

Generating Anomaly Data for ICS Environments: Comparative Study

JuHyeon Lee^{1*}, Ilhwan Ji^{1*}, Seungho Jeon^{1†} and Jung Taek Seo^{1‡}

¹Gachon University, Seongnam-daero 1342, Seongnam-si, Republic of Korea
noncaptain@gmail.com, {ilhwan1013, shjeon90, seojt}@gachon.ac.kr

Abstract

A security system is used for cybersecurity in the ICS environment. Security system development and testing in an ICS environment requires anomaly data. However, there are often no anomaly symptom data in an actual ICS environment. In this case, research on anomaly data generation is necessary. In this paper, we analyze anomaly data generation research that can be applied to the ICS environment and compare existing anomaly data generation research based on comparison items. Comparison items can be derived from each research and can be judged intuitively. Based on the comparative results of anomaly data research, we discussed anomaly data generation research that can be utilized according to specific situations and purposes.

Keywords: Industrial Control System, Anomaly Data, Cyber Security, Cybersecurity, Data Generation

1 Introduction

Industrial control systems (ICS) are essential in controlling field devices such as energy, transportation, and manufacturing. The scale of ICS has increased due to the Fourth Industrial Revolution, and the size and complexity of the current ICS environment are increasing further as it is combined with the latest ICT technology. However, these environmental changes increase potential cyber threats to ICS. Cyber-attack entry points have increased as ICT technology is incorporated into the ICS environment. Additionally, because ICS was designed a long time ago, it does not have built-in security features, and system patches can cause problems with ICS availability, so some systems have not been patched to the latest version[1]. Cyber threats from ICS can be responded to by introducing security systems such as intrusion detection systems (IDS) and anomaly detection systems.

The 7th International Conference on Mobile Internet Security (MobiSec'23), Dec. 19-21, 2023, Okinawa, Japan, Article No. 25

*Corresponding author: Global Campus College of IT Convergence, Gachon University, Seongnam-si, 13120, Republic of Korea, seojt@gachon.ac.kr

A data set for development and testing is essential to introduce a security system for ICS. However, there is a problem: the data set for security system development and testing contains only normal data. Because availability is essential for ICS, it is challenging to generate abnormal data because it is impossible to conduct a cyber attack in an actual power plant environment and conduct an experiment to create anomaly data.

Developing and testing security systems requires leveraging cyber attack data or anomaly data to make them more robust and test performance. If security system development and experiments are performed using only normal data or a small number of abnormal data, performance may deteriorate, and performance evaluation cannot be performed. To solve these problems, research is actively being conducted to generate anomaly data. In this paper, we conduct analysis and comparison by investigating anomaly data generation studies applicable to the ICS environment. For existing research to be compared correctly, comparative items that can be analyzed objectively are needed. Comparison items were selected that could be extracted from existing research, and that could be judged intuitively. Comparison items consist of Data Types, Number of anomaly data types generated, Data generation methods, Purpose of generating anomaly data, and Reflect Cyberattack Attributes. Based on the compared results, we discuss which research is appropriate depending on the environment and purpose for which anomaly data is needed. The contributions of this paper are as follows.

- Analyze anomaly data generation research that can be used in the ICS environment and perform comparisons based on comparison items.
- Analyze which anomaly data generation research is appropriate depending on the security system development environment and purpose.

The structure of the paper is as follows. Section 2 analyzes security technologies that require anomaly data in the ICS environment. Section 3 analyzes and compares existing anomaly data generation research. Section 4 analyzes which existing research is appropriate for various security system development environments and purposes. Finally, Section 5 describes the results and future research directions.

The Research Question (RQ) is as follows.

- RQ1: What techniques are used to generate anomaly data?
- RQ2: What technologies can generate cyberattack data?
- RQ3: What types of anomaly data are suitable for use in security systems applied to ICS?

2 ICS Security System Analysis Requiring Anomaly Data

Section 2 analyzes security systems that require abnormal data during development among the security systems used in the ICS environment. IDS and anomaly detection systems are representative ICS security technologies that require abnormal data. IDS and anomaly detection systems are analyzed.

2.1 Intrusion Detection System

In 1980, intrusion detection technology was proposed in Anderson JP's research[2], and intrusion detection technology used data to detect anomaly behavior. IDS detects abnormal behavior that violates the security policies established by each organization. IDS can monitor network traffic to detect potential cyber threats to a network or system. An intrusion prevention system (IPS) is a security system like IDS. IPS blocks traffic that violates security policies. In an ICS environment, IDS is used instead of IPS because there is a risk of disruption in operation by IPS blocking normal traffic. In the ICS environment, IDS is structured as shown in Figure 1.

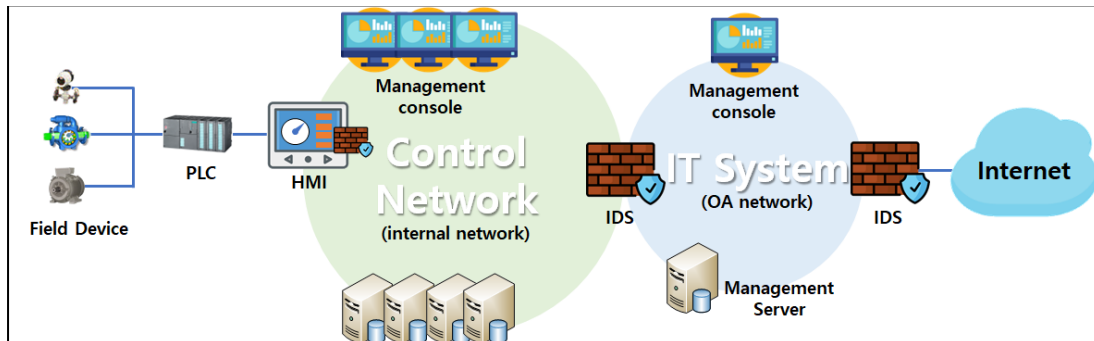


Figure1: ICS network structure and IDS locations

IDS is divided into HIDS (Host-Based Intrusion Detection System) and NIDS (Network-Based Intrusion Detection System), depending on the installation location and purpose. HIDS is installed and operated as an addition to the operating system or installed on a general client. HIDS cannot detect intrusions into the network but can detect when it becomes the target of an attack. NIDS operates independently and can monitor the entire network[3].

The IDS attack detection method generally inputs a known attack pattern and detects the corresponding pattern. The false positive rate is low because it detects using already known information, but it is difficult to detect new attacks[4].

The IDS performance evaluation method requires setting a security policy, injecting normal and malicious traffic, and evaluating whether it detects malicious traffic. By repeating this process, IDS performance can be improved. For IDS performance to be improved and evaluated, various anomaly data are needed, and research is being conducted. CIC-IDS2017 collected eight cyberattack data to improve IDS performance[5]. As in this research, various anomaly data are needed to improve the performance of IDS.

2.2 AI-based Anomaly Detection System

The anomaly detection system in ICS is being developed using machine learning and deep learning algorithms for data preprocessing, feature extraction, and data learning[6]. An anomaly detection system for cybersecurity identifies potential cyber threats by judging them as anomalies when they do not match normal patterns or when anomaly patterns are detected[7]. For example, an anomaly detection model learned from normal data monitors the data and performs anomaly detection on data that exceeds the normal threshold. One research that operates this way is the E-SFD study[8]. This research study proposed an LSTM-based anomaly detection system for ICS. The LSTM anomaly detection system learned normal data, calculated the difference between the predicted and actual values, and detected an anomaly when the error exceeded the threshold. The design and configuration of the AI-based anomaly detection system are shown in Figure 2.

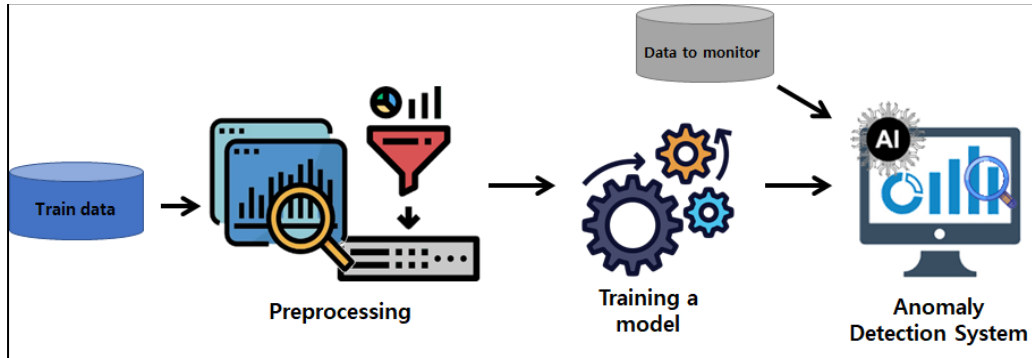


Figure 2: Anomaly Detection System Design and Structure

The quality of data determines a high-performing anomaly detection system. Suppose there is a problem with data imbalance. In that case, the normal category of the data may be incorrectly learned when learning data, and accurate performance cannot be derived when evaluating performance. Therefore, the ratio of normal data to anomaly data is essential, and the abnormal data must be diverse for the anomaly detection model to have good performance and accurate testing. However, securing abnormal data caused by cyber-attacks is challenging in environments such as actual power plants. Therefore, research on anomaly data generation is actively underway.

3 Analyze and Compare Anomaly Data Generation Research

Compare each research by analyzing existing anomaly data generation research and deriving the characteristics of each analysis. The above data generation study selected for comparison was a study that could utilize the data generated by ICS (network packets, operational information, logs). Methods for generating anomaly data in the selected research are divided into methods for generating anomaly data using AI and methods for generating anomaly data using testbeds.

Items for comparison were selected as items that could be derived universally in each research and items that could determine the characteristics of the anomaly data generation method. The comparison items are Data Types, Number of anomaly data types generated, Data generation methods, Purpose of generating anomaly data, and Reflect Cyberattack Attributes. Table 1 is a table that analyzes and compares existing research for each item.

index	Research	Data types	Number of generated anomaly data types	Data generating methods	Purpose of generating anomaly data	Reflect cyberattack attributes
1[9]	AL Perales, 2019	Network Packet	7	Testbed (SCADA system)	Improve anomaly detection system performance	O
2[11]	T.Schlegl, 2019	Image	1	Using AI(GAN)	Resolve data unbalancing	N
3[12]	Pourreza M., 2021	Image	1	Using AI(GAN)	Anomaly detection	N
4[13]	XuL., 2022	Network Packet, Operational information	Depends on the dataset	Using AI(GAN)	Anomaly detection	Depends on the dataset
5[14]	DanL., 2019	Operational information	Depends on the dataset	Using AI(GAN)	Anomaly detection	Depends on the dataset
6[15]	A.Goh, 2017	Operational information	36	Testbed (Water Treatment)	Cybersecurity data generation	O
7[16]	AhmedC., 2017	Operational information	15	Testbed (Water Distribution)	Cybersecurity data generation	O
8[17]	ChuahryM., 2021	Operational information	4	Testbed (Power Grid)	Cybersecurity data generation	O
9[18]	DataHAI, 2023	Operational information	55	Testbed (Hardware-In-the-Loop)	Cybersecurity data generation	O
10[19]	AlirezaD., 2023	Network Packet	7	Testbed (Sample bottle filling plant)	Evaluate IDS	O
11[20]	ShishirK., 2023	Network Packet	9	Testbed (VLAN network)	Cybersecurity data generation	O
12[21]	Moustafa N., 2015	Network Packet	9	Testbed (Hypothetical)	Cybersecurity data generation	O

Table1: Comparison and analysis of existing research on abnormal data generation.

Á. L. Perales Gómez et al. proposed a methodology for generating anomaly datasets for ICS environments[9]. The research methodology consists of four steps: attack selection, attack deployment, traffic capture, and feature calculation. The method aims to create a reliable data set. Attack selection is selecting an attack to perform on the testbed. This research selected Reconnaissance attacks, False data injection attacks, and Replay attacks. Attack deployment is the task of carrying out a selected attack. Traffic capture captures malicious traffic generated after an attack on a testbed. Function calculation is the step of selecting functions for anomaly detection. This method created a new ICS dataset called Electra[10]. The research was conducted on a testbed of industrial devices such as programmable logic controllers (PLC) and a SCADA system using S7Comm and Modbus.

f-AnoGAN was announced in 2019 and presents a method to detect anomalies in data using GAN (Generative Adversarial Network)[11]. f-AnoGAN performs anomaly detection in medical images, but

since it is challenging to analyze anomalies in medical images and determine labels, anomaly detection based on unsupervised learning was proposed. f-AnoGAN uses an encoder to create latent variables for query data and uses the generator as a decoder to restore data. This model calculates an anomaly score using query data, latent variables, and restored data and determines that there is an anomaly in the data if the abnormality score of the generated data is higher than the threshold. f-AnoGAN uses GAN to detect anomaly behavior in unlabeled data by learning the data's normal pattern and calculating the pattern's deviation by assigning an anomaly score.

G2D reduced the anomaly detection problem to binary classification to improve anomaly detection performance and simplify detection model implementation[12]. G2D uses GAN, which trains two deep neural networks (generator and discriminator) using only normal data. Cases in which the generator failed to generate normal data during the learning stage were divided into the initial learning stage and the stage before complete optimization, and these stages were regarded as irregular generators. G2D adopts irregularities at this stage to generate anomaly (irregular) data. This method generates anomaly data that deviates from the distribution of normal data. Data before the generator generates data in the normal category is considered an anomaly.

TGAN-AD proposed a Transformer-based GAN for anomaly detection in time series data[13]. The generator of TGAN-AD can improve performance because it can extract contextual features of time series data, and the discriminator is used to determine anomaly data. TGAN-AD used a Transformer to capture context information of time series data for the GAN framework. For the experiment of TGAN-AD, Secure Water Treatment (SWaT), Water Distribution (WADI), and KDD Cup 1999 were used as data sets. TGAN-AD anomaly detection performance showed a higher level of performance than the existing model.

MAD-GAN proposed GAN-based unsupervised multivariate anomaly detection using LSTM as the base model[14]. The research pointed out that unsupervised machine learning fails to utilize correlations and other dependencies between various variables in the system when anomaly detection. Instead of processing each data stream independently, MAD-GAN considers the entire set of variables simultaneously and analyzes potential interactions between variables. The research utilized SWaT and WADI data and fully utilized GAN using an anomaly score called DR-score.

SWaT is a testbed for water treatment that produces 5 US gallons/hr of filtered water[15]. SWaT functions as a water treatment model and provides a platform for studying potential cyberattacks and their impacts and devising new countermeasures. The SWaT testbed consists of six stages, labeled P1 to P6: P1) Raw water intake, P2) Chemical disinfection, P3) Ultrafiltration, P4) Dechlorination using ultraviolet lamps, P5) Purification by reverse osmosis, P6) Ultrafiltration membrane backwash and cleaning. Each stage is controlled by its own set of dual PLCs, with one PLC acting as the primary and the other PLC acting as a backup in case of primary failure. Normal data is data generated by the normal operation of the SWaT test bed. Anomaly data is data generated by a cyber attack on the SWaT testbed. An attack scenario using the SWaT testbed includes attacker A, who has access to the local factory communication network; attacker B, who is physically nearby but not directly at the site; and attacker C, who is at the site and has physical access to the device. SWaT attack scenarios consist of a system reconnaissance scenario, an infiltration scenario through a wireless network, and an infiltration scenario through physical access.

The WADI (Water Distribution) testbed consists of a water distribution testbed, a water treatment system like SWaT, and is used to secure the water distribution network[16]. WADI operates by PLC and RTU. WADI control network is connected to the SCADA workstation through a wired or wireless network, and anomaly data is data generated after a cyberattack is performed through the control network. An example of a cyberattack is spoofing sensor readings to cut off water supply to a consumer tank.

EPIC (Electric Power and Intelligent Control) is a power grid testbed consisting of four stages: generation, transmission, microgrid, and smart home[17]. EPIC uses the IEC 61850 protocol for substations and automation systems. The anomaly data is generated after a cyber attack on the EPIC

testbed. EPIC attack scenarios consist of Power supply interruption attacks, Nuisance tripping attacks, Physical damage attacks, and Attacks related to economic advantage.

HAI (HIL-based Augmented ICS) is a HIL (Hardware-in-the-Loop) based augmented ICS testbed[18]. The HAI testbed consists of a boiler, turbine, water treatment process, and Hardware-In-the-Loop (HIL) simulator to simulate various scenarios and environments. The HAI testbed is designed to simulate various scenarios and environments, such as thermal and hydroelectric power generation. The process architecture of the HAI testbed consists of four main processes: boiler process (P1), turbine process (P2), water treatment process (P3), and HIL simulator (P4). The anomaly data is generated after a cyber attack was performed on the HAI testbed. HAI data attack scenario consists of two I/O point and internal point attack scenarios. Based on HAI 23.05, there are 47 attack scenarios targeting I/O points, and there are 8 attack scenarios targeting internal points.

ICS-Flow is a testbed consisting of a control system that controls factory equipment such as pipes, valves, conveyor belts, and water tanks to fill empty bottles with water from the tanks[19]. ICS-Flow built a simulator using ICSSIM, which provides an ICS testbed production function. The anomaly data was used to carry out a cyber attack and use the data generated by exploiting the fact that the protocol used in the control system is vulnerable due to a lack of authentication, communication encryption, and integrity checks. ICS-Flow considered five attack types that scan the network, sniff or modify packets, and disrupt ICS operations to exploit vulnerabilities. Attack scenarios consist of a Reconnaissance attack, DDoS attack, man-in-the-middle attack, command injection attack, Replay attack, and others. To carry out the attack, ICS-Flow developed a Python script that automates the execution of the attack on the control system, and using this script, the attacker node quickly launched an attack according to a predefined scenario.

Cyberattack evaluation research was conducted on a virtual testbed composed of VLANs and hosts using Docker to perform various attack scenarios[20]. The anomaly data is the network traffic of virtual machines performing various attack scenarios and configured using Wireshark and tcpdump. Attack scenarios include Unpatched Hosts, Network Scanning, Service Exploitation, Web-Based Exploitation, Distributed Denial of Service, and ARP Spoofing.

UNSW-NB15 was developed including a new attack scenario (synthetic attack) and was developed in a virtual testbed to solve the benchmark problem (availability problem) of the data set[21]. UNSW-NB15 data was developed using the IXIA PerfectStorm tool. IXIA PerfectStorm generates normal and anomaly network traffic[22]. The anomaly data was used after performing an attack scenario. UNSW-NB15 attack scenario consists of Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode, and Worms.

4 Discussion

This paper analyzed anomaly data generation research that can be used in the ICS environment and performed analysis and comparison to select anomaly data generation studies according to the situation and purpose. In order to utilize existing research, three perspectives were discussed. Research suitable for the content discussed was mapped using the index in Table 1.

RQ1) What techniques are used to generate anomaly data? There are two main ways to generate anomaly data. There is research to generate anomaly data based on AI [2-5]. Another method is to configure a test bed and generate anomaly data [1, 6-12]. Research using AI generates anomaly data based on GAN. Use GAN to augment anomaly data or use incomplete data generated while GAN is learning as anomaly data. Research that generates anomaly data using testbeds simulates various environments such as water treatment systems, HIL, power, and virtual networks and generates anomaly data.

RQ2) What technologies can generate cyberattack data? In existing research, research that generates anomaly data reflecting cyberattack characteristics includes using cyberattack data and performing a cyberattack on a testbed. Research on generating anomaly data using GAN reflects the characteristics of cyberattacks because if cyberattack data exists in the data set used, anomaly data like cyberattack data is generated [4, 5]. Research using testbeds builds testbeds, performs cyberattacks such as Reconnaissance attacks, DDoS attacks, Man-in-the-middle attacks, Command injection attacks, and Replay attacks, and captures the data that occurs [1, 6-12].

RQ3) What types of anomaly data are suitable for use in security systems applied to ICS? If the IDS security policy consists of a blacklist, data that violates the security policy must be used. Anomaly data used in this situation should utilize Anomaly data generated by a cyberattack rather than aspects of data other than normal data. Measuring the performance of an IDS can only be evaluated using data generated from cyberattacks [1, 6-12]. When evaluating an anomaly detection system, abnormal data that falls outside the normal category, including cyberattack data, can be used. Because the anomaly detection system learns normal or anomaly data based on AI, it is possible to determine normal/anomaly data. When an anomaly detection system learns normal data, data that falls outside the scope of normal data is judged to be an anomaly, so anomaly data generated in existing research can be used regardless of whether the anomaly data reflects cyberattack characteristics [1-12].

In order to utilize anomaly data for cyber security in ICS, cyberattack characteristics must be reflected in the anomaly data to use it efficiently. The way to generate data reflecting cyberattack characteristics is to generate it based on AI using existing data or perform a cyberattack in a test bed. However, these methods have two constraints. Generating anomaly data using cyberattack data is relatively easy but requires actual cyberattack data. Methods for performing cyberattacks on testbeds require cost and effort to imitate a specific domain. Most testbeds do not ideally mimic the domain because they only implement part of the actual domain environment. Therefore, it is believed that research is needed to generate anomaly data reflecting cyberattack characteristics using normal data without building a separate test bed.

5 Conclusion and Future Work

This paper analyzed and compared research on anomaly data generation that can be used in the ICS environment. Using the comparison results, we analyzed research that can be applied when using a security system in ICS. The analysis and comparison items for the anomaly data generation research consist of Data Types, Number of anomaly data types generated, Data generation methods, Purpose of generating anomaly data, and Reflect Cyberattack Attributes. A total of 12 anomaly data generation research were analyzed, and anomaly data generation research that could be utilized considering the situation and purpose were analyzed. Through this paper, you can utilize the data generation method in the ICS environment according to the situation and purpose. Future research will develop an anomaly data generation method that reflects cyberattack characteristics based on normal data and AI.

Acknowledgements

This work was supported by the Nuclear Safety Research Program through the Korea Foundation of Nuclear Safety (KoFONS) using the financial resource granted by the Nuclear Safety and Security Commission (NSSC) of the Republic of Korea (No. 2106058, 50%) and this work was partly supported by the Korea Institute of Energy Technology Evaluation and Planning(KETEP) grant funded by the Korea government(MOTIE) (20224B10100140, 50%)

References

- [1] Ashwin K. (2023, January). *Top 10 ICS cybersecurity threats and challenges*. Retrieved from <https://www.techtarget.com/searchsecurity/tip/Top-10-ICS-cybersecurity-threats-and-challenges>
- [2] Anderson JP (1980) Computer security threat monitoring and surveillance. Tech Rep James P Anderson Co 56. Retrieved from <https://seclab.cs.ucdavis.edu/projects/history/papers/ande80.pdf>
- [3] Kaouk M, Flaus, J. -M. Potet, M. -L. and Groz, R. A Review of Intrusion Detection Systems for Industrial Control Systems. In *2019 6th International Conference on Control, Decision and Information Technologies (CoDIT)* (pp. 1699-1704). IEEE.
- [4] Khraisat, A. Gondal, I, Vamplew, P. et al. (2019). *Survey of intrusion detection systems: techniques, datasets and challenges*. Cybersecurity, 20. Retrieved from <https://doi.org/10.1186/s42400-019-0038-7>
- [5] Iman, S. Arash, H. and Ali, A. G. (2018, January) Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization, In *4th International Conference on Information Systems Security and Privacy (ICISSP)*. INSTICC
- [6] Jeffrey, N. Tan, Q. Villar, JR. (2023) *A Review of Anomaly Detection Strategies to Detect Threats to Cyber-Physical Systems*. Electronics. (pp. 3283-3317). Retrieved from <https://doi.org/10.3390/electronics12153283>
- [7] Nassif A. B. Talib, M. A. Nasir, Q. Dakalbab, F. M. (2021). Machine Learning for Anomaly Detection: A Systematic Review, *IEEE Access*, (pp. 78658-78700). Retrieved from <https://doi.org/10.1109/ACCESS.2021.3083060>.
- [8] Hwang, C. Lee, T. (2021) E-SFD: Explainable Sensor Fault Detection in the ICS Anomaly Detection System, *IEEE Access*, (pp. 140470-140486), Retrieved from doi: 10.1109/ACCESS.2021.3119573.
- [9] Perales Gómez, Á, L. et al. (2019) On the Generation of Anomaly Detection Datasets in Industrial Control Systems. *IEEE Access*, (pp. 177460-177473), Retrieved from doi: 10.1109/ACCESS.2019.2958284.
- [10] Electra dataset. (2016). Electra dataset: Anomaly detection ICS dataset. Retrieved from <http://perception.inf.um.es/ICS-datasets/>
- [11] Schlegl, T. et al. (2019) f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks, *Medical Image Analysis*. (pp. 30-44). Retrieved from <https://doi.org/10.1016/j.media.2019.01.010>
- [12] Pourreza, M. et al. (2021) G2D: Generate to Detect Anomaly, In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE. (pp. 2002-2011).
- [13] Xu, L. Xu, K. Qin, Y. Li, Y. Huang, X. Lin, Z. Ye, N. Ji, X. (2022). TGAN-AD: Transformer-Based GAN for Anomaly Detection of Time Series. *Applied Sciences*. (pp. 8085-8097). Retrieved from <https://doi.org/10.3390/app12168085>
- [14] Dan, L. Dacheng, C. Baihong, J. Lei, S. Jonathan, G. See-Kiong, Ng. (2019). MAD-GAN: Multivariate Anomaly Detection for Time Series Data with Generative. In *ICANN 2019: Text and Time Series: 28th International Conference on Artificial Neural Networks*, (pp. 17-19). Retrieved from https://doi.org/10.1007/978-3-030-30490-4_56
- [15] Goh J. Adepu S. Junejo K. N. Mathur A. (2017). A Dataset to Support Research in the Design of Secure Water Treatment Systems. In *The 11th International Conference on Critical Information Infrastructures Security*.
- [16] Ahmed, C, M. Palleti, V, R. Mathur, A, P. (2017). WADI: A water distribution testbed for research in the design of secure cyber physical systems. In *Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks*. ACM. (pp. 25-28).
- [17] Chuadhry, M, A. Nandha, K, K. (2021). A Comprehensive Dataset from a Smart Grid Testbed for Machine Learning Based CPS Security Research. In *CPS4CIP 2020: Cyber-Physical Security for Critical Infrastructures Protection*. (pp. 123-135)

- [18] HAI (HIL-based Augmented ICS) Security Dataset. (2023). Retrieved from GitHub: <https://github.com/icsdataset/hai>
- [19] Alireza, D, G. Ali, B. Mahshid, H, M. Hans H. Mauro, C. (2023). ICSSIM — A framework for building industrial control systems security testbeds, *Computers in Industry*, Retrieved from <https://doi.org/10.1016/j.compind.2023.103906>.
- [20] Shishir, K, S. Chirag, G. Ivan, I. Atulya, K, Nagar. (2023). Cyber attack evaluation dataset for deep packet inspection and analysis. *Data in Brief*. Retrieved from <https://doi.org/10.1016/j.dib.2022.108771>.
- [21] Moustafa, N. Slay, J. (2015). UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In *2015 Military Communications and Information Systems Conference (MilCIS)*, (pp. 1-6). Retrieved from doi: 10.1109/MilCIS.2015.7348942.
- [22] PerfectStorm. (2023). KEYSIGHT. Retrieved from KEYSIGHT: <https://www.keysight.com/us/en/products/network-test/network-test-hardware/perfectstorm.html>