sensors

MDPI

*Article*

# Data-Driven Detection of Stealthy IA Attacks in Industrial Cyber-Physical Systems via DGM Quantification †

**Jingzhao Chen [1,2], Bin Liu [2,*] and Zhiqun Jiang [1]**

[1] Henan Engineering Research Center of Intelligent Manufacturing and Digital Twin, SIAS University, Zhengzhou 451150, China; jzh.chen@sias.edu.cn (J.C.); zhiqun.jiang@foxmail.com (Z.J.)
[2] Engineering Research Center of Metallurgical Automation and Measurement Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China
* Correspondence: liubin@wust.edu.cn
† A preliminary and significantly shorter version of this research was presented in Chen, J.; Fan, Y.; Li, T. QDGM: A Framework of Detection and Performance Evaluation for Industrial CPS Security. In Proceedings of the 43rd Chinese Control Conference, Kunming, China, 28–31 July 2024; pp. 5178–5183.

**Abstract**

Industrial cyber-physical systems face increasing security threats from sophisticated cyber attacks. Traditional anomaly detectors generally require exact system models and noise statistics, which are often unavailable in practical industrial environments. To address this, this paper proposes a data-driven security detection framework based on the quantification of differences in generalized models (DGM). Utilizing only accessible operational data from the control layer, the method employs closed-loop subspace orthogonal projections to construct two detection variables: a static innovation sequence estimator and an extended dynamic Markov matrix estimator. The static estimator identifies fundamental model mismatches caused by standard denial-of-service and essential false data injection attacks. Meanwhile, the dynamic estimator successfully captures the structural distortions induced by advanced stealthy attacks that typically deceive Kullback–Leibler divergence detectors. The proposed methods were validated using a hardware-in-the-loop platform featuring a two degree-of-freedom robot and a DC servo motor. Experimental results confirm that the DGM framework effectively detects multiple types of stealthy integrity and availability (IA) attacks without relying on system parameters or degrading optimal control performance.

**Keywords:** industrial cyber-physical systems; data-driven detection; anomaly detection; subspace methods; stealthy attacks; false data injection



Academic Editor: Jorn Mehnen

Received: 30 July 2026
Revised: 3 September 2026
Accepted: 8 September 2026
Published: 10 September 2026



## 1. Introduction

Industrial cyber-physical systems (ICPSs) represent a class of intelligent systems characterized by the deep integration and real-time interaction between computing units and physical entities across diverse network environments [1–3]. By fundamentally transforming traditional manufacturing architectures, ICPSs have been widely deployed in critical national infrastructures, including smart grids, energy distribution, and advanced manufacturing [4,5]. However, this high degree of integration inherently expands the attack surface, allowing cyber threats to propagate directly from the network layer to the physical operational layer, thereby causing severe safety hazards. Recent real-world incidents, ranging from the Triton malware attack on safety instrumented systems [6] and the notorious Stuxnet manipulations [7] to the Florida water treatment plant breach [8] and sophisticated

cross-regional attacks via Starlink satellite links [9], underscore the urgency and criticality of robust security defense mechanisms for ICPS.

To counteract these escalating threats, security defense strategies are generally classified into three paradigms: security prevention, resilient control, and detection–isolation mechanisms [10–12]. While prevention and resilient control are effective, they often fail when adversaries successfully bypass network firewalls or launch destructive attacks targeting the actuator channels. Consequently, anomaly detection acting as the core of the detection–isolation paradigm has emerged as the last and most vital line of defense [13]. Traditional detection methods, such as the widely applied $\chi^2$ and Kullback–Leibler divergence (KLD) detectors [14], typically rely on state estimators to analyze the statistical anomalies of the residual sequences [15–17]. However, these model-based methodologies suffer from a fundamental limitation: they mandate the exact knowledge of the physical plant's mathematical model matrices and the precise statistical properties of process and measurement noises.

In practical industrial applications, especially those employing model-free, fuzzy, or non-linear control algorithms [18,19] alongside advanced privacy-preserving and unknown-input filtering schemes for complex physical environments [20], fulfilling such stringent modeling assumptions is highly unrealistic.

Furthermore, attackers continually evolve their strategies to evade detection, developing sophisticated stealthy attacks capable of bypassing residual-based monitors. To defend against advanced stealthy manipulations, such as replay and zero-dynamics attacks [21], researchers have proposed active defense techniques like physical watermarking [22,23] and random coding. Unfortunately, injecting exogenous noise signals or modifying hardware architectures inherently compromises the optimal control performance of the physical plant. Moreover, the deployment of watermarking techniques often creates coupling conflicts when multiple detectors are required to operate simultaneously.

To address the limitations of active defense, understanding and modeling the attack behaviors becomes paramount. In our previous work [24], we systematically investigated a multiplicative integrity and availability (IA) attack model. Unlike conventional additive false data injection (FDI) attacks that require continuous online signal generation, the multiplicative attack framework parameterizes the malicious actions as static deviations in the generalized system matrices. This model significantly reduces the adversary's hardware costs and exposure risks, while satisfying the essential stealthiness constraint (i.e., avoiding obvious anomalies in the control layer monitoring). The emergence of such advanced attack frameworks necessitates the development of novel detection schemes that are independent of nominal system models and do not degrade the inherent control performance.

Driven by these challenges, data-driven security detection techniques leveraging historical operational data have gained significant attention. Our preliminary explorations [18,25] have demonstrated the feasibility of detecting attacks by evaluating the modal mismatches using subspace methods. Building upon the foundational conceptual framework proposed in our preliminary conference paper [26], this article systematically investigates a completely data-driven detection methodology and establishes a mature security defense framework. The primary similarity between this study and our previous explorations lies in the shared mathematical basis of utilizing historical operational data for subspace evaluations. However, the current manuscript introduces substantial theoretical advancements and empirical differences. The unique academic contributions of this study compared to previous works are summarized as follows:

1. We propose a unified data-driven anomaly detection framework that strictly utilizes accessible operational data sequences, effectively eliminating the restrictive dependency on exact physical plant parameters and noise statistical knowledge. Furthermore,

the proposed DGM detectors do not require the injection of any exogenous signals, thereby ensuring zero degradation of the ICPS control performance.

2. Advancing beyond the basic static evaluations in our preliminary work, we derive an extended dynamic Markov matrix estimator and introduce a novel geometric mechanism analysis. This formal analysis elucidates how the dynamic estimator uniquely captures the rotational and stretching distortions induced by sophisticated FDI-II attacks that successfully deceive traditional KLD detectors.
3. We elevate the validation methodology from previous pure numerical simulations to comprehensive hardware-in-the-loop experiments. By developing a physical platform featuring a two degree-of-freedom robot and a DC servo motor, empirical results systematically validate the practicability and superiority of the proposed framework against denial-of-service attacks, essential FDI attacks, and highly stealthy FDI scenarios.

In summary, the primary theoretical contribution of this research lies in establishing a direct mathematical relationship between physical attack behaviors and the structural variations in the data subspace. Rather than simply providing an alternative numerical algorithm, the proposed framework explicitly quantifies how stealthy attacks alter the inherent system dynamics using strictly operational data. This provides a rigorous analytical foundation for designing security detectors in industrial cyber-physical systems where exact mathematical models are unavailable.

The remainder of this article is organized as follows: Section 2 formulates the system model and the multiplicative IA attacks. Section 3 details the data-driven DGM detection methodologies, including the derivations of the static and dynamic detection variables along with their geometric interpretations. Section 4 presents the experimental setup and analyzes the evaluation results. Finally, Section 5 concludes this paper.

## 2. Problem Formulation

As illustrated in Figure 1, the architecture of the considered ICPS primarily consists of three interacting components. These components are the control layer, the communication network, and the physical plant. From the defender's perspective at the control layer, the physical plant and the communication network including potential attackers are collectively treated as a generalized controlled system. Specifically, $y_r(k)$ is the reference trajectory, $y_c(k)$ and $u_c(k)$ denote the received output and control input at the control layer, respectively. $\Phi_a(k)$ and $\Phi_s(k)$ represent the attacks injected into the communication network, while $u_p(k)$ and $y_p(k)$ are the actual input and output of the physical plant subjected to process noise $w_x(k)$ and measurement noise $w_y(k)$. To facilitate readability, the main mathematical symbols used in this section are summarized in Table 1.

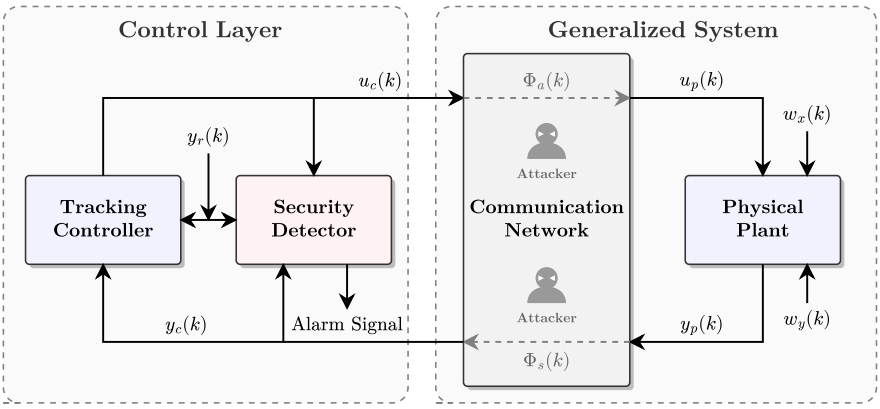


**Figure 1.** The tracking control ICPS deployed with a security detector.

**Table 1.** Nomenclature of main symbols used in Section 2.

| Symbol | Description |
|---|---|
| $x_p(k), u_p(k), y_p(k)$ | State vector, control input, and measurement output of the physical plant |
| $A_p, B_p, C_p$ | Model parameter matrices of the physical plant |
| $w_x(k), w_y(k)$ | Process noise and measurement noise |
| $n_x, n_u, n_y$ | Dimensions of the state, input, and output variables |
| $x_c(k), u_c(k), y_c(k)$ | State, output, and input variables of the tracking controller |
| $y_r(k), y_e(k)$ | Desired output trajectory and output tracking error |
| $\Phi_a(k), \Phi_s(k)$ | Time-varying attack matrices on the actuator and sensor channels |

### *2.1. System Description*

The underlying physical plant is located within the physical layer, and its dynamics can be described by a discrete-time linear time-invariant stochastic state-space model given by

$$\begin{cases} x_p(k+1) = A_p x_p(k) + B_p u_p(k) + w_x(k), \\ y_p(k) = C_p x_p(k) + w_y(k), \end{cases} \tag{1}$$

where $x_p(k) \in \mathbb{R}^{n_x}$, $u_p(k) \in \mathbb{R}^{n_u}$, and $y_p(k) \in \mathbb{R}^{n_y}$ represent the system state, control input, and measurement output, respectively. The matrices $A_p$, $B_p$, and $C_p$ denote the model parameters, while $w_x(k)$ and $w_y(k)$ represent the noise signals.

For the controlled physical plant defined in Equation (1), we consider the following two assumptions that distinguish our work from the existing literature in the security detection field.

**Assumption 1.** *The exact values of the system parameters $A_p$, $B_p$, and $C_p$ cannot be precisely known or are completely unknown to the defender. However, the dimensions of the system variables, namely $n_x$, $n_u$, and $n_y$, are known.*

**Assumption 2.** *The process noise $w_x(k)$ and measurement noise $w_y(k)$ in Equation (1) are independent Gaussian white noises, but their corresponding statistical parameters are unknown.*

**Remark 1.** *Systems satisfying the aforementioned two assumptions are common in practice, such as industrial control systems that utilize non-model-based controllers. Existing security detection methods are generally inapplicable under these two assumptions because they typically require exact knowledge of both the model parameters and the statistical properties of the noise. This reality motivates us to investigate a more practical data-driven security detection method. In this approach, the detector is constructed and deployed directly using the system's operational data stored at the control layer, without relying on the physical plant's model parameters.*

The inherent controller of the ICPS is an output feedback tracking controller formulated as

$$\begin{cases} x_c(k+1) = A_c x_c(k) + B_c y_e(k), \\ u_c(k) = C_c x_c(k) + D_c y_e(k), \end{cases} \tag{2}$$

where $y_e(k) \triangleq y_c(k) - y_r(k)$ denotes the output tracking error, with $y_r(k)$ being the desired output trajectory. The variables $y_c(k)$ and $u_c(k)$ represent the physical layer output received by the control layer and the control input to be sent to the physical layer, respectively.

### *2.2. Integrity and Availability Attack Model*

Suppose the attacker injects integrity and availability attacks into the network layer according to the model expressed as

$$\begin{cases} u_p(k) = \Phi_a(k)u_c(k), \\ y_c(k) = \Phi_s(k)y_p(k), \end{cases} \tag{3}$$

where $\Phi_a(k)$ and $\Phi_s(k)$ denote the time-varying attack matrices on the actuator and sensor channels, respectively. For an ideal communication network, when no attack is initiated, the attack matrices are identity matrices, meaning $u_p(k) = u_c(k)$ and $y_c(k) = y_p(k)$.

To provide a clearer physical interpretation, the multiplicative attack framework adopted herein reflects realistic operational vulnerabilities in networked industrial environments. Specifically, the actuator attack matrix $\Phi_a(k)$ characterizes malicious gain modifications or command corruptions introduced by compromised field control units or malicious firmware inside programmable logic controllers. Simultaneously, the sensor attack matrix $\Phi_s(k)$ represents sophisticated data spoofing and bias manipulations executed within vulnerable communication links or remote telemetry units. These adversarial models directly correspond to practical attack scenarios such as stealthy firmware tampering in industrial servo drives and covert measurement corruption in distributed water distribution networks.

Regarding the specific forms of the integrity and availability attacks, the following conditions are assumed.

**Assumption 3.** *The attack space formulated by the attacker contains four specific types of attacks. These include DoS, FDI satisfying essential stealthiness, FDI stealthy against the $\chi^2$ detector denoted as FDI-I, and FDI stealthy against the KLD detector denoted as FDI-II. The specific real-time attack type chosen by the attacker is unknown to the defender.*

To facilitate the study of the stealthy attack types FDI-I and FDI-II, which inherently assume system observability, we add the following general assumption for the physical system defined as Equation (1).

**Assumption 4.** *The physical system defined in Equation (1) is state-observable. Furthermore, valid $\chi^2$ and KLD detectors are assumed to exist for this system if the model parameters were known.*

To prevent attacks from causing severe damage to the physical system, the defender intends to deploy a security detector in the control layer, as shown in Figure 1. The alarm signal output by this detector will trigger preset isolation and protection mechanisms to secure the ICPS. The primary objective of this paper is to study a data-driven security detection method capable of handling the aforementioned attacks under Assumptions 1, 2, and 4.

## 3. Data-Driven Detection Mechanisms Based on DGM Quantification

To overcome the inherent limitations of traditional model-dependent anomaly detectors heavily relying on the exact knowledge of system matrices $A_p$, $B_p$, and $C_p$, this section details a comprehensive data-driven detection framework (see Figure 2). This mechanism evaluates the security status of the ICPS by quantifying the DGM using solely the historical and real-time operational data collected at the control layer.

In Figure 2, the cyan, yellow, and pink boxes represent the operational data windows under security, transitional, and attack modes, respectively. P1, P2, and P3 denote the three main phases of the framework: baseline acquisition, online detection variable computation,

and security alarm generation. In addition to the basic system parameters defined previously, the newly introduced symbols related to the data-driven detection methods in this section are listed in Table 2.

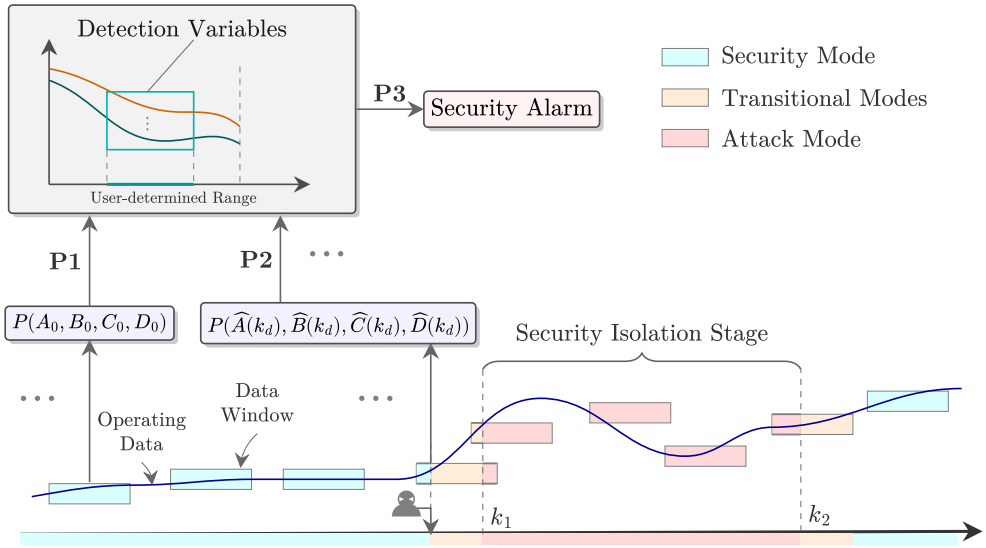


**Figure 2.** Schematic diagram of the security detection framework.

**Table 2.** Nomenclature of main symbols introduced in Section 3.

| Symbol | Description |
|---|---|
| $\alpha(k), \alpha_{th}$ | Real-time detection variable and the predefined detection threshold |
| $Q$ | Positive definite weighting matrix |
| $s$ | Length of the sliding data window |
| $U_f, Y_f, W_f$ | Block-Hankel matrices of past inputs, past outputs, and combined past data |
| $U_b, Y_b$ | Block-Hankel matrices of recent inputs and recent outputs |
| $Z_f, Z_b, z(k)$ | Block-Hankel matrices and the vector of the innovation sequence |
| $p, q$ | Number of block rows and block columns in the Hankel matrices |
| $\Gamma$ | Extended observability matrix |
| $\hat{C}_p$ | Estimated observation matrix under the subspace projection |
| $\hat{M}(k)$ | Consistent estimate of the extended Markov matrix |

### *3.1. Overall Detection Strategy and Output Estimation*

When the ICPS is subjected to an integrity and availability attack, the adversarial interventions fundamentally alter the generalized system model comprising the physical plant and the compromised network layer. Specifically, substituting the attack model into the nominal plant yields the compromised generalized model defined by

$$\begin{cases} x(k+1) = A_p x(k) + B_p \Phi_a(k) u_c(k) + w_x(k) \\ y_c(k) = \Phi_s(k) C_p x(k) + \Phi_s(k) w_y(k) \end{cases} \tag{4}$$

where the adversarial matrices $\Phi_a(k)$ and $\Phi_s(k)$ force the generalized control and measurement matrices to drift from their nominal states.

In conventional control systems, state estimation residuals are widely adopted for anomaly detection. A standard Kalman filter estimates the system state $\hat{x}(k)$ using the innovation gain $K(k)$ and relies on the precise mathematical models to generate the residual

signal $z(k) = y_c(k) - C_p\hat{x}(k)$. However, this conventional paradigm becomes invalid under our established assumptions, where model parameters remain entirely unknown.

To bridge this gap, we transform the state estimation problem into a generalized parameter estimation framework. For the unknown generalized system, we consider an input–output subsystem parameterized by an unknown vector $\theta$ and a regression vector $\varphi(k)$ containing historical inputs and outputs. During the secure operation mode, a recursive data-driven output estimator is constructed to iteratively approximate the subsystem dynamics. The detailed mathematical derivations and the specific iterative equations for this estimator, which is designed to minimize a predefined cost function utilizing historical operational data, have been comprehensively established in our preliminary work [25]. Consequently, the redundant formula expansions are omitted here.

The estimated parameter converges to a baseline value $\hat{\theta}_0$ at a specific steady-state instance $k_0$. The corresponding baseline estimation error $v_0 = y_c(k_0) - \varphi^\top(k_0)\hat{\theta}_0$ serves as the secure threshold metric.

When attacks trigger structural mismatches in the generalized model, the real-time estimation error $v(k)$ will deviate from its secure baseline $v_0$. To formalize this DGM strategy and enhance the detection sensitivity against sophisticated stealthy attacks, we present two specific subspace-based quantification methods in the following subsections encompassing static quantification via innovation sequences and dynamic quantification via extended Markov matrices.

The unified anomaly detector is evaluated using a weighted Euclidean distance metric formulated as

$$\alpha(k) = \|v(k) - v_0\|_Q^2 \underset{S_0}{\overset{S_1}{\gtrless}} \alpha_{th} \tag{5}$$

where $v(k)$ is the real-time detection variable constructed from the subspace data; $Q$ is a positive definite weighting matrix; $S_0$ and $S_1$ correspond to the secure and alarm states, respectively; and $\alpha_{th}$ is a predefined threshold configured based on an acceptable false alarm rate $\beta^*$.

Regarding the detailed estimation of the threshold $\alpha_{th}$ in Equation (5), it is computed empirically from a finite set of secure operational data collected during the baseline acquisition phase. Given a finite secure monitoring window of length $N$, the empirical maximum upper bound of the background fluctuations is recorded. The acceptable false-alarm rate $\beta^*$ is specified as a design parameter based on industrial tolerance limits, which determines the specific empirical safety margin added above the maximum bound. When the industrial plant undergoes significant operating-condition changes or shifts its reference trajectories, the nominal baseline and threshold require a brief recalibration during the transitional mode to maintain robust false-alarm resistance.

### *3.2. Static Quantification via Innovation Sequence Estimation*

To mathematically formulate the subspace-based detection variables, we first introduce several fundamental definitions regarding long time-series data matrices and orthogonal projections.

**Definition 1.** *For a long time-series data (denoted by LTD) matrix with a sequence length $l$, the mathematical correlation is defined by the LTD expectation expressed as*

$$\overline{E}_l[\mathcal{X}\mathcal{Y}^\top] \triangleq \lim_{l\to\infty} \frac{1}{l}\mathcal{X}\mathcal{Y}^\top \tag{6}$$

*where the matrices $\mathcal{X}$ and $\mathcal{Y}$ are block-Hankel matrices constructed from random variables over $l$ sampling instances.*

**Definition 2.** *For two matrices $\mathcal{M}$ and $\mathcal{N}$ with compatible dimensions, the orthogonal projection of the row space of $\mathcal{M}$ onto the row space of $\mathcal{N}$ is denoted by $\mathcal{M}/\mathcal{N}$ and computed by*

$$\mathcal{M}/\mathcal{N} \triangleq \mathcal{M}\mathcal{N}^\top(\mathcal{N}\mathcal{N}^\top)^\dagger\mathcal{N} \tag{7}$$

*where the superscript † represents the Moore–Penrose pseudo-inverse.*

**Definition 3.** *To guarantee numerical stability in practical ICPS setups and circumvent the severe truncation errors caused by the pseudo-inverse operation on massive data matrices, the orthogonal projection is strictly implemented via QR decomposition. Let the augmented matrix of $\mathcal{N}$ and $\mathcal{M}$ be decomposed as*

$$\begin{pmatrix}\mathcal{N}\\ \mathcal{M}\end{pmatrix} = \begin{bmatrix}R_{11} & 0\\ R_{21} & R_{22}\end{bmatrix}\begin{pmatrix}Q_1^\top\\ Q_2^\top\end{pmatrix} \tag{8}$$

*where $Q_1$ and $Q_2$ are orthogonal matrices and $R_{11}$ is a lower triangular matrix. The orthogonal projection is thereby robustly obtained by evaluating $\mathcal{M}/\mathcal{N} = R_{21}R_{11}^{-1}\mathcal{N}$.*

Building upon these definitions, we construct the subspace equations. Assuming the physical plant is observable according to Assumption 4, there exists a stabilizing observer gain $L$ rendering the matrix $A_o \triangleq A_p - LC_p$ strictly Schur stable. Under secure conditions where the attack matrices equal the identity matrix $I$, the system transforms into its innovation form described by

$$\begin{cases}x(k+1) = A_o x(k) + B_p u_c(k) + L y_c(k)\\ y_c(k) = C_p x(k) + z(k)\end{cases} \tag{9}$$

where $x(k)$ denotes the unmeasurable state vector and $z(k) \triangleq y_c(k) - C_p x(k)$ denotes the innovation sequence.

We define a sliding data window of length $s$. Using the operational data sequences $u_c(k)$ and $y_c(k)$, we formulate the block-Hankel matrices $U_f$, $Y_f$, $U_b$, and $Y_b$. Each block-Hankel matrix features $p$ block rows and $q$ columns satisfying the dimension constraint $q + 2p - 1 = s$. The subscripts $f$ and $b$ separate the data horizon into the past and the recent blocks, respectively.

**Theorem 1.** *Given the control layer operational data sequences $u_c(k)$ and $y_c(k)$ of length $s$, a consistent estimate of the innovation sequence $Z(k) = \{z(k-p+1), \cdots, z(k-p+q+2)\}$ corresponding to the recent data window can be obtained by*

$$\hat{Z}_b^- = Y_b^- - Y_b^-/W_f \tag{10}$$

*where $W_f \triangleq [U_f^\top, Y_f^\top]^\top$, and the block-Hankel matrices $Y_f$ and $Y_b$ are, respectively, given by*

$$Y_f = \begin{bmatrix}y_c(0) & y_c(1) & \cdots & y_c(q-1)\\ y_c(1) & y_c(2) & \cdots & y_c(q)\\ \vdots & \vdots & \ddots & \vdots\\ y_c(p-1) & y_c(p) & \cdots & y_c(p+q-2)\end{bmatrix}, \tag{11}$$

$$Y_b = \begin{bmatrix}y_c(p) & y_c(p+1) & \cdots & y_c(p+q-1)\\ y_c(p+1) & y_c(p+2) & \cdots & y_c(p+q)\\ \vdots & \vdots & \ddots & \vdots\\ y_c(2p-1) & y_c(2p) & \cdots & y_c(2p+q-2)\end{bmatrix}. \tag{12}$$

*The number of block rows satisfies* $n_x < p \ll s$, *and the number of columns* $q$ *satisfies* $q = s - 2p + 1$, *with* $y_c(0) \triangleq y_c(k-s+1), \ldots, y_c(2p+q-2) \triangleq y_c(k)$. *The structures of* $U_f$ *and* $U_b$ *are identical to those of* $Y_f$ *and* $Y_b$.

**Proof.** Based on the innovation model defined previously and the data sequences, the subspace equations can be formulated by recursive expansion

$$\begin{aligned} Y_f &= \Gamma X_f + H_u U_f + H_y Y_f + Z_f, \\ Y_b &= \Gamma X_b + H_u U_b + H_y Y_b + Z_b, \\ X_b &= A_o^p X_f + F_u U_f + F_y Y_f. \end{aligned} \tag{13}$$

In Equation (13), $Z_f$ and $Z_b$ are block-Hankel matrices constructed from the innovation sequence $z(k)$. The augmented state variables $X_f$ and $X_b$ and the extended observability matrix $\Gamma$ are, respectively, expressed as

$$X_f = \begin{bmatrix} x(0) & x(1) & \cdots & x(q-1) \end{bmatrix}, \tag{14}$$

$$X_b = \begin{bmatrix} x(p) & x(p+1) & \cdots & x(p+q-1) \end{bmatrix}, \tag{15}$$

$$\Gamma = \begin{bmatrix} C_p^\top & (C_p A_o)^\top & (C_p A_o^2)^\top & \cdots & (C_p A_o^{p-1})^\top \end{bmatrix}^\top. \tag{16}$$

$H_u$ and $H_y$ are lower triangular block-Toeplitz matrices with identical structures. Taking $H_u$ as an example

$$H_u = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ C_p B_p & 0 & 0 & \cdots & 0 \\ C_p A_o B_p & C_p B_p & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_p A_o^{p-2} B_p & C_p A_o^{p-3} B_p & C_p A_o^{p-4} B_p & \cdots & 0 \end{bmatrix}. \tag{17}$$

$F_u$ and $F_y$ also share identical structures, as shown by

$$F_u = \begin{bmatrix} A_o^{p-1} B_p & A_o^{p-2} B_p & \cdots & A_o B_p & B_p \end{bmatrix}. \tag{18}$$

Consider the subspace state equation in Equation (13). Since the system is observable and $A_o = A_p - LC_p$ is strictly stable, for a sufficiently large $p$, we have

$$\begin{aligned} \overline{E}_p[X_b] &= \overline{E}_p[A_o^p X_f + F_u U_f + F_y Y_f] \\ &= \overline{E}_p[F_u U_f + F_y Y_f]. \end{aligned} \tag{19}$$

Denoting $\hat{X}_b = F_u U_f + F_y Y_f$, $H \triangleq H_u U_b + H_y Y_b + Z_b$ and substituting it into the expression of $Y_b$ in Equation (13) to eliminate $X_b$, we obtain

$$\hat{\Gamma}(F_u U_f + F_y Y_f) = Y_b - \hat{H}, \tag{20}$$

and the corresponding estimation error is $\overline{E}_p[\Gamma A_o^p X_f] = 0$.

Extracting the first block row of Equation (20) (denoted by the superscript "$-$") and noting that the first block rows of the strictly lower triangular block-Toeplitz matrices $H_u$ and $H_y$ are identically zero matrices (i.e., $H_u^- = 0$ and $H_y^- = 0$), we deduce

$$\hat{C}_p(F_u U_f + F_y Y_f) = Y_b^- - \hat{Z}_b^-, \tag{21}$$

where $Y_b^-$ and $Z_b^-$ denote the first block rows of the matrices $Y_b$ and $Z_b$ (i.e., the first $n_y$ rows), respectively.

Defining the combined past data matrix $W_f = [U_f^\top, Y_f^\top]^\top$, Equation (21) can be rewritten as

$$Y_b^- = \hat{C}_p \begin{bmatrix} F_u & F_y \end{bmatrix} W_f + \hat{Z}_b^-. \tag{22}$$

Applying the orthogonal projection operation onto the row space of $W_f$ on both sides of Equation (22) yields

$$Y_b^- / W_f = \hat{C}_p \begin{bmatrix} F_u & F_y \end{bmatrix} W_f + \hat{Z}_b^- / W_f. \tag{23}$$

Crucially, the innovation sequence $Z_b^-$ behaves as white noise and is statistically uncorrelated with the past data matrix $W_f$, meaning $\overline{E}_s[\hat{Z}_b^- / W_f] = 0$. Therefore, substituting this property into Equation (23) and combining it with Equation (22), we obtain the consistent estimate of the innovation sequence

$$\hat{Z}_b^- = Y_b^- - \hat{C}_p \begin{bmatrix} F_u & F_y \end{bmatrix} W_f = Y_b^- - Y_b^- / W_f. \tag{24}$$

The estimation error of $\hat{Z}_b^-$ is $\overline{E}_s[\hat{Z}_b^- / W_f] = 0$. Since $Z_b^-$ consists of the first $n_y$ rows of $Z_b$, according to Equation (12), for the operational data generated from time $k-s+1$ to time $k$, we have $Z_b^- = Z(k) = \{z(k-p+1), \cdots, z(k-p+q+2)\}$. Thus, Equation (24) is a consistent estimate of the innovation sequence $Z(k)$. This completes the proof. □

**Remark 2.** *The core essence of Theorem 1 lies in separating the predictable system baseline from unpredictable anomalies using pure data-driven orthogonal projections. By projecting the recent operational data onto the mathematical subspace spanned by the historical observation data, the predictable functional relationships are effectively isolated. The remaining orthogonal residual directly constitutes the innovation sequence representing the novel information or potential disturbances entering the physical plant. Building upon this theoretical decoupling mechanism, the estimated innovation vector transformed into a column sequence $v_s(k) \triangleq vec(\hat{Z}_b^-)$ serves as the primary static detection variable. Any numerical deviations observed in $v_s(k)$ relative to its nominal secure baseline $v_{s,0}$ effectively capture the steady-state discrepancies imposed by essential attacks.*

**Remark 3.** *It is worth noting from a finite-sample perspective that because the subspace matrices are constructed using finite operational data rather than infinite realizations, the empirical estimation error of the innovation sequence is bounded by the persistent excitation level of the input data and the singular value distribution of the Hankel matrices. As the data window sample size increases, the empirical distribution of the finite-sample estimation converges reliably to its theoretical counterpart.*

### *3.3. Dynamic Quantification via Extended Markov Matrix Estimation*

While the static innovation sequence strictly evaluates the amplitude variations representing the steady-state model mismatch, capturing the temporal transition behaviors is universally required for mitigating advanced stealthy anomalies. To accomplish dynamic quantification, we leverage the Markov parameters of the generalized system.

**Definition 4.** *For a typical discrete LTI system governed by matrices $A$, $B$, $C$, and $D$, the Markov parameter sequence denoted by $J(p)$ represents the impulse response data over $p$ sampling steps formulated as $J(p) = \{D, CB, CAB, \ldots, CA^{p-2}B\}$. The matrix constructed sequentially by these block elements constitutes the complete Markov matrix identifying the dynamic transient properties of the system.*

**Theorem 2.** *For the ICPS consisting of the physical plant and the tracking controller, and given the control layer operational data sequences $u_c(k)$ and $y_c(k)$ of length $s$, when the system input satisfies the persistent excitation condition, a consistent estimate of the extended Markov matrix $M = C_p \begin{bmatrix} F_u & F_y \end{bmatrix}$ from the augmented input $\zeta(k) \triangleq [u_c^\top(k), y_c^\top(k)]^\top$ to the system output $y_c(k+1)$ is given by*

$$M(k) = Y_b^- W_f^\top (W_f W_f^\top)^{-1} \tag{25}$$

**Proof.** The rigorous mathematical derivations regarding the consistent estimation of the extended Markov matrix have been presented in our preliminary work [26]. The process primarily relies on the intermediate algebraic result $Y_b^- / W_f = M W_f$ derived from the subspace projection operations. When the control inputs fulfill the persistent excitation condition, the unique optimal solution can be obtained by applying the standard least squares estimation method, which results in Equation (25). This completes the proof. □

**Remark 4.** *Theorem 2 further advances the static security evaluation by providing a real-time consistent estimator for the dynamic matrix features. This theorem mathematically guarantees that even under severe unknown measurement noises, the inherent dynamic characteristics of the physical plant can be accurately tracked through the block-Hankel matrices. The calculated extended Markov matrix continuously reflects the authentic dynamic mapping relationship between the inputs and outputs. Consequently, the dynamic detection variable is subsequently constructed using the vectorization operation, resulting in $v_d(k) \triangleq vec(\hat{M}(k))$. Substituting $v_d(k)$ back into the generalized distance metric framework expressed in Equation (5) fully establishes the dynamic anomaly detector targeting sophisticated temporal manipulations.*

### *3.4. Geometric Mechanism Analysis*

To better understand why certain detection variables are effective against advanced stealthy attacks while others fail, it is necessary to analyze the geometric mechanism of the proposed DGM method in the subspace domain.

As illustrated in Figure 3a, which shows the orthogonal relationship between the past data $W_f$ and the innovation $Z_b^-$ under secure conditions, the row space of the past data matrix $W_f$ (represented horizontally) is strictly orthogonal to the row space of the innovation matrix $Z_b^-$ (represented vertically). The geometric interpretation of Theorem 1 is that the nominal output $Y_b^-$ is decomposed into two orthogonal components: the projection onto the past data subspace ($M_0 W_f$) and the orthogonal residual ($Z_b^-$).

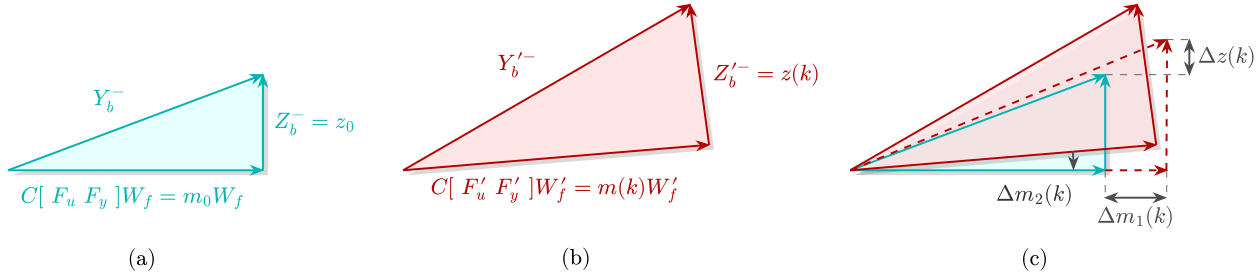


**Figure 3.** Geometric interpretations of two types of detection variables. (**a**) The orthogonal relationship between the past data $W_f$ and the innovation $Z_b^-$ under secure conditions. (**b**) The subspace distortion under an integrity and availability attack. (**c**) Comparison of deviations in the static and dynamic detection variables.

When an IA attack occurs, it alters the inherent structures of the system matrices. Geometrically, as shown in Figure 3b (illustrating the subspace distortion under an integrity and availability attack) and Figure 3c (displaying the comparison of deviations in the static

and dynamic detection variables), the attack stretches and rotates the original row space $M_0 W_f$ into a distorted space denoted by $M(k)W_f'$. To facilitate a direct comparison between the secure and attacked states, Figure 3c aligns the two geometric representations. The dashed lines serve purely as auxiliary lines to visualize the dimensional changes in the subspace matrices. Specifically, the gray arrows denoted by $\Delta z(k)$ and $\Delta m_1(k)$ represent the magnitude variations in the innovation sequence (the right edge) and the base subspace matrix (the bottom edge) before and after the attack, respectively. Meanwhile, the gray arrow $\Delta m_2(k)$ indicates the angular rotational shift of the subspace caused by the attack.

In an FDI-II attack, the attacker carefully designs the injected noise signals to maintain the orthogonal relationship between the compromised outputs and the historical data. Consequently, the rotation angle $\Delta m_2(k)$ of the row space is significantly restricted. This causes the variation in the innovation sequence, $\Delta z(k)$, to be extremely small and often fall below the detection threshold $\alpha_{th}$. As a result, the static innovation detector fails to detect the attack.

In contrast, the extended Markov matrix $\hat{M}(k)$ evaluates both the linear stretching and rotational shifts of the subspace, represented by $\Delta m_1(k)$ and $\Delta m_2(k)$. By capturing the dynamic characteristics and generalized impulse responses of the system, this dynamic detection mechanism is highly sensitive to the structural changes caused by attacks, thereby successfully detecting stealthy attacks.

To provide a clearer intuitive illustration of the theoretical derivations and heavy subspace notations, consider a practical operational instance where the sliding data window receives continuous input–output samples from the control layer. As historical data accumulate into the block-Hankel matrices $U_f$ and $Y_f$, the subspace projection operation isolates the predictable functional relationship without requiring explicit knowledge of the physical system matrices $A_p$ and $B_p$. When a stealthy attack occurs, the data flow entering the Hankel matrices causes an immediate structural rotation, which is quantitatively captured by the projection residual or the extended Markov matrix estimator. This step-by-step data processing workflow ensures that abstract mathematical formulations are directly translated into intuitive, real-time security indicators.

## 4. Hardware Experiments and Analysis of Results

Before delving into the hardware implementation, it is worth noting that the preliminary validation of the proposed data-driven detection methodology has been conducted through numerical simulations on a Western States Coordinated Council (WSCC) 9-bus power system, which is detailed in our previous work [25]. Building upon those fundamental simulation results, this section is dedicated to further evaluating the practical performance and robustness of the proposed detectors against sophisticated stealthy attacks using a physical testbed.

In this section, a two degree-of-freedom (2-DoF) robot and a DC servo motor (both manufactured by Quanser Inc., Markham, ON, Canada) are selected as the controlled physical plants of the ICPS to experimentally validate and evaluate the performance of the two detection methods proposed in Section 3. The experiments are divided into two parts. The first experiment aims to verify the detection performance of the two detectors against DoS and essential FDI attacks, respectively. The second experiment is designed to evaluate their performance against the first and second types of stealthy FDI attacks (i.e., FDI-I and FDI-II).

The established hardware-in-the-loop (HIL) experimental platform is shown in Figure 4. The defender's controller and detector are deployed in a dedicated HIL environment (Quarc 2019 SP1, Quanser Inc., Markham, ON, Canada) on the defender's PC, while the attacker implements the four attack strategies based on the HIL environment on another

PC to simulate practical IA attack scenarios. UDP communication protocol is used for data exchange between the two PCs to construct the network layer of the ICPS.

The 2-DoF robot consists of two DC servo motors and a rigid four-bar linkage mechanism, as shown in Figure 5. Here, points $A, B, C$, and $D$ denote the rotational joints, and point $E$ is the end-effector. $L_b$ represents the length of the rigid links, while $q_A$ and $q_B$ indicate the rotational angles controlled by the two servo motors. The end-effector is point $E$, and the control objective is to make point $E$ track a predefined trajectory. The servo motor in the robot system has the same structure as the standalone servo motor module, and its armature circuit and gear transmission components are shown in Figure 6. The output of the standalone servo motor system is the motor rotation angle, and the objective is angle tracking control. The inherent controllers for the systems are all discrete PID controllers. The desired trajectory coordinates for the end-effector $E$ of the 2-DoF robot are defined as $y_r(k) \triangleq [y_{rx}(k), y_{ry}(k)]^\top$, where $y_{rx}(k) = 5 + \sin(0.05\pi k)$ and $y_{ry}(k) = 5 + \sin(0.1\pi k)$. The desired output angle for the servo motor module is $y_{rm}(k) = 10\sin(0.1\pi k)$.

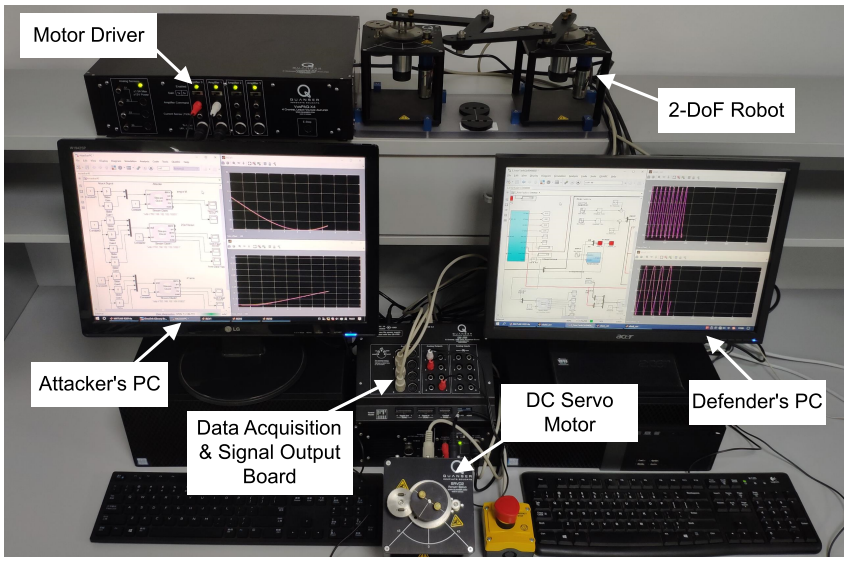


**Figure 4.** The HIL experimental platform for 2-DoF robots and a servo motor.

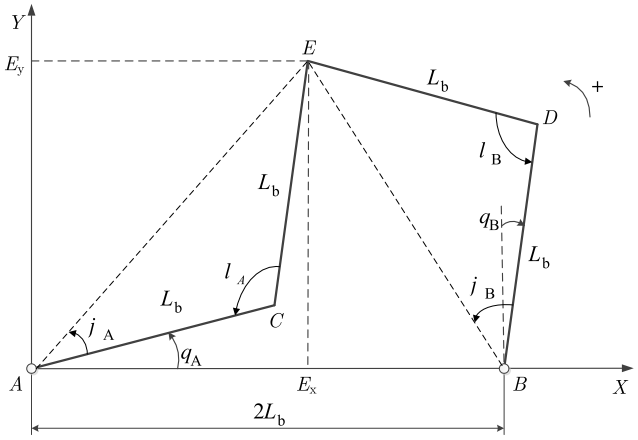


**Figure 5.** Linkage mechanism of the 2-DoF robot.

### *4.1. Performance Evaluation Under DoS and Essential FDI Attacks*

The purpose of this experiment is to verify the detection performance of the innovation sequence estimator and the extended Markov detector against DoS and essential FDI attacks, respectively. HIL experiments are conducted on the established platform with a sampling time of 0.02 s and a total experimental duration of 200 s. The attacker launches DoS and

essential FDI attacks against the servo motor and the 2-DoF robot subsystems, respectively, starting from 100 s until the end of the experiment. The two proposed detectors are deployed in the control layer to monitor the security status of the ICPS.

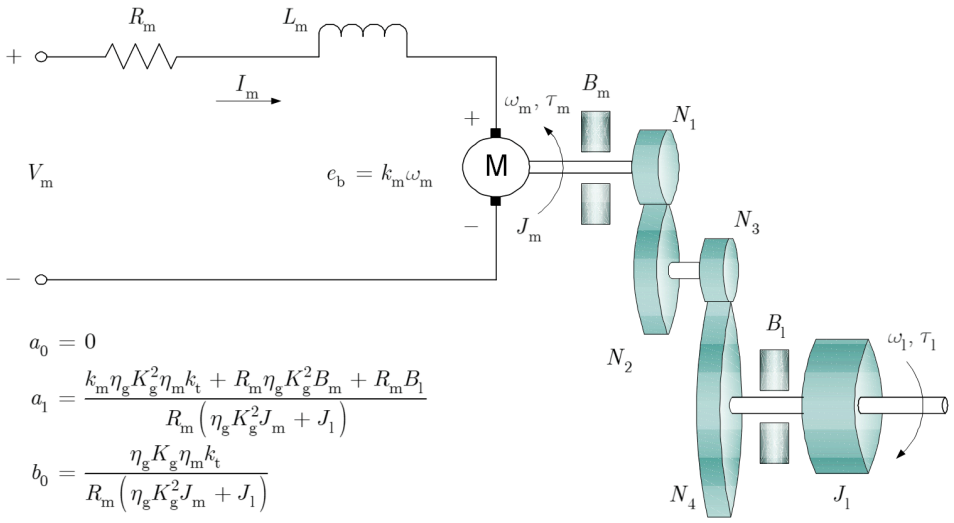


**Figure 6.** The DC servo motor armature circuit and its driven components.

Regarding the attack parameter settings, the DoS attack targets both the sensor and actuator channels of the servo motor subsystem, and the corresponding packet loss probabilities are both set to 35% to avoid hardware damage to the experimental system. The parameters for the essential FDI attack are set as shown in Equation (26):

$$\begin{cases} \Phi_a(k) = 0.0167k' \,\mathrm{diag}\{1, 1.2\}, \\ \Phi_s(k) = I + S(k)\,\mathrm{diag}\{y_{p1}(k), y_{p2}(k)\}^{-1}, \\ S(k) = \mathrm{diag}\{-0.5 + 1.2\sin(0.05\pi k), 0.5 - 1.2\sin(0.05\pi k)\}. \end{cases} \tag{26}$$

where $k' = k - k_0$, with $k_0$ representing the attack onset time, and $S(k)$ representing the FDI attack bias signal. In terms of the detector parameter settings, the number of block rows of the subspace matrix is set to $p = 5$, the data window length is $s = 1009$, the data length of the innovation sequence (i.e., the number of columns in the subspace matrix) is $q = 1000$, and the weighting matrices $Q$ for both the innovation estimation detector and the extended Markov detector are set as identity matrices.

The assignment of these specific values is jointly determined by the theoretical prerequisites of the subspace matrix processing techniques and the practical operational conditions of the physical system. Specifically, the number of block rows is chosen as five to guarantee that the extended observability matrix achieves full column rank, which is necessary to fully capture the underlying mechanical dynamics of the robot without inducing excessive computational delay. The column dimension is set to one thousand to ensure the statistical consistency of the data-driven estimation, effectively smoothing out the non-Gaussian noises inherent in the hardware sensors. The total sliding window length of one thousand and nine is subsequently dictated by the strict mathematical construction requirements of the block-Hankel matrices. Finally, the weighting matrices are configured as identity matrices to evaluate all measurement channels uniformly without introducing subjective baseline bias.

In the HIL experimental environment, the predefined detection threshold is determined empirically during the initial baseline acquisition phase to ensure robustness against false alarms. Because pure theoretical thresholds often fail to account for unmodeled mechanical friction and non-Gaussian measurement noises inherent in physical servo motors,

we operate the 2-DoF robot under normal secure conditions to observe the steady-state responses. By recording the maximum fluctuation upper bound of the detection variable during this secure phase and incorporating a slight empirical margin, the threshold is decisively established. This practical calibration effectively absorbs the unmodeled systemic disturbances and forms a reliable baseline for the subsequent online security monitoring.

Figure 7 presents the output curves of the servo motor system under DoS attack. To highlight the changes in system tracking performance before and after the attack, only 60 s of data are shown. It can be observed that under a relatively low packet loss rate, the system can still track the predefined trajectory but with noticeable performance degradation. The reason is that both the control layer and physical layer of the system contain zero-order holds, which is common in practice. Moreover, as seen from Figure 7, the DoS attack does not satisfy essential stealthiness, but since it is stealthy to the $\chi^2$ detector, researching effective detection methods for this type of attack remains valuable.

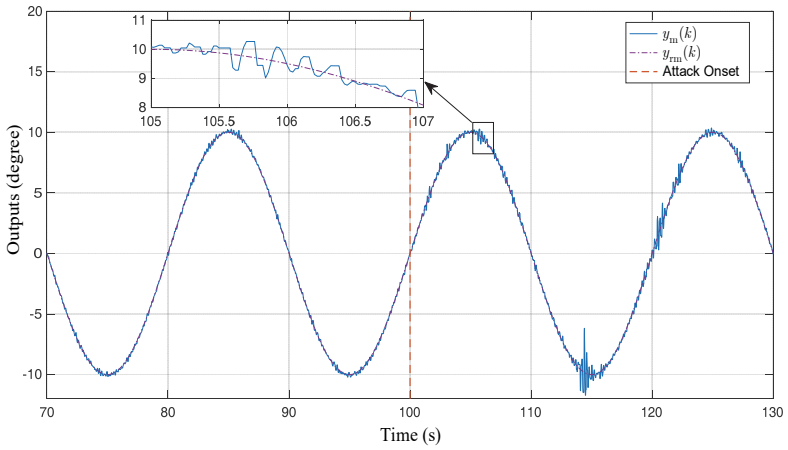


**Figure 7.** Servo motor output tracking control curve under DoS attack.

The operating trajectory of the 2-DoF robot system under essential FDI attack is given in Figure 8. In this figure, the main central plot illustrates the actual operational trajectory of the physical plant. Here, the solid blue line represents the tracking control curve of the robot under normal secure conditions, the solid orange line with dot markers indicates the actual compromised trajectory under attack, and the dashed purple line represents the predefined reference trajectory. The inset in the upper-right corner provides a local magnified view of the main plot at the attack onset. Meanwhile, the sub-figure in the lower-left corner shows the output trajectory as displayed from the perspective of the control layer. It is clear that this control layer curve shows no obvious anomalies before and after the attack, indicating that the selected attack parameters satisfy essential stealthiness. On the other hand, the actual operating trajectory of the physical layer completely deviates from the predefined trajectory under the attack, demonstrating that the attacker has achieved the intended target.

In terms of detection performance, Figure 9 and Figure 10 show the detection results of the two types of detectors against the two types of attacks, respectively, which are the main focus of this experiment. Figure 9 shows the output curves of the innovation estimation detector for DoS and essential FDI attacks, with detection thresholds $\alpha_{1th} = 16.5$ and $\alpha_{2th} = 52$. Note that although the three motors in the physical layer have identical parameters, the physical meanings of the outputs for the servo motor and the 2-DoF robot subsystems are different (angle vs. position), leading to distinct $\alpha_{1th}$ and $\alpha_{2th}$. The subscript $k_d$ in the detection variables $\alpha_1(k_d)$ and $\alpha_2(k_d)$ denotes the detection instant, meaning the detection period is set independently of the ICPS operating period; in our experiments, the

detection period is uniformly set to 0.2 s. From Figure 9, it can be seen that the proposed innovation estimation detector is effective against both DoS and essential FDI attacks.

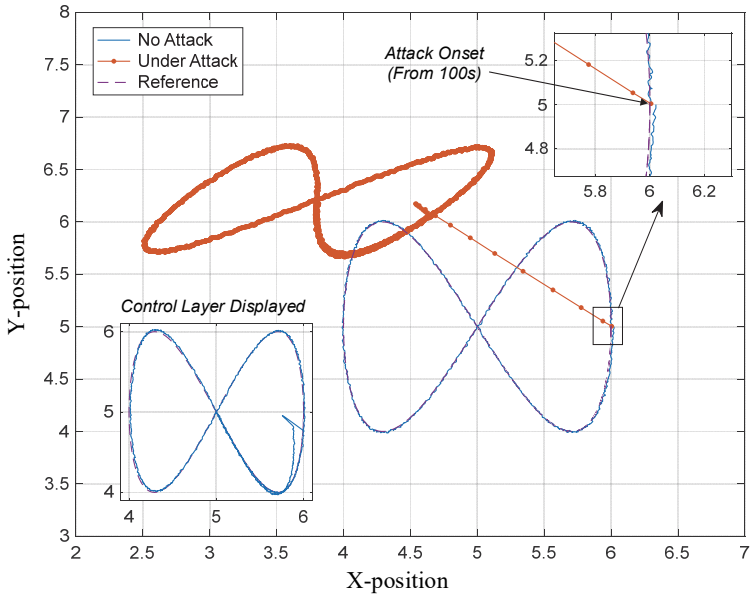


**Figure 8.** Trajectory tracking curves of the 2-DoF robot under FDI attacks.

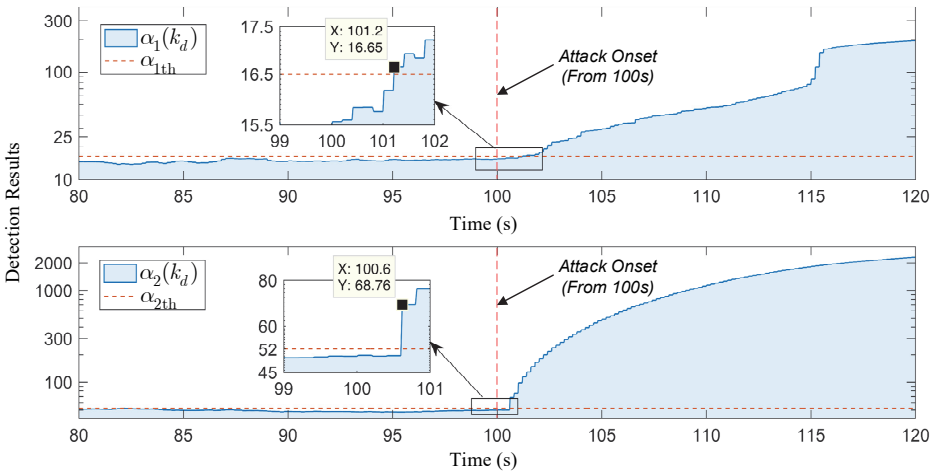


**Figure 9.** Detection results of the innovation detector for DoS and essential FDI attacks.

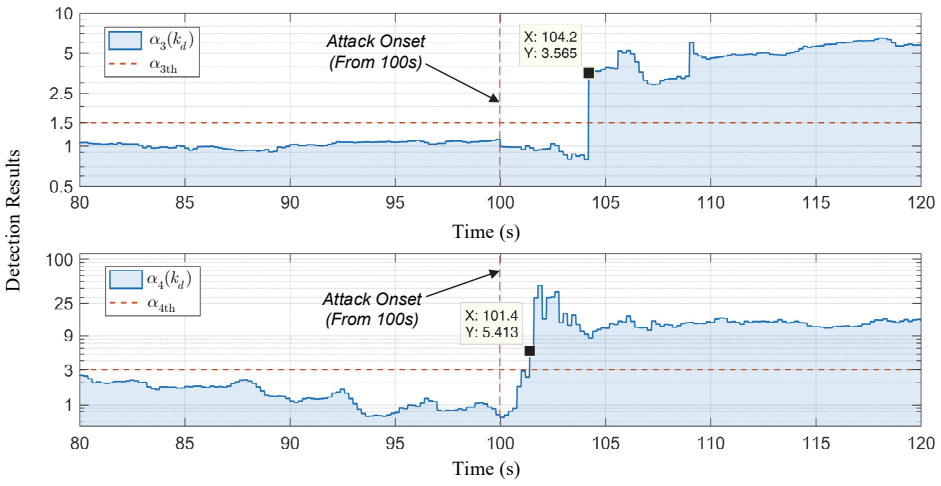


**Figure 10.** Detection results of the extended Markov detector for DoS and essential FDI attacks.

Figure 10 presents the detection results of the extended Markov detector proposed in Section 3 against DoS and essential FDI attacks, where the detection thresholds are $\alpha_{3th} = 1.5$ and $\alpha_{4th} = 3$, respectively. It can be seen that this detector is equally effective against both attack types.

### *4.2. Performance Evaluation Under FDI-I and FDI-II Attacks*

The purpose of this experiment is to verify the detection performance of the innovation estimation detector and the extended Markov detector against two types of stealthy attacks (i.e., FDI-I and FDI-II), where FDI-I is stealthy against the $\chi^2$ detector and FDI-II is stealthy against the KLD detector. HIL experiments are conducted on the established platform with a sampling time of 0.02 s and an experimental duration of 200 s. The attacker initiates FDI-I and FDI-II attacks on the servo motor and 2-DoF robot subsystems, respectively, starting from the 100th second until the end of the experiment. Both proposed detectors are deployed in the control layer to monitor the ICPS security status.

Regarding the attack parameter settings, the FDI-I attack adopts the strategy shown by

$$\begin{cases} \Phi_a(k) = I, \\ \Phi_s(k) = I + S(k)\,\mathrm{diag}\{y_{p1}(k), y_{p2}(k)\}^{-1}, \\ S(k) = -C_p A_o^k x_a(0), \end{cases} \tag{27}$$

where $x_a(0)$ is the initial state of the attack bias signal, set as $x_a(0) = [5, 5]^\top$. In terms of stealthiness, the attack bias signal $S(k)$ causes identical deviations in both the output $y_c(k)$ received by the control layer and the output estimate $C_p\hat{x}(k)$ based on state observation. Therefore, the state estimation residual term remains unchanged before and after the attack, ensuring stealthiness against the $\chi^2$ detector. Note that the discrete parameters for the servo motor model are given by

$$A_p = \begin{bmatrix} 1 & 0.0159 \\ 0 & 0.6169 \end{bmatrix}, \quad B_p = \begin{bmatrix} 0.0063 \\ 0.5863 \end{bmatrix}, \quad C_p = \begin{bmatrix} 1 & 0 \end{bmatrix}. \tag{28}$$

The matrix $A_p$ has eigenvalues 1 and 0.6169, which causes the attack bias signal to eventually converge to a constant value, thereby achieving the intended attack goal.

The parameters for the FDI-II attack are formulated as

$$\begin{cases} \Phi_a(k) = \mathrm{diag}\{a_1(k), a_2(k)\}, \\ \Phi_s(k) = I + S(k)\,\mathrm{diag}\{y_{p1}(k), y_{p2}(k)\}^{-1}, \\ S(k) = z_a(k) + C_p\hat{x}(k), \end{cases} \tag{29}$$

where $a_1(k), a_2(k) \sim \mathcal{N}(1, 0.06)$, and $z_a(k)$ is independent and identically distributed with the secure residual signal $z_0(k)$, generated online by the attacker. It can be seen that the actuator attack parameter $\Phi_a(k)$ is used to compromise the physical system's control performance, while the sensor channel circumvents the KLD detector by generating $z_a(k)$ with the same distribution as $z_0(k)$.

It is worth noting that traditional residual-based detectors including the $\chi^2$ and KLD detectors are purposefully not included as control groups in the following experimental evaluations. Because the formulated FDI-I and FDI-II attack strategies mathematically guarantee the strict preservation of the statistical properties of the innovation sequences, the theoretical failure of these traditional detectors is a mathematical certainty rather than an empirical hypothesis. Consequently, displaying their non-responsive detection curves would yield redundant information. The experimental evaluations herein focus strictly

on demonstrating how the proposed DGM framework successfully breaks this theoretical undetectability.

Figure 11 illustrates the output tracking curves of the servo motor for 20 s before and after the attack. Under the FDI-I attack, the actual output of the physical system deviates from the predefined trajectory, yet no obvious anomaly appears in the physical layer output curve (displayed on the control layer), indicating that both the attack goal and stealthiness meet expectations. The trajectory tracking curve of the 2-DoF robot under FDI-II attack is given in Figure 12. Comparing the physical layer and control layer data clearly shows that this attack strategy also achieves the intended target while ensuring stealthiness.

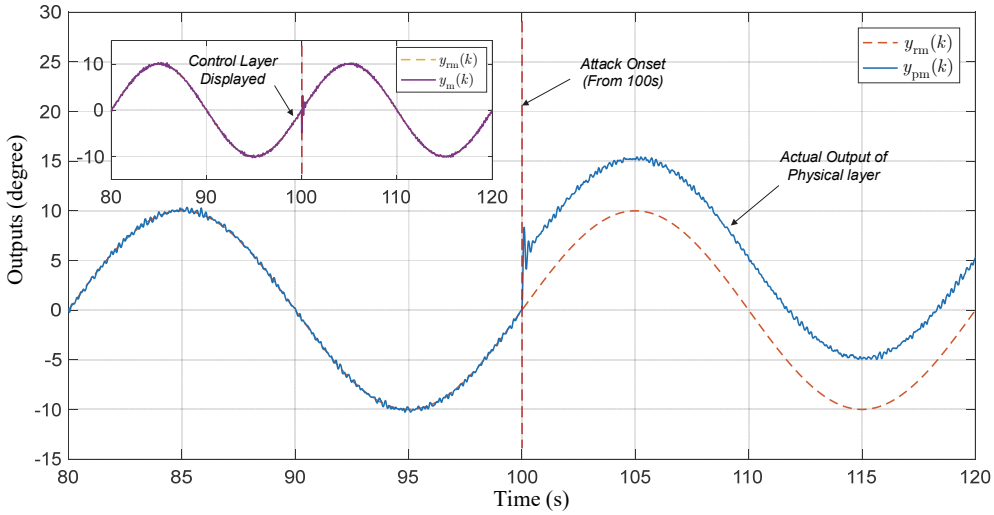


**Figure 11.** Output tracking control curve of the servo motor under FDI-I attacks.

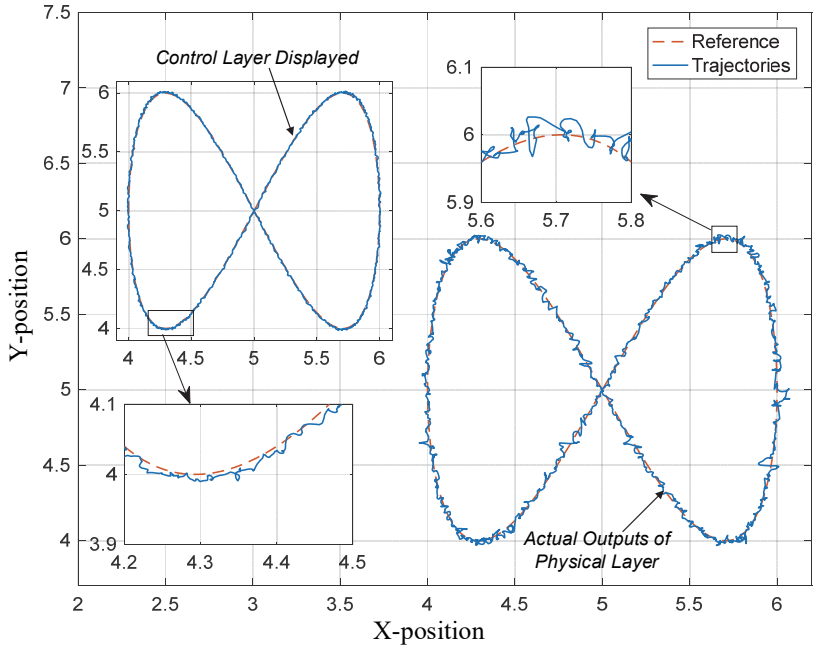


**Figure 12.** Trajectory tracking curves of the 2-DoF robot under FDI-II attacks.

Regarding detection performance, Figure 13 shows the detection results of the innovation estimation detector against the two types of stealthy attacks. With the detection period set to 0.2 s, the detection variable $\alpha_5(k_d)$ exceeds its threshold $\alpha_{5th} = 16.5$ starting from 103.8 s, demonstrating that the detector is effective against the FDI-I attack on the

servo motor system. However, the detection results of the innovation estimation detector for FDI-II are shown by $\alpha_6(k_d)$ and $\alpha_{6th} = 52$ in Figure 13. It can be observed that the detection variable shows no significant change before and after the attack, indicating that this detector fails to detect the FDI-II attack. The reason for this failure can be analyzed using the geometric interpretation in Section 3: on one hand, before and after the FDI-II attack, the subspace of the innovation sequence always maintains orthogonality with $M_0 W_f$, ensuring that the estimation of $Z_b^-$ remains consistent (i.e., the variable $\alpha_6(k_d)$ always converges); on the other hand, the innovation term is numerically small, meaning that the base and hypotenuse of the triangle in the geometric space are nearly equal. Thus, the perturbation caused by FDI-II produces a minimal change in the orthogonal residual, ultimately rendering the detector blind to FDI-II attacks.

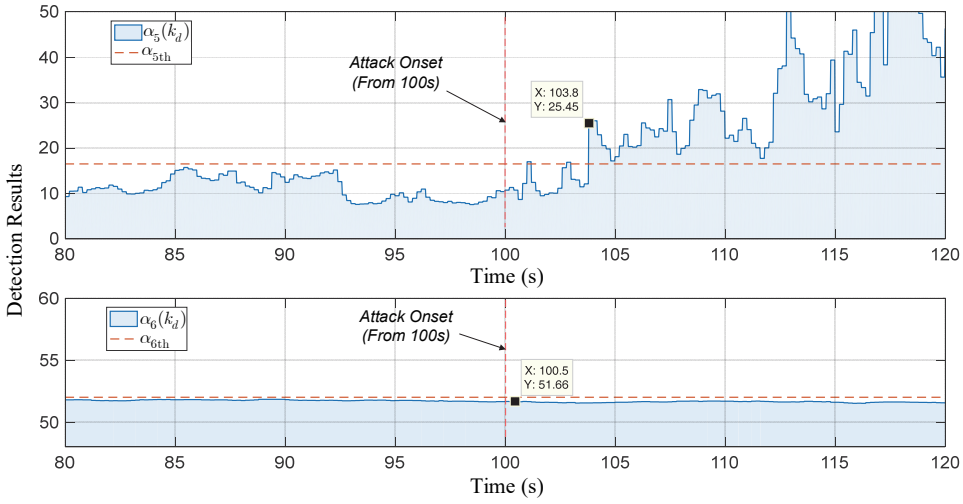


**Figure 13.** Detection results of the innovation detector for FDI-I and FDI-II attacks.

For the two types of stealthy FDI attacks, the detection performance of the extended Markov detector is presented in Figure 14. Evidently, the FDI-I and FDI-II attacks are detected at 101 s and 101.8 s, respectively, and maintain high detection variable outputs. These results indicate that the extended Markov detector is effective against both types of stealthy attacks. Comparing Figures 13 and 14, it is clear that by introducing the dynamic quantification of the DGM, the extended Markov matrix reflects the causality of the process data, thereby maintaining detection capability even against attack strategies constructed using independent and identically distributed data.

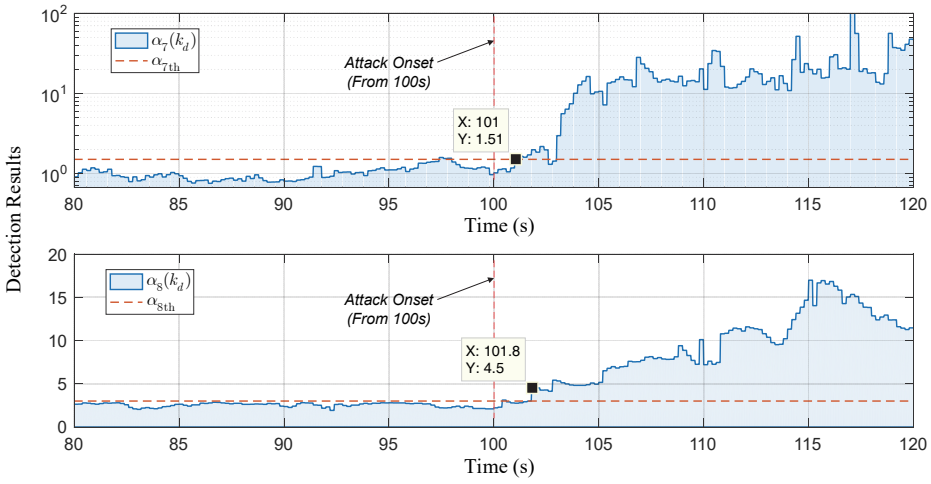


**Figure 14.** Detection results of the extended Markov detector for FDI-I and FDI-II attacks.

## 5. Conclusions and Future Work

In this paper, a novel data-driven security detection framework based on DGM is proposed to safeguard ICPS against stealthy IA attacks. Building upon the multiplicative IA attack model presented in our previous work, this study relaxes the assumptions of known physical plant parameters and noise statistical properties that traditional residual-based detectors often rely on.

Using subspace matrix processing techniques, we construct two types of detection variables strictly from the operational data in the control layer. The static innovation sequence captures steady-state mismatches, showing high sensitivity against standard DoS and essential FDI attacks, as well as FDI-I attacks that can bypass conventional $\chi^2$ detectors. Furthermore, to address the more advanced FDI-II attacks that preserve the orthogonal relationship of the data subspace (and thus deceive KLD detectors), an extended Markov matrix estimator is proposed. Our geometric analysis explains that this dynamic detection variable effectively captures the rotational and stretching distortions of the subspace caused by malicious sequences. HIL experiments based on a 2-DoF robot and a DC servo motor verify the effectiveness of the proposed detection framework.

Future research will focus on extending the current linear DGM framework to nonlinear and time-varying industrial processes. Additionally, while this study demonstrates the binary capability of breaking the undetectability of stealthy attacks via a custom HIL platform, evaluating precise quantitative statistical metrics including the false alarm rate, missed detection rate, and average detection delay remains challenging due to the current lack of universally standardized benchmark datasets in the ICPS domain. Therefore, establishing a comprehensive benchmarking methodology and investigating adaptive thresholds under severe exogenous disturbances, alongside exploring the distributed deployment of DGM detectors in large-scale networked CPS, represent promising directions to further improve industrial control security.

**Author Contributions:** Conceptualization, J.C. and B.L.; methodology, J.C. and B.L.; software, J.C. and Z.J.; validation, J.C. and Z.J.; investigation, J.C. and B.L.; resources, B.L.; data curation, J.C. and Z.J.; writing—original draft preparation, J.C.; visualization, J.C. and Z.J.; supervision, B.L.; project administration, B.L.; funding acquisition, B.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Guangdong Basic and Applied Basic Research Foundation under grant 2023A1515110877, the project ZR2026QC0818 supported by Shandong Provincial Natural Science Foundation, and the Guangdong Basic and Applied Basic Research Foundation under grant 2023A1515110414.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

**Acknowledgments:** This article is a revised and expanded version of a paper entitled "QDGM: A Framework of Detection and Performance Evaluation for Industrial CPS Security," which was presented at the 43rd Chinese Control Conference, Kunming, China, 28–31 July 2024 [26]. The authors would like to acknowledge the Qiaoshi Electronic Engineering Research Institute and the Key Discipline of Software Engineering at SIAS University for their administrative and technical support. During the preparation of this manuscript, the authors used Gemini 3.1 Pro for the purposes of English language polishing and LaTeX formatting. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| 2-DoF | Two Degree-of-Freedom |
| CPS | Cyber-Physical Systems |
| DC | Direct Current |
| DGM | Differences in Generalized Models |
| DoS | Denial-of-Service |
| FDI | False Data Injection |
| HIL | Hardware-in-the-Loop |
| IA | Integrity and Availability |
| ICPS | Industrial Cyber-Physical Systems |
| KLD | Kullback–Leibler Divergence |
| LTD | Long Time-series Data |
| LTI | Linear Time-Invariant |
| PID | Proportional Integral Derivative |
| UDP | User Datagram Protocol |
| WSCC | Western States Coordinated Council |

## References

1. Rani, S.; Kataria, A.; Kumar, S.; Karar, V. A new generation cyber-physical system: A comprehensive review from security perspective. *Comput. Secur.* **2025**, *148*, 104095. [CrossRef]
2. Singh, N.; Panigrahi, P.K.; Zhang, Z.; Jasimuddin, S.M. Cyber-physical systems: A bibliometric analysis of literature. *J. Intell. Manuf.* **2025**, *36*, 2335–2371. [CrossRef]
3. Sun, Z.; Chen, G.; Ding, Y.; Yang, S. Joint safety and security risk analysis in industrial cyber-physical systems: A survey. *IET Cyber-Phys. Syst. Theory Appl.* **2024**, *9*, 334–349. [CrossRef]
4. Yuan, S.; Yang, M.; Reniers, G. Integrated process safety and process security risk assessment of industrial cyber-physical systems in chemical plants. *Comput. Ind.* **2024**, *155*, 104056. [CrossRef]
5. Kure, H.I.; Islam, S.; Ghazanfar, M.; Raza, A.; Pasha, M. Asset criticality and risk prediction for an effective cybersecurity risk management of cyber-physical system. *Neural Comput. Appl.* **2022**, *34*, 493–514. [CrossRef]
6. Kumar, R.; Kela, R.; Singh, S.; Trujillo-Rasua, R. APT attacks on industrial control systems: A tale of three incidents. *Int. J. Crit. Infrastruct. Prot.* **2022**, *37*, 100521. [CrossRef]
7. Aanonsen, C.E. Stuxnet, revisited (again): Producing the strategic relevance of cyber operations. *J. Cyber Policy* **2025**, *10*, 68–84. [CrossRef]
8. Cervini, J.; Rubin, A.; Watkins, L. Don't drink the cyber: Extrapolating the possibilities of Oldsmar's water treatment cyberattack. In Proceedings of the International Conference on Cyber Warfare and Security, Albany, NY, USA, 17–18 March 2022; pp. 19–25.
9. Gan, J.; Luo, C.; Shi, W.; Liu, Y.; Liu, X.; Tian, Z. An attack exploiting cyber-arm industry. *IEEE Trans. Dependable Secur. Comput.* **2025**, *22*, 1686–1702. [CrossRef]
10. Duo, W.; Zhou, M.; Abusorrah, A. A survey of cyber attacks on cyber physical systems: Recent advances and challenges. *IEEE/CAA J. Autom. Sin.* **2022**, *9*, 784–800. [CrossRef]
11. Yu, Z.; Gao, H.; Cong, X.; Wu, N.; Song, H.H. A survey on cyber-physical systems security. *IEEE Internet Things J.* **2023**, *10*, 21670–21686. [CrossRef]
12. Teixeira, A.; Shames, I.; Sandberg, H.; Johansson, K.H. A secure control framework for resource-limited adversaries. *Automatica* **2015**, *51*, 135–148. [CrossRef]
13. Lian, Z.; Shi, P.; Chen, M. A Survey on Cyber-Attacks for Cyber-Physical Systems: Modeling, Defense and Design. *IEEE Internet Things J.* **2025**, *12*, 1471–1483. [CrossRef]
14. Teixeira, A.M.H.; Tosun, F.E.; Dong, J.; Ahlén, A.; Dey, S. Kullback-Liebler Divergence-Based Observer Design Against Sensor Bias Injection Attacks in Single-Output Systems. *IEEE Trans. Inf. Forensics Secur.* **2025**, *20*, 2763–2777. [CrossRef]
15. Cai, F.; Koutsoukos, X. Real-time detection of deception attacks in cyber-physical systems. *Int. J. Inf. Secur.* **2023**, *22*, 1099–1114. [CrossRef]
16. Ye, D.; Zhang, T.Y. Summation detector for false data-injection attack in cyber-physical systems. *IEEE Trans. Cybern.* **2019**, *50*, 2338–2345. [CrossRef] [PubMed]
17. Dong, Y.; Gupta, N.; Chopra, N. False data injection attacks in bilateral teleoperation systems. *IEEE Trans. Control Syst. Technol.* **2020**, *28*, 1168–1176. [CrossRef]

18. Liu, B.; Chen, J.; Hu, Y. Mode division-based anomaly detection against integrity and availability attacks in industrial cyber-physical systems. *Comput. Ind.* **2022**, *137*, 103609. [CrossRef]
19. Li, T.F.; Ding, L.; Xiong, J.; Li, H. Quantized dissipative control for fuzzy hot strip mill cooling system under input constraint via dynamic output feedback approach. *Appl. Math. Comput.* **2026**, *508*, 129615. [CrossRef]
20. Zhu, K.; Wang, Z.; Zheng, X.; Cui, Z.; Li, Z.; Li, K. Fusion filtering for RIS-assisted vehicle localization with unknown inputs: A privacy-preserving output-mask strategy. *IEEE Trans. Ind. Inform.* **2026**, *22*, 5909–5919. [CrossRef]
21. Zhu, M.; Martinez, S. On the performance analysis of resilient networked control systems under replay attacks. *IEEE Trans. Autom. Control* **2021**, *59*, 804–808.
22. Wang, Y.A.; Wang, Z.; Zou, L.; Shen, B.; Dong, H. Detection of perfect stealthy attacks on cyber-physical systems subject to measurement quantizations: A watermark-based strategy. *IEEE/CAA J. Autom. Sin.* **2025**, *12*, 114–125. [CrossRef]
23. Liu, H.; Mo, Y.; Yan, J.; Xie, L.; Johansson, K.H. An online approach to physical watermark design. *IEEE Trans. Autom. Control* **2020**, *65*, 3895–3902. [CrossRef]
24. Chen, J.; Liu, B.; Li, T.; Hu, Y. Multiplicative Attacks with Essential Stealthiness in Sensor and Actuator Loops against Cyber-Physical Systems. *Sensors* **2023**, *23*, 1957. [CrossRef] [PubMed]
25. Liu, B.; Chen, J.; Hu, Y. A Simple Approach to Data-driven Security Detection for Industrial Cyber-Physical Systems. In Proceedings of the 34th Chinese Control and Decision Conference (CCDC), Hefei, China, 15–17 August 2022; pp. 5440–5445.
26. Chen, J.; Fan, Y.; Li, T. QDGM: A Framework of Detection and Performance Evaluation for Industrial CPS Security. In Proceedings of the 43rd Chinese Control Conference, Kunming, China, 28–31 July 2024; pp. 5178–5183.