Identifying Anomalies in Industrial Networks: A Proposed Testbed for Experimental Evaluation

Karel Kuchar* and Petr Blazek
Brno University of Technology,
Faculty of Electrical Engineering and Communications, Dept. of Telecommunications,
Technicka 12, 616 00 Brno, Czech Republic
*karel.kuchar@vut.cz

***Abstract*—Not only because of the convergence of Information Technology (IT) and Operational Technology (OT) networks, the pass-through network environment needs to be monitored and adequate security implemented. Due to the occurrence of different types of anomalies and their inconsistency in the literature, three main types of anomalies have been identified in this paper and a testbed has been proposed to serve for further experimental testing. This testbed was created using an anemometer and in the current state using an accelerometer. From the results so far, a correlation between the normal condition and the induced operational anomaly can be observed.**

***Index Terms*—Anemometer, Anomaly detection, Network security, Testbed**

I. Introduction

Industrial Control Systems (ICS) are used to control and automate industrial processes such as manufacturing, production, and more. These systems can be found in a wide range of industries, including energy, transportation, water treatment, and manufacturing. It is thus a combination of hardware resources and software components in order to control and monitor industrial processes. The main objective of ICS is to ensure the efficiency, reliability, process safety, and productivity of industrial processes. Along with the development of industries, another industrial revolution called Industry 4.0 has taken place. There has also been a convergence of Information Technology (IT) and Operational Technology (OT), which has enabled the use of many tools from the IT environment. On the other hand, these technologies have also been exposed to IT threats. Common in OT networks until this time was mainly provided by physical separation, but this has disappeared with the convergence of IT and OT networks and so it is necessary to develop new tools using modern technologies to ensure security in these networks.

To be able to ensure sufficient security level in these types of networks it is needed not only to use new approaches but also to use appropriate methods using all available data. These data might be taken from industrial devices and industrial processes but also from redundant sensors. Redundant sensors might be used to monitor both, industrial processes and devices to get the most possible overview and use correlation evaluate data to evaluate the current state. Current industrial testbeds and dataset [1] mostly focus on incident detection using developed tools mainly via machine-learning and deep neural networks [2]. These approaches are mostly capable of detecting and distinguishing individual attacks using all available data. In other words, it is necessary to provide the maximum amount of information to the incident detection system in order to increase the potential of the tool and to enable not only the correct detection of anomalies but also their classification into defined classes. The data can be divided into main three groups (1) data taken from a transmission medium (protocol data, medium usage, etc.), (2) data taken from the end devices (masters, slaves, servers, actuators, sensors, etc.), and (3) data taken from additional sensors which main purpose is not to control and measure industrial process but to control and measure the state of individual devices.

The purpose of this paper is to present a proposal for the creation of a new industrial testbed that will generate data only from the industrial process but will also use redundant sensors and sensors that will monitor the status of the testbed. The purpose is to further provide proof of work that the use of additional data can be used for early-fault detection and classification of anomalies. Also, the aim of the paper is to introduce the division of anomalies into three categories, namely: security, operational, and service anomaly.

II. Current State-of-the-Art

Due to the heterogeneous approach of anomaly detection and classification in the industrial network, the current state of the art is not consistent from the view of individual anomaly classes. For this reason, we have included in this section relevant resources dealing with the identification of security, operational, and service anomalies. Goran Jurišić, et al focused in their paper [3] on creating of testbed created for analysis of fault detection reaction times. In their approach, they use industrial protocol GOOSE to test the time of reactions. Due to the fact, that the testbed simulates power-grid distribution and impacts of failure, the testbed is also capable of generating different values based on the chosen scenario. In paper [4] Jhonathan Julián Gallego Rojas, et al. introduced real-time Small-Scale Wind Turbine Emulator. This system is used for static behavior analysis of the whole system. The testbed is so not specifically, in this stage, focused on specific anomaly detection. From the view of output data, the testbed is capable to generate diverse data due to the different behavior of the whole system in dependence on wind speed (rotor movement).

Michael Sinner, et al. [5] focused on the development of a model predictive controller for blade pitch control of wind turbines. In their approach, they use as the only one from the selected papers, data from redundant sensors as an additional input of data. They focus on blade pitch control, so the data are not used for anomaly identification/classification. Due to the variable dataset, the output data might vary from the settings of the scenario. Daijiry Narzary and Kalyana Chakravarthy Veluvolu introduced in their paper [6] sensor fault detection method. Their presented method uses data that is normally processed by the system and does not use additional sensors, thus focusing only on early-fault detection. The testbed handles the drive, so the output might be different for each scenario. In paper [7] Jing Wang, et al. described data processing analysis implementation of processing data taken from a penicillin simulator for fault detection. In their approach, they focus on the identification of operational anomalies (fault detection of machines) but this approach is not capable early-fault detection/prediction. The used testbed is able to generate various data.

Paper [8] written by Jinrui Nan, et al. focused on Big data-based early fault detection in relation to batteries. The presented paper describes a method for data extraction for early-fault detection based on big data processed via machine learning approaches. On the other hand, their approach does not detect the existence of attacker and anomaly classification. Paper [9] created by Sinil Mubarak, et al. presents industrial datasets with an ICS testbed. The approach describes a method for data evaluation and anomaly detection presented on industrial protocol Modbus. The data they used do not contain data taken from other external sensors or external machines and their state. Asuka Terai, et al. present in paper [10] cyber-attack detection based on the monitoring system and data evaluation. The approach is presented on a water distribution testbed using industrial protocol OPC. The presented approach is focused only on security incidents. In paper [11] Jonathan Goh, et al. presented a secure water treatment testbed/dataset. The dataset is highly used for data evaluation purposes and detection of cyber anomalies in the industrial system due to the precisely presented dataset consisting of many industrial parts (physical one). The EtherNet/IP protocol is used as a transmission protocol.

Paper [12] described Franck Sicard, et al. developed a physical testbed for naval defense security. In their paper, they describe four possible attacks on that testbed including its detection. In their testbed, they use S7 communication protocol. The resulting dataset contains heterogeneous data due to the heterogeneity of the test environment. Jehn-Ruey Jiang and Yan-Ting Lin presented in their paper [13] anomaly detection technique via deep learning methods. The approach is based on powergrid dataset that uses the Modbus communication protocol. In their paper, they focused only on the detection of security anomalies. Jehn-Ruey Jiang and Yan-Ting Chen presented in their paper [14] anomaly detection and its classification using network traffic. In their paper, they performed the classification of a total of six attacks, i.e. they performed the classification of security anomalies. Their research was based on two publicly available datasets.

Table I provides a summary of the mentioned papers in terms of the parameters assessed. From the table, it is visible that most papers are focused on security anomalies (incidents). Based on the analysis, the approach of evaluating all types of anomalies is not yet used in the current literature. Similarly, the use of additional/redundant sensor data sources and their processing is not frequent. There is also no intersection between early-fault detection and security anomaly detection within industrial networks. There is also no clear definition of the different types of anomalies within the current literature. For this reason, the individual data had to be obtained by classifying the different approaches of the compared papers into defined sections. Values that could not be derived from the content of the paper are marked with a question mark. From the state-of-the-art analysis, it was also found that many papers do not deal with operational anomaly detection and if they do it is typically in combination with early-fault detection. However, many times it is difficult to deduce from the text of the work whether it is possible to perform not only early-fault detection but also anomaly detection at the current time and vice versa.

TABLE I
CURRENT STATE OF THE ART

Paper	Year	RS	Protocol	Anomaly			EFD	Variable data
				Sec.	Oper.	Serv.		
[3]	2018	No	GOOSE	Yes	No	Yes	No	Yes
[4]	2019	No	?	No	No	No	No	Yes
[5]	2021	Yes	?	No	No	No	No	Yes
[6]	2022	No	?	No	Yes	No	Yes	Yes
[7]	2022	No	?	No	Yes	No	No	Yes
[8]	2022	No	?	No	Yes	No	Yes	Yes
[9]	2021	No	Modbus	Yes	No	No	No	Dataset
[10]	2017	No	OPC	Yes	No	No	No	Yes
[11]	2017	No	EtherNet/IP	Yes	No	No	No	Yes
[12]	2022	No	S7	Yes	No	No	No	Yes
[13]	2022	No	Modbus	Yes	No	No	No	Dataset
[14]	2022	No	Modbus/S7	Yes	No	No	No	Dataset

The abbreviation Sec. stands for Security; Oper.: Operational; Serv.: Service; EFD: Early Fault Detection.

III. INDUSTRY 4.0 AND ANOMALY DETECTION APPROACHES

The scope of the fourth industrial revolution called Industry 4.0 is very wide from the perspective of all possible components and features, due to the integration of several digital technologies and concepts in the manufacturing industry, such as the Internet of Things (IoT), artificial intelligence (AI), and big data analytics. Also the convergence of IT and OT technologies opens hitherto completely isolated systems to new threats (from the perspective of OT networks). OT networks were not prepared for this convergence without the effort of additional security features from the perspective of the often-used legacy industrial environment (typically old devices running over an insecure industrial protocol). Therefore, state-of-the-art approaches must be used for the early detection and classification of anomalies within industrial networks. Only

through early detection of an intrusion (preferably just an attempted intrusion) can potential damages such as production downtime, data theft/manipulation, equipment damage, safety risk (from the view of the industrial process), data exfiltration (ransomware), malware infiltration (for later usage/backdoor), etc. be avoided. Thus, one of the key aspects appears to be the use of a digital forensic investigation architecture using the maximum amount of data [15].

Anomalies, not only in industrial networks, define a condition where an unusual or unexpected event has occurred within an observed entity (object) that deviates from expected states/values/patterns/behaviour. Based on the purpose and combination of conditions, the anomalies can be divided into three main categories, namely on: security anomaly, operational anomaly, and service anomaly. A security anomaly includes any condition in which a security breach is attempted and can be further divided into subcategories such as Denial of Service (DoS), Man in the Middle (MitM), unauthorized access, and Malware. This category of anomaly can cause targeted network damage, data leakage, denial of service, etc. Thus, it includes all cyber-attack related behavior. Typically this kind of security breach can be detected by analyzing the transmission medium using appropriate intrusion detection and prevention methods and deep analysis such as machine learning techniques or neural networks.

Operational anomalies include all conditions in which an anomaly occurs from the perspective of individual devices as such without the need for targeted intervention or damage. Typically, these can be various types of faults on individual pieces of equipment caused, for example, by material wear or manufacturing defects. The consequence of such anomalies may be deviations in production, and delays in transportation or logistics operations. Operational anomalies can also include early-fault prediction, or the use of data nuances to detect a potential fault early and perform a timely service action without the need to shut down the production process.

Service anomaly refers to all conditions in which an anomaly occurs due to service interventions caused by improper station/element configuration. This can include packet dropping, overloading of some stations, lack of security features, wrong cipher suite, communication errors, slow network connection, excessive network traffic, etc. Some of these events can be detected by collating data obtained from redundant sensors, network traffic and values obtained from individual stations.

IV. Industrial Testbed for Anomaly Classification

In order to perform validation testing of the proposed method, i.e., using redundant sensors to classify the detected anomaly, it is necessary to create an experimental site. Industrial control systems typically work with mechanical devices, their processes or states are monitored or controlled within the process. End devices, i.e. controlled or monitored devices, can be divided into actuators and sensors according to the activity performed. An actuator represents an active device in terms of the operation being performed, its purpose is typically to perform an action (e.g. to close a tap) by means of mechanical work. Sensors, on the other hand, are used to monitor this quantity, i.e., for example, to monitor the flow of a fluid through a monitored pipe by converting the observed phenomenon into an electrical phenomenon (evaluable by software). The resulting correlation, however, can be made using data obtained from the aforementioned actuators and sensors (controlling process) and additional sensors (which are not designed to monitor the observed quantity within the process, but to monitor quantities related to the state of the equipment – for example, vibration). Fig. 1 shows the actuator being controlled through the control station (green line) while data from additional sensors are acquired (orange line). The data is then evaluated at the monitoring and evaluation station using machine learning structures.

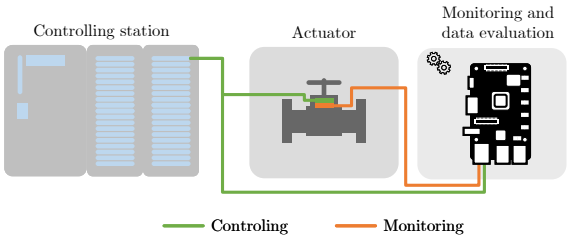

Fig. 1. The principle of advanced data evaluation using redundant sensory sources.

There are certain criteria that need to be taken into account when creating this workplace. The criteria are namely: repeatability, reproducibility, extensibility, based on a real processes, and data variability. The requirement for repeatability focuses on the observed process and the possibility to perform the same experiment repeatedly assuming identical or very similar results. The requirement of reproducibility determines the possibility to create an identical testbed based on the knowledge of the monitored process. The data generated is thus not dependent on a specific workplace, but on the process being monitored. Another requirement is extensibility, which defines the possibility to use other monitored processes or to integrate existing workplaces into another structure, ensuring mutual operability. Furthermore, it is necessary to take into account the requirement for the veracity of the data, or the use of real processes that are monitored and controlled in order to present the possible use in practice and at the same time to minimize the differences from the testbed and real use. Related to this is the requirement for data variability, which defines the variety of data within a process. In other words, it is possible to generate different data for different scenarios, and these data correspond to reality.

A. Testbed description – current state

Based on the established criteria, a testbed was designed. This testbed is based on an anemometer that communicates using the industrial Modbus RTU protocol. This anemometer thus uses the Modbus protocol to acquire individual sensor data such as wind direction and wind speed. It is thus a

representation of a sensor that converts the mechanical rotation of the blades relative to the base into an electrical signal, which is then stored in the protocol memory blocks. These register values are then read by the master station where the individual values are stored. Additional data are acquired using a 3-axis accelerometer. Using this data it is possible to identify sensor position, individual vibrations, and acceleration. The Raspberry Pi is connected via a USB cable to the RS-485 converter to achieve Modbus RTU messages. The acceleration sensor (MMA7660FC) is capable to distinguish changes ± 1.5g (1 g = 9.8 m/s^2). The structure is shown in Fig. 2, where the sensor data taken from the anemometer are marked blue and the additional data taken from the additional accelerometer are marked purple. The accelerometer is rigidly attached to the anemometer structure where the induced shocks are acquired. Data is acquired from the accelerometer through the GrovePi+ hat, which is used to connect I2C and other analog ports.

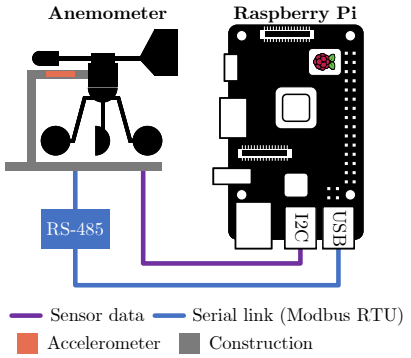

Fig. 2. Mutual interconnection of anemometer and raspberry Pi.

B. Testbed Goals

To create a more purpose-built testbed, further expansion of the workstation with Pulse Width Modulation (PWM) controlled fans is planned. The aim of these fans is to simulate a real environment. In this way, a high-quality dataset covering different states will be obtained and, in particular, this dataset will be diverse in terms of long-term measurements. In addition, other sensors will be placed within the workstation, such as an ultrasonic ranger to measure the vibrations generated by the ultrasonic signal. Subsequently, a data correlation of the measured values from the testbed and the data obtained from the additional sensors is planned. To perform anomaly classification, scenarios focusing on security, operational, and service scenarios will be developed. Based on their analysis, patterns will then be identified and the data will be processed by neural networks to automatically classify the data.

V. Proof of Work

Using the current state of the mentioned testbed, proof of the work is given in this section. With the usage of an acceleration sensor (MMA7660FC) and data taken via Modbus RTU, the two scenarios were described. The first scenario is focused on the legitimate/normal behavior of the testbed. The second scenario is focused on the operational anomaly, so the anomaly behavior is simulated by adding material to one of the rotor blades. The change in wind direction was not considered in this experiment. The wind force was simulated using compressed air. The purpose of this proof of work is only to demonstrate the ability of the basic correlation between the values obtained from the observation of the process (obtaining anemometer readings) and the values obtained from the anemometer. This experiment will be further used to modify the proposed testbed and the use of individual sensors.

Fig. 3 shows the correlation between the individual vibrations as a function of the measured speed from the anemometer in the case of normal operation with no anomalies occurring. In the figure, the speed is shown in red, and the measured vibration is in blue. The average measured velocity was 9.62 m/s, the average vibration rate was 3.24 m/s^2, the median measured velocity was 10.80 m/s and the vibration rate was 2.45 m/s^2. Thus, the ratio of the median values of wind speed and vibration in this case is 2.97 s.

Fig. 4 shows the correlation between the vibration and the measured speed from the anemometer in the case of an operational anomaly. The color coding is identical to Fig. 3. The average measured velocity was 10.85 m/s, the average vibration rate was 6.05 m/s^2, and the median measured velocity and vibration rate were 11.50 m/s and 5.20 m/s^2, respectively. Thus, the median wind speed and vibration ratio, in this case, is 2.21 s. Compared to the normal condition, the ratio was thus reduced by 0.76 s. The operating anomaly was simulated by placing a 2 g load in one of the blades of the anemometer. This is to simulate the clogging of the blade by e.g. a layer of ice. The different scenarios were simulated to achieve the smallest possible differences between scenarios, but deviations still occurred. Therefore, the proposed testbed needs to be refined to ensure data and state consistency.

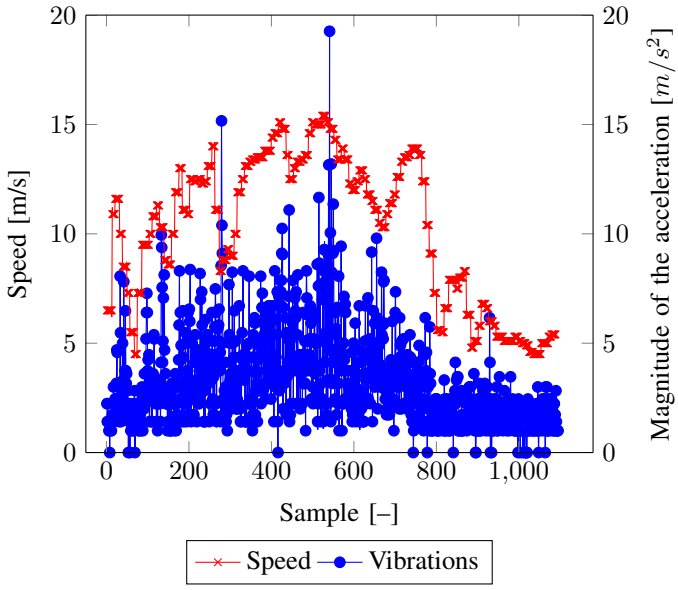

Fig. 3. Values obtained from anemometer and accelerometer - normal behaviour.

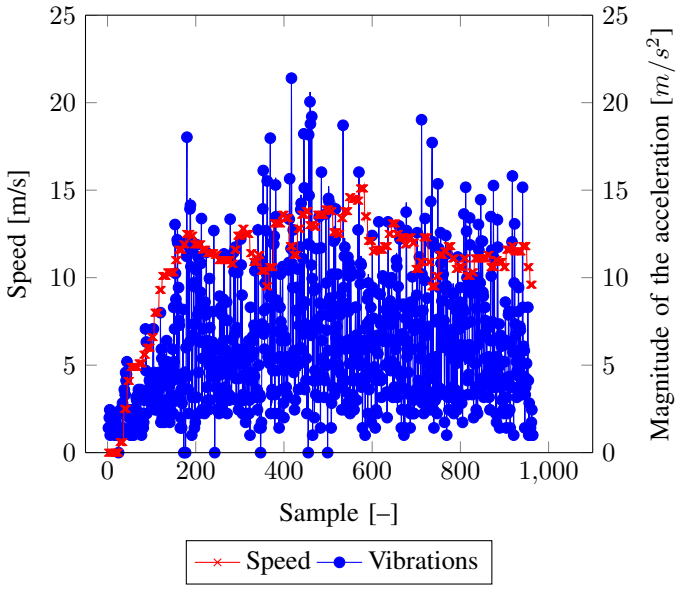

Fig. 4. Values obtained from anemometer and accelerometer - operational anomalies.

VI. Conclusion

This paper focused on presenting three identified classes of anomalies in industrial networks using additional sensor data. Distinguishing these types of anomalies can help in anomaly identification, maintenance planning, and activation of appropriate response measures. A combination of available network data as well as sensor data is needed for appropriate resolution.

Subsequently, an industrial testbed was designed and partially implemented to present these classes of anomalies to obtain a robust dataset. The industrial testbed produced two sensors where the deviations and correlation between measured wind speed and acceleration were observed. The wind speed was obtained through the industrial Modbus RTU protocol and the acceleration was obtained using an accelerometer mounted on the base of the anemometer. The first scenario focused on the correlation during normal running, data was acquired and the ratio of these values was evaluated.

The second scenario simulated an operating anomaly by placing a load in one of the rotor blades presenting, for example, the blade is covered by a layer of ice. From the experimental test, anomalies were identified where the ratio of wind force to total acceleration was 2.97 s in the normal operation and 2.21 s in the second test. Thus, the difference in values is 0.76 s. Thus, by using the designed testbed and the available sensor gauges, it is already possible to differentiate these conditions from each other in this condition and thus identify the operating anomaly. This is despite the limitation that the scenarios were not identical in terms of testing (same wind speed at the same time for both scenarios) and thus cannot be directly compared without incurring additional errors. Thus, the proposed testbed could be used to acquire and evaluate additional data using more types of sensory gauges and to expose additional types and kinds of anomalies.

Acknowledgment

The presented research is a part of the project reg. no. FW06010490, financially supported by the Technology agency of the Czech Republic under the 6. TREND programme.

References

[1] M. Conti, D. Donadel, and F. Turrin, "A survey on industrial control system testbeds and datasets for security research," vol. 23, no. 4, pp. 2248–2294, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9471765/

[2] S. Olugbade, S. Ojo, A. L. Imoize, J. Isabona, and M. O. Alaba, "A review of artificial intelligence and machine learning for incident detectors in road transport systems," *Mathematical and Computational Applications*, vol. 27, no. 5, 2022. [Online]. Available: https://www.mdpi.com/2297-8747/27/5/77

[3] G. Jurišić, J. Havelka, T. Capuder, and S. Sučić, "Laboratory test bed for analyzing fault-detection reaction times of protection relays in different substation topologies," *Energies*, vol. 11, no. 9, 2018. [Online]. Available: https://www.mdpi.com/1996-1073/11/9/2482

[4] J. J. G. Rojas, D. A. Z. Prada, and W. A. L. M. O. Lopez-Santos, "Real-time small-scale wind turbine emulator for a hybrid microgrid laboratory testbed," in *2019 IEEE Workshop on Power Electronics and Power Quality Applications (PEPQA)*, 2019, pp. 1–6.

[5] M. Sinner, V. Petrović, A. Langidis, L. Neuhaus, M. Hölling, M. Kühn, and L. Y. Pao, "Experimental testing of a preview-enabled model predictive controller for blade pitch control of wind turbines," *IEEE Transactions on Control Systems Technology*, vol. 30, no. 2, pp. 583–597, 2022.

[6] D. Narzary and K. C. Veluvolu, "Multiple sensor fault detection using index-based method," *Sensors*, vol. 22, no. 20, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/20/7988

[7] J. Wang, J. Zhou, and X. Chen, *Fault Identification Based on Local Feature Correlation*. Singapore: Springer Singapore, 2022, pp. 119–146. [Online]. Available: https://doi.org/10.1007/978-981-16-8044-1_8

[8] J. Nan, B. Deng, W. Cao, J. Hu, Y. Chang, Y. Cai, and Z. Zhong, "Big data-based early fault warning of batteries combining short-text mining and grey correlation," *Energies*, vol. 15, no. 15, 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/15/5333

[9] S. Mubarak, M. H. Habaebi, M. R. Islam, A. Balla, M. Tahir, E. A. A. Elsheikh, and F. M. Suliman, "Industrial datasets with ics testbed and attack detection using machine learning techniques," vol. 31, no. 3, pp. 1345–1360, 2022. [Online]. Available: https://www.techscience.com/iasc/v31n3/44856

[10] A. Terai, S. Abe, S. Kojima, Y. Takano, and I. Koshijima, "Cyber-attack detection for industrial control system monitoring with support vector machine based on communication profile," pp. 132–138, 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7966982/

[11] J. Goh, S. Adepu, K. N. Junejo, and A. Mathur, "A dataset to support research in the design of secure water treatment systems," *Critical Information Infrastructures Security*, pp. 88–99, 2017. [Online]. Available: http://link.springer.com/10.1007/978-3-319-71368-7_8

[12] F. Sicard, E. Hotellier, and J. Francq, "An industrial control system physical testbed for naval defense cyber-security research," pp. 413–422, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9799429/

[13] J.-R. Jiang and Y.-T. Lin, "Deep learning anomaly classification using multi-attention residual blocks for industrial control systems," *Sensors*, vol. 22, no. 23, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/23/9084

[14] J.-R. Jiang and Y.-T. Chen, "Industrial control system anomaly detection and classification based on network traffic," *IEEE Access*, vol. 10, pp. 41 874–41 888, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9758754/

[15] V. R. Kebande, "Industrial internet of things (iiot) forensics," *Forensic Science International: Reports*, vol. 5, 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2665910722000032