Syst. Man Cybern. Syst.*, **54** (2024), 445–456. https://doi.org/10.1109/TSMC.2023.3311853
23. A. Kazemy, J. Lam, X. Zhang, Event-triggered output feedback synchronization of master-slave neural networks under deception attacks, *IEEE Trans. Neural Networks Learn. Syst.*, **33** (2022), 952–961. https://doi.org/10.1109/TNNLS.2020.3030638

24. X. Zhang, Q. Han, X. Ge, A novel approach to H_∞ performance analysis of discrete-time networked systems subject to network-induced delays and malicious packet dropouts, *Automatica*, **136** (2022), 110010. https://doi.org/10.1016/j.automatica.2021.110010

25. Y. Yang, S. Bi, R. Dai, Q. Shen, Self-triggered predefined-time cooperative control against DoS attacks for multiagent systems with uncertain powers, *IEEE Trans. Cybern.*, **55** (2025), 5202–5212. https://doi.org/10.1109/TCYB.2025.3574331

26. J. Guo, H. Liu, Hammerstein system identification with quantised inputs and quantised output observations, *IET Control Theory Appl.*, **11** (2017), 593–599. https://doi.org/10.1049/iet-cta.2016.1113

27. J. Guo, J. Zhang, Y. Zhao, Adaptive tracking of a class of first-order systems with binary-valued observations and fixed thresholds, *J. Syst. Sci. Complex.*, **25** (2012), 1041–1051. https://doi.org/10.1007/s11424-012-1257-0

28. G. Qian, L. Shen, J. Qian, S. Wang, Y. Chien, Robust adaptive Hammerstein filtering for censored regression under contaminated Gaussian noise, *IEEE Trans. Aerosp. Electron. Syst.*, **61** (2025), 18248–18261. https://doi.org/10.1109/TAES.2025.3611923

29. S. Li, M. Tan, I. W. Tsang, J. T. Kwok, A hybrid PSO-BFGS strategy for global optimization of multimodal functions, *IEEE Trans. Syst. Man Cybern. Part B*, **41** (2011), 1003–1014. https://doi.org/10.1109/TSMCB.2010.2103055

30. S. Mirjalili, SCA: A sine cosine algorithm for solving optimization problems, *Knowl.-Based Syst.*, **96** (2016), 120–133. https://doi.org/10.1016/j.knosys.2015.12.022

AIMS Press