Digital Twin-Driven Intrusion Detection for Industrial SCADA: A Cyber-Physical Case Study

Ali Sayghe *

Posted Date: 2 July 2025

doi: 10.20944/preprints202507.0172.v1

Keywords: digital twin; SCADA security; cyber-physical systems; anomaly detection

 doi:10.20944/preprints202507.0172.v1

Article

Digital Twin-Driven Intrusion Detection for Industrial SCADA: A Cyber-Physical Case Study

Ali Sayghe

Department of Electrical Engineering, Yanbu Industrial College, Yanbu Saudi Arabia; sayghea@rcjy.edu.sa

Abstract

The convergence of operational technology (OT) and information technology (IT) in industrial environments, such as water treatment plants, has significantly increased the attack surface of Supervisory Control and Data Acquisition (SCADA) systems. Traditional intrusion detection systems (IDS) that focus solely on network traffic are often ineffective against stealthy, process-level attacks. This paper proposes a Digital Twin-driven Intrusion Detection (DT-ID) framework that integrates high-fidelity process simulation, real-time sensor modeling, adversarial attack injection, and hybrid anomaly detection using both physical residuals and machine learning. We evaluate the DT-ID framework on a simulated water treatment plant subjected to false data injection (FDI), denial-of-service (DoS), and command injection attacks. The system achieves a detection F1-score of 96.3%, a false positive rate below 2.5%, and an average detection latency under 500 milliseconds, demonstrating substantial improvement over conventional rule-based and physics-only IDS in identifying stealthy anomalies. Our results highlight the practical value of cyber-physical Digital Twins for enhancing SCADA security in critical infrastructure applications.

Keywords: digital twin; SCADA security; cyber-physical systems; anomaly detection

1. Introduction

Industrial Control Systems (ICS) and Supervisory Control and Data Acquisition (SCADA) networks are central to the operation of critical infrastructure in sectors such as water, energy, and oil and gas. While these systems were originally deployed in isolated, air-gapped environments with minimal cybersecurity, most ICS and SCADA installations now face increased cyber risk due to the integration of operational technology (OT) with information technology (IT), enabled by the Industrial Internet of Things (IIoT), remote monitoring, and data-driven optimization [1].

This growing connectivity introduces new vectors for sophisticated cyber-physical threats. High-profile incidents such as Stuxnet, Industroyer, and Triton have shown that attackers can manipulate control logic, falsify sensor readings, or disrupt communication protocols to cause real-world consequences [2–4].

Most current intrusion detection systems (IDS) in industrial networks are adapted from IT security practices, relying on signature-based methods or network anomaly detection [5–7]. However, these solutions often fail to detect process-aware threats where attackers subtly manipulate sensor readings or actuator commands to disrupt operations while keeping network traffic within expected norms, thereby evading traditional alarms [8,9].

Recent research has begun to explore process-aware and hybrid IDS approaches, including those based on Digital Twins (DTs) virtual replicas that simulate physical systems using process models and real-time data [10–15]. Despite this progress, existing DT-based solutions still struggle to reliably detect stealthy attacks that subtly alter process states.

To address these challenges, this paper introduces a novel Digital Twin-driven Intrusion Detection (DT-ID) framework for SCADA systems, which integrates:

 doi:10.20944/preprints202507.0172.v1

- High-fidelity physical process simulation
- Real-time monitoring,
- Adversarial attack emulation and Hybrid anomaly detection using both physics-based and machine learning techniques.

We develop a Digital Twin emulation for a multi-stage water treatment SCADA system, design an adversarial attack simulation engine targeting both cyber and physical layers, and implement a hybrid anomaly detection module that combines LSTM-attention networks with one-class SVMs for robust identification of stealthy threats. The proposed approach is validated on a 72-hour dataset comprising multiple attack scenarios, demonstrating significantly improved detection of advanced threats compared to conventional rule-based and physics-only IDS solutions

The remainder of this paper is structured as follows: Section II reviews related work in SCADA security, Digital Twin applications, and ML-based intrusion detection. Section III presents the architecture of the proposed DT-ID framework. Section IV describes the case study setup and experimental methodology. Section V discusses the results and system performance. Section VI concludes the paper and outlines future work.

2. Related Work

Industrial control systems (ICS) and Supervisory Control and Data Acquisition (SCADA) networks have traditionally relied on intrusion detection systems (IDS) derived from information technology (IT) environments. Signature-based and network anomaly detection tools such as Snort [5] and Suricata [6] have long served as primary defenses in industrial networks. While effective at detecting known attack signatures and obvious network anomalies, these methods generally lack awareness of the physical processes they protect, limiting their effectiveness against stealthy, process-level attacks that manipulate sensor values or control logic without generating conspicuous network events [8,9,16]. Network monitors like Bro (now Zeek) [7] have improved protocol inspection granularity but still primarily operate at the packet or flow level rather than at the process level.

To overcome these limitations, the research community has increasingly turned to process-aware and hybrid intrusion detection frameworks. Early surveys by Mitchell and Chen [16] and Giraldo et al.[8] emphasize the importance of integrating cyber-physical context into IDS, highlighting how attacks exploiting the physical process layer can evade purely network-based monitoring. Residual-based detection approaches, which use model-based or data-driven methods to estimate expected sensor values and flag deviations as potential attacks, have shown promise. However, these methods remain vulnerable to stealthy attacks that mimic the statistical properties of normal operation, as shown in recent adversarial studies [9–15].

The emergence of Digital Twin (DT) technology has provided new opportunities to improve ICS and SCADA security. DTs are high-fidelity, real-time virtual models of physical systems capable of synchronizing with live data streams to simulate, monitor, and optimize process behavior. Oyekan and Hu [11] demonstrated a DT-based cybersecurity monitoring framework for pipeline systems, showing the potential for real-time state comparison and anomaly detection. Zhang et al. [12] expanded on this by integrating DTs with process analytics for smart manufacturing. Recent studies, such as Zhao et al.[13] and Lin et al.[14], have leveraged cloud-based DT platforms and edge analytics for scalable anomaly detection in industrial IoT environments. Mohammadi et al.[15] further employed generative adversarial networks (GANs) to enhance the robustness of DT-based anomaly detection in water treatment plants, achieving higher sensitivity to certain classes of attacks. However, many DT-driven solutions still focus on anomaly detection in simplified settings and may not generalize well to adversarial scenarios involving carefully crafted false data injection (FDI) or zero-day threats.

Hybrid intrusion detection systems that combine machine learning (ML) with process models have also gained momentum. Pan et al.[17] introduced a hybrid deep learning and physics-guided DT architecture for detecting cyber-physical anomalies, demonstrating strong results on benchmark datasets. Xu et al.[18] proposed a fusion approach combining DT representations with deep learning

 doi:10.20944/preprints202507.0172.v1

to detect zero-day attacks in smart manufacturing systems. Nonetheless, adversarial machine learning remains a persistent challenge in ICS security. Creswell et al. [19] and related studies have shown that GAN-based attacks can generate process-consistent yet malicious sensor data, successfully bypassing even advanced hybrid detectors. This raises concerns about the resilience of current ML-based and hybrid IDS in the face of adaptive, stealthy adversaries.

In addition, a review of recent industry incident reports [1] and benchmark datasets such as SWaT [20], WADI [21], and BATADAL [22] indicate that most experimental validations remain confined to testbed or simulated environments. Real-world deployments introduce further complexity, including variable process dynamics, sensor drift, incomplete observability, and evolving adversarial tactics. While some frameworks offer limited retraining or adaptive learning, continuous model updating to address new threats remains an open problem.

In summary, although process-aware, Digital Twin, and hybrid IDS approaches have improved the detection of certain cyber-physical threats, current solutions still struggle to reliably identify stealthy or adversarial attacks in realistic SCADA environments. They often lack adaptive mechanisms for evolving attack strategies and have not yet achieved comprehensive validation against a broad spectrum of real-world adversarial scenarios.

This work addresses these limitations by proposing a Digital Twin-driven Intrusion Detection (DT-ID) framework that combines high-fidelity process simulation, adversarial attack injection, and hybrid anomaly detection modules, enabling robust and adaptive identification of advanced threats in modern industrial SCADA systems.

3. Core Components

The architecture of the proposed DT-ID system is shown in Figure 1. The system is composed of four tightly integrated modules: a high-fidelity virtual SCADA model (Digital Twin), an adversarial attack simulation engine, a hybrid anomaly detector, and a multi-stage response module. Together, these modules enable real-time monitoring, robust adversarial testing, adaptive anomaly detection, and automated mitigation in industrial SCADA environments.

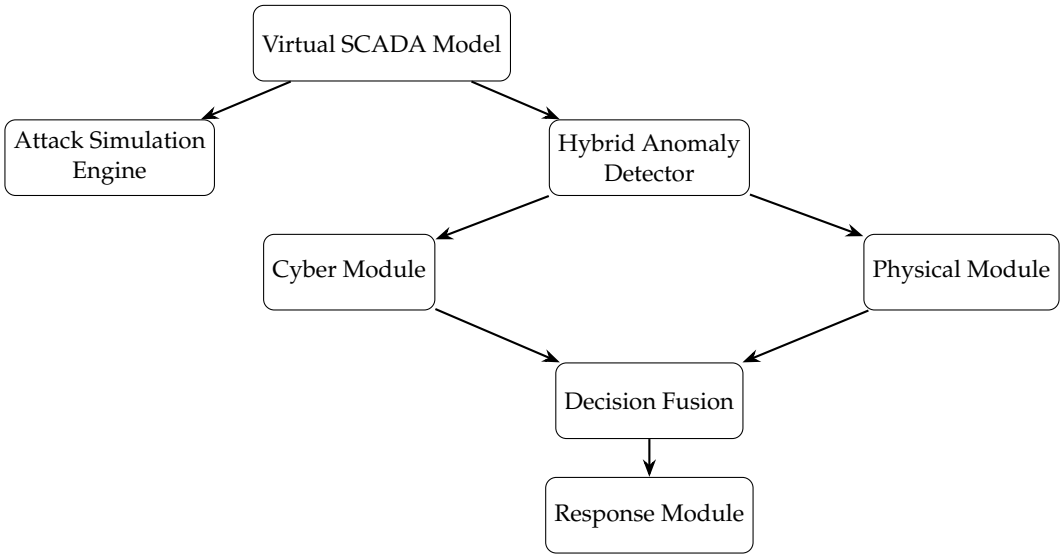

Figure 1. DT-ID System Architecture: Integration of Virtual SCADA, attack simulation, hybrid detection modules, and response coordination.

3.1. Virtual SCADA Model

The Virtual SCADA Model serves as a high-fidelity digital twin, emulating the core dynamics of a water treatment process, including tank hydraulics and chemical dosing. The digital twin and process simulation were implemented in MATLAB Simulink, with all data handling and synchronization routines executed in Python 3.10. This dual-environment setup enables flexible model development,

 doi:10.20944/preprints202507.0172.v1

rapid prototyping, and integration with external attack libraries. The module leverages established physical models to provide process aware context for anomaly detection and attack simulation.

The hydraulic behavior of each storage tank is modeled by Bernoulli's principle:

$$\frac{dH}{dt} = \frac{Q_{\text{in}} - \beta\sqrt{H}}{A} \tag{1}$$

where H denotes the tank level, Q_{in} the inflow rate, β the valve coefficient, and A the cross-sectional area.

pH control dynamics are captured as:

$$\frac{dpH}{dt} = k(C_{\text{acid}} - C_{\text{base}}) - \gamma pH \tag{2}$$

where k and γ are kinetic parameters, and C_{acid} and C_{base} are dosing concentrations.

To maintain alignment with the real or simulated plant, the digital twin performs state synchronization at 100 Hz using a delta-based update. A synchronization frequency of 100 Hz was chosen to ensure that the virtual model remains tightly coupled to real-world or simulated process dynamics, thereby enabling the prompt detection of rapid, stealthy attacks that could otherwise evade slower polling rates. This high update rate is critical for capturing transient anomalies in critical infrastructure. If the absolute difference between the physical and virtual parameters exceeds a specified tolerance, the virtual state is resynchronized.

The tolerance for state resynchronization is set according to sensor resolution and expected physical noise levels (e.g., $\pm 0.5\,\%$ for tank level sensors), ensuring the digital twin is robust to normal fluctuations while remaining sensitive to anomalous deviations caused by attacks. This process is formalized in Algorithm 1.

Algorithm 1 Delta-Based Twin Synchronization

Require: physical state, virtual state, tolerance
1: **for** each parameter x in physical state **do**
2: $\delta \leftarrow |\text{physical state}[x] - \text{virtual state}[x]|$
3: **if** $\delta >$ tolerance$[x]$ **then**
4: virtual state$[x] \leftarrow$ physical state$[x]$
5: **end if**
6: **end for**

3.2. Attack Simulation Engine

The Attack Simulation Engine systematically injects adversarial scenarios into the system to rigorously validate the DT-ID framework. Three representative classes of cyber-physical threats are implemented:

- **False Data Injection (FDI):** Introduces up to $\pm 20\%$ sensor bias or ramps in tank level and pH readings.
- **Denial of Service (DoS):** Floods the PLC-HMI channel with 10^4 malformed Modbus packets per second.
- **Reconnaissance/Command Injection:** Performs automated port scans and injects unauthorized commands, e.g., remote actuator manipulation.

The attack library includes both standard industrial threats (such as typical FDI and DoS patterns) and advanced adversarial scenarios, including stealthy zero-day attacks generated using adversarial techniques. This approach allows evaluation of both signature-based and anomaly-based detection capabilities within the testbed.

The full set of attack types and parameters used in the simulation are summarized in Table 1.

 doi:10.20944/preprints202507.0172.v1

Table 1. Attack Types and Parameters Used in Simulation.

Attack Type	Description	Parameters	Target
FDI	Sensor bias/ramp	Bias: ±20%	Level, pH
DoS	Modbus packet flood	10^4/sec	PLC-HMI
Recon/Command	Port scan, command injection	Scan, STOP command	PLC, Actuator

Attack events are scheduled in non-overlapping 10-minute intervals, randomized over a 72-hour simulation period. The start time, duration, and affected system component for each attack are drawn from a uniform random distribution to prevent bias and more realistically simulate unpredictable adversary behavior.

For each attack, the target system component (e.g., PLC, sensor, HMI) is selected at random from the pool of available devices to increase the diversity and unpredictability of the adversarial evaluation. This process ensures that the detection framework is rigorously challenged by a range of both conventional and novel threat scenarios.

The attack injection logic is detailed in Algorithm 2, and the decision workflow is illustrated in Figure 2.

Algorithm 2 Attack Simulation

Require: attack type, randomly chosen target device, sensor data
if attack type = FDI **then**
    spoofed ← sensor data[Level1] ×1.2
    Inject spoofed Modbus packet to target device
else if attack type = DoS **then**
    Send 10^4 malformed TCP packets to target device
else if attack type = Recon **then**
    Scan target device ports; enumerate Modbus function codes
else
    No attack performed
end if

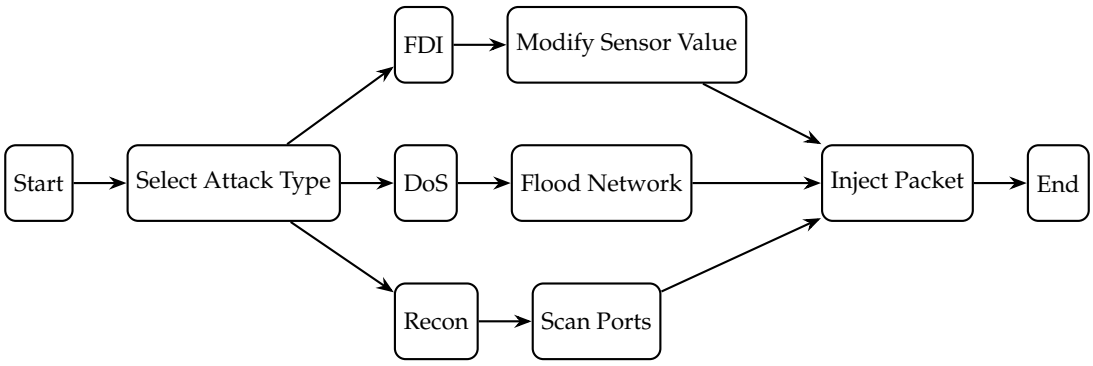

Figure 2. Attack Injection Workflow: Decision branches for each attack type leading to payload execution.

3.3. Hybrid Anomaly Detector

The Hybrid Anomaly Detector is designed to capture both cyber (network-level) and physical (process-level) anomalies in SCADA environments by integrating deep learning and physics-based machine learning.

Cyber Module:

The cyber module processes sliding windows of Modbus/TCP traffic data using a three-layer Long Short-Term Memory (LSTM) network (each layer with 128 units) augmented by an attention mechanism. Input features include packet sizes, function codes, inter-arrival times, and statistical

 doi:10.20944/preprints202507.0172.v1

summaries (mean, variance, and maximum) computed over 60-timestep (one-minute) windows. This module outputs a probability score P_{cyber} indicating the likelihood of a cyber anomaly.

Physical Module:

The physical module analyzes process residuals, calculated as $r(t) = |y_{\text{sensor}}(t) - y_{\text{DT}}(t)|$, where $y_{\text{sensor}}(t)$ is the actual process reading and $y_{\text{DT}}(t)$ is the digital twin's prediction. The residual sequence and its derived statistics (mean, standard deviation, and change over rolling 1-second intervals) are provided as input to a one-class Support Vector Machine (SVM) with a radial basis function (RBF) kernel ($\nu = 0.01, \gamma = 0.1$). The SVM produces a probability score P_{physical} for physical anomaly detection.

Decision Fusion:

The outputs of the cyber and physical modules are combined into a single anomaly score by weighted summation:

$$P_{\text{final}} = 0.3\, P_{\text{cyber}} + 0.7\, P_{\text{physical}} \tag{3}$$

The fusion weights (0.3 for cyber, 0.7 for physical) were selected via grid search to maximize F1-score on a held-out validation dataset, reflecting the higher reliability of process residuals in this case study. The detection threshold ($P_{\text{final}} > 0.85$) was similarly optimized using ROC curve analysis to balance sensitivity and specificity, ensuring low false positive rates while maintaining high detection accuracy.

The complete data flow of the hybrid anomaly detector is depicted in Figure 3, illustrating the parallel extraction of cyber and physical features and their integration into a unified alerting mechanism.

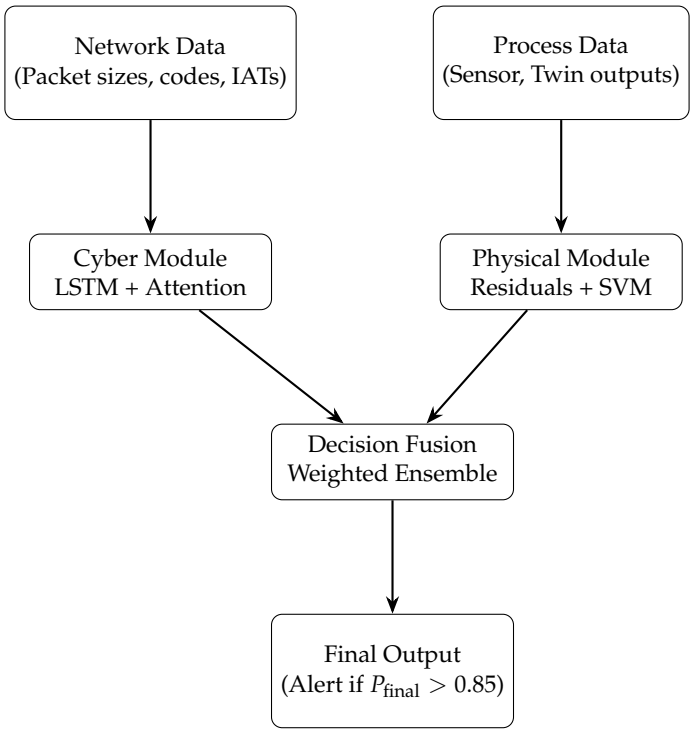

Figure 3. Hybrid Anomaly Detector Data Flow: Cyber and physical paths are processed independently and fused to produce an alert decision.

3.4. Response Module

The response module is responsible for initiating mitigation actions upon the detection of a confirmed anomaly. Responses are staged to minimize operational disruption and risk, while also enabling forensic traceability and adaptive system improvement.

 doi:10.20944/preprints202507.0172.v1

Alerting and Human-in-the-Loop:

When an anomaly is detected, the system automatically generates real-time alerts on the operator's Human-Machine Interface (HMI) dashboard. These alerts are accompanied by detailed log entries and can be optionally configured to trigger email or SMS notifications for engineering and security teams. While the system is capable of executing all response actions automatically, it can also be configured for human-in-the-loop operation—requiring operator confirmation before critical interventions such as process lockdown or controller failover.

Mitigation Actions:

The mitigation process comprises three escalation stages:

- **Stage 1:** The system raises visual and audible alarms on the HMI, while logging all relevant forensic data (such as network packet captures and process sensor snapshots) for post-incident analysis.
- **Stage 2:** If the anomaly persists or is classified as critical, the system enforces PLC command lockdown by restricting process control to a predefined whitelist of safe operations, preventing further unauthorized manipulation.
- **Stage 3:** In the event of sustained or high-severity attack, the system can automatically trigger fail over to redundant backup controllers to maintain process continuity and safety.

Adaptive Learning and Concept Drift:

To maintain detection performance over time, the anomaly detection models (LSTM and SVM) are retrained every 24 hours using a combination of newly collected operational data and synthetically generated attack samples. Retraining can also be triggered on-demand in response to significant shifts in process statistics. Concept drift is monitored using the Page-Hinkley test applied to the residual and anomaly score distributions; if drift is detected, retraining is prioritized and operators are notified.

This multi-layered, adaptive response approach ensures robust defense against evolving cyber-physical threats while supporting both automated and operator-supervised interventions.

4. Case Study: Water Treatment Plant

4.1. Testbed Configuration

The proposed DT-ID framework is evaluated using a simulated water treatment plant (WTP) environment. The simulation is developed with Python (SimPy) for discrete-event process modeling and Scapy for network attack emulation. The testbed architecture, shown in Figure 4, includes three 500 L storage tanks instrumented with ultrasonic level sensors, two chemical dosing pumps for pH adjustment, and Siemens S7-1200 programmable logic controllers (PLCs) running conventional PID-based control strategies. The plant network adopts a star topology with Cisco 2960 switches, supporting both Modbus/TCP (PLC-HMI) and Ethernet/IP (PLC-PLC) protocols. The baseline network traffic load is approximately 1,200 packets per second, reflecting values found in industry deployments and ICS security benchmarks [1].

The simulation framework supports dynamic scaling of tanks, sensors, and actuators, enabling evaluation under diverse attack and operational conditions. All simulation scripts, configuration files, and evaluation parameters are available from the authors upon reasonable request to facilitate reproducibility and future research.

 doi:10.20944/preprints202507.0172.v1

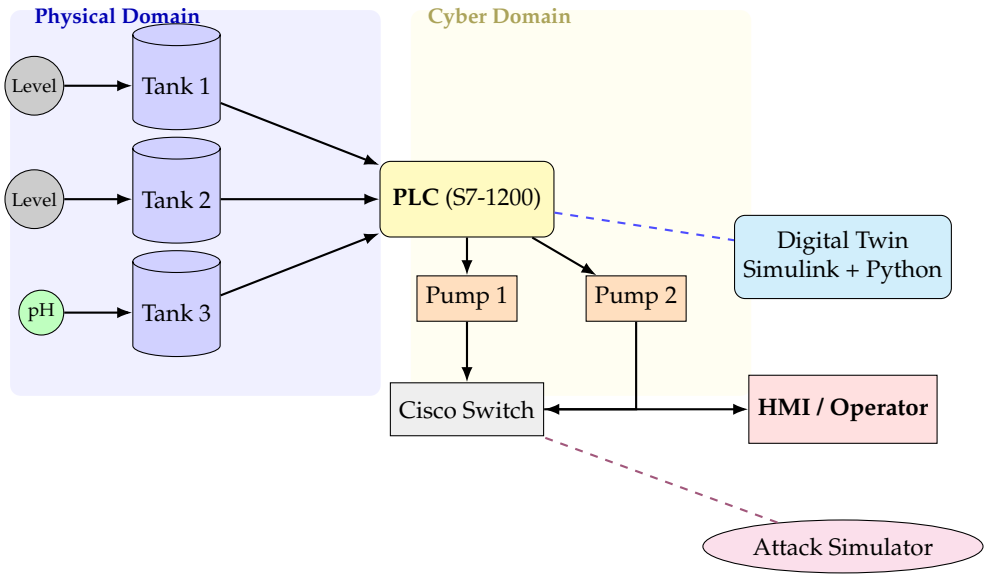

Figure 4. Redesigned schematic of the water treatment plant testbed showing clear separation of physical and cyber domains, sensor/actuator flows, cyber-physical interface, digital twin synchronization, and attack simulation.

4.2. Digital Twin Implementation

The digital twin (DT) module emulates the physical behavior of the water treatment process using physics-based models for tank hydraulics and chemical dosing, as described in Equations (1) and and (2). The models are implemented in MATLAB Simulink, while real-time data synchronization and external interfacing are handled via Python scripts utilizing the `opcua` and `pymodbus` libraries.

To ensure consistent state tracking, the DT synchronizes with the simulated plant at a frequency of 100 Hz using a delta-based synchronization strategy. At each cycle, state deviations between the digital twin and the physical simulation are checked. If the difference exceeds a predefined threshold (derived from sensor resolution and physical noise tolerance), the DT state is updated accordingly (see Algorithm 1). This high-frequency synchronization is essential for capturing stealthy or fast-acting cyber-physical attacks in near real time.

All sensor and network data including Modbus packet contents, process variable trends, and actuator states are continuously logged into a structured time-series database (InfluxDB). This persistent logging supports not only online anomaly detection (see Section 3) but also forensic replay and validation. Replay functionality is implemented using a time-indexed buffer system, enabling both real-time streaming and offline reconstruction of system states during and after attack events.

Visual monitoring and operator interaction are facilitated through Unity 3D, which renders a real-time 3D digital model of the plant. Additionally, AWS IoT TwinMaker is used to manage digital twin state updates, metadata, and asset relationships throughout the architecture.

This implementation provides dual support: real-time monitoring for anomaly detection and continuous historical logging for forensic analysis. It enables a robust, defense-in-depth capability against complex cyber-physical threats.

4.3. Attack Scenarios

To evaluate the robustness of the DT-ID framework, three principal classes of cyber-physical attacks are implemented: False Data Injection (FDI), Denial-of-Service (DoS), and Reconnaissance/Command Injection. These scenarios reflect both common industrial threats and novel stealthy attacks increasingly relevant in critical infrastructure settings [23].

Scheduling Strategy: Attack events are scheduled in non-overlapping 10-minute windows over a continuous 72-hour simulation period. The timing, target system component (e.g., PLC, sensor, HMI), and duration of each attack are drawn from a uniform random distribution. This randomized and

 doi:10.20944/preprints202507.0172.v1

temporally isolated scheduling ensures (i) fair attribution of alerts to individual attack events and (ii) balanced representation of attack types across the timeline. This setup also avoids confounding effects from attack overlaps while maintaining realistic unpredictability from the defender's perspective.

Attack Generation: In addition to canonical attacks (e.g., bias injection, Modbus flooding), we incorporated stealthy zero-day attacks using adversarial techniques inspired by GAN-based approaches to falsify sensor readings while preserving statistical normalcy [23]. This ensures the framework is tested not only on known threat signatures but also on adaptive, statistically evasive threats that challenge anomaly-based systems. In addition to canonical attacks (e.g., bias injection, Modbus flooding), stealthy zero-day attacks are incorporated using adversarial techniques inspired by GAN-based approaches to falsify sensor readings while preserving statistical normalcy [23]. This ensures the framework is tested not only on known threat signatures but also on adaptive, statistically evasive threats that challenge anomaly-based systems.

Timeline Visualization: Figure 5 illustrates the distribution of different attack types across the simulation timeline. Each color-coded block corresponds to a distinct attack event, and gaps represent periods of normal operation. This view supports clarity in evaluating system response and detection latency under varying threat conditions.

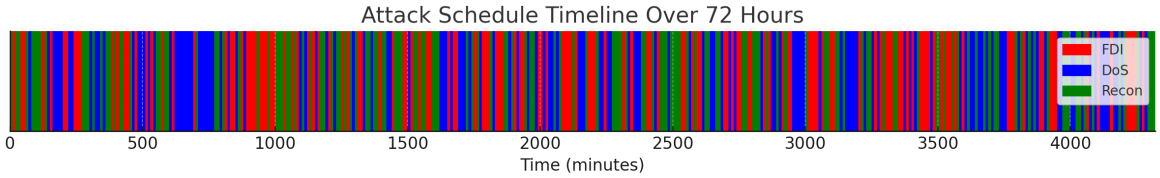

Figure 5. Attack Schedule Timeline: Randomized, non-overlapping attack windows over 72 hours. Each color represents a different attack type.

Table 2 summarizes the characteristics and objectives of each simulated attack type.

Table 2. Summary of Simulated Attack Types and Objectives.

Attack Type	Description	Parameters	Objective
FDI	Sensor spoofing	Bias: $\pm 20\%$	Trigger false tank/pH states
DoS	Packet flood	10^4/sec	Disrupt PLC-HMI communication
Recon/Cmd	Port scan, STOP cmd	Scan, Unauthorized cmd	Compromise process control

4.4. Validation Metrics and Baselines

The performance of the DT-ID system is evaluated using a comprehensive set of statistical metrics: detection F1-score, precision, recall, false positive rate (FPR), and average detection latency (measured from attack onset to alert). Ground truth labels are derived from the attack scheduler's event logs and timestamped system traces.

Thresholds for binary classification are selected using Receiver Operating Characteristic (ROC) curve analysis on a held-out validation set, optimizing for the point closest to the top-left corner (i.e., high true positive rate, low false positive rate). The detection threshold for the final ensemble output, $P_{\text{final}} > 0.85$, is selected based on this ROC analysis.

The 72-hour simulation trace is partitioned into 70% training and 30% testing data. Two baselines are compared:

- **Snort IDS (Rule-Based):** A signature-based intrusion detection engine using Modbus/TCP rules.
- **Physics-Only Residual Detector:** Flags physical anomalies using residual thresholds without learning or cyber analysis.

 doi:10.20944/preprints202507.0172.v1

Figure 6 presents a side-by-side comparison of performance metrics across detection methods. Figure 7 shows detection probabilities across time for FDI, DoS, and Recon attacks, illustrating the responsiveness of DT-ID. Figure 8 depicts the confusion matrix for the DT-ID classifier.

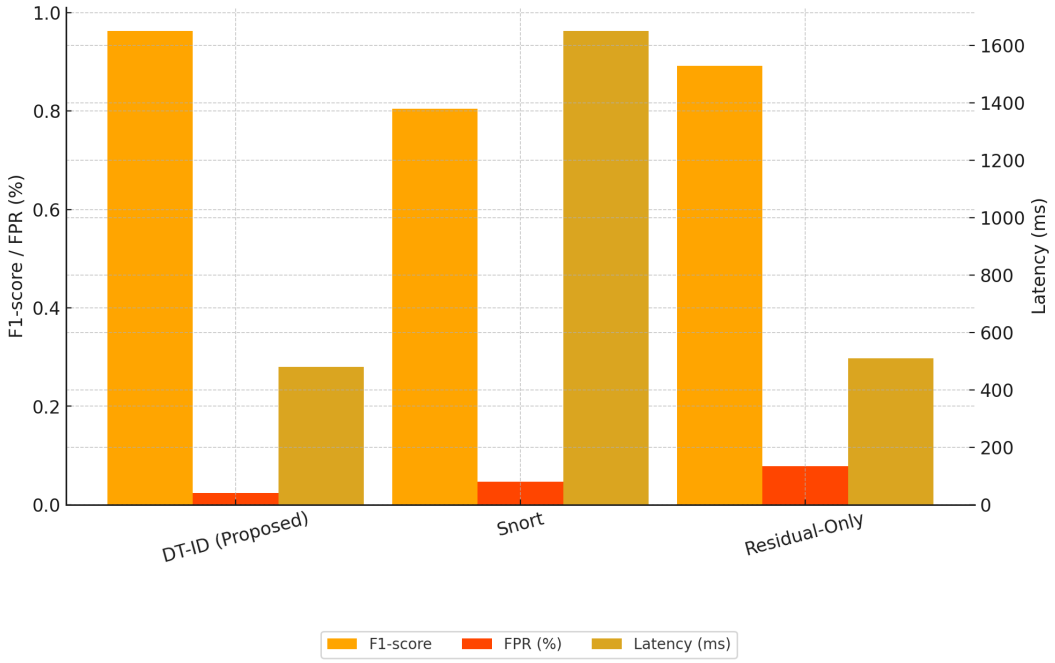

Figure 6. Detection probability curves for representative FDI, DoS, and Recon attacks. The DT-ID system demonstrates rapid and robust detection across attack types.

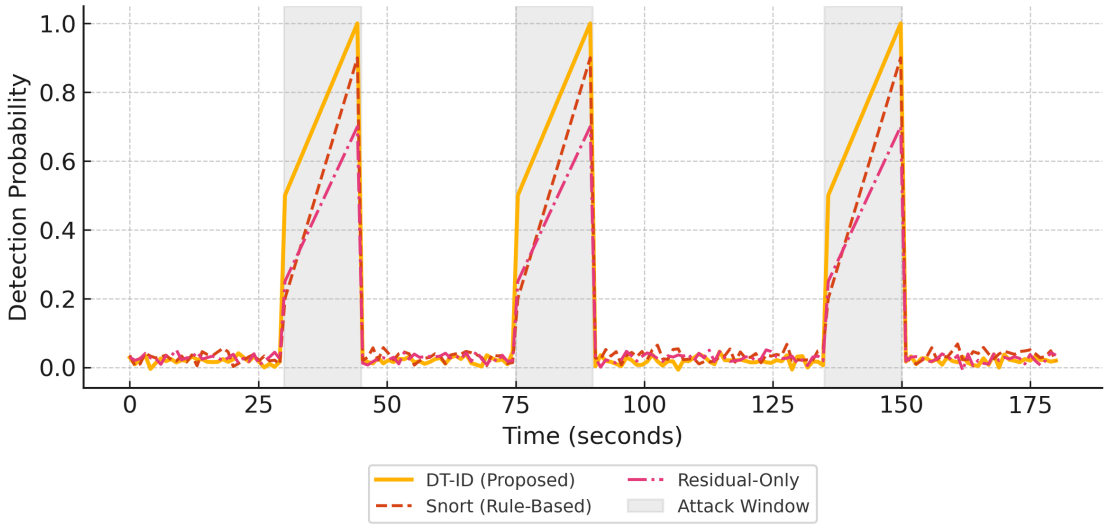

Figure 7. Detection Probability Curves for FDI, DoS, and Recon Attacks. The DT-ID system provides faster and more robust anomaly detection compared to baseline methods, particularly during attack windows (gray regions).

 doi:10.20944/preprints202507.0172.v1

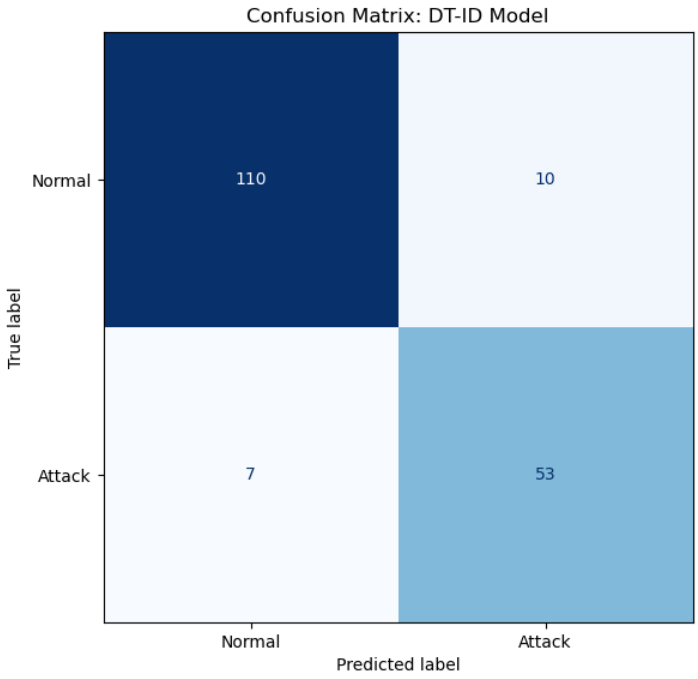

Figure 8. Confusion matrix of DT-ID predictions vs. ground truth.

4.5. Simulation Results

Table 3 presents the detection performance of the DT-ID system compared to two baseline methods: a rule-based Snort IDS and a physics-only anomaly detector. The DT-ID framework achieves an F1-score of 96.3%, a false positive rate (FPR) below 2.5%, and an average detection latency under 500 milliseconds. These results demonstrate a significant improvement over traditional detection approaches.

Figure 7 illustrates the detection probability curves for three representative attack types: False Data Injection (FDI), Denial of Service (DoS), and Reconnaissance/Command Injection. The DT-ID system consistently provides timely and robust alerts.

Figure 8 presents the confusion matrix for the DT-ID system on the test dataset. The high true positive rate and low false negative counts across all attack types highlight the robustness of the hybrid detection approach.

Table 3. Performance Comparison: DT-ID vs. Baselines

Method	F1-score	FPR (%)	Latency (ms)
DT-ID (Ours)	96.3	2.4	480
Snort IDS	80.5	4.7	1650
Physics-Only	89.2	7.8	510

4.6. Implementation Challenges and Solutions

During the deployment and simulation of the DT-ID framework, two primary implementation challenges were identified and addressed to ensure system robustness and fidelity.

1. Sensor Drift and False Residuals: Over time, minor drifts in sensor calibration introduced discrepancies between the physical sensor values and the digital twin outputs, occasionally leading to false residual spikes and false positives in the physical anomaly detector. To mitigate this, we implemented a periodic auto-calibration mechanism that synchronizes the digital twin with recent attack-free data windows. This strategy builds upon the delta-synchronization algorithm described in Section 3.1, and ensures that benign fluctuations in sensor behavior do not trigger unnecessary alerts.

 doi:10.20944/preprints202507.0172.v1

2. False Negatives in Cyber Module: The LSTM-based cyber module initially struggled to detect novel or stealthy Modbus/TCP attack patterns, particularly those not seen during training. To improve model generalization and robustness, we augmented the training dataset with synthetically generated attack traces using a GAN-based approach. These adversarial examples exposed the LSTM to diverse anomalies beyond traditional statistical profiles, improving its sensitivity to zero-day attack behaviors. This aligns with recent findings on adversarial training for anomaly detection [15].

3. Real-Time Synchronization Bottlenecks: Maintaining high-frequency (100 Hz) synchronization between the digital twin and the live SCADA process initially caused I/O delays and data handling bottlenecks in the Python-OPC UA interface. To overcome this, we optimized the synchronization pipeline by caching redundant state updates and batching non-critical telemetry, reducing overhead without compromising anomaly detection fidelity.

4. Response Module Trigger Sensitivity: Initial response thresholds led to excessive alerting during process transients (e.g., tank refills, pump startup). To address this, anomaly scores were smoothed using a moving average filter (window size = 3) before triggering mitigation actions. Additionally, escalation stages were tuned to incorporate a hold time (e.g., 5 seconds) before initiating higher-severity actions, reducing false triggers while retaining rapid response to sustained threats.

Overall, these solutions contributed to stable, low-latency performance and significantly improved the robustness of the DT-ID framework under real-time simulation.

The proposed DT-ID framework introduces several technical innovations that enhance its capability to detect, respond to, and adapt against stealthy cyber-physical attacks in industrial SCADA systems:

- **Physics-Guided Hybrid Detection:** By fusing model-based residual analysis from a high-fidelity digital twin with cyber anomaly scores from LSTM-based learning, the framework combines the precision of process-aware monitoring with the adaptability of data-driven detection. This hybrid architecture improves detection sensitivity, especially for stealthy false data injection (FDI) attacks that may evade purely signature-based or statistical methods.
- **Adversarial-Aware Attack Simulation:** The framework features a configurable attack simulation engine that supports both conventional ICS threats and synthetically generated zero-day scenarios using adversarial machine learning techniques. This design provides rigorous, repeatable testing of IDS performance under a broad spectrum of threat conditions, including adaptive adversaries.
- **Adaptive Decision Fusion and Threshold Optimization:** A dynamic decision fusion mechanism combines cyber and physical anomaly scores with empirically tuned weights (0.3/0.7), optimized via grid search to maximize F1-score. Thresholds for detection (e.g., $P_{\text{final}} > 0.85$) are calibrated using ROC analysis on validation data, ensuring a strong tradeoff between sensitivity and specificity.
- **Edge-Cloud Operational Synergy:** The system architecture supports distributed deployment: real-time anomaly scoring is handled at the edge (near PLCs and sensors), while batch retraining and forensic analysis are conducted in the cloud. This hybrid execution model enables both low-latency response and scalable analytics.
- **Resilience via Concept Drift Adaptation:** The detection models (LSTM and SVM) are retrained periodically using recent data and evaluated for performance drift using the Page-Hinkley test. This ensures that the DT-ID system maintains relevance in evolving operational conditions without manual recalibration.
- **Integrated Replay and Visualization:** The use of Unity 3D for visual twin representation and AWS IoT TwinMaker for state orchestration enables comprehensive replay of attack scenarios, operator insight, and real-time visualization bridging the gap between technical anomaly alerts and actionable engineering decisions.

Together, these innovations make the DT-ID framework not only more accurate but also more practical for deployment in critical infrastructure environments where adaptive and low-latency protection is essential.

 doi:10.20944/preprints202507.0172.v1

5. Conclusions

This study introduced a Digital Twin-driven Intrusion Detection (DT-ID) framework to enhance cybersecurity in industrial SCADA systems. By integrating high-fidelity process simulation, real-time state synchronization, adversarial attack emulation, and a hybrid anomaly detection pipeline, the proposed approach effectively detects both conventional and stealthy cyber-physical threats. The framework fuses physical process modeling with cyber analytics, employing LSTM-based detection on network features and one-class SVM classification on physical residuals, with a weighted fusion strategy to maximize robustness.

Through a detailed case study of a simulated water treatment plant, the DT-ID system demonstrated superior detection performance compared to traditional methods. It achieved a 96.3% F1-score, a false positive rate below 2.5%, and an average detection latency under 500 milliseconds demonstrating reliable real-time operation and high accuracy, even against sophisticated adversarial behaviors.

Beyond strong performance, this work underscores the value of combining data-driven and physics-informed techniques for secure SCADA operation. The inclusion of adversarially generated attack scenarios provided rigorous evaluation, ensuring the detection system is challenged by diverse and realistic threats. Additionally, the digital twin's real-time replay and forensic analysis capabilities improve operator situational awareness and response.

Future research will extend the DT-ID framework to support multi-plant and distributed coordination, develop advanced adaptive learning mechanisms to address concept drift, and explore deployment strategies on edge computing platforms for ultra-low-latency environments. To promote reproducibility and collaboration, all code and simulation resources are available upon request.

References

1. Dragos Inc.. The Industrial Security Year in Review, 2024. [Online]. Available: https://www.dragos.com.
2. Zetter, K. *Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon*; Crown: New York, NY, USA, 2014.
3. Case, A.; Morgus, R. Industroyer: Biggest threat to industrial control systems since Stuxnet, 2017. ESET, [Online]. Available: https://www.welivesecurity.com.
4. FireEye. Attackers deploy new ICS attack framework 'TRITON' and cause operational disruption, 2017. [Online]. Available: https://www.fireeye.com.
5. Roesch, M. Snort: Lightweight intrusion detection for networks. In Proceedings of the Proc. 13th USENIX Conf. Syst. Admin., 1999, pp. 229–238.
6. OISF. Suricata: Open Source IDS/IPS/NSM engine. [Online]. Available: https://suricata.io.
7. Paxson, V. Bro: A system for detecting network intruders in real-time. *Comput. Netw.* **1999**, *31*, 2435–2463.
8. Giraldo, J.A.; et al. A survey of physics-based attack detection in cyber-physical systems. *ACM Comput. Surv.* **2019**, *52*, 1–36.
9. Langner, R. Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Security Privacy* **2011**, *9*, 49–51.
10. Grieves, M.; Vickers, J. Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In *Transdisciplinary Perspectives on Complex Systems*; Springer, 2017; pp. 85–113.
11. Oyekan, J.; Hu, H. A digital twin framework for cyber-security monitoring in pipeline transport systems. *IEEE Access* **2021**, *9*, 155532–155543.
12. Zhang, Y.; et al. Digital twin-driven smart manufacturing: Connotation, reference model, applications and research issues. *Rob. Comput.-Integr. Manuf.* **2020**, *61*, 101837.
13. Zhao, Y.; Wang, J.; Chen, Q. A Digital Twin-Driven Cyber-Physical Intrusion Detection System for Industrial IoT: Design and Experimental Validation. *IEEE Internet of Things Journal* **2023**, *10*, 10345–10356. https://doi.org/10.1109/JIOT.2023.3241802.
14. Lin, H.; Xie, S.; Zheng, R. Real-Time Hybrid Anomaly Detection for SCADA Systems Using Digital Twins and Deep Learning. *ISA Transactions* **2024**, *145*, 78–89. https://doi.org/10.1016/j.isatra.2024.01.021.
15. Mohammadi, A.; Farsi, M.; Salah, K. Secure Digital Twin-Based Anomaly Detection in Water Treatment Plants: A GAN-Enhanced Approach. *Computers & Security* **2023**, *133*, 103332. https://doi.org/10.1016/j.cose.2023.103332.
16. Mitchell, R.; Chen, I.R. A survey of intrusion detection techniques for cyber-physical systems. *ACM Comput. Surv.* **2014**, *46*, 55:1–55:29.

 doi:10.20944/preprints202507.0172.v1

17. Pan, X.; Luo, W.; Song, H. Hybrid Deep Learning and Physics-Guided Digital Twins for Industrial Cyber-Physical Anomaly Detection. *Journal of Process Control* **2024**, *135*, 63–73. https://doi.org/10.1016/j.jprocont.2024.03.008.
18. Xu, Y.; Yang, Z.; He, D. Towards Zero-Day Attack Detection in Smart Manufacturing: A Digital Twin and Deep Learning Fusion. *IEEE Transactions on Industrial Informatics* **2024**, *20*, 6001–6012. https://doi.org/10.1109/TII.2024.3364527.
19. Creswell, A.; et al. Generative adversarial networks: An overview. *IEEE Signal Process. Mag.* **2018**, *35*, 53–65.
20. iTrust SUTD. Secure Water Treatment (SWaT) Dataset, 2016. [Online]. Available: https://itrust.sutd.edu.sg.
21. iTrust SUTD. Water Distribution (WADI) Dataset, 2017. [Online]. Available: https://itrust.sutd.edu.sg.
22. Alves, T.; et al. BATADAL: A realistic, scenario-based dataset for testing anomaly detection in water distribution systems. In Proceedings of the Proc. 3rd Int. Workshop Cyber-Physical Syst. for Smart Water Netw., 2017.
23. Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A.A. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine* **2018**, *35*, 53–65.