*electronics*

MDPI

*Article*

# Fusing Design and Machine Learning for Anomaly Detection in Water Treatment Plants

**Gauthama Raman [1,*] and Aditya Mathur [1,2]**

[1] iTrust, Centre for Cyber Security Research, Singapore University of Technology and Design, Singapore 487372, Singapore; aditya_mathur@sutd.edu.sg
[2] Department of Computer Science, Purdue University, West Lafayette, IN 47907, USA
* Correspondence: gauthama_mani@sutd.edu.sg

**Abstract:** Accurate detection of process anomalies is crucial for maintaining reliable operations in critical infrastructures such as water treatment plants. Traditional methods for creating anomaly detection systems in these facilities typically focus on either design-based strategies, which encompass physical and engineering aspects, or on data-driven models that utilize machine learning to interpret complex data patterns. Challenges in creating these detectors arise from factors such as dynamic operating conditions, lack of design knowledge, and the complex interdependencies among heterogeneous components. This paper proposes a novel fusion detector that combines the strengths of both design-based and machine learning approaches for accurate detection of process anomalies. The proposed methodology was implemented in an operational secure water treatment (SWaT) testbed, and its performance evaluated during the Critical Infrastructure Security Showdown (CISS) 2022 event. A comparative analysis against four commercially available anomaly detection systems that participated in the CISS 2022 event revealed that our fusion detector successfully detected 19 out of 22 attacks, demonstrating high accuracy with a low rate of false positives.

**Keywords:** process anomaly; critical infrastructure protection; industrial control systems; Critical Infrastructure Security Showdown (CISS) 2022; fusion of design and data-centric approaches; water treatment plants

check for updates

**Citation:** Raman, G.; Mathur, A. Fusing Design and Machine Learning for Anomaly Detection in Water Treatment Plants. *Electronics* **2024**, *13*, 2267. https://doi.org/10.3390/electronics13122267

Academic Editors: Rosario Catelli and Giovanni Paragliola

Received: 7 April 2024
Revised: 25 May 2024
Accepted: 3 June 2024
Published: 9 June 2024

**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Industrial control systems (ICSs) seamlessly integrate computational elements with physical components to monitor and control the processes underlying critical infrastructure (CI), such as water treatment and distribution plants. The increase in connectivity with corporate internet technology and the incorporation of proprietary software and devices have led to a more dynamic and challenging threat environment for ICSs. Traditional threat intelligence, primarily developed for IT systems, often falls short in addressing the unique requirements of ICSs. This is highlighted by recent successful security breaches against public infrastructure, underscoring the need for robust, ICS-specific anomaly detection systems.

The methods for building detectors for ICSs generally fall into two main categories: design-centric and data-centric. Design-centric approaches leverage the plant's design to create invariants, i.e., rules that govern normal plant operation. These invariants explicitly represent the interactions among plant components, such as pumps, valves, circuit breakers, and sensors. An example of this approach is distributed anomaly detection (DAD) [1], a fully functional anomaly detector for the secure water treatment (SWaT) testbed. Conversely, data-centric approaches involve creating representative models by applying machine learning to operational data collected during normal plant operations. These models are then used to predict the behavior of each component or a group of components, and anomalies are detected by comparing these predictions against the actual

behavior of the plant. Detectors such as MLP_CUSUM [2], PbNN [3], and AICrit [4], operational on SWaT, exemplify data-centric approaches.

While the aforementioned detectors have demonstrated their effectiveness independently in numerous cybersecurity exercises within the SWaT and larger-scale water treatment facilities in Singapore, significant challenges remain in developing such systems for system-wide monitoring. These challenges arise from factors such as inadequate design knowledge, the dynamic nature of operational conditions, and the complex interdependencies among the components. An illustrative example is shown in Figure 1. In stage 2 of SWaT, the AIT202 sensor monitors the pH level after chemical dosing. Following ultrafiltration in stage 3 the AIT302 sensor measures the pH level. However, due to the absence of accurate mathematical models for the ultrafiltration process, its impact on the water's pH cannot be converted into invariants by DAD. Conversely, the data-centric approach can effectively forecast the AIT302 readings based on the chemical parameters in stage 2 by treating the ultrafiltration process as a black box. In another scenario, DAD can explicitly represent the interactions among sensors and valves in stage 3 (see Figure 1) as invariants; however, data-centric approaches fail to capture such relationships due to the heterogeneity in the nature of the data.

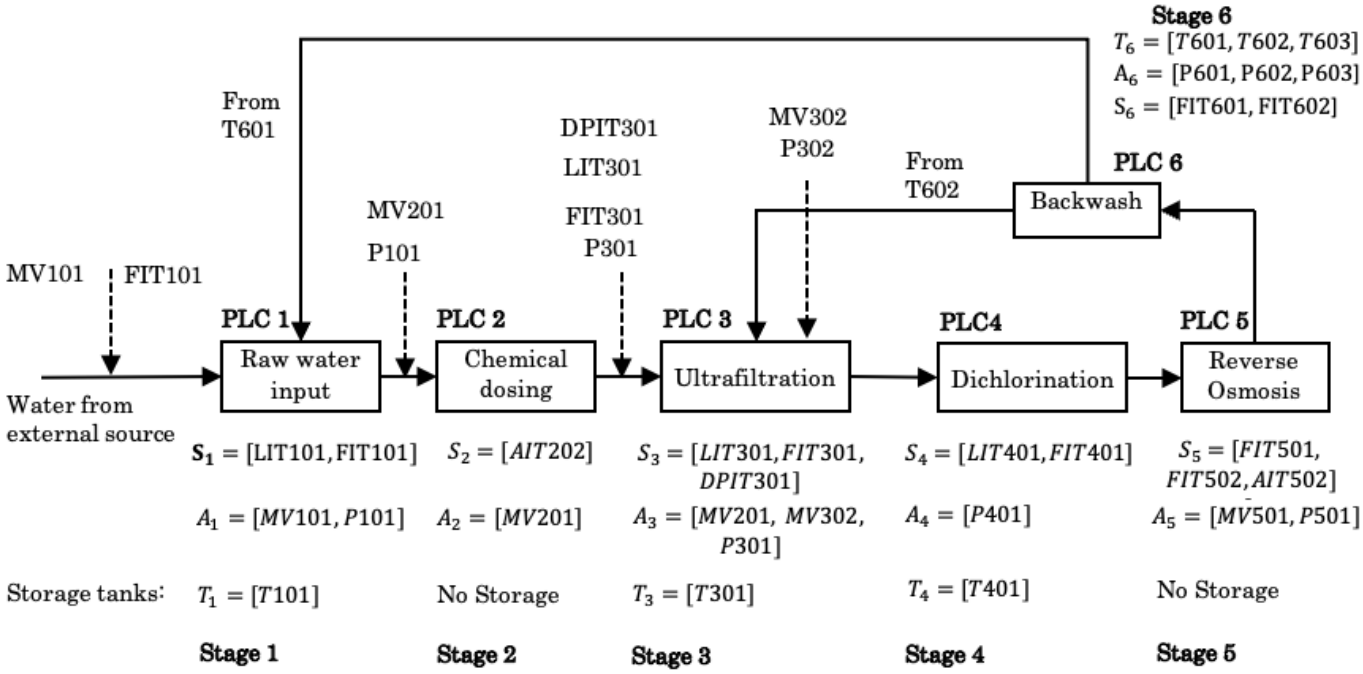



**Figure 1.** SWaT architecture; LITxxx: level sensor; Pxxx: pump; AITxxx: chemical property sensor; FITxxx: flow sensor; DPITxxx: differential pressure indicator.

The scenarios outlined above highlight the challenges of independently modeling design-centric and data-centric approaches to build an effective anomaly detection system. In light of these insights, we propose a fusion detector that synergistically combines the strengths of both methodologies for comprehensive system-wide monitoring, aiming to achieve an enhanced detection rate. The proposed approach leverages the design-centric method's capability to monitor interactions among heterogeneous components, supplemented by the data-centric approach in scenarios where explicit design knowledge is lacking. Thus, the invariants, that monitor interactions among sensors and actuators, are combined with AICrit's representative models that monitor sensor behavior across and within the stages of SWaT, and the resulting combination is referred to as the fusion detector.

The fusion detector was implemented in SWaT and its effectiveness assessed against a series of attacks launched by iTrust's red teams during the Critical Infrastructure Security Showdown 2022 (CISS2022). Furthermore, the performance of the fusion detector

was benchmarked against four state-of-the-art commercial intrusion detection systems (IDSs) that were also used in CISS2022, thus providing a comprehensive assessment of its capabilities in a real-world scenario.

The following research questions are the focus of this paper with respect to the Fusion detector:

- **RQ 1:** *How effective is the Fusion detector in detecting the cyber-physical attacks that have an adverse impact on the physical processes of underlying SWaT*?
- **RQ 2:** *How efficient is the Fusion detector in describing threats and supporting the plant operators towards active incident response*?

Our contributions are listed below.

1. A list of attacks, and their description, launched by the iTrust red team during CISS2022. This list is considered valuable for evaluating anomaly detectors when using the SWaT dataset [5].
2. Summary of the benefits of CISS2022 to the community involved in safeguarding critical infrastructure in water treatment systems.
3. A comparison of the performance of Fusion detector and four commercially available intrusion detection systems.

Organization : The remainder of this paper is organized as follows. Section 2 summarizes the recent research carried out in safeguarding ICS. Information related to the SWaT testbed and CISS2022 is provided in Section 3. Section 4 provides a detailed description of each module in the Fusion detector. Preparation for CISS2022 and the attacks launched by the iTrust Red Team are described in Section 5. Performance evaluation of the Fusion detector is given in Section 6. In Section 7 we return to the research questions mentioned above and answers them along with the future directions. Conclusions based on CISS2022 are given in Section 8.

## 2. Related Work

In addition to the iTrust detectors mentioned in the above section, there are several techniques for safeguarding ICSs. In [6], the techniques to detect process anomalies by ensuring the correctness of sensor measurements and control commands of the actuators are broadly classified as (i) design-based and (ii) data-based. This section provides a summary of recent work carried out in these three categories.

### *2.1. Design-Based Approach*

In the design-based approach, process anomalies are detected by the state estimator. The state estimator operates over the physical or chemical properties of the processes controlled by one or more PLCs to estimate the state of the actuators or sensor measurements for each time step. DAD is one such model-based approach. SCADMAN [7] provides a control code generation and verification mechanism to ensure the correctness of the behavior of individual PLCs. It was implemented and evaluated on the SWaT testbed. Orpheus [8] operates over the FSM to monitor the behavior of the control codes and check whether it is legitimate for a given context. The authors of [9] proposed a residual-based anomaly detection system to monitor the sensor measurements in the water treatment process. A similar approach was proposed by Ghaeini et al. [10], wherein the CUSUM approach was utilized to improve the attack detection rate and to minimize the false alarms. SafeCI [11] is an anomaly detection and prevention approach that validates the control commands issued by PLCs before they reach the target actuators. Using invariants, the state of each actuator is estimated using the current state of the plant, and when the estimated state does not match with the control commands issued by PLCs, alerts are raised.

### *2.2. Data-Based Approach*

In the data-based approach, process anomalies are detected based on statistical techniques and machine learning algorithms. From the historical data obtained from the plant,

the patterns in the behavior of the plant components are learned and modeled. Further, the trained model is utilized to detect any anomalies in the behavior due to process anomalies. AICrit is one of the data-based approaches. The authors of [12] proposed a distributed anomaly detection architecture based on autoencoders, Transformer, and Fourier mixing sublayer. The proposed unsupervised approach was evaluated on several benchmark datasets like SWaT, HAI, gas pipeline, power demand, etc., and proven to achieve better detection precision and F1 score. Tang et al. [13] proposed a multivariate time series prediction approach to ensure the reliable operation of ICS. In this work, both neural graph networks and gated recurrent units are integrated to accurately model the dependencies among the sensors for the anomaly detection process. This approach was evaluated on SWaT and WADI datasets and proven to achieve a better detection rate compared with nine state-of-the-art algorithms. Das et al. [14] proposed a supervised learning approach named logical analysis of data (LAD) to extract the rules from the historical data for several operating conditions of the plant. The efficiency of the proposed approach was evaluated on the sensor measurements of SWaT data compared with other anomaly detection approaches. Although the proposed approach can effectively localize the anomalies, it can detect anomalies that target only the sensors. The authors of [15] proposed a physics-informed gated recurrent graph neural network for anomaly detection in industrial cyber-physical systems. Initially, the dependencies among the variables are modeled using direct graphs, and their interactions are learned using the recurrent graph neural network. The proposed approach was evaluated on SWaT and WADI datasets and proved to be better in detection ability compared with ten state-of-the-art methods. Shuaiyi et al. [16] introduced an attributed heterogeneous graph analyzer based on graph neural networks (GNNs) for SWaT and WADI datasets. Their approach aims to identify anomalous patterns at the device level by employing a comprehensive process-oriented associativity learning method.

To summarize, unlike DAD and AICrit, the abovementioned works are evaluated on a benchmark dataset or on the simulated environment. There is no guarantee that these techniques operate in a similar manner by overcoming the real-time implementation issues while deploying them on the operational plants.

## 3. Preliminaries and Background

The general architecture of SWaT [17] is presented next, followed by an overview of the CISS2022 event.

### *3.1. SWaT—Architecture*

The SWaT is a six-stage water treatment plant used by researchers to evaluate their defense mechanisms and examine the effects of cyber-attacks on CI. The water treatment process in the SWaT is outlined below. For additional information related to the network architecture, refer to [17].

The SWaT consists of six interconnected stages labeled stage 1 through stage 6. Each stage consists of a set of sensors and actuators controlled by a programmable logic controller (PLC). In stage 1, the incoming water is stored in the raw water tank T101. Water from T101 is transferred via a chemical dosing station in stage 2 to the ultra-filtration (UF) unit in stage 3. Using the UF feed pump (P301), water is passed to stage 4 for the removal of free chlorine. Next, the inorganic impurities are removed from the dechlorinated water using a 2-stage reverse osmosis (RO) process in stage 5. Treated water is stored in stage 6 (T601) and recycled to stage 1 (tank T101). The rejected water from RO (tank T602) is used to clean the UF unit in stage 3 via a backwash process.

### *3.2. CISS2022*

CISS2022 (https://itrust.sutd.edu.sg/ciss-2022/ (accessed on 1 February 2024)) was held in iTrust from 12 to 20 September 2022, where participants were able to launch attacks from their respective remote locations. The goal of CISS2022 was to (i) validate and assess the effectiveness of defense mechanisms designed for iTrust testbeds, (ii) develop

capabilities for safeguarding CI against cyberattacks, (iii) understand the composite tactics, techniques, and procedures (TTP) for enhanced operational security, and (iv) practice the approaches for compromising and defending CI. The participants in CISS2022 were divided into the following categories.

1. Red teams: Up to 10 local and international teams from government organizations, the private sector, and academia.
2. Blue teams: Commercial vendors were invited based on their past performance in similar events and nominations by the Singapore Government agencies.
3. IHL anomaly detector teams: Anomaly detectors from iTrust.
4. CI blue teams: CI operators and regulators.
5. Observers: Singapore Government agencies and their invitees.

During the event, red teams from several organizations launched attacks on the three testbeds in iTrust. An 8-h slot was assigned to each blue team for responding to the attacks launched. Detailed information regarding the red and blue teams, online exercise platform, and attack launch procedures are available in [18]. In this paper, we focus only on the attacks launched on the SWaT testbed.

## 4. Fusion Detector: Architecture and Building Blocks

Figure 2 shows the infrastructure used for evaluating the Fusion detector considered in this study. As shown, the infrastructure contains four key modules: (i) data feed and validation, (ii) threat lookup, (iii) logging, and (iv) plant visualizer. The operation of each module is coordinated by a reconfigurable digital twin [19] that allows the plant management to create and configure new modules or reconfigure the existing modules based on their requirements. The primary goal of the twin is to provide operators with an easy-to-use interface for managing the operation of multiple modules through an interactive graphical user interface. A description of each module in Figure 2 follows.

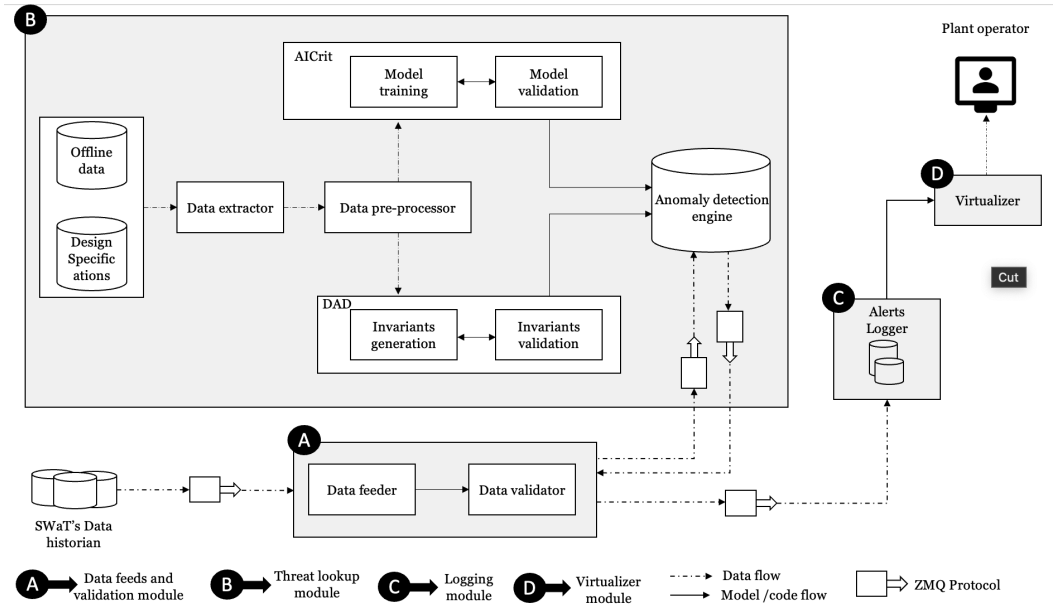



**Figure 2.** Infrastructure used in the performance assessment of Fusion detector and commercial IDS.

### *4.1. Data Feeds and Validation*

Data are extracted from the SWaT Historian once every second and streamed using the OPC-UA protocol to the data validation and feeds module. Data are received in the form of a Python dictionary containing sensor measurements and actuator states. The received data are validated to ensure correctness before distribution. Additionally, the validator module adds a new tag name called "bad input", which is usually an empty list if all tag values are in the correct format. In cases of communication error or a temporal glitch, the

tag values received may be considered noisy data. In such instances, a list of tags consisting of the noisy data is added to the "bad input" key. This information is useful for anomaly detectors that disregard the tags with noisy data. Validated data are published using the ZMQ protocol and are available to the subscribers in the infrastructure.

### *4.2. Threat Lookup Module*

The core part of the fusion detector is the threat lookup module that performs real-time process monitoring for anomaly detection. This module comprises representative models and invariants derived from the training processes in AICrit and DAD, respectively. By using the plant design and historical data corresponding to normal plant operations as inputs, this module initiates the training process for both detectors, and their outcomes are stored in the anomaly detection engine for the detection of anomalies. Next, we will provide an overview of each detector.

1. DAD: Distributed anomaly detection (DAD) [1] detects anomalies in industrial processes by utilizing mathematical relationships across device states known as "invariants". The invariants serve as a means of monitoring the system state and are executed in a cyclical manner. In each cycle, sensor readings and actuator status are obtained, and the invariants are checked. Violation of an invariant is considered an anomalous event and an alert is generated.
2. AICrit: AICrit presents a unified framework aimed at real-time process monitoring, emphasizing the preservation of the ICS's control behavior integrity. Through the application of machine learning algorithms (data-centric approach) and a substantial amount of design knowledge (design-centric approach), the framework accurately learns the normal spatiotemporal relationships among a group of correlated components. The development process of AICrit, which involves unsupervised learning, consists of two primary steps.

In the first step, the framework models the normal behavior of continuous-valued state variables, such as water level sensors, by utilizing temporal dependencies to forecast their behavior with minimal error. In the subsequent second step, it captures the higher-order and nonlinear correlations among both discrete and continuous type state variables, including cross-correlations among sensors and actuators. By combining these two sets of models, AICrit enables continuous monitoring of the functional dependencies of the sensors and actuators, enhancing confidence in detecting and reporting a wide range of process anomalies. For detailed information related to the training process, the reader can refer to [4]. AICrit employs several deep learning algorithms to accurately model interrelationships, encountering challenges such as overfitting, model and data drift, and hyperparameter fine-tuning. A comprehensive case study, detailed in [20,21], explores these challenges and the methodologies used to address them during the translation to an industrial-grade environment.

As discussed in Section 1, the anomaly detection engine in the fusion detector utilizes DAD's invariants, which monitor interactions among sensors and actuators, along with the representative models from AICrit, which monitor the spatiotemporal dependencies of sensors within and across the system. During this integration process, DAD's invariants are converted to Python code, and separate data pipelines are created for both detectors to facilitate the anomaly detection process.

### *4.3. Logging Module*

The Fusion detector logging module records and archives raw data from the plant, as well as the status of the threat lookup module, in InfluxDB2 (https://docs.influxdata.com/influxdb/v2.6/ (accessed on 1 February 2024)), an open-source time-series database. As shown in Figure 2, the twin receives the data from the plant and the alerts from the threat lookup module, and combines and publishes the information in a specified port. The logging module retrieves this information using the ZMQ client and pushes it to the InfluxDB2 database using the Python client library (https://github.com/influxdata/

influxdb-client-python (accessed on 1 February 2024)). At the end of each red team session, the logging module generates three logs in the .csv format discussed below.

1. Detectors log: A separate .csv file is generated each for DAD and AICrit, containing the alerts generated by the detectors and the corresponding timestamps. During the post-CISS analysis, this log is compared against the list of attacks launched to evaluate the performance of each detector.
2. Data log: This log contains sensor readings and actuator status recorded at each second during the exercise, together with the attack details. These data are made available to the public for offline evaluation of their defense mechanisms.
3. Statistics log: This log contains statistical parameters, including the minimum, maximum, and mean values of all sensors, as well as the transition time of each valve. This information will be used for predictive analytics, such as component failure and drift detection.

### *4.4. Virtualizer Module*

The virtualizer module serves as a bridge between the SWaT and the plant operators. It features an interactive GUI built with Grafana [22], an open-source data-visualization platform. This dynamic web application provides real-time visibility into the state of SWaT, including any anomalies reported by the detectors. The virtualizer receives consolidated information from the logging module in real time and presents it through the following three distinct dashboards.

1. **SWaT—State:** This dashboard organizes components by plant stage and displays their current state.
2. **AICrit Statistics**: This dashboard presents the minimum and maximum values of each model in AICrit.
3. **SWaT—Overall State**: This dashboard displays individual alerts from DAD and AICrit and provides a comprehensive overview of SWaT's state at each second.

## 5. Preparations for the CISS2022 and iTrust Red Team Exercises

As part of the CISS2022 preparation, two months prior to the event, each red team was given the technical details of iTrust testbeds, including their respective network architecture, communication protocols used, and the physical devices present. Further, they were informed of the devices (refer to Table 1) that could serve as attack targets. The detector owners were provided with five days of IT and OT traffic collected from SWaT operated under normal conditions for the baselining process. Since both DAD and AICrit are already trained to monitor SWaT, these five days of data were used to test for false positives. If DAD generated false alarms, the corresponding threshold values were recomputed. Similarly, if AICrit generated false alarms, the corresponding models and threshold values were updated through an incremental learning process. Finally, the updated detectors were integrated with the twin and deployed on a virtual machine with 32 GB of RAM running Ubuntu OS, enabling it to receive real-time data from the SWaT historian via an OPC-UA client.

**Table 1.** Attack targets in CISS2022.

| Target * | Description |
|---|---|
| Valve | Open or close motorized valves. |
| Pump | Stop or run a pump. |
| Pressure | Spoof the pressure measurement. |
| Water level | Spoof the water level in a tank. |
| Chemical dosing | Change the amount of chemicals dosed. |
| Historian | Alter the data in the historian; launch DoS attacks on historian server. |

**Table 1.** *Cont.*

| Target * | Description |
| --- | --- |
| HMI/SCADA | Alter the sensor measurements, actuator states at HMI; SCADA DoS attacks on HMI or SCADA. |
| PLC | Alter the control code in a PLC; launch DoS attack on a PLC; alter the commands and values the PLC receives and sends, respectively. |
| RIO/Display | Control the RIO by disconnecting analogue input/output pin. |

* TMultiple components can be targeted in an attack.

During the event, timestamped alerts generated by the iTrust detectors and other IDSs were logged in a separate system, the Alerts Logger. The performance of the detector was validated against the logs from the Alerts Logger and the attacks launched.

### *5.1. Scope of the Fusion Detector*

The Fusion detector is designed to detect process anomalies. The term "process anomalies" refers to a set of actions that deviate from the plant's physical processes and do not conform to its design specifications. Therefore, any attacks that impact the water treatment process of the SWaT will be detected by Fusion detector. However, some passive attacks do not cause process anomalies. An example of such an attack is the crashing of the historian server. Additionally, attacks that cause process anomalies but are removed before detection, such as distributed denial of service (DDoS) attacks, are also not detected by Fusion detector.

### *5.2. iTrust Red Team Exercises*

All attacks launched by the iTrust red team are enumerated in Table 2. Of the 22 successful attacks, 18 were launched remotely via the SWaT network, while the remaining 4 were insider attacks. The insider attacks comprised two physical attacks within SWaT and two attacks that were a combination of physical and network-based methods.

**Table 2.** Attacks launched by the iTrust red team during CISS2022.

| Attack ID | Method | Objective(s) | Target | Description |
| --- | --- | --- | --- | --- |
| 1 | Network | Drain the tank T-401 | MV201 and P401 | Close MV201 and run P401 |
| 2 | Network | Drain the tank T-301 | LIT301 | Set the value of LIT301 to 1200 mm |
| 3 | Network | Damage MV201 | MV201 | Continuously alternate the state of MV201 between open and close |
| 4 | Network | Overflow the tank T301 | P101, MV201, P301 and P302 | Run P101, open MV201, stop both P301 and P302 |
| 5 | Network | Poison the water | P101, MV201, and P203 | Run P101, open MV201 and run P203 |
| 6 | Network | Drain the tank T-602 | LIT602 | Set the value of LIT602 to 800 mm |
| 7 | Network | Drain the tank T-101 | LIT101 | Set the value of LIT101 to 1000 mm |
| 8 | Network | Damage the pump P601 | P601 | Continuously alternate the state of P601 between run and stop |
| 9 | Network | Damage the valve MV304 | MV304 | Continuously alternate the state of MV304 between open and close |
| 10 | Network | Poison the water | AIT202 | Set the AIT202 value to 10 |
| 11 | Network | Poison the water | AIT202, MV201 | Open MV201 and set the value of AIT202 to 10 |
| 12 | Network | Prevent the backwash process | MV301, MV302, and MV303 | Open both MV301 and MV302 and close MV303 |

**Table 2.** *Cont.*

| Attack ID | Method | Objective(s) | Target | Description |
|---|---|---|---|---|
| 13 | Network | Drain the untreated water | MV303 and MV304 | Open both MV303 and MV304 |
| 14 | Network | Drain the untreated water | MV504 | Open MV504 |
| 15 | Network | Stop the water treatment output | MV501 and MV503 | Close MV501 and open MV503 |
| 16 | Network | Stop the water treatment process | LIT401 | Set the value of LIT401 less than 250 mm |
| 17 | Network | Stop the water treatment process | LIT601 | Set the value of LIT601 less than 250 mm |
| 18 | Network/Physical | Drain the tank T101 | T101 (Drain valve), P101, MV201 | Open the drain valve, run P101 and open MV201 |
| 19 | Network/Physical | Drain the tank T101 | T101 (Drain valve), P101, MV201 | Open the drain valve, run P101 and open MV201 |
| 20 | Physical | Drain the tank T301 | T301 (Drain Valve) | Open the drain valve |
| 21 | Physical | Swap sensor values to overflow tanks T301 and T401 | LIT301 and LIT401 | - |
| 22 | Network | Change the setpoints of LIT101 | LIT101 | Change both L and LL values to 700 mm |

Prior to launch, the responsible red team member was asked to specify the attack targets and objectives to the judges. Conversely, the plant operators were responsible for monitoring the plant's behavior while an attack was active to ensure its safety. For example, attack ID 3 involved the attacker continuously toggling the state of valve MV201. Such an attack, when active for an extended period, could reduce the valve's lifespan or damage it. Thus, once the attack succeeded and its impact was observed, the plant operator removed it within a preset time. Similarly, for the tank overflow attack (ID 4), to prevent overflow, the plant operator removed the attack when the water level reached 1250 mm. A ten-minute gap was introduced between each attack to restore normal plant operation and avoid cascading effects.

## 6. Results

In this section, we analyze how well the Fusion detector performs in accurately detecting attacks launched by the iTrust red team during CISS2022 on the operational SWaT system. Additionally, we demonstrate how the Fusion detector effectively explains the underlying semantics of the detected anomalies to plant operators. For comparison, we present the results of commercially available IDSs that participated in the CISS2022 event, without disclosing their names due to the nondisclosure agreement signed between iTrsut and external IDSs.

Data in Table 3 indicate that the Fusion detector successfully detected 19 out of 22 attacks, leaving three attacks undetected. The Fusion detector also generated alerts for each detected attack, providing a clear explanation of its impact on the physical processes in the SWaT together with the components targeted. For instance, consider the single-stage single-point attack on LIT301 (Attack ID: 2) to drain the tank by setting its value to 1200 mm in PLC3. Consequently, valve MV201 remained closed while the outlet pump P301 kept running. The Fusion detector detected the sudden change in the value of LIT301 and generated an alert to notify the plant operator. Similar scenarios apply to attack IDs 6 and 7.

Furthermore, consider the multistage multipoint attack (ID: 22) intended to drain Ttnk T101 by physically opening its drain valve while keeping the valve MV201 open and

running the outlet pump P101. Detecting this attack was a challenging task since the state of the drain valve connected to the tanks in SWaT was not reported to the PLCs. Despite this challenge, the Fusion detector accurately detected the attack since it has learned the relationship between LIT101 and P101. The Fusion detector models the rate at which LIT101 varies when P101 is running. Since the value of LIT101 drops faster due to the opening of the drain valve, the Fusion detector detected the attack. Next, we compare the performance of the Fusion detector with other external detectors that participated in the CISS2022 event.

Next, we discuss the reasons why certain attacks were not detected by the Fusion detector. First, we consider attack ID 1, where the attacker's intention was to drain tank T401 by keeping valve MV302 open and P401 running continuously. It should be noted that tanks T101, T301, and T401 have four water level markers, namely, LL, L, H, and HH, as discussed in the previous section. During normal SWaT operation, MV302 is kept open until the water level in tanks T401 and T301 are above "H" and below "L", respectively. Similarly, P401 continues to run until the water level in tank T401 is above "L". However, when the attack was launched, the actual water level in tank T401 was between "H" and "L". As a result, the water level gradually decreased, but the attack was removed before it reached "L". Although the impact of the attack was realized, the physical process did not deviate from the actual design specification; therefore, no process anomalies existed, which allowed the attack to remain undetected.

Moving on to attack ID 15, the attacker's intention was to stop the flow of treated water into tank T601 by closing the inlet valve MV501 and opening the backwash valve MV503. Under normal circumstances, MV503 is opened and MV501 is closed to initiate the RO backwash process. However, the attacker opened MV503 when there was no backwash, and the attack did not affect the physical process except for changing the valve status. This change in valve status was not detected by the Fusion detector. Finally, attack ID 17 involved the attacker setting the value of LIT601 to less than 250 mm. Prior to the attack, the RO backwash process had been initiated; hence, the water level in tank T601 was already below 250 mm. As the attack did not cause any impact on the physical process, it remained undetected.

Subsequently, we conducted a comparative analysis of the Fusion detector's performance against other commercially available IDSs that participated in the CISS2022 event. As demonstrated in Table 3, the Fusion detector outperformed other IDSs by detecting the maximum number of attacks while generating the least number of false alarms. However, we note that it is not sufficient to evaluate an anomaly detector solely based on its rate of detection rate and false alarms. The ability to explain the semantics of the detected anomalies, especially in terms of the affected plant components, is a critical factor that should also be considered. Such insights are essential for plant operators to rapidly initiate recovery actions. As reported in Table 4, the Fusion detector's alerts for each attack provide a clear explanation of their impact on the physical process of SWaT. In contrast, alerts raised by other IDSs are purely IT-based.

**Table 3.** Performance of anomaly detectors during CISS2022.

| Detector | Detected ** | False Alarms * | Total Alerts |
|---|---|---|---|
| External detector 1 | 16 | NP | 468 |
| External detector 2 | 2 | NP | 5751 |
| External detector 3 | 17 | 7 | 10,823 |
| External detector 4 | 9 | NP | 5899 |
| External detector 5 | 18 | NP | 14,625 |
| Fusion detector | 19 | 3 | 11,466 |

* NP: Information not provided by the detector owner; ** A total of 22 attacks were launched successfully.

**Table 4.** Alerts generated by the Fusion detector in response to the attacks listed in Table 2.

| Attack ID | Detected? | Alerts |
|---|---|---|
| 1 | Not detected | - |
| 2 | Detected | Abnormal change in the LIT301 from 1003 mm to 1200 mm |
| 3 | Detected | Chatter attack on MV201 |
| 4 | Detected | MV201 is OPEN but LIT301 is greater than 1000 mm |
| 5 | Detected | AIT201 is below the L marker |
| 6 | Detected | Abnormal change in the LIT602 from 547 mm to 800 mm |
| 7 | Detected | Abnormal change in the LIT101 from 820 mm to 1000 mm |
| 8 | Detected | Chatter attack on P601 |
| 9 | Detected | Chatter attack on MV304 |
| 10 | Detected | AIT202 is out of range |
| 11 | Detected | AIT202 is out of range |
| 12 | Detected | Both MV301 and MV302 are open |
| 13 | Detected | Both MV303 and MV304 are open |
| 14 | Not Detected | Both MV502 and MV504 are open |
| 15 | Detected | - |
| 16 | Detected | Abnormal change in LIT401 from 834 mm to 250 mm |
| 17 | Not detected | - |
| 18 | Detected | Abnormal change in LIT101 |
| 19 | Detected | Abnormal change in LIT101 |
| 20 | Detected | Abnormal change in LIT301 |
| 21 | Detected | Abnormal change in LIT301 and LIT401 |
| 22 | Detected | MV101 is open but LIT101 is above 800 mm |

For instance, consider attack ID 2. The attacker manipulated the value of LIT301 in PLC3 to 1200 mm via the network. Upon detecting this abnormal change, Fusion detector generated an alert stating **"Abnormal change in LIT301"** . The attack was also detected by external detector 2, which generated the message in Figure 3. Although the generated alert implies abnormal traffic from device 192.168.1.210 to PLC3, it fails to clearly explain to a plant operator the semantics of the detected attack.

**"Sep 20 2022 CEF:0| |CTD|4.5.2|Event/Policy Violation|Policy. Violation|7|src=192.168.1.210 dst=192.168.1.30 smac=00:1a:a0:eb:19:68 shost=nmap dmac=00:1d:9c:c7:b0:70 externalId=559652**
**cat=Integrity/Policy Violation start=Sep 15 2022 02:23:34**
**msg=Violation between (192.168.1.210 → 192.168.1.30),**
**Protocol: CIP, Category: 2, Access: 2 deviceExternalId=CISS cs1Label=SourceAssetType cs1=HMI cs3Label=SourceZone**
**cs3=Endpoint: Rockwell cs4Label=DestZone cs4=PLC: Rockwell cs6Label=CTDlink cs6=https://192.168.4.3/detection/alert/106718-1 cn1Label=IndicatorScore cn1=90 cn2Label=AlertID cn2=106718"**

**Figure 3.** Alert message generated by external detector 2.

In conclusion, the Fusion detector outperforms all commercially available external detectors that participated in the CISS2022 event in terms of rates of detection and false alarms and is able to point to the device that participates in the creation of the anomaly. Devices affected by the cascading effects of an attack are also made visible to the plant operator.

## 7. Discussion

### *7.1. Future of the Fusion Detector*

As part of future enhancement, the threat lookup module of the Fusion detector will be upgraded to operate in a distributed manner to handle multiple data sources. The SWaT system follows the NIST standards, which offer various levels of data extraction and analysis. Implementing this approach will enable early detection of anomalies in sensor measurements or actuator states, thereby preventing the plant from moving towards an abnormal state. For instance, during network-based attacks such as those in CISS2022, an attacker can manipulate sensor measurements and actuator control commands in PLCs by sending invalid or rogue network packets. Analyzing the data at level 1 helps in detecting inconsistencies in sensor measurements or invalid control commands to actuators before they reach the physical devices and cause process anomalies.

Similarly, when a PLC is compromised, as in the CISS2021 attacks [23], invalid control commands from the PLCs can be detected by analyzing the level 0 data to avoid process anomalies. Therefore, the proposed future enhancement can significantly improve the system's ability to detect potential threats and prevent process anomalies.

### *7.2. Benefits of CISS2022*

The CISS event serves as a valuable platform for organizers, participants, and researchers to gain insights into the planning, design, and execution of cyber-attacks on critical infrastructure, particularly those that remain undetected until their objectives are met. Below we list the benefits of CISS2022.

1. Enhanced understanding of the various components that constitute an ICS, as well as their interactions and interdependencies critical to controlling and managing physical processes in the plant.
2. Increased awareness of the potential impact of cyber-physical attacks on ICS components, including any cascading effects.
3. Opportunity for industry and academic partners to evaluate their technologies, identify gaps, and collaborate on ways to improve their security posture, including through benchmarking against external IDSs.
4. Observing the operation of the Fusion detector, a technology underlying the event, may lead to its adoption by plant management seeking to enhance their cyber-resilience.

### *7.3. Research Questions*

The research questions formulated in Section 1 are revisited in the context of the experimental results presented above.

**RQ 1:** *How effective is the Fusion detector in detecting the cyber-physical attacks that have an adverse impact on the physical processes of underlying SWaT*?

As reported in Table 3, theFusion detector was successful in detecting 19 out of the 22 cyber-physical attacks. The remaining three attacks were not detected as they did not have any impact on the physical processes of SWaT. In addition, the Fusion detector generated three false alarms during the event, which were caused by the transition of SWaT from anomalous to normal mode during the attack removal process. Overall, the experimental results demonstrate the ability of the Fusion detector to accurately detect cyber-physical attacks that lead to process anomalies.

**RQ 2:** *How efficient is the Fusion detector in describing threats and supporting the plant operators towards active incident response*?

The Fusion detectorexhibits a notable advantage in comparison to the other existing IDSs that participated in CISS2022, which is its ability to explain the semantics of the detected anomalies. As reported in Table 4, for each detected attack, the Fusion detector can explain its impact on the physical process of the SWaT, along with the targeted components. Thus, it is evident that the Fusion detector can accurately localize the anomalies and identify

and report the components that are being targeted. As a result, the plant operator can initiate recovery measures promptly, before the attacker is successful in achieving their objectives.

## 8. Conclusions

There are plenty of technologies such as firewalls and IDSs available to safeguard ICS from cyber-attacks. However, the methodology, named Fusion detector, used in this work is different from commercially available firewalls and IDSs. The Fusion detector integrates design knowledge and machine learning algorithms to operate over the physics that govern the water treatment processes of SWaT. Therefore, one could consider the Fusion detector as the last line of defense. The traces of the attacks launched by the iTrust red team during the CISS2022 event in both IT and OT traffic are publicly available. Other researchers can use them to assess the effectiveness of their attack detection mechanisms. From the study reported here, the Fusion detector was found to be effective in detecting all cyber-physical attacks that lead to process anomalies, compared with other commercially available IDSs. The Fusion detector supports the plant operator in localizing the anomalies, identifying the area of impact, and suggesting recovery actions to restore normal plant operation, as observed from Table 4. As discussed in Section 4, the threat look module comprises two anomaly detection methodologies, namely, DAD and AICrit. It is interesting to observe that there are a few attacks that are detected only by either DAD or AICrit. For instance, the attack on P601 and MV304 (attack IDs 8 and 9 in Table 2) are detected only by DAD, and, similarly, the attack on LIT601 (attack ID 17 in the Table 2) is detected only by AICrit.

As a fusion detector integrates both data-centric and design-centric anomaly detection approaches, when applied in industrial environments, these methods may independently trigger alerts. However, a significant limitation of this work is the challenge of distinguishing between genuine attacks and false alarms, prompting an open-ended research question: What actions should an operator take when only one detection method raises an alert?

**Author Contributions:** Conceptualization, G.R.; Methodology, G.R. and A.M.; Validation, G.R.; Writing—original draft, G.R.; Writing—review & editing, A.M.; Supervision, A.M.; Funding acquisition, A.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is supported in part by the National Research Foundation, Singapore, under its National Satellite of Excellence Programme "Design Science and Technology for Secure Critical Infrastructure: Phase II" (Award No: NRF-NCR25-NSOE05-0001). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore.

**Data Availability Statement:** Data is available on request

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

## References

1. Adepu, S.; Mathur, A. Distributed Attack Detection in a Water Treatment Plant: Method and Case Study. *IEEE Trans. Dependable Secur. Comput.* **2021**, *18*, 86–99. [CrossRef]
2. MR, G.; Somu, N.; Mathur, A. A multilayer perceptron model for anomaly detection in water treatment plants. *Int. J. Crit. Infrastruct. Prot.* **2020**, *31*, 100393. Available online: https://www.sciencedirect.com/science/article/pii/S1874548220300573 (accessed on 1 February 2024).
3. Raman, M.R.G.; Mathur, A.P. A Hybrid Physics-Based Data-Driven Framework for Anomaly Detection in Industrial Control Systems. *IEEE Trans. Syst. Man Cybern. Syst.* **2021**, *52*, 6003–6014. [CrossRef]
4. M.R., G.R.; P. Mathur, A. AICrit: A unified framework for real-time anomaly detection in water treatment plants. *J. Inf. Secur. Appl.* **2022**, *64*, 103046. [CrossRef]
5. Goh, J.; Adepu, S.; Junejo, K.N.; Mathur, A. A dataset to support research in the design of secure water treatment systems. In Proceedings of the International Conference on Critical Information Infrastructures Security, Paris, France, 10–12 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 88–99.
6. Gauthama Raman, M.; Dong, W.; Mathur, A. Deep autoencoders as anomaly detectors: Method and case study in a distributed water treatment plant. *Comput. Secur.* **2020**, *99*, 102055. [CrossRef]

7. Adepu, S.; Brasser, F.; Garcia, L.; Rodler, M.; Davi, L.; Sadeghi, A.R.; Zonouz, S. Control behavior integrity for distributed cyber-physical systems. In Proceedings of the 2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPS), Sydney, Australia, 22–24 April 2020; IEEE: New York, NY, USA, 2020; pp. 30–40.
8. Cheng, L.; Tian, K.; Yao, D.D. Orpheus: Enforcing Cyber-Physical Execution Semantics to Defend Against Data-Oriented Attacks. In Proceedings of the 33rd Annual Computer Security Applications Conference (ACSAC '17), Orlando, FL, USA, 4–8 December 2017; pp. 315–326. [CrossRef]
9. Urbina, D.; Giraldo, J.; Tippenhauer, N.O.; Cardenas, A. Attacking fieldbus communications in ICS: Applications to the SWaT testbed. In Proceedings of the Singapore Cyber-Security Conference (SG-CRC), Singapore, 14–15 January 2016; IOS Press: Amsterdam, The Netherlands, 2016; pp. 75–89.
10. Ghaeini, H.R.; Antonioli, D.; Brasser, F.; Sadeghi, A.R.; Tippenhauer, N.O. State-aware anomaly detection for industrial control systems. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing, Pau, France, 9–13 April 2018; pp. 1620–1628.
11. Mathur, A. SafeCI: Avoiding process anomalies in critical infrastructure. *Int. J. Crit. Infrastruct. Prot.* **2021**, *34*, 100435.
12. Truong, H.T.; Ta, B.P.; Le, Q.A.; Nguyen, D.M.; Le, C.T.; Nguyen, H.X.; Do, H.T.; Nguyen, H.T.; Tran, K.P. Light-weight federated learning-based anomaly detection for time-series data in industrial control systems. *Comput. Ind.* **2022**, *140*, 103692. [CrossRef]
13. Tang, C.; Xu, L.; Yang, B.; Tang, Y.; Zhao, D. GRU-Based Interpretable Multivariate Time Series Anomaly Detection in Industrial Control System. *Comput. Secur.* **2023**, *127*, 103094. [CrossRef]
14. Das, T.K.; Adepu, S.; Zhou, J. Anomaly detection in industrial control systems using logical analysis of data. *Comput. Secur.* **2020**, *96*, 101935. [CrossRef]
15. Wu, W.; Song, C.; Zhao, J.; Xu, Z. Physics-informed gated recurrent graph attention unit network for anomaly detection in industrial cyber-physical systems. *Inf. Sci.* **2023**, *629*, 618–633. [CrossRef]
16. L(y)u, S.; Wang, K.; Zhang, L.; Wang, B. Process-Oriented heterogeneous graph learning in GNN-Based ICS anomalous pattern recognition. *Pattern Recognit.* **2023**, *141*, 109661. [CrossRef]
17. Mathur, A.P.; Tippenhauer, N.O. SWaT: A water treatment testbed for research and training on ICS security. In Proceedings of the 2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater), Vienna, Austria, 11 April 2016; pp. 31–36.
18. CISS2022-OL. Critical Infrastructure Security Showdown 2021—Online (CISS2022-OL) Technical Report. 2023. Available online: https://itrust.sutd.edu.sg/ciss-2022/ (accessed on 14 January 2023).
19. Mathur, A.P. Reconfigurable Digital Twin to Support Research, Education, and Training in the Defense of Critical Infrastructure. *IEEE Secur. Priv.* **2023**, *21*, 51–60. [CrossRef]
20. MR, G.; Ahmed, C.; Mathur, A. Machine learning for intrusion detection in industrial control systems: Challenges and lessons from experimental evaluation. *Cybersecurity* **2021**, *4*, 27.
21. Ahmed, C.; MR, G.; Mathur, A. Challenges in machine learning based approaches for real-time anomaly detection in industrial control systems. In Proceedings of the 6th ACM on Cyber-physical System Security Workshop, Taipei, Taiwan, 6 October 2020; pp. 23–29.
22. Chakraborty, M.; Kundan, A.P. Grafana. In *Monitoring Cloud-Native Applications: Lead Agile Operations Confidently Using Open Source Software*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 187–240.
23. CISS2022-OL. Critical Infrastructure Security Showdown 2021—Online (CISS2021-OL) Technical Report. 2022. Available online: https://itrust.sutd.edu.sg/ciss/ciss-2021-ol/ (accessed on 22 May 2022).


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.