Security through Digital Twin-Based Intrusion Detection: A SWaT Dataset Analysis

Mehmet Bozdal
Electrical and Electronics Engineering
Abdullah Gül University
Kayseri, Türkiye
0000-0002-2081-7101

2023 16th International Conference on Information Security and Cryptology (ISCTürkiye) | 979-8-3503-9399-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISCTrkiye61151.2023.10336137

Digital twin, as a virtual replica of physical entity, offer valuable insights into Industrial Control System (ICS) behavior and characteristics. Leveraging the convergence of digital twins and cybersecurity, this research explores its role in securing critical infrastructure, using the Secure Water Treatment (SWaT) system as a case study. Existing intrusion detection systems (IDS) for SWaT encounter challenges related to requiring huge amounts of a dataset for training, being unable to adopt high data dimensionality, and adaptability to emerging threats. To address these issues, a hybrid digital twin model is proposed, combining physics-based models and data-driven approaches. This model facilitates precise attack localization and explainable IDS outcomes. The method exhibits promising capabilities for enhancing critical infrastructure security and adapting to evolving cyber threats. Experimental results demonstrate the ability to detect eight out of nine attack types.

Keywords—intrusion detection, digital-twin, cybersecurity.

I. INTRODUCTION

The adoption of the Internet of Things (IoT) technology has revolutionized industrial control systems, enabling seamless communication, real-time data sharing, and synchronized operations. This transformation generates massive amounts of data. This data presents new opportunities for leveraging big data analytics and artificial intelligence techniques to extract valuable insights. One emerging concept that harnesses this potential is the digital twin.

A digital twin is a virtual representation or replica of a physical entity, such as a product, process, or system[1]. It encompasses various aspects, including its structure, behavior, and characteristics. Initially conceived in the manufacturing industry to model and simulate complex processes, the concept has expanded to diverse domains including healthcare, transportation, energy, and infrastructure. Due to its potential to improve efficiency, reliability, and safety.

In recent years, the convergence of digital twin technology with cybersecurity has opened up new possibilities for safeguarding systems against cyber threats. As digital twin captures and models the behavior of their physical counterpart, it can serve as a powerful tool for intrusion detection and security enhancement. By monitoring and analyzing the digital twin's data deviations from normal behavior can be identified, signaling potential cyber-attacks or security breaches.

Cyber-attacks have become a significant concern for industrial processes and critical infrastructure, with the potential to cause severe damage and catastrophic consequences. Recent years have seen numerous attacks on industrial control systems, such as ShadowPad [2] and Stuxnet [3], as well as critical infrastructure, including railways [4], irrigation systems [5], and power grids [6]. These attacks highlight the urgency of implementing effective countermeasures.

Addressing these concerns, this research investigates the application of digital twin to enhance security, specifically as an Intrusion Detection System (IDS), with a focus on the Secure Water Treatment (SWaT) [7] system. The SWaT system, developed by the iTrust research group, is a scaled-down representation of a real-world water treatment facility. Serving as a vital testbed, it mirrors the operational dynamics of a water treatment facility.

Yazdinejad et al. [9] proposed Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) to understand the long-term dependencies of the SWaT system. The proposed method demonstrated enhanced detection capabilities compared to conventional machine learning algorithms. However, a notable drawback is a substantial increase in training time due to the complexity of RNN and LSTM architectures. Furthermore, the supervised methods [10], including Binary Logistic Regression (BLR) [11], and Logical Analysis of Data [12], encounter challenges in obtaining realistic and diverse attack datasets for training.

Additionally, some of these supervised learning methods are limited to detecting only known attacks, impeding their ability to identify new and emerging threats. Consequently, these IDSs require substantial amounts of data to be effectively trained, and the complexity of generating accurate datasets that represent real-world cyber-attacks hampers their overall effectiveness.

Inoue et al. [13] introduced a machine-learning approach utilizing deep neural networks for intrusion detection. One of the notable advantages of their method is that it operates in an unsupervised manner, meaning it only requires normal data without any labeled attack samples for training. However, the detection performance was low with an F score of 0.80281. This is the result of the method being insensitive to subtle variations in data and actuators which have on-and-off situations.

Similarly, Boateng et al. [14] presented an unsupervised one-class neural network. The method offers the advantage of having very few hyperparameters, making it easier to train and tune the model. Despite its computational simplicity, it achieves a good detection performance. However, it should be noted that the overall performance is not as high as that of supervised deep learning methods. Additionally, the method exhibits some detection delay in certain attack scenarios.

Securing the SWaT system presents unique challenges due to its complexity and high-dimensional data, like many industrial control systems. Machine learning approaches often struggle to handle such data efficiently, encountering the curse of dimensionality, leading to increased computational complexity, longer training times, and diminished performance. Although dimensionality reduction techniques

exist, research [8] indicates that they can negatively impact performance on the SWaT dataset. These limitations highlight the need for more robust and adaptable approaches to securing industrial control systems.

Digital twin technology has emerged as a promising avenue for bolstering security in smart grids [15] and industrial control systems [16]. Xu et al. [17], for instance, have leveraged the power of digital twin by employing curriculum learning on the SWaT dataset. Their data-driven implementation eliminates the need for domain knowledge, but the training phase can become complex. As the digital twin model is constructed using a timed automaton that relies on the timing of system state changes, it can understand only known behavior from the training data. Therefore, its evolution continues with real-time data, allowing for adaptability to new scenarios and potential cyber threats. A limitation of this method is its inability to localize anomalies.

Cheong et al. [18] presented a valuable contribution by developing a physics-aware system that utilizes system state changes and considers state dependency. While their method shares a similar objective with our research, it involves a laborious task of extracting interdependent state variables. In contrast, our proposed hybrid digital twin architecture effectively models the system dynamics without the need for considering interdependent variables. The digital twin rules are derived directly from the automation system, which minimizes any significant overhead if the digital twin is designed during the physical twin's development. Furthermore, the digital twin approach allows for a more granular attack localization narrowed down to the component level rather than the stage level.

This paper introduces a novel IDS for the SWaT dataset, employing a hybrid digital twin. The hybrid approach combines the fundamental principles of the physics model with valuable insights from real-world data, creating a more comprehensive and coherent representation of the system's dynamics. The main contributions of this research are as follows:

- Enhanced anomaly detection by leveraging inherent system dynamics and real-world data insights.
- The granular attack localization feature allows for pinpointing security threats at the component level.
- Real-time monitoring capabilities for swift detection and response to potential cyber threats.

The paper is organized as follows: Section II introduces the SWaT dataset followed by the methodology, describing the construction of the digital twin and intrusion detection system in Section III. Section IV discusses the results and future directions. Finally, Section V concludes the paper, by highlighting the significance of the hybrid digital twin for securing industrial control systems.

II. SWaT Dataset

The SWaT dataset is a time-series dataset that spans 11 days and consists of 51 attributes. These attributes encompass 26 continuous values that represent sensor values, alongside 25 actuators with discrete states, including "on," "off," and transitional states. The dataset contains vital information about various process variables, including water flow rates, pH levels, temperature, pressure, and valve positions, all sampled at one-second intervals. The dataset is categorized into two distinct splits: one devoid of any attacks and the other one has 36 attack implementations. It consists of 890,298 normal samples and 54,621 attack samples, revealing an inherently unbalanced distribution.

The water treatment system comprises six staged processes: P1 for receiving and storing raw water, P2 for adding chemicals to improve water quality, P3 for ultra-filtration, P4 for dechlorination using ultraviolet lamps, P5 for reverse osmosis filtration, and P6 for storing and distributing water.

Table I provides an overview of the attacks targeting P1, encompassing various types, such as Single Stage Single Point Attacks (1, 2, 3, 34, 36), Single Stage Multi Point Attacks (21, 35), Multi-Stage Single Point Attacks (26), and Multi-Stage Multi Point Attacks (30). As the digital twin representation is currently limited to the first stage of the SWaT system, the evaluation of the proposed digital twin-based IDS is focused solely on the effect of stage one attacks.

TABLE I. Attacks Targeting P1

No	Start Time	End Time	Attack Point	Start State	Attack
1	28/12/2015 10:29:14	10:44:53	MV101	MV101 is off	Open MV101
2	28/12/2015 10:51:08	10:58:30	P102	P101 is on and P102 is off	Open P102
3	28/12/2015 11:22:00	11:28:22	LIT101	Water level between L and H	+ 1 mm every second
21	29/12/2015 18:30:00	18:42:00	MV101 & LIT101	MV101 is on; LIT101 between L and H	Open MV101 & set LIT101 to 700 mm
26	30/12/2015 17:04:56	17:29:00	P101, LIT301	P101 is off; P102 is on; Stage 3	Open P101
30	31/12/2015 15:47:40	16:07:10	LIT101 & P101 & MV101	P-101 is off; MV-101 is off; Stage 2	Open P101, MV101 & set LIT101 to 700 mm
34	1/01/2016 17:12:40	17:14:20	P101	P-101 is on	Turn P101 off
35	1/01/2016 17:18:56	17:26:56	P101 & P102	P-101 is on; P-102 is off	Turn P101 off & keep P102 off
36	1/01/2016 22:16:01	22:25:00	LIT101	Water level between L and H	Set LIT101 to less than LL

III. Methodology

A. Tank Dynamics

The first stage of the SwaT is modeled according to its Piping and Instrumentation Diagram(P&ID) [19]. A simplified version of the system is presented in Figure 1. The tank system consists of a single-compartment T-101 with a capacity of 1800 m^3, a height of 1.36 m, and a diameter of 1.38 m. It is controlled by a motorized valve (MV101) to feed the tank and a pump (P101) to drain the tank. The system includes

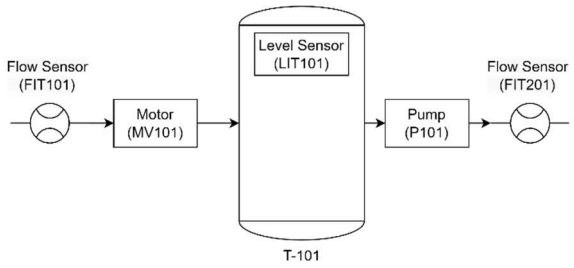

Fig. 1. Simplified overview of the initial stage in the SWaT system.

level sensor LIT101 to monitor the tank level, and flow sensors FIT101 and FIT201 to measure the inflow and outflow rates, respectively.

The tank dynamics are governed by the principle of mass conservation, where the difference between inflow and outflow rates determines the rate of level change. When the inflow rate exceeds the outflow rate, the tank level increases, and conversely, when the outflow rate surpasses the inflow rate, the tank level decreases. The rate of the water level change in the tank ($dH(t)/dt$) can be expressed mathematically as follows:

$$dH(t)/dt = (Inflow(t) - Outflow(t)) * (1/A) \quad (1)$$

where $H(t)$ represents the water level in the tank at time t, A is the cross-sectional area of the tank, $Inflow(t)$ and $Outflow(t)$ denotes the rate of water flowing into and out of the tank at time t, respectively.

The physics model also considers the different states of the MV101 and P101, directly influencing the inflow and outflow rates, respectively. In specific scenarios, the following outcomes arise:

- When MV101 is off and P101 is off, there is no inflow or outflow, resulting in $dH(t)/dt = 0$.
- When MV101 is off and P101 is on, only water flows out of the tank, influencing the change in water level based on the outflow rate.
- When MV101 is on and P101 is off, the change in water level is determined solely by the inflow rate.
- When both MV101 and P101 are on, the change in water level is determined by the difference between the inflow and outflow rates, as expressed in (1).

While the physics model can simulate the behavior of the tank, real-world scenarios often introduce additional factors and uncertainties, such as motor transitions and leakage, leading to discrepancies between the model's predictions and observed data. Over time, the error accumulates and results in deviation of the model from the actual system.

To address these discrepancies and enhance the accuracy of the model, a data-driven approach is applied in conjunction with the physics model. Equation (1) is reformulated as (2), incorporating calibration and scaling factors:

$$dH(t)/dt = (Inflow(t) - Outflow(t) * Calibration\ Factor) * Scaling\ Factor \quad (2)$$

Although the calibration factor directly impacts the outflow rate, it effectively equalizes any discrepancies between the inflow and outflow sensors. This ensures that the model accurately represents the real-world behavior of the system by fine-tuning the dynamics to match observed data. On the other hand, the scaling factor is responsible for proportionally adjusting the overall change in the water level. It allows the model to be appropriately scaled to match the physical system's characteristics.

To determine the parameter values for the calibration and scaling factors, the hybrid digital twin model is fed with physical twin data without any attacks (SWaT Data Normal). Only the first 25,000 data samples are used to obtain calibration and scaling factors. The data includes one transition state and two normal steady states of the system. Thanks to the physics model, the system can be fine-tuned effectively with limited data.

The convergence between the digital twin and physical twin is assessed with a correlation coefficient. The correlation coefficient serves as a measure to assess the similarity, quantifying the strength of the linear relationship between the variables on a scale ranging from -1 to 1. The physics-only model, utilizing (1), achieves a correlation coefficient of 0.8241, and the hybrid model, implementing (2), significantly outperforms with a correlation coefficient of 0.9999. By combining the fundamental principles of the physics model with insights gained from real-world data, the hybrid model offers a more comprehensive and coherent representation of the system's dynamics. The exceptional correlation in the hybrid model validates its capability to capture and represent the dynamics of the physical twin.

B. Digital Twin of Control Logic

The dynamics of the tank ensure that sensory data remains unmodified. However, the security of the system is not limited to preventing unauthorized modification to sensory data; control logic can also be targeted by hacking the programmable logic controller. To accurately simulate and respond to these attack scenarios, the digital twin should also incorporate the control logic of the system.

Consider a scenario where a system with a predefined high-level setting of 800m for the water level. In an attack scenario, MV101 (a valve) is compromised and causes an influx of water into the system even when the water level is already at the high limit. As a result, the compromised MV101 fails to respond appropriately, allowing the tank to overflow. In such a situation, the digital twin should accurately reflect the overflow that would occur in the physical system. To ensure the system operates within desired parameters, appropriate checks and control mechanisms must be implemented within the digital twin, just as they would exist in the physical system.

In the first stage of the SWaT system, crucial controls are in place for MV101 and P101 based on the water level. When the water level is below the low level, P101 should be turned off to prevent tank underflow and damage to the pump. Conversely, when the water level is high, MV101 should be turned off to avoid potential overflow. Additionally, when P101 from stage one is turned on, P102 in stage two should also be activated to maintain the proper water flow and prevent water pipe bursts. These control rules are already pre-coded within the Programmable Logic Controllers (PLCs) of the system, ensuring the proper functioning and safety of the water treatment process. Therefore, digital twin will not require new design considerations.

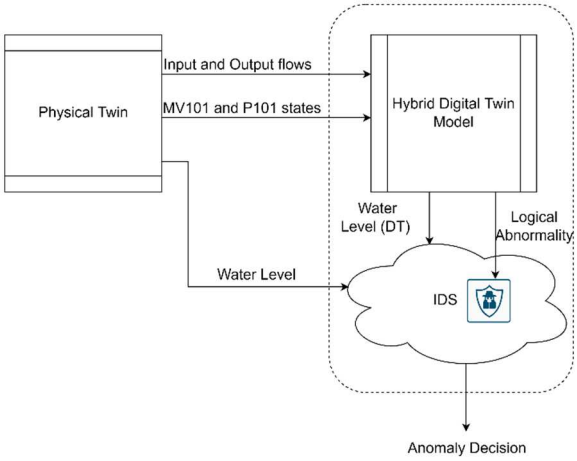

Fig. 2. The intrusion detection capability of the digital twin and data flow.

C. Intrusion Detection

The integration of tank dynamics and control logic in the digital twin creates a comprehensive and accurate virtual replica of the first stage of the SWaT system. As illustrated in Figure 2, the physical system provides sensory data and actuator information to the digital twin. The digital twin processes this information based on tank dynamics and control logic. The outputs of both the physical and digital twins are then assessed by an IDS. To demonstrate the digital twin's security capabilities, a simple thresholding mechanism is implemented.

The intrusion detection capability of the digital twin is leveraged by comparing the water level output of the model with the physical water level using thresholding. A predetermined relationship is established between the real water level value and the digital twin's water level output. If the deviation between these two values exceeds the threshold of experimentally determined 2%, it is flagged as an anomaly, indicating a potential attack on the system. Additionally, the digital twin employs control logic rules acquired from process design behavior and safety considerations to verify the system's specifications. It ensures that the control logic in the digital twin is identical to that of the physical system. Each situation is assigned an anomaly score based on a geometric sequence, which starts with a score of 1 and has a common ratio of ½. These scores are then aggregated for each occurrence of an anomaly. This approach enables easy decomposition of the total value, providing the ability to precisely pinpoint the specific location of anomalies within the system.

IV. Results & Discussions

A. Results

The hybrid digital twin with IDS capability was tested on both the normal and attacked datasets. As the digital twin for the SWaT system is currently only implemented for the first stage, the focus of intrusion detection is limited to attacks that specifically affect this stage as summarized in Table I. The first three of these attacks and anomaly scores are presented in Figure 3.

The first attack involves the malicious opening of MV-101, even though the water level is already high. This attack is successfully detected by the digital twin of the control logic,

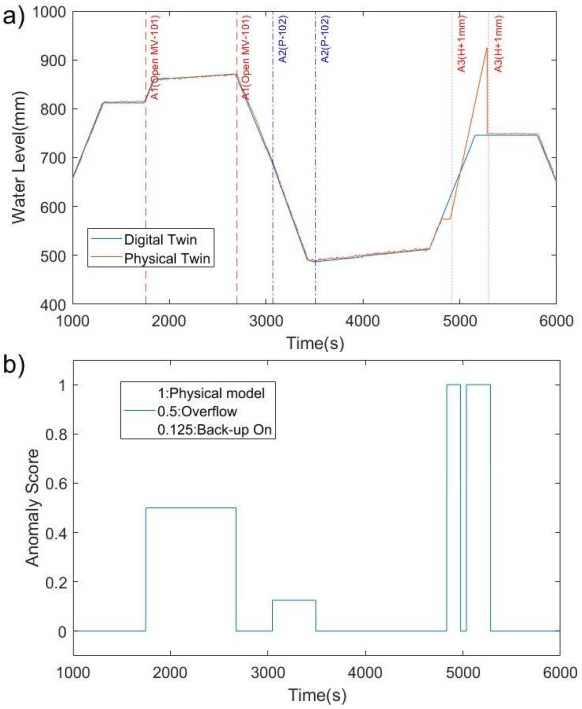

Fig. 3. **a)**The first three attacks on the SWaT dataset and **b)** the anomaly score output of the digital twin-based intrusion detection system.

as it identifies the inconsistency in the system's behavior. Similarly, the second attack, where the pipes between the first and second stages are burst by pumping water at the same time through P101 and backup pump P102, is also detected by the digital twin of control logic in the same manner. The third attack aims to underflow the tank and damage pump P101 by gradually increasing the water level by 1 mm every second. This attack is effectively detected by the digital twin's IDS capabilities, which observe the deviation of the digital twin model from the reported anomalous physical value.

Each of these attacks is assigned a specific anomaly score, allowing for precise localization of the attack points. The proposed IDS leverages the geometric sequence value for each attack and aggregates individual anomaly scores. Consequently, the IDS can even highlight multiple attack points, providing enhanced detection capabilities for identifying complex attack scenarios.

Table II below presents the confusion matrix for the IDS outcome for each attack type. The IDS demonstrates accurate detection capabilities with high TP and low FP for most attacks, achieving Recall and Precision values of 1 for attacks 2, 26, and 34.

However, for attack 3, the IDS shows 83 FP. This is due to the initial behavior of the attack, where the water level was lower than the digital twin (DT). As the attack progresses and the water level rises, the DT eventually converges with the physical twin, leading to a temporary lack of warning. However, the DT subsequently detects the attack as the water level continues to rise.

For attacks 21, 30, and 36, which involve manipulating the water level, the recall values are low. This is partially attributed to the periodic synchronization of the DT with the physical twin, causing the DT to converge with the system's behavior. If significant water level changes occur during an

attack, the DT may deviate, affecting performance metrics but still detecting the attack's start. While this impacts the recall value, it does not hinder real-life effectiveness, as the DT successfully identifies the attack starts, allowing supervisors to take appropriate actions and sync the DT with the physical twin accordingly.

The IDS fails to detect attack 35, where the system's outflow is stopped by turning P101 off. This attack scenario occurred while the system was in a stable state, thereby leading to the absence of any anomalies signaled by the digital twin. If continuous water outflow is anticipated within the system, the digital twin can be suitably modified to address this scenario.

TABLE II. CONFUSION MATRIX FOR THE PROPOSED IDS OUTCOME

Attack No	TP	FP	FN	Recall	Precision
1	925	0	18	0.9809	1
2	444	0	0	1	1
3	449	83	0	1	0.8440
21	720	0	2299	0.2385	1
26	1643	0	0	1	1
30	1161	0	3792	0.2344	1
34	98	0	0	1	1
35	0	480	0	0	0
36	515	0	4737	0.0981	1

TABLE III. RECALL PERFORMANCE COMPARISON WITH PRIOR RESEARCH

Attack No	SVM [13]	OC [14]	1D-CNN [20]	DIF [21]	This paper
1	0	0	**0.99**	0.01	0.9809
2	0	0.08	**1.00**	0.29	**1.00**
3	0	0.59	0.23	1.00	**1.00**
21	0.0167	0.32	**1.00**	0.17	0.2385
26	0	**1.00**	0.30	**1.00**	**1.00**
30	0.003	**1.00**	0	0.95	0.2344
34	0	**1.00**	0.91	**1.00**	**1.00**
35	0	**1.00**	1.00	**1.00**	0
36	0.119	**1.00**	0.64	0.63	0.0981

Table III presents a comparative analysis between our proposed method and established intrusion detection techniques, namely Support Vector Machine (SVM) [13], One-Class Neural Network[14], 1D-CNN[20], and Dual Isolation Forest (DIF) [21]. Despite our method being a work in progress, it demonstrates competitive performance through a straightforward thresholding approach.

The performance of the proposed method on the normal dataset is also evaluated in terms of the false positive rate. Among the 495,000 samples in the dataset, only 923 were labeled as anomalous. However, upon closer inspection of the data, it was observed that the majority of false positives (920 out of 923) occurred during the initial stages of the system when the tank was filling. This particular portion of the data is often excluded from most IDS testing due to the system's instability during this phase. However, one of the advantages of the digital twin employed in the proposed method is it can precisely capture the complete spectrum of system behavior, including transition states. As most of the false positives are related to safety warnings triggered by the system's instability during initialization, the supervisor can confidently disregard safety-related warnings during the tank initialization phase.

B. Discussions and Future Directions

The physics-based model, built on domain knowledge and fundamental principles, ensures a strong foundation for accurately simulating the system's behavior. However, it relies on precise system design information, which may not always be readily available or difficult to measure, introducing uncertainties in the model's accuracy. Therefore, digital twin should be built early design stages of the physical counterpart to easily adapt the design knowledge.

On the other hand, the data-driven approach excels at handling uncertainties and gaps in system knowledge by leveraging real-world data. It creates a model purely based on data, reducing the reliance on explicit knowledge of all system parameters. This adaptability allows the data-driven approach to capture complex interactions that may not be fully represented in the physics-based model. However, it requires extensive data, and it is not always possible to gather datasets with anomalies. Additionally, integrating data from diverse sources may cause data quality and compatibility issues.

The hybrid digital twin overcomes these limitations by integrating both approaches. It offers attack localization and explainability of the IDS decisions, allowing supervisors to take appropriate actions and understand the nature of attacks for preventative measures. Moreover, the virtual environment of the hybrid digital twin allows for iterative improvements and experimentation without risking the integrity of the physical system.

For future directions, research efforts should be directed towards automating the creation of physics-based models to reduce the dependency on system design information and make the process more efficient. Exploring advanced machine learning techniques to optimize and combine both the physics-based and data-driven approaches can lead to even more accurate and efficient digital twins.

It is crucial to acknowledge the potential vulnerabilities of the digital twin itself in the context of cybersecurity. For instance, attackers could manipulate the digital twin by providing it with stable historical data, making it believe that the physical system is secure when it is actually under attack. This highlights the importance of ensuring secure communication between the digital twin and the physical system to prevent such manipulations.

Furthermore, the digital twin contains valuable information, including system design, intellectual properties, and sensitive data. If attackers gain access to this information, they can exploit it to bypass security measures or devise more targeted attacks. Therefore, it is essential to implement robust access control mechanisms and obfuscation techniques to safeguard the digital twin from unauthorized access.

V. CONCLUSION

This research presents a novel and comprehensive approach to enhancing the security of industrial control systems through the implementation of a hybrid digital twin-

based IDS. The hybrid digital twin integrates both physics-based modeling and data-driven techniques, offering a dynamic and accurate virtual replica of the physical system.

The digital twin-based IDS can even highlight multiple attack points, providing enhanced detection capabilities for identifying complex attack scenarios. This granular attack localization and multi-point detection contribute to the effectiveness of the proposed method in safeguarding the SWaT system against potential cyber threats.

Experimental evaluation of the SWaT dataset demonstrates the effectiveness of the hybrid digital twin-based IDS in detecting various attack scenarios. While it successfully identifies attacks on the first stage of the SWaT system, it is essential to extend the implementation to cover all stages for comprehensive intrusion detection and proper comparison with other state-of-the-art methods. Additionally, we aim to enhance the digital twin's capabilities further by integrating advanced smart algorithms, moving beyond simple thresholding techniques.

REFERENCES

[1] C. Alcaraz and J. Lopez, "Digital Twin: A Comprehensive Survey of Security Threats," *IEEE Commun. Surv. Tutorials*, vol. 24, no. 3, pp. 1475–1503, 2022, doi: 10.1109/COMST.2022.3171465.

[2] N. Evancich and J. Li, "Attacks on Industrial Control Systems," 2016, pp. 95–110.

[3] D. Kushner, "The real story of stuxnet," *IEEE Spectr.*, vol. 50, no. 3, pp. 48–53, 2013, doi: 10.1109/MSPEC.2013.6471059.

[4] T. Balmforth, "Belarusian group claims hack on railway system after Russian troop moves | Reuters," Jan. 25, 2022. https://www.reuters.com/legal/litigation/belarusian-group-claims-hack-railway-system-after-russian-troop-moves-2022-01-24/ (accessed Jul. 21, 2023).

[5] E. Kovacs, "Irrigation Systems in Israel Disrupted by Hacker Attacks on ICS - SecurityWeek," *SecurityWeek.Com*, Apr. 13, 2023. https://www.securityweek.com/irrigation-systems-in-israel-disrupted-by-hacker-attacks-on-ics/ (accessed Jul. 21, 2023).

[6] R. M. Lee, M. J. Assante, and T. Conway, "Analysis of the Cyber Attack on the Ukrainian Power Grid Defense Use Case," *Electr. Inf. Shar. Anal. Cent.*, p. 36, 2016, Accessed: Jan. 11, 2023. [Online]. Available: https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf.

[7] J. Goh, S. Adepu, K. N. Junejo, and A. Mathur, "A Dataset to Support Research in the Design of Secure Water Treatment Systems," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10242 LNCS, 2017, pp. 88–99.

[8] M. Bozdal, K. Ileri, and A. Ozkahraman, "Comparative analysis of dimensionality reduction techniques for cybersecurity in the SWaT dataset," *J. Supercomput.*, pp. 1–21, Jul. 2023, doi: 10.1007/s11227-023-05511-w.

[9] A. Yazdinejad, M. Kazemi, R. M. Parizi, A. Dehghantanha, and H. Karimipour, "An ensemble deep learning model for cyber threat hunting in industrial internet of things," *Digit. Commun. Networks*, vol. 9, no. 1, pp. 101–110, Feb. 2023, doi: 10.1016/j.dcan.2022.09.008.

[10] K. N. Junejo and D. Yau, "Data driven physical modelling for intrusion detection in cyber physical systems," *Cryptol. Inf. Secur. Ser.*, vol. 14, pp. 43–57, 2016, doi: 10.3233/978-1-61499-617-0-43.

[11] C. Charilaou, C. I. Ioannou, and V. Vassiliou, "System for Operational Technology Attack Detection in Industrial IoT," in *20th Mediterranean Communication and Computer Networking Conference (MedComNet)*, Jun. 2022, pp. 84–93, doi: 10.1109/MedComNet55087.2022.9810446.

[12] T. K. Das, S. Adepu, and J. Zhou, "Anomaly detection in Industrial Control Systems using Logical Analysis of Data," *Comput. Secur.*, vol. 96, p. 101935, 2020, doi: 10.1016/j.cose.2020.101935.

[13] J. Inoue, Y. Yamagata, Y. Chen, C. M. Poskitt, and J. Sun, "Anomaly detection for a water treatment system using unsupervised machine learning," in *IEEE International Conference on Data Mining Workshops, ICDMW*, 2017, vol. 2017-Novem, pp. 1058–1065, doi: 10.1109/ICDMW.2017.149.

[14] E. Aboah Boateng, J. W. Bruce, and D. A. Talbert, "Anomaly Detection for a Water Treatment System Based on One-Class Neural Network," *IEEE Access*, vol. 10, pp. 115179–115191, 2022, doi: 10.1109/ACCESS.2022.3218624.

[15] W. Danilczyk, Y. L. Sun, and H. He, "Smart Grid Anomaly Detection using a Deep Learning Digital Twin," in *52nd North American Power Symposium (NAPS)*, Apr. 2021, pp. 1–6, doi: 10.1109/NAPS50074.2021.9449682.

[16] M. Eckhart and A. Ekelhart, "Towards Security-Aware Virtual Environments for Digital Twins," in *Proceedings of the 4th ACM Workshop on Cyber-Physical System Security*, May 2018, pp. 61–72, doi: 10.1145/3198458.3198464.

[17] Q. Xu, S. Ali, and T. Yue, "Digital Twin-based Anomaly Detection with Curriculum Learning in Cyber-physical Systems," *ACM Trans. Softw. Eng. Methodol.*, vol. 1, Feb. 2023, doi: 10.1145/3582571.

[18] D. Cheong Lien Sung, G. R. M.R., and A. P Mathur, "Design-knowledge in learning plant dynamics for detecting process anomalies in water treatment plants," *Comput. Secur.*, vol. 113, p. 102532, Feb. 2022, doi: 10.1016/j.cose.2021.102532.

[19] iTrust Centre for Research in Cyber Security, "Secure Water Treatment (SWaT) Testbed," *iTrust lab.*, no. October, 2018, [Online]. Available: https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/#swat.

[20] M. Kravchik and A. Shabtai, "Detecting cyber attacks in industrial control systems using convolutional neural networks," in *Proceedings of the ACM Conference on Computer and Communications Security*, 2018, pp. 72–83, doi: 10.1145/3264888.3264896.

[21] M. Elnour, N. Meskin, K. Khan, and R. Jain, "A dual-isolation-forests-based attack detection framework for industrial control systems," *IEEE Access*, vol. 8, pp. 36639–36651, 2020, doi: 10.1109/ACCESS.2020.2975066.