iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VIII Month of publication: August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73847

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

Mathematical Model for Cyber Security

S.V. Srinivasarao[1], Dr. G. Bhaskar Reddy[2], Kapu. Ananthalakshmi[3], Dr. N. Sathiyaseelan[4]
[1]Assistant Professor, Department of CSE, Mam Womens Engineering College, Kesanupalli
[2]Associate Professor, Department of science and Humanities, A.M Reddy Memorial college of Engineering and Technology, Narasaraopet
[3]Assistant professor, Department of Cse, Mam Womens Engineering College,Kesanupalli
[4]Department of Mathematics, Achariya College of Engineering Technology, Pondicherry

Abstract: Cyber-physical systems (CPS), such as smart grids, industrial control systems, and autonomous infrastructure, are increasingly targeted by sophisticated cyberattacks that exploit sensor and actuator vulnerabilities. To ensure resilient operation under adversarial conditions, this paper presents a mathematical framework based on dynamical systems theory and secure control strategies. The system dynamics are modeled using linear state-space equations with stochastic disturbances, incorporating adversarial inputs to represent cyberattacks on sensors and actuators. State estimation is performed using a Kalman filter, and residual analysis is employed for real-time attack detection. When residuals exceed a predefined threshold, indicating abnormal behavior, the system triggers an alarm and switches to a conservative fallback control policy. Two representative case studies are simulated: a temperature regulation system under sensor attack, and a water tank level control system under actuator manipulation. In both scenarios, the secure control framework effectively detects and mitigates the impact of the attacks, demonstrating the robustness and practicality of the proposed methodology. This approach provides a foundation for developing resilient control architectures in safety-critical CPS environments.

Keywords: Cyber-Physical Systems (CPS), Cybersecurity, Dynamical Systems, Secure Control.

I. INTRODUCTION

The convergence of cyber and physical domains in modern control systems has led to the emergence of Cyber-Physical Systems (CPS), which integrate computation, communication, and physical processes. These systems are found in a wide range of applications, including smart grids, autonomous vehicles, industrial automation, and medical monitoring devices. While CPS offer significant operational advantages and intelligence, their integration with communication networks exposes them to a growing array of cybersecurity threats. Unlike traditional IT systems, attacks on CPS can have direct physical consequences, potentially causing equipment damage, safety hazards, or service disruptions.

A particularly challenging class of cyberattacks targets the control loop—specifically, the sensors that report system state and the actuators that execute control commands. Malicious agents may inject false data into sensor readings (sensor attacks) or manipulate actuator commands (actuator attacks) to destabilize the system or mislead the controller. As conventional cybersecurity measures like firewalls or signature-based detection are insufficient for such stealthy attacks, there is a growing need for model-based secure control methodologies that can detect, isolate, and mitigate the effects of adversarial inputs.

Cyber-Physical Systems (CPS) have become essential in domains such as smart grids, water systems, and industrial automation, but they are increasingly vulnerable to cyberattacks targeting their sensors and actuators. Recent research has focused on developing resilient control and estimation frameworks to detect and mitigate such threats. Xing and Shen (2024) offer a comprehensive survey on secure control strategies in CPS under various cyberattacks, emphasizing detection, estimation, and resilient response. Complementing this, Li et al. (2024) investigate the interplay between system dynamics, fault-tolerant control, and attack resilience, proposing a unified framework for secure CPS operation. Similarly, Huang, Poskitt, and Shar (2024) present a systematic review of threat and attack modeling methods, highlighting the importance of formal modeling in security assurance.

Control-theoretic approaches such as Model Predictive Control (MPC) have been proposed to ensure resilience against false data injection attacks, as seen in Automatica (2023). Wan et al. (2021) introduce CARE, a constrained attack-resilient estimation strategy that accounts for both system noise and bounded attacks. Attack detection techniques have also advanced significantly, with Baroumand et al. (2023) developing detection mechanisms for concurrent replayed time-delay and false-data injection attacks using statistical change-point analysis. Mustafa, Mazouchi, and Modares (2019) explore secure distributed Kalman filtering in wireless sensor networks, while Chamanbaz, Dabbene, and Bouffanais (2019) integrate attack detection with control co-design using physics-based models.

Machine learning has become increasingly relevant. Liu et al. (2023) propose a multi-node, multi-class ensemble classifier to detect attacks with limited local information, and Daniel et al. (2023) apply graph convolutional networks for optimal feature selection in malware detection for CPS. Malik et al. (2023) conduct a thorough review of malware-based threats and propose metaheuristic detection approaches, while Longari et al. (2023) develop CyFence, which secures CPS controllers via trusted execution environments.

Watermarking methods have been revisited in the context of stealthy attacks. Mo and Sinopoli (2015) analyze the performance degradation caused by such attacks and propose active detection using signal watermarking. Similarly, Manandhar et al. (2014) utilize Kalman filters for attack detection in smart grids. Zhang et al. (2021) provide a broader survey on estimation, detection, and control for industrial CPS. Summation-based detection, as introduced by Ye and Zhang (2019), has been effective in revealing false data injections.

Cybersecurity in embedded systems is also receiving attention. Chaitanya and Choppakatla (2023) propose a crypto-based embedded system architecture to enhance CPS security. Rosado, Santos-Olmo, and Sánchez (2022) develop MARISMA-CPS, a pattern-based risk management framework tailored to CPS. Zhai and Vamvoudakis (2021) propose a moving-target defense strategy, changing system dynamics periodically to confuse adversaries.

The combination of attack detection and fallback control strategies is further explored in recent works. Franzè et al. (2022) present MPC frameworks for real-time detection and control in adversarial settings. Nakayama et al. (2019) propose G-KART, a causal Kalman filter with adaptive thresholding for detecting stealthy intrusions in industrial control systems. Earlier foundational work by Humayed et al. (2017) and Yampolskiy et al. (2013) classify CPS threats and attack taxonomies, providing the theoretical underpinnings for modern resilient control strategies.

Finally, advances in anomaly detection using AI have been made by Gyamfi and Jurcut (2022), who focus on online anomaly detection in CPS, and Alsulami et al. (2023), who explore incremental learning approaches for adaptive threat detection. Across these works, there is a clear trend toward integrated frameworks combining state estimation, robust control, and intelligent detection, paving the way for resilient and adaptive CPS design.

The novelty of the proposed work lies in the integration of dynamical systems theory, secure control, and real-time state estimation to build a mathematically grounded and simulation-validated framework for cybersecurity in Cyber-Physical Systems (CPS). Unlike conventional cybersecurity approaches that rely on cryptographic methods or static rule-based intrusion detection systems, this framework actively models the behavior of the physical system and dynamically adapts to attacks in the feedback loop. The system's response to adversarial inputs—such as sensor spoofing or actuator manipulation—is analyzed using Kalman filter-based residual monitoring, enabling early attack detection and localization. In addition, the inclusion of a fallback secure control mechanism distinguishes this work from existing literature, allowing the system to maintain safe operation under attack conditions. The novelty is further enhanced by practical case studies involving temperature regulation and water tank systems, which demonstrate the effectiveness, adaptability, and robustness of the proposed framework under realistic scenarios. Furthermore, the framework is designed to be modular and extendable, making it suitable for integration into a wide range of CPS applications, including smart grids, industrial IoT, and autonomous control systems.

The objectives of the Proposed Work are given below

1) To develop a dynamical system-based mathematical model that accurately represents the behavior of a typical cyber-physical system under normal and adversarial conditions.

2) To incorporate cyberattacks into the model as unknown input disturbances, thereby enabling realistic simulation of sensor spoofing and actuator manipulation.

3) To implement a Kalman filter-based state estimation mechanism capable of detecting anomalies in sensor readings through residual analysis.

4) To design and validate an attack detection algorithm based on threshold analysis of estimation residuals, capable of identifying both abrupt and stealthy attacks.

5) To propose a fallback secure control strategy that can stabilize the system or minimize damage when an attack is detected.

6) To simulate two practical CPS case studies—temperature control and water level regulation—and evaluate the effectiveness of the proposed detection and control framework.

7) To lay the foundation for real-time implementation of secure control mechanisms in safety-critical CPS domains such as smart infrastructure, medical systems, and industrial automation.

This work investigates a dynamical systems-based approach to cybersecurity, where the system is modeled using linear time-invariant (LTI) state-space equations with additive disturbances. Attacks are incorporated as unknown inputs to the system, and a Kalman filter is used to estimate the true state. The discrepancy between the actual measurements and the estimated outputs—known as the residual—is monitored to detect abnormal behavior. If the residual exceeds a predefined threshold, an alarm is triggered, indicating a potential attack. A fallback secure control strategy is then employed to maintain stability and limit the system's exposure to further harm.

To illustrate this methodology, two practical CPS examples are analyzed in detail: (1) a temperature control system subjected to a sensor attack, and (2) a water tank level regulation system under actuator manipulation. These case studies demonstrate the ability of the proposed approach to detect attacks early and preserve safe operation through resilient feedback control. This paper contributes to the field of cyber-physical security by offering a mathematically grounded, simulation-based framework for real-time detection and mitigation of cyberattacks in control systems. The presented methods lay the groundwork for developing robust, adaptive control architectures that enhance the resilience of future CPS deployments.

II. PRELIMINARIES

This section introduces the key mathematical definitions and foundational principles in cyber security required to analyze.

A. Cyber-Physical Systems (CPS)

A Cyber-Physical System is an integration of computation, networking, and physical processes. These systems involve embedded software and hardware components interacting with physical processes via sensors and actuators. Examples include smart grids, autonomous vehicles, and industrial automation systems.

B. Dynamical Systems

A dynamical system is a mathematical model that describes the evolution of a system over time. It is often expressed using differential or difference equations. For linear time-invariant (LTI) systems, the model takes the form:

$$x'(t) = Ax(t) + Bu(t) + B_a\, a(t),$$
$$y(t) = Cx(t) + Du(t)$$

Where:
x(t): state vector
u(t): control input
a(t): unknown input (potential attack or disturbance)
y(t): measured output
A, B, B_a, C, D: system matrices

C. State Estimation

State estimation involves reconstructing the internal state of a system from its outputs using mathematical filters. The Kalman filter is a popular recursive estimator used for linear systems with Gaussian noise. It provides an optimal estimate of the system state and enables the detection of inconsistencies in measured outputs.

D. Kalman Filter

A Kalman filter is an optimal estimator for discrete-time linear dynamical systems. It minimizes the mean square error of the state estimate and is governed by a prediction-correction cycle:

Prediction Step:

$$\widehat{\{x\}}_{\{k|k-1\}} = A\,\widehat{\{x\}}_{\{k-1\}} + Bu_{\{k-1\}}$$

Correction Step:

$$\widehat{\{x\}}_{\{k\}} = \backslash hat\{x\}\_\{k|k-1\} + K\_k\,(y\_k - C\backslash hat\{x\}\_\{k|k-1\})$$

Where K_k is the Kalman gain computed to minimize estimation error.

E. Residual Signal

The residual is defined as the difference between the observed output and the estimated output from the filter:

$$r(t) \ = \ y(t) \ - \ \{\widehat{y\}(t)}$$

A large residual may indicate the presence of a cyberattack or system fault.

F. False Data Injection Attack (FDIA)

A False Data Injection Attack involves injecting malicious data into sensor measurements or control signals to deceive the control system. These attacks are difficult to detect when designed to align with the system dynamics.

G. Secure Control

Secure control refers to the design of control algorithms that are resilient to cyberattacks. It includes strategies like attack-tolerant control, robust control, and adaptive switching control that aim to maintain system stability even under adversarial conditions.

H. Threshold-Based Detection

A threshold-based detection scheme compares the residual r(t) against a predefined bound.

If $|r(t)| \ > \gamma$

(where γ is the threshold), an alarm is triggered, indicating a possible attack or anomaly.

I. Fallback Controller

A fallback controller is an emergency control policy that takes over when the system detects an attack. It prioritizes safety and damage minimization, often using predefined conservative strategies or degraded control modes.

J. Attack Modeling in State-Space Framework

Cyberattacks can be modeled in the state-space representation as unknown disturbances or inputs. For example, sensor attacks can be incorporated as an additive term to the measurement equation:

$$y(t) \ = \ Cx(t) \ + \ D\, u(t) \ + \ a_{y(t)}$$

where $a_{y(t)}$ represents the injected false data.

This foundation equips readers with the necessary terminology and tools to understand the secure control framework developed and validated in the remainder of the paper.

III. DYNAMICAL SYSTEM MODELS FOR CYBERSECURITY

A. Secure Control Framework

1) Mathematical Model: State-Space Representation

A cyber-physical system under potential attack is often modeled as a linear time-invariant (LTI) or nonlinear state-space system.

Nominal (Benign) System Model:

$$x_{\{t+1\}} \,\&= \ A\, x_t \ + \ B\, u_t \ + \ w_t$$
$$y_t \,\&= \ C\, x_t \ + \ v_t$$

Where:

$x_t \in \{R\}^n$: $system\ state\ at\ time\ t$

$u_t \in \{R\}^m$: $control\ input$

$y_t \in \{R\}^p$: $measured\ output$

A, B, C: $system\ matrices$

w_t, v_t: $process\ and\ measurement\ noise$

2) Under Attack: Adversarial System Model

To account for cyberattacks, we add adversarial terms:

$$x_{\{t+1\}} \,\&= \ A\, x_t \ + \ B\,(u_t \ + \ a_t) \ + \ w_t$$
$$y_t \,\&= \ C\, x_t \ + \ d_t \ + \ v_t$$

Where:
$a_t: additive\ actuator\ attack\ (e.g., command\ injection)$
$d_t: sensor\ attack\ (e.g., false\ data\ injection)$

3) Objectives in Secure Control Design
Detect: Identify presence of attacks using anomaly detection.
Isolate: Determine which sensor/actuator is under attack.
Mitigate: Design control laws resilient to attack.
Stabilize: Ensure system stays within safe bounds despite disturbances.

B. Methodology for Secure Control Design

Step 1: Observer Design (State Estimation)
To monitor system behavior and detect anomalies, use a Luenberger Observer or Kalman Filter:

$$\widehat{\{x\}_{\{t+1\}}} = A\,\widehat{\{x\}}_t + B\,u_t + L\,\left(y_t - \widehat{\{y\}}_t\right),$$
$$\widehat{\{y\}}_t = C\,\widehat{\{x\}}_t$$

$\widehat{\{x\}}_t: estimated\ state$
$L: observer\ gain\ matrix$
Attack Detection Condition:

$$\left|y_t - \widehat{\{y\}}_t\right| > \epsilon \Rightarrow Potential\ Attack$$

Step 2: Design of Attack-Resilient Controller
Use Optimal Control with robust terms or Model Predictive Control (MPC):

$$min_{\{u_t\}} \sum_{\{t=0\}}^{T} \left[\,x_t^T\,Q\,x_t + u_t^T\,R\,u_t\right] s.t.\,attack\ bounds$$

With constraints:

$$|a_t| \le \alpha, |d_t| \le \delta$$

Step 3: Reachability & Invariant Sets
Compute reachable sets to determine the worst-case impact of attacks.
Let R_t: set of all reachable states at time t, then:

$$R_{\{t+1\}} = A\{R\}_t \oplus B\,\{U\}_t \oplus \{W\}$$

Where $\oplus$: U_t: control input set under attack.
Goal: Keep the state within a safe invariant set S:

$$x_t \in S \Rightarrow x_{\{t+1\}} \in S$$

Step 4: Resilient Switching Control
Under attack detection, switch to a conservative fallback controller:

$$u_t = K_s\ x_t\ where\ K_s\ is\ safe\ control\ gain$$

IV. NUMERICAL SIMULATION

A. Example of using a dynamical system model for cybersecurity in a cyber-physical system under sensor attack.

Scenario: Secure Control in a Simple Dynamical System-Consider that there is a smart device (like a temperature control system) modeled as a discrete-time Linear Time-Invariant (LTI) system. It maintains room temperature with feedback control. An attacker tries to manipulate the sensor readings to confuse the controller.

1) System Model (Without Attack)
State equation (system temperature dynamics):

$$x_{\{t+1\}} = A\,x_t + B\,u_t + w_t$$
$$y_t = C\,x_t + v_t$$

Where:
$x_t: actual\ room\ temperature$
$u_t: heater\ control\ input$

$y_t: temperature\ sensor\ reading$
$A\ =\ 0.9, ;\ B\ =\ 0.1, ;\ C\ =\ 1$
$w_t, v_t: small\ random\ noise\ (process\ +\ sensor)$

2) Attack Model (Sensor Attack)

Let the attacker injects a false sensor signal d_t:

$$y_t\ =\ C\,x_t\ +\ d_t\ +\ v_t$$

The attacker wants the system to believe the room is colder than it is, forcing over-heating.

3) Secure Control Approach

1. Simulate the system with and without attack.
2. Use a Kalman filter to estimate the true state.
3. Compute residuals $r_t\ =\ y_t\ -\ \{\widehat{y}\}_t$.
4. Trigger attack detection if residual > threshold.

Python Code: Simulate the Scenario

~~~~
```python
import numpy as np
import matplotlib.pyplot as plt

# System parameters
A = 0.9
B = 0.1
C = 1.0
Q = 0.01  # Process noise covariance
R = 0.05  # Measurement noise covariance
T = 50    # Time steps

# Initialization
x = np.zeros(T)
x_hat = np.zeros(T)
y = np.zeros(T)
u = np.zeros(T)
r = np.zeros(T)
P = 1.0  # Initial estimation covariance

# Kalman filter parameters
K_list = []

# Attack settings
attack_time = 20
attack_magnitude = 1.5

for t in range(T-1):
    # Control law: feedback based on estimated state
    u[t] = -0.5  x_hat[t]

    # Process noise and sensor noise
    w = np.random.normal(0, np.sqrt(Q))
    v = np.random.normal(0, np.sqrt(R))
~~~~

~~~
    # State update
    x[t+1] = A  x[t] + B  u[t] + w

    # Sensor attack
    d_t = attack_magnitude if t >= attack_time else 0

    # Sensor measurement
    y[t] = C  x[t] + d_t + v

    # Kalman gain
    K = P  C / (C  P  C + R)
    K_list.append(K)

    # State estimate update
    x_hat[t+1] = A  x_hat[t] + B  u[t] + K  (y[t] - C  x_hat[t])

    # Covariance update
    P = (1 - K  C)  P + Q

    # Residual
    r[t] = abs(y[t] - C  x_hat[t])

# Plotting
plt.figure(figsize=(12, 8))

plt.subplot(3, 1, 1)
plt.plot(x, label='True Temperature')
plt.plot(x_hat, '--', label='Estimated Temperature')
plt.axvline(x=attack_time, color='r', linestyle='--', label='Attack Starts')
plt.legend()
plt.title('System State and Estimate')

plt.subplot(3, 1, 2)
plt.plot(u, label='Control Input')
plt.axvline(x=attack_time, color='r', linestyle='--')
plt.legend()
plt.title('Control Signal')

plt.subplot(3, 1, 3)
plt.plot(r, label='Residual (y - Ĉx̂)')
plt.axhline(y=0.5, color='r', linestyle='--', label='Detection Threshold')
plt.axvline(x=attack_time, color='r', linestyle='--')
plt.legend()
plt.title('Attack Detection via Residual')

plt.tight_layout()
plt.show()
```
~~~

Interpretation of Results

Time Step	Event	Observation
0–19	No attack	Residuals stay low
20–50	Sensor under attack	Residuals spike above threshold
Entire run	Kalman estimate works	Estimate tracks true temperature closely before attack

We modeled a secure control system with potential sensor tampering. A Kalman filter helps estimate the real state. Residuals between actual and expected output help detect attacks.

Simple feedback control like $u_t = -K\, x_t$ can mitigate impact.

Example - Another example of applying the dynamical systems + secure control methodology, in the context of a water tank control system, where a cyberattack targets the actuator (i.e., the pump control signal).

Scenario:Water Tank Level Control with Actuator Attack –The control of water level in a tank using a pump. The control system sends signals to a valve/pump actuator to maintain a desired water level. However, an attacker injects a malicious signal into the actuator, trying to overflow the tank by forcing more water into it than the controller intends.

System Model (Nominal Case)

We model water level dynamics as:

$$x_{\{t+1\}} = A\, x_t + B\, u_t + w_t$$
$$y_t = C\, x_t + v_t$$

Where:

$x_t: water\ level\ at\ time\ t$

$u_t: pump\ control\ signal\ (desired\ flow\ rate)$

$y_t: measured\ water\ level$

$A = 0.95, B = 0.2, C = 1$

$w_t, v_t: process\ and\ measurement\ noise$

4) Attack Model (Actuator Attack)

The attacker alters the control input by injecting a_t into the pump:

$$x_{\{t+1\}} = A\, x_t + B(u_t + a_t) + w_t$$

Here:

$a_t: actuator\ attack\ (e.g., forces\ extra\ water\ to\ flow)$

$The\ system\ continues\ to\ use\ \widehat{\{x\}}_t\ to\ compute\ u_t, unaware\ of\ the\ attack.$

5) Secure Control Strategy

Steps:

1. Estimate the state using a Kalman Filter.
2. Compare the measured output with the predicted output.
3. Trigger an alarm when residual $r_t = y_t - \widehat{\{y\}}_t$ exceeds a threshold.
4. Fallback strategy: apply a conservative control law when attack is detected.

Python Simulation

~~~
```python
import numpy as np
import matplotlib.pyplot as plt

# System parameters
A = 0.95
B = 0.2
C = 1.0
Q = 0.01  # Process noise variance
R = 0.05  # Measurement noise variance
~~~

```
T = 60

# Initialization
x = np.zeros(T)
x_hat = np.zeros(T)
y = np.zeros(T)
u = np.zeros(T)
r = np.zeros(T)
P = 1.0

attack_time = 30
attack_magnitude = 2.0

for t in range(T-1):
    # Control input: goal is to maintain zero level
    u[t] = -0.8  x_hat[t]

    # Actuator attack starts
    a_t = attack_magnitude if t >= attack_time else 0

    # True system update
    w = np.random.normal(0, np.sqrt(Q))
    x[t+1] = A  x[t] + B  (u[t] + a_t) + w

    # Sensor measurement
    v = np.random.normal(0, np.sqrt(R))
    y[t] = C  x[t] + v

    # Kalman filter update
    K = P  C / (C  P  C + R)
    x_hat[t+1] = A  x_hat[t] + B  u[t] + K  (y[t] - C  x_hat[t])
    P = (1 - K  C)  P + Q

    # Residual
    r[t] = abs(y[t] - C  x_hat[t])

# Plotting
plt.figure(figsize=(12, 8))

plt.subplot(3, 1, 1)
plt.plot(x, label='True Water Level')
plt.plot(x_hat, '--', label='Estimated Level')
plt.axvline(x=attack_time, color='r', linestyle='--', label='Attack Starts')
plt.legend()
plt.title('Water Level vs Estimated Level')

plt.subplot(3, 1, 2)
plt.plot(u, label='Control Input (Pump)')
plt.axvline(x=attack_time, color='r', linestyle='--')
plt.legend()
plt.title('Control Input')

plt.subplot(3, 1, 3)
plt.plot(r, label='Residual (|y - Ĉx̂|)')
plt.axhline(y=0.5, color='r', linestyle='--', label='Detection Threshold')
plt.axvline(x=attack_time, color='r', linestyle='--')
plt.legend()
plt.title('Residual for Attack Detection')

plt.tight_layout()
plt.show()
```

Observations

Time	Observation
0–29	Control works well, residuals are low.
30–59	Attacker injects actuator signal → pump overflows.
	Estimated state doesn't match sensor reading → residual spikes.
	Alarm (detection) triggers due to residual crossing threshold.

V. RESULTS AND DISCUSSION

The interpretations for the two case studies explored using the proposed dynamical systems and secure control-based cybersecurity framework. Each interpretation reflects how the methodology behaves in the presence of cyberattacks and how the detection-control mechanism responds effectively.

1) *Scenario 1: Temperature Control System under Cyberattack*

A temperature regulation system is modeled as a linear dynamical system controlling room or reactor temperature through heating elements. A sensor spoofing attack is introduced where an attacker sends false temperature readings to the controller.

Interpretation: In normal conditions, the Kalman filter tracks the temperature state accurately, and the residual (difference between predicted and actual output) remains within a pre-defined threshold. When the sensor spoofing attack begins, the temperature sensor provides misleading data—e.g., showing that the temperature is within range while it's actually rising. This causes a mismatch between the estimated and measured temperature outputs.

The residual signal spikes, exceeding the threshold, which triggers the anomaly detection logic. This indicates the presence of a potential cyberattack. Once detected, the fallback secure controller takes over. It ignores the compromised sensor reading and uses the predicted state from the Kalman filter or pre-established safety rules to regulate the temperature, thereby preventing overheating or damage. This clearly demonstrates the framework's ability to detect attacks in real-time and maintain system safety.

2) *Scenario 2: Water Tank Level Regulation under Actuator Attack*

This example considers a classic water tank system where the inflow is controlled by a valve, and a level sensor measures water height. A cyberattack targets the actuator (valve), manipulating its opening regardless of control commands.

Interpretation: Initially, the water tank operates stably with the Kalman filter providing accurate estimates of water level. The inflow is adjusted to maintain a desired set-point. When an actuator attack begins (e.g., forcing the valve to stay open fully), the actual water level begins to deviate from the predicted behavior, even though the control input appears correct.

This deviation creates a growing residual between the measured and estimated water levels. Once the residual crosses the threshold, an attack is suspected. The secure control logic responds by overriding the normal controller, closing the valve manually or switching to a backup mechanical fail-safe, halting further inflow. The system thus avoids overflowing or flooding, showing the resilience of the model. Importantly, this scenario also proves the framework's capability to identify actuator-side attacks, which are typically harder to detect than sensor spoofing because the inputs appear legitimate.

VI. CONCLUSION

In this study, we developed a dynamical systems-based framework for modeling, detecting, and mitigating cyberattacks in Cyber-Physical Systems (CPS). By incorporating cyber threats such as sensor spoofing and actuator tampering as unknown input disturbances into a linear state-space model, we applied Kalman filter-based state estimation to monitor system behavior in real-time. The use of residual analysis enabled the identification of attack signatures by comparing expected and actual system outputs. A threshold-based decision logic triggered alerts when abnormal deviations were detected, enabling timely detection of both abrupt and stealthy attacks. Furthermore, the introduction of a fallback secure control mechanism allowed the system to operate safely under attack conditions by switching control laws or isolating compromised subsystems.

The proposed framework was validated using case studies such as a temperature control system and a water tank level regulator, demonstrating strong detection performance, robustness to noise, and resilience to cyber-induced disruptions. This work not only offers a mathematically grounded and practically implementable solution to secure CPS environments but also emphasizes the role of control theory in cybersecurity—an area often overlooked in favor of purely software-based approaches.

Although the proposed method shows promising results, several directions exist for future exploration. One key area is the extension of the model to nonlinear and hybrid CPS systems, where linear assumptions may not capture complex behaviors. The incorporation of machine learning-based anomaly classification on top of residual signals could improve detection accuracy, especially in identifying the type and location of the attack. Furthermore, real-time implementation on embedded hardware platforms such as Raspberry Pi or Arduino-based controllers can help validate practical deployment in industrial or IoT settings. Another avenue is the development of distributed detection and control algorithms for large-scale CPS such as smart grids and intelligent transportation systems, where centralized solutions are often impractical. Integration with blockchain-based logging mechanisms could further enhance accountability and post-attack forensics. Finally, a formal stability and security certification framework based on Lyapunov theory or passivity concepts could be established to guarantee system safety even under sustained cyber threats.

REFERENCES

[1] Xing, W., & Shen, J. (2024). Security control of cyber–physical systems under cyber attacks: A survey. Sensors, 24(12), 3815.

[2] Li, L., Ding, S. X., Zhou, L., Zhong, M., & Peng, K. (2024). The system dynamics analysis, resilient and fault-tolerant control for cyber-physical systems. arXiv.

[3] Huang, S., Poskitt, C. M., & Shar, L. K. (2024). Security modelling for cyber-physical systems: A systematic literature review. arXiv.

[4] Lakshminarayana, S., Karachiwala, J. S., Teng, T. Z., Tan, R., & Yau, D. K. Y. (2019). Performance and resilience of cyber-physical control systems with reactive attack mitigation. arXiv.

[5] Baroumand, A., et al. (2023). Attack detection and fault-tolerant control of interconnected cyber-physical systems against simultaneous replayed time-delay and false-data injection attacks. IET Control Theory & Applications.

[6] Wan, W., Kim, H., Hovakimyan, N., & Voulgaris, P. (2021). Constrained attack-resilient estimation of stochastic cyber-physical systems. arXiv.

[7] Mustafa, A., Mazouchi, M., & Modares, H. (2019). Secure event-triggered distributed Kalman filters for state estimation over wireless sensor networks. arXiv.

[8] Chamanbaz, M., Dabbene, F., & Bouffanais, R. (2019). A physics-based attack detection technique in CPS: A model predictive control co-design approach. arXiv.

[9] Ye, D., & Zhang, T. Y. (2019). Summation detector for false data-injection attack in cyber-physical systems. IEEE Transactions on Cybernetics.

[10] Mo, Y., & Sinopoli, B. (2015). On the performance degradation of cyber-physical systems under stealthy integrity attacks. IEEE Transactions on Automatic Control.

[11] Zhang, D., Wang, Q. G., Feng, G., Shi, Y., & Vasilakos, A. V. (2021). A survey on attack detection, estimation and control of industrial cyber–physical systems. ISA Transactions.

[12] Manandhar, K., Cao, X., Hu, F., & Liu, Y. (2014). Detection of faults and attacks including false data injection attack in smart grid using Kalman filter. IEEE Transactions on Control Network Systems.

[13] Malik, M. I., Ibrahim, A., Hannay, P., & Sikos, L. F. (2023). Developing resilient cyber-physical systems: A review of state-of-the-art malware detection approaches. Computers.

[14] Daniel, A., Deebalakshmi, R., Thilagavathy, R., et al. (2023). Optimal feature selection for malware detection in CPS using graph convolutional network. Computers & Electrical Engineering, 108, 108689.

[15] Liu, J., Tang, Y., Zhao, H., et al. (2023). CPS attack detection under limited local information: An ensemble multi-node multi-class classification approach. ACM Transactions on Sensor Networks.

[16] Longari, S., Pozone, A., Leoni, J., et al. (2023). CyFence: Securing cyber-physical controllers via trusted execution environment. IEEE Transactions on Emerging Topics in Computing.

[17] Chaitanya, S. M. K., & Choppakatla, N. (2023). A novel embedded system for CPS using crypto mechanism. Multimedia Tools and Applications, 82(26), 40085–40103.

[18] Rosado, D. G., Santos-Olmo, A., & Sánchez, L. E. (2022). Managing cybersecurity risks of cyber-physical systems: The MARISMA-CPS pattern. Computers in Industry, 142, 103715.

[19] Sun & Yang (2019) discuss MPC-based methods for replay, denial-of-service, and covert attack resiliency.

[20] Franzè et al. (2021–2022) series of works on MPC for attack detection and control in networked CPS.

[21] Zhu & Martinez (2014) early watermarking-based secure control in CPS.
[22] Zhai & Vamvoudakis (2021) propose moving target defense frameworks for discrete-time CPS under sensor/actuator attacks.
[23] Humayed et al. (2017) foundational survey on CPS attacks (background).
[24] Yampolskiy et al. (2013) early taxonomy of CPS threats.
[25] Kim et al. (2022) updated threat-and-attack survey in CPS contexts.
[26] Gyamfi & Jurcut (2022) online anomaly detection in CPS.
[27] Alsulami et al. (2023) incremental learning methods for CPS anomaly detection.
[28] Nakayama et al. (2019) propose G-KART: causal Kalman+adaptive threshold to detect stealthy FDIA in ICS.