Securing SCADA Systems: A Protocol-Based Intrusion Detection Approach with Shapley Analysis

Serpil Ustebay
dept. computer engineering
Istanbul Medeniyet University)
Istanbul, Türkiye
serpil.ustebay@medeniyet.edu.tr
https://orcid.org/0000-0003-0541-0765

Berhan Berk Akgün
dept. computer engineering
Istanbul Medeniyet University)
Istanbul, Türkiye
https://orcid.org/0009-0009-0539-1622

Piotr Gaj
dept. distributed systems and IT devices
Silesian University of Technology
Gliwice, Poland,
piotr.gaj@polsl.pl,
https://orcid.org/0000-0002-2291-7341

***Abstract*—This paper presents an approach to enhancing the security of Supervisory Control and Data Acquisition (SCADA) systems through the development of Explainable Intrusion Detection Systems (X-IDS). SCADA systems play a crucial role in managing industrial processes and critical infrastructure, yet they are increasingly targeted by cyberattacks, posing significant risks to operational integrity. Leveraging Machine Learning (ML) techniques, particularly focused on network protocols involved in communication with SCADA HMI, this study proposes a protocol-based layered IDS framework to mitigate potential threats. Evaluation of the proposed models demonstrates promising results, achieving high accuracy rates surpassing previous studies. The performance metrics, including accuracy, precision, recall, and F1 scores, along with Brier scores, are comprehensively analyzed to gauge the effectiveness of the models. Additionally, Explainable Machine Learning (XAI) techniques such as SHAP values provide transparent insights into the model's decision-making process. The study highlights the importance of securing HMIs within SCADA systems and offers valuable insights for future research directions in enhancing overall network security.**

Keywords—IDS, SCADA (Supervisory Control and Data Acquisition), Human Machine Interface, Machine Learning, Industrial Control Systems (ICS), Explainable Intrusion Detection Systems (x-ids)

I. INTRODUCTION

SCADA component of ICS is essentially a supervisory and visualization system that aids in the efficient management of industrial processes or smart grids. SCADA makes it possible to gather information and create a comprehensive and interactive graphic user interface on an industrial facility or process in real-time. Energy, water, oil refineries, natural gas distribution, manufacturing, communications, distribution of electric power, transportation, and other Industrial Control Systems (ICS) all make extensive use of SCADA systems [1]. Industry 4.0 has experienced several innovations and developments such as Connectivity and the Internet of Things (IoT), Big Data and Data Analytics, Real-Time Monitoring and Control, and Flexibility and Scalability [2].

Critical infrastructure security is being seriously threatened by cyberattacks on SCADA systems. These attacks have the potential to do significant harm and frequently target the victim's operations or process control systems. Possibly the most well-known SCADA assault is Stuxnet [3]. This intricate and clever malware was created especially to attack Iran's nuclear program. In 2015 and 2016, there were two distinct attacks against Ukraine's energy sector [4]. Large-scale power outages were caused by the attacks, which were directed toward SCADA systems and affected the electrical distribution of power facilities. Cyberattacks on SCADA systems were launched against a German pool and an American water treatment plant [5]. Dragonfly (Aviation Industry, 2011–2014) attack tried to affect power grids, aviation facilities, and SCADA systems [6]. Sometimes Scada systems are not the final target of attacks, but become middleware that has privileged access to the ICS infrastructure, thus enabling the creation of an effective attack vector against controllers and other key components.

II. RELATED WORKS

Authors of [7] processed pcap files containing network flood information and converted the data obtained into a format that ML algorithms can understand. The data set they use has a balanced distribution with 44814 normal and 41220 attack records. In the study, Trained Logistic Regression, KNN, Naive Bayes, Decision Tree, Random Forest, ANN and SVM models were used with 80% training data. With the test data, the highest accuracy was achieved by Random Forest with 81%, while Logistic Regression obtained the lowest accuracy with 65%. [8] devised an innovative SDN-based SCADA architecture utilizing RNN and SVM techniques to identify DDoS attacks aimed at SDN-based SCADA systems. Their dataset encompasses TCP, UDP, and ICMP packets across four scenarios, comprising non-attack and malicious packets. The proposed parallel RNN-based model achieved an accuracy of 96.67%, surpassing the LSTM method by 1%. A novel Few-Step Learning-Based Intrusion Detection System (FS-IDS) [9] has been developed to identify cyber-attacks targeting SCADA networks, particularly in scenarios where

979-8-3503-7943-3/24/$31.00 ©2024 IEEE

traditional approaches may be limited by data availability or scalability constraints. FS-IDS was trained using the gas pipeline dataset and achieved an average accuracy of 84%. [10] devised a novel hybrid ensemble model (ELM) approach to identify hostile intrusions that have bypassed conventional firewalls and typical IDS systems. Their methodology was evaluated on two distinct datasets: one obtained from a gas pipeline system provided by Mississippi State University (MSU), and the other from a water system sourced from University of New South Wales-NB 2015 (UNSW-NB15) data. Data preprocessing involved employing the unity normalization method, while feature extraction from the high-dimensional datasets was conducted using Principal Component Analysis (PCA). Their approaches resulted in metrics, including an accuracy of 99%, precision of 100%, recall of 100%, and a detection rate of 99.90%.

[11] employed supervised machine learning algorithms to detect Modbus protocol Denial-of-Service (DoS) attacks in industrial control system networks. They integrated network flood data from three testbeds to create a more resilient model capable of generalizing effectively when utilizing data from similar networks. The dataset comprises 127,758 records, with 114,700 benign instances and 13,058 attack instances. Experimental findings demonstrate that the proposed mechanism, leveraging the XGBoost algorithm, achieves a remarkable 99% accuracy in identifying this type of attack. [12] proposes a novel architecture model to enhance IoT cybersecurity by employing embeddings, commonly used in Natural Language Processing, to transform indices into vector spaces for words or characters. The study achieved 98.91% accuracy in the Modbus Binary dataset.

III. DATASET DESCRIPTION

This research utilized the CIC Modbus Dataset [13], generated through the analysis of Wireshark captures from a simulated testbed (depicted in Figure 1). Docker containers were created to simulate Intelligent Electronic Devices (IEDs) and SCADA HMIs. IEDs are used to periodically change voltage levels randomly or in response to a request from the SCADA HMI. The SCADA HMI, in turn, is programmed to respond to overvoltage or undervoltage by opening or closing, and it adjusts taps based on values received from the IED.

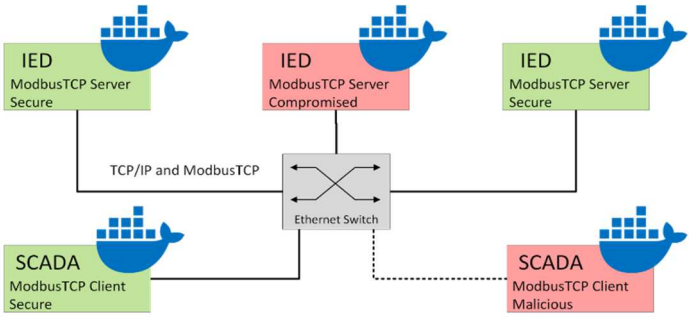

Fig. 1. Simulation testbed architecture

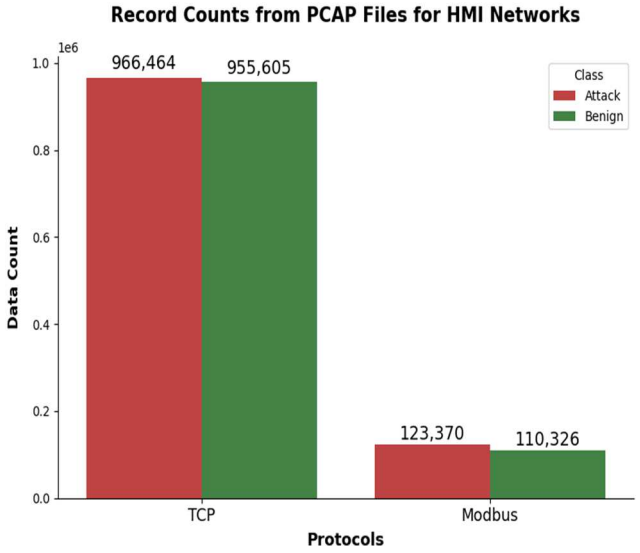

Fig. 2. The number of records parsed from pcap files for HMI networks.

Fig. 3. Example pcap data packet contents captured by Wireshark.

Dataset encompasses attacks from three distinct scenarios: attacks from compromised IEDs, attacks from devices external to the network, and attacks from exploited SCADA HMI. Specific logs documenting the associated assault activities were generated for each scenario. Since SCADA HMI can respond to overvoltage or undervoltage situations by adjusting taps based on information received from IED in this scenario, it has a critical role. A weakened SCADA HMI may jeopardize the system's integrity and safety. We designed a machine learning-based IDS system by using network captures of SCADA HMI to prevent cyber threats. Captured network floods were archived in the pcap file format and categorized into two distinct files for attacks and normal traffic. The dataset includes 966,464 TCP attack records and 955,605 TCP benign records. Additionally, it contains 123,370 Modbus attack records and 110,326 Modbus benign records. The details of these records are depicted in Figure 2. Communication in the SCADA HMI network involves the utilization of two protocols, namely TCP and Modbus/TCP. Figure 3 illustrates sample .pcap

records related to these protocols. Each protocol possesses distinct features.

IV. PROPOSED FRAMEWORK FOR IDS SYSTEM

SCADA systems utilize a variety of communication protocols to facilitate seamless information exchange among diverse system components. These protocols play a crucial role in ensuring effective process control and communication within sectors such as manufacturing, energy, and infrastructure. Each protocol is tailored to meet the specific requirements, goals, and characteristics of the systems or applications it supports, thereby contributing to the overall functionality of SCADA systems. However, each protocol introduces its own unique set of security challenges. These complexities can pose challenges for a central IDS. To address this issue, we have developed an IDS structure based on communication protocol layers, aiming to simplify the complexity associated with these diverse protocols and enhance defense against potential attacks.

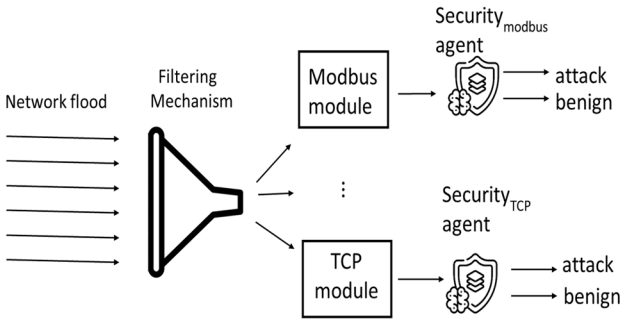

Fig. 4. Proposed IDS architecture

TABLE I. LIST OF EXTRACTED FEATURES AND DEFINITIONS USED BY THE TCP AGENT

Feature Name	Definition
Length	Total length of the packet.
SrcPkt	Total number of packets sent from the source IP and port.
SrcBytes	Total number of bytes sent from the source IP and port.
DstPkt	Total number of packets reaching the destination IP and port.
Sport	Source port number.
Dport	Destination port number.
IP_id	Identifier for the IP packet.
Sequence	The sequence number of the transmitted data in the TCP stream.
Acknowledgment	The sequence number of the received data in the TCP stream.
Window	Window size in the TCP stream.
Flags_encoded	Contains encoded flags of the TCP connection.

TABLE II. LIST OF EXTRACTED FEATURES AND DEFINITIONS USED BY THE MODBUS AGENT

Feature Name	Definition
Transaction_ID	Identifier for the message in Modbus communication.
Protocol_ID	Identifier for the protocol used in Modbus communication.
Length	Indicates the length of the data packet.
Unit_ID	A number indicating which device or unit the message is targeting in Modbus communication.
Function_Code	In Modbus protocol, a specific function code is assigned to each operation. This code determines how the message content will be interpreted.
Ip_id	Identifier for the IP packet.
Sport	Source port number.
Chksum	A checksum is a type of error-checking method used to ensure the accuracy of data in this communication.

Selection of network communication protocols for a SCADA system varies on its structural design. Since network communication protocols adhere to distinct rules governing data formatting, transmission, and reception within a network, we have devised a security model grounded in protocol specifications. The proposed model is depicted visually in Figure 4. In this structure, the network flow is monitored and the incoming network flood is parsed according to the protocol used. Incoming floods are analyzed and routed to their corresponding protocol modules for processing. In this study, both TCP and Modbus TCP protocols were used for SCADA HMI. Additional modules can be added to accommodate diverse testing environments. For TCP, the parsed features are detailed in Table 1, while Table 2 provides a breakdown of features parsed for Modbus along with their explanations. Extracted features are then relayed to the corresponding security agent. This security agent incorporates an AI-driven Intrusion Detection System (IDS) tasked with determining whether the transmitted features indicate benign activity or an impending attack. Based on the prediction outcome, the IP address posing a security threat is promptly blocked.

V. EVALUATION PERFORMANCE METRICS

When evaluating the performance of our classifiers, we used the most popular [14] accuracy, precision, recall, and F1 scores, in addition to the Brier score. Before delving into the evaluation metrics for assessing the classifiers, it's essential to clarify the following terms:

- True Positives (TP): The count of actual positives correctly predicted.
- True Negatives (TN): The count of actual negatives correctly predicted.
- False Positives (FP): The count of actual negatives predicted incorrectly as positives.
- False Negatives (FN): The count of actual positives predicted incorrectly as negatives.
- Accuracy = (TP+TN) / (TP+TN+FP+FN)
- Precision = TP/ (TP + FP)
- F1_Score $= (2 * TP)/(2 * TP + FP + FN)$
- Brier Score= $1/N(\Sigma (ft-ot)^2)$
    - N = the number of samples for calculating a Brier score,
    - ft is the forecast probability (i.e. 25% chance),
    - ot is the outcome (1 if it happened, 0 if it didn't).

Accuracy serves as a gauge for the overall correctness of a predictive model, calculated by determining the ratio of correctly predicted instances to the total instances. Precision focuses on the accuracy of positive predictions. The F1-Score acts as a unified metric, blending precision and recall into a singular value. Represented as the harmonic mean of precision and recall, it strikes a balance between the two aspects. Recall, or sensitivity, assesses a model's proficiency in capturing all relevant instances. This metric is determined by the ratio of true positive predictions to the sum of true positives and false negatives. The Brier Score [15] is calculated as the mean squared difference between the predicted probabilities and the actual outcomes. It considers the entire probability distribution for each prediction. Lower scores are indicative of superior performance. We utilized an XAI technique, such as Shapley Additive Explanations (SHAP) [16] values, to provide human-understandable explanations or justifications for the predictions made by our ML model. SHAP values assign a value to each feature's contribution, explaining the model outputs transparently. By considering all potential feature combinations, SHAP values enhance predictability and transparency, offering insights into intricate model behaviour.

VI. RESULTS

In this research, artificial intelligence models were employed using three distinct approaches. Decision trees were selected to assess the tree-based model, while Random Forest represented an ensemble-based approach, and an MLP artificial neural network was utilized for neural network-based analysis. We utilized a training set consisting of 70% of the data for model training, reserving the remaining 30% for testing purposes.

GridSearchCV methodology was used for hyperparameter tuning. For the Random Forest model, the parameter search space included 'n_estimators' (50, 100, 200), 'max_depth' (None, 10, 20, 30), and 'min_samples_split' (2, 5, 10). The Decision Tree model's search space consisted of 'criterion' ('gini', 'entropy'), 'max_depth' (None, 10, 20, 30), and 'min_samples_split' (2, 5, 10). The MLP model was tuned with 'hidden_layer_sizes' ((50, 50, 50), (50, 100, 50), (100,)), 'activation' ('tanh', 'relu'), 'solver' ('sgd', 'adam'), 'alpha' (0.0001, 0.05), and 'learning_rate' ('constant', 'adaptive'). The performance metrics of the models trained by selecting the parameters that give the highest accuracy values within the parameter space of the algorithms are shown in Table 3 and Table 4.

TABLE III. RESULTS OF MODBUS AGENT

Classification Algorithm	Performance Metrics				
	Accuracy	Precision	F1-Score	Recall	Brier-Score
Decision Tree	0.99	0.99	0.99	0.99	0.007
Random Forest	0.98	0.99	0.98	0.97	0.016
MLP	0.86	0.91	0.86	0.82	0.133

TABLE IV. RESULTS OF THE TCP AGENT

Classification Algorithm	Performance Metrics				
	Accuracy	Precision	F1-Score	Recall	Brier-Score
Decision Tree	0.99	0.99	0.99	1.0	1.73e-06
Random Forest	0.99	0.99	0.99	1.0	1.73e-06
MLP	0.50	0.50	0.66	0.99	0.49

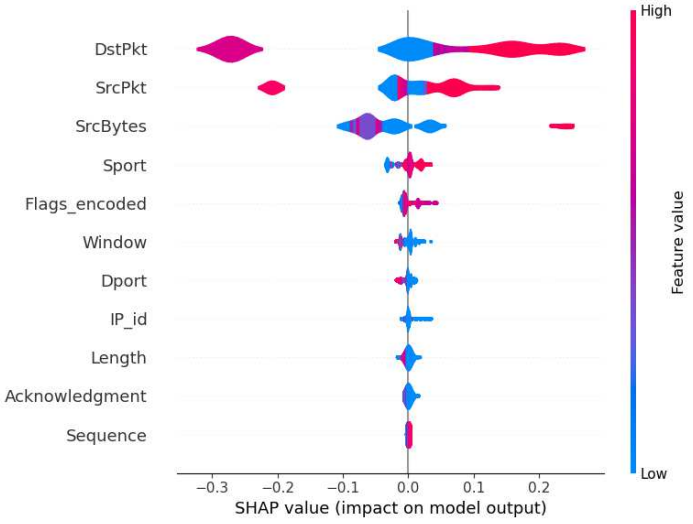

Fig. 5. TCP agent SHAP values

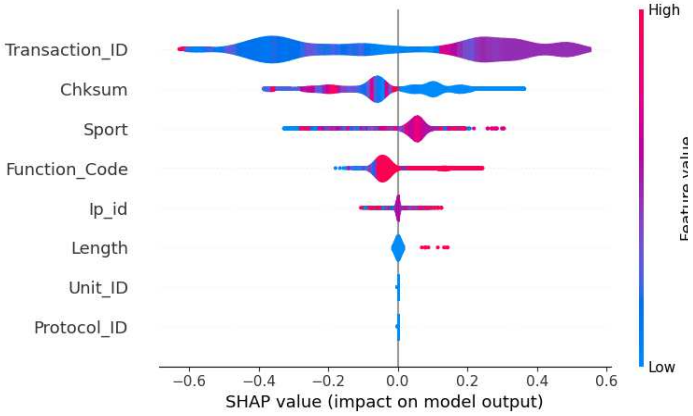

Fig. 6. Modbus agent SHAP values

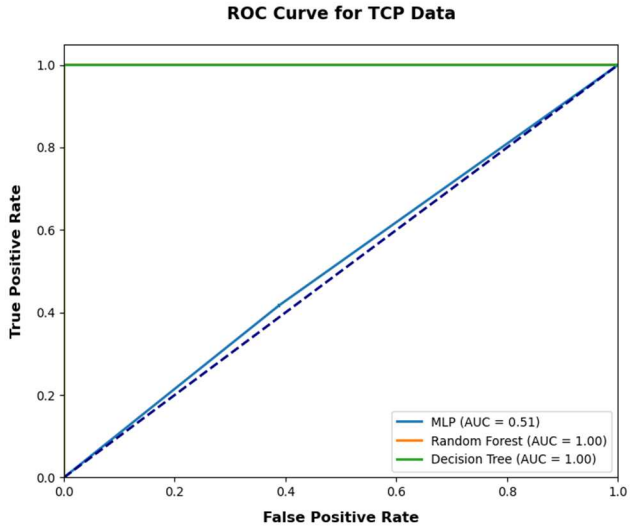

Fig. 7. ROC curves for TCP data

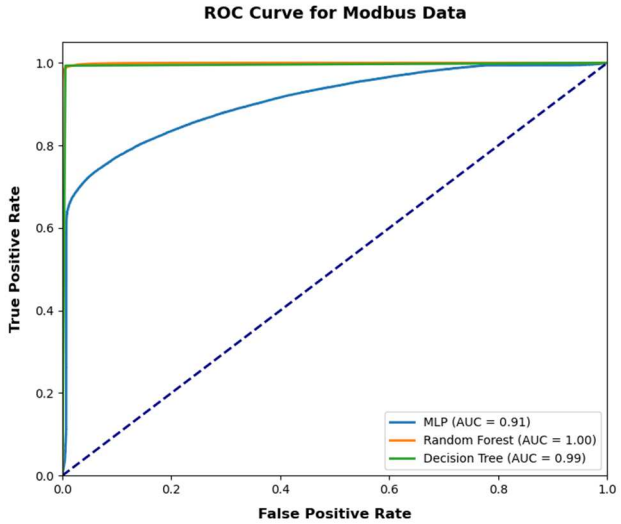

Fig. 8. ROC curves for Modbus data

The experimental design utilized Python 3.10 for coding and was executed on a computer equipped with an i5 Intel microprocessor. Among the models tested, the random forest achieved the highest accuracy, while the MLP demonstrated the lowest accuracy. While the highest accuracy was obtained with random forest, the lowest accuracy was obtained with MLP. SHAP values were computed to provide a deeper

understanding of the random forest model's performance. These values are depicted in Figure 5 and Figure 6 for the Modbus and TCP models, respectively. Additionally, the ROC curves for the models trained on TCP and Modbus data are shown in Figure 7 and Figure 8, respectively, highlighting the performance of each model across these two protocols.

VII. DISCUSSION AND CONCLUSION

In our study, we focused on the security of the SCADA HMI, which is an important part of the SCADA system in the test environment we examined. The SCADA HMI plays a critical role, especially with the task of quickly responding to voltage situations received from the IED. SCADA HMI works with communication protocols using TCP/IP stack as well as other protocols e.g., Modbus, COTP, Ethernet/IP, etc. , and within this framework, we have developed models trained with 3 different machine-learning approaches. In our study, we achieved an accuracy of 0.99, which is a higher accuracy rate than previous studies such as [7,8,9]. Additionally, we achieved similar accuracy values to studies [10-12] with a simpler ML approach. In our proposed approach, we proposed to provide security with a protocol-based layered approach, using a simpler ML model. These simple models make predictions faster than more complex models regarding the speed of mathematical operations. Particularly, considering the SCADA transaction volume, we can say that speed is an important performance criterion.

We analysed the SHAP values of the models which have the highest accuracy. We observed that the most important features are DstPkt, and SrcPkt to understand whether the TCP communication, in the study, is threatening or secure. Low values of both features indicated a negative contribution to the model's prediction. Although the prediction effect of the Sport and Sport features was not high, their effects showed a heterogeneous distribution according to their high and low values. Ports allow communication to be directed to a specific target, and cyber attackers often launch their attacks by targeting open ports on the target system. This distribution is likely due to the complex use of open ports in the test environment, such as attacks and safe use. For a secure system, detecting and closing unused ports with port Scan techniques will reduce the risk of attack [17].

When we examined the Shap values calculated for the Modbus protocol (Fig. 6), we observed that the most effective feature was Transaction ID. Transaction ID helps uniquely identify each transaction or message exchanged between the Modbus master and slave devices. This unique ID ensures that each message can be effectively tracked and managed. The results show that Transaction ID in Modbus communication plays a vital role in ensuring message integrity, error detection and recovery, security, and proper sequencing of communications. Results indicate that checksum stands as the second most crucial factor in Modbus communication, facilitating error detection, including transmission noise, interference, or device malfunctions. Upon examining Fig. 6, it becomes apparent that Chksum is the second important feature. High values have a negative impact, while low values have both negative and positive impacts on the model. This pertains to the feature's contribution to reducing the predicted value of the target variable.

In the future study, we plan to develop an IDS (Intrusion Detection System) that can ensure security by monitoring the entire system's network traffic, in addition to SCADA HMI.

REFERENCES

[1] S. V. B. Rakas, M. D. Stojanović and J. D. Marković-Petrović, "A Review of Research Work on Network-Based SCADA Intrusion Detection Systems," in IEEE Access, vol. 8, pp. 93083-93108, 2020, doi: 10.1109/ACCESS.2020.2994961.

[2] A. Nechibvute & H. D. Mafukidze (2023) Integration of SCADA and Industrial IoT: Opportunities and Challenges, IETE Technical Review, DOI: 10.1080/02564602.2023.2246426.

[3] Al-Rabiaah, S. (2018, April). The “Stuxnet” virus of 2010 as an example of a “APT” and its “Recent” variances. In 2018 21st Saudi Computer Society National Computer Conference (NCC) (pp. 1-5). IEEE.

[4] 4-Case, D. U. (2016). Analysis of the cyber attack on the Ukrainian power grid. Electricity Information Sharing and Analysis Center (E-ISAC), 388(1-29), 3.

[5] Hassanzadeh, A., Rasekh, A., Galelli, S., Aghashahi, M., Taormina, R., Ostfeld, A., & Banks, M. K. (2020). A review of cybersecurity incidents in the water sector. Journal of Environmental Engineering, 146(5), 03120003.

[6] Hemsley, Kevin E., & E. Fisher, Dr. Ronald. History of Industrial Control System Cyber Incidents. United States. https://doi.org/10.2172/1505628.

[7] Mubarak, S., Habaebi, M. H., Islam, M. R., Rahman, F. D. A., & Tahir, M. (2021). Anomaly Detection in ICS Datasets with Machine Learning Algorithms. Computer Systems Science & Engineering, 37(1).

[8] Polat, H., Türkoğlu, M., Polat, O., & Şengür, A. (2022). A novel approach for accurate detection of the DDoS attacks in SDN-based SCADA systems based on deep recurrent neural networks. Expert Systems with Applications, 197, 116748.

[9] Ouyang, Y., Li, B., Kong, Q., Song, H., & Li, T. (2021, June). FS-IDS: a novel few-shot learning based intrusion detection system for scada networks. In ICC 2021-IEEE International Conference on Communications (pp. 1-6). IEEE.

[10] Saheed, Y. K., Abdulganiyu, O. H., & Tchakoucht, T. A. (2023). A novel hybrid ensemble learning for anomaly detection in industrial sensor networks and SCADA systems for smart city infrastructures. Journal of King Saud University-Computer and Information Sciences, 35(5), 101532.

[11] Abubakar Sadiq Mohammed, Eirini Anthi, Omer Rana, Neetesh Saxena, Pete Burnap,Detection and mitigation of field flooding attacks on oil and gas critical infrastructure communication, Computers & Security,Volume 124,2023,103007, ISSN 0167-4048, https://doi.org/10.1016/j.cose.2022.103007.

[12] Gueye, T., Wang, Y., Rehman, M., Mushtaq, R. T., & Zahoor, S. (2023). A novel method to detect cyber-attacks in IoT/IIoT devices on the modbus protocol using deep learning. Cluster Computing, 1-27.

[13] Kwasi Boakye-Boateng, Ali A. Ghorbani, and Arash Habibi Lashkari, "Securing Substations with Trust, Risk Posture and Multi-Agent Systems: A Comprehensive Approach," 20th International Conference on Privacy, Security and Trust (PST), Copenhagen, Denmark, August. 2023.

[14] Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., Chen, S., Liu, D., & Li, J. (2020). Performance comparison and current challenges of

using machine learning techniques in cybersecurity. Energies, 13(10), 2509.

[15] Park, S. Y., Park, J. E., Kim, H., & Park, S. H. (2021). Review of statistical methods for evaluating the performance of survival or other time-to-event prediction models (from conventional to deep learning approaches). Korean Journal of Radiology, 22(10), 1697.

[16] Chen, H., Covert, I. C., Lundberg, S. M., & Lee, S. I. (2023). Algorithms to estimate Shapley value feature attributions. Nature Machine Intelligence, 5(6), 590-601.

[17] Anitha, A. A., & Arockiam, L. (2022). A review on intrusion detection systems to secure IoT networks. International Journal of Computer Networks and Applications, 9(1), 38-50.