Ensemble common features technique for lightweight intrusion detection in industrial control system.

OTOKWALA, U.J. and PETROVSKI, A.

2023

Ensemble Common Features Technique for Lightweight Intrusion Detection in Industrial Control System

1st Uneneibotejit J. Otokwala
School of Computing
Robert Gordon University
Aberdeen, UK
u.otokwala@rgu.ac.uk

2nd Andrei Petrovski
School of Computing
Robert Gordon University
Aberden, UK
a.petrovski@rgu.ac.uk

***Abstract*—The integration of the Industrial Control System (ICS) with corporate intranets and the internet has exposed the previously isolated SCADA system to a wide range of cyber-attacks. Interestingly, the vulnerabilities in the Modbus protocol, with which the ICS communicates, make data obfuscation and communication between component entities less secure. In this work, we propose a Common Features Technique (CFT) for Lightweight Intrusion Detection based on an ensembled feature selection approach. Our Common Features Technique, which used fewer features, was able to detect intrusion at the same level as models using information gain, Chi-Squared, and Gini Index feature selection techniques datasets after fitting Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbour (KNN) models. More importantly, when p-values were computed, the CFT model computation time and memory usage were statistically significantly different at 95% and 90% Confidence Interval (CI) when compared to the model on the other techniques.**

***Index Terms*—Intrusion Detection, Industrial Control Systems, Machine Learning, Feature selection, SCADA**

TABLE I
NOTATION TABLE

Abbreviations	
ICS	Industial Control System
SCADA	Supervisory Control and Data Acquisition
CFT	Common Features Technique
NIDS	Network Intrusion Detection System
RF	Random Forest
SVM	Support Vector Machine
KNN	K-Nearest Neighbour
Ci	Confidence Interval
RTU	Remote Terminal Unit
HMI	Human Machine Interface
DCS	Distributed Control Systems
PLC	Programmable logic Controller
Inf Gain	Information Gain

I. Introduction

Supervisory Control and Data Acquisition (SCADA) systems, Distributed Control Systems (DCSs), and Programmable Logic Controllers (PLCs) are all examples of Industrial Control Systems (ICSs) [1]. They are legacy systems that run on proprietary control protocols. The SCADA system was designed to collect data from Remote Terminal Units (RTU) and send it to the Human Machine Interface (HMI) station via a communication channel. Previously, these systems were typically isolated from corporate network connections [2], and the isolation made the systems less vulnerable to cyber security issues except for deployment errors and security flaws. However, as technology advanced and improved, the SCADA system transitioned from serial network communication to TCP/IP-based network communication, particularly with its integration into the corporate network [3]. As a result of this change, the system has become vulnerable to a slew of cyber attacks, jeopardising its effectiveness. Interestingly, the Modbus protocol, which is widely used in SCADA systems, has been found to be deficient in terms of data authentication, ineffective intrusion detection, data encryption, authorization, and effective threat identification [4]. Because of the numerous flaws in ICS firmware and software, an effective lightweight intrusion detection mechanism based on the proper feature selection approach is required, especially given that some constituent devices, such as edge devices and PLCs, have memory constraints and limited computational resources.

A lightweight intrusion detection technique based on feature reduction is one that is concerned with selecting the appropriate features while also filtering out redundant features. There are two approaches to achieving effective feature reduction: (a) the dimensionality reduction approach and (b) the feature selection method. In the dimensionality reduction approach, t he entire feature space of a dataset is transformed and a set of new features is created from the original features [5]. The feature selection approach, on the other hand, is made up of three methods: (i) the filter method, (ii) the wrapper method, and (iii) the embedded method. According to Kuncheva [6], the filter feature selection method is classifier independent and resistant to overfitting. Furthermore, the filter method employs statistical techniques to investigate the relationship between the independent and dependent variables, with the statistical score derived from the evaluation being used to select the relevant features [7]. In this paper, we used filter selection level aggregation to

identify a subset of common features among the results of various feature selection techniques applied to a dataset. Following that, a model is fitted to the subset to classify the data. Because a large volume of data is frequently generated, and learning algorithms typically take a long time to train, generalise, and classify attacks, the filter feature selection approach is therefore, very important.

The remainder of this paper is organised as follows. Section II discusses additional related issues, studies, and proposals on the subject. It also discusses the work's motivation and contribution. Section III describes the dataset and the methodology we used to implement our approach. Section IV discusses models application as well as the evaluation's results. Section V summarises the work and highlights areas that need further research.

II. Related issues and work

Because the Modbus protocol lacks data encryption and authentication, it is difficult for the SCADA system to provide an identity verification mechanism for session hijacking or IP spoofing if a device's IP is hijacked and malicious programmes are injected. McLaughlin et al. [10] proposed a design of a PLC malware that can dynamically generate a malicious payload based on observations of internal processes in an Industrial Control Systeme. However, Mayor et al. [11], had proposed the use of the Metasploit framework for the identification of vulnerabilities such as those enumerated by McLaughlin and the detection of intrusion in ICS. Although Mayor's approach appears lofty, it is incapable of detecting the attacks highlighted by McLaughlin. This is premised on the fact that PLCs and other Edge devices have resource limitation which therefore necessitates the use of lightweight intrusion detection. To this end therefore, Aljawarneh et al., [12] proposed A hybrid of vote algorithm and information gain for the selection relevant features that could positively aid in effective lightweight label classification. In a similar vein, Ayo et al., [13] proposed a NIDS based on a deep learning model that is optimised using a hybrid of rule-based feature selection techniques. According to the authors, the architecture is organised into three stages: rule-based hybrid feature selection, rule evaluation, and detection. Following that, search algorithms and feature evaluators are then combined to improve feature selection and model fitting.

Similarly, Zhou et al., [14] propose a model based on feature selection and ensemble learning techniques. The CFS-BA heuristic algorithm is used to reduce dimensionality and select an optimal subset of features based on feature correlation. Thereafter, an ensembled model comprised of C4.5, Random Forest (RF), and Forest by Penalizing Attributes (Forest PA) is then used to classify the labels. Similarly, Li et al., [15], proposed a model in which the Gini index is used to select the best subset of features, and the gradient boosted decision tree (GBDT) model is fitted and optimised using swarm optimization (PSO) for effective classification.

While all of the approaches listed above have demonstrated a high classification rate, it is worth noting that, according to Zhang et al., [8], using a single feature selection method may result in the generation of a sub-optimal feature subset, compromising the learning algorithm's performance. This may also explain why some learning algorithms perform better with some datasets and poorly with others. As a result, [9] suggests that combining multiple feature ranking techniques to select a subset of common features is a more dependable approach for improving classification and overall accuracy.

A. Motivation

ICS are prone to attacks due to the vulnerabilities in their structure. The vulnerabilities revolve around the lack of any mechanisms for authentication of the communicating entities, as well as the lack of basic security and data protection mechanisms. To that end, we are motivated, among other things, to discover:

- How to make intrusion detection effective in Industrial Control Systems. This is because of the fact that successful breaches have the possibility of annihilating the System and its constituent devices.
- How to achieve intrusion detection with minimal computation cost in minimal time and at a lower resource overhead.

B. Contribution

- We propose a lightweight intrusion detection technique based on a shared subset of features from different feature selection techniques.
- We demonstrated that our approach can achieve the same overall accuracy, sensitivity, and specificity with fewer features as it can with more.
- We demonstrated that our approach is capable of significantly lowering CPU computational costs and time.

III. Dataset and Methodology

The datasets used in this work are a laboratory simulated cyber-attacks on Industrial Control System Network Traffic on gas pipeline and water storage tank. Basically, two datasets were used and they are in [16], [17].

A. Gas pipeline dataset

This dataset is made up of 8 instant classes consisting of 7 different kinds of malicious classes and 1 benign traffic. The distribution of the classes are displayed in Table II.

TABLE II
Instant classes distribution for Gas pipeline dataset

Classes	Normal	NMRI	CMRI	MSCI	MPCI	MFCI	DOS	Recon
Rows	6671	335	1664	93	842	41	189	783

From Table II, we could observe that the distribution of the instant classes is highly skewed towards the benign traffic. This is because the number of observations of the other classes

are lower than the number of observations in the majority class. As a result, we augmented the dataset by oversampling the minority classes to the number of the majority class by using the approach by [18]. The number of observations in the benign class was used as the baseline for the oversampling of the minority classes. Furthermore, prior to the oversampling of minority classes, the dataset contained 10,618 observations and 27 features. The total number of observations in the dataset increased to 53,359 after the minority classes were augmented. A min - max normalization approach was used to scale the values in order to eliminate bias and to ensure that all the features contribute equally (see equation 1).

$$(x - min(x))/(max(x) - min(x)) \tag{1}$$

B. Water Tank dataset

This is another dataset originating from laboratory simulation, and it contains seven different types of malicious as well as benign traffic. As shown in Table III, the distribution is heavily skewed towards the Normal class, resulting in a highly imbalanced dataset. In light of this, we augmented the minority classes using the approach proposed by [18]. Also, the number of observations in the benign class formed the threshold for oversampling the minority classes. Furthermore, prior to the oversampling of minority classes, the dataset contained 27,199 observations and 23 features. However, after the minority classes were augmented, the total number of observations in the dataset increased to 155,675. and scaling was done using the min - max normalisation approach in order to eliminate bias and ensure that all features contribute equally. (see equation 1)

TABLE III
INSTANT CLASSES DISTRIBUTION FOR WATER TANK DATASET

Classes	**CMRI**	**DoS**	**MFCI**	**MPCI**	**MSCI**	**NMRI**	**Normal**	**Recon**
Rows	1457	135	155	410	209	1198	19503	4132

C. Methodology

The following features selection techniques were used: Information Gain, Chi-Squared, and Gini-Index. Using all of the approaches, each dataset's features were ranked in order of importance. The cumulative variance of the values returned by each technique for the features was then used to eliminate redundant and less contributing features. A threshold is assumed at the point where cumulative variance is saturated, with further addition not resulting in further increment. This resulted in three (3) datasets being generated from the Gas pipeline and water tank datasets. The datasets obtained are as follows: (a) the Information Gain dataset, (b) Chi-Squared dataset, and (c) Gini-Index dataset. However, from these three (3) datasets, a fourth (4th) (Common Features Technique) dataset is generated. This is achieved by selecting a subset of the features held in common by the three other datasets.

Based on the threshold, 18 features were retained for each of the Information Gain, Chi-Squared, and Gini-Index datasets for the Gas pipeline dataset after ranking and elimination of redundant features. However, for the Common Features Technique (CFT) dataset, 14 features were common and they formed the CFT dataset's features. Similarly, after removing the redundant features from the water tank dataset, 17 features were retained for modeling of information Gain, Chi-Squared, and Gini-Index approaches. The Common Features Techniques had 14 features which were the common features, and they constituted the CFT dataset.

Algorithm For Feature Selection and CFT Approaches

Load dataset
For dataset, T
  rank features $(v_i, v_{i+1}, , , v_n\ )$
  order $i \leftarrow a_i, , , ,$: magnitude in descending order
  compute $Cum_{var} \leftarrow Var(a_{i+1}, ...a_n)$
  exit if $Cum_{var} + a_i \not> Cum_{var}$
  save $A_i \leftarrow v_i$: save selected features based on Cum_{var}
Repeat $k \leftarrow 1$ to N
  Do: step 2 - 7 with new F_selector
If $x \in (\ A_i \cap A_{ii} \cap A_{iii})$
  select $x_i, ..., x_n$
  save $Y \leftarrow x_i$
  repeat step 9 - 11
  end if
  Return (Y)

IV. Model Fitting and Discussion

Upon completion of data augmentation and sub-dataset generation, we proceeded to fit RF, SVM, and KNN models on the 4 datasets. Fitting a variety of models was done to ensure consistency across models and also to test the effectiveness of our technique. However, before the fitting of the models, we split the dataset 80:20% for training and validation and then ran 20 iterations with a script. 70% of the training dataset was used without replacement for each iteration. For each iteration, metrics such as overall accuracy, sensitivity, specificity, computation time, and memory consumed were recorded. The models were evaluated on datasets containing Information Gain, Chi-Squared, Gini-index, and Common Features Technique (CFT) feature selection techniques datasets. In all, on each table, 80 iterations (20 on each dataset) were performed. The average performance metrics values for each model on each dataset were computed and summarised in tabular form for comparison. Tables IV through VI display the average values of the metrics for Random Forest, Support Vector Machine and K-Nearest Neighbour models. Similarly, Tables VII displays the summary of the average values of the metrics from the fitting of a random forest model on the sub-datasets of the Water tank.

TABLE IV
SUMMARY OF AVERAGE METRICS VALUES WITH SVM MODEL ON GAS PIPELINE DATASETS

	Accuracy	Sensitivity								Specificity								Time (sec)	Memory
		CMRI	DoS	MFCI	MPCI	MSCI	NMRI	Normal	Recon	CMRI	DoS	MFCI	MPCI	MSCI	NMRI	Normal	Recon		
CFT	87.1	99.5	60.7	95.5	98.55	92.08	96.6	57.34	100	99.7	100	100	99.95	100	88.68	94.47	100	5.01	27.57
Info_Gain	88.7	99	68.12	95.3	99.03	93.39	96.49	56.64	100	99.79	100	100	99.93	99.99	92.47	94.91	100	6.04	30.26
Chi_Sqd	89.83	99.3	77.82	95.4	99.11	93.33	97.38	57.15	100	99.73	100	100	99.93	99.99	93.42	95.35	100	5.97	29.85
Gini_Index	88.72	99.1	71.31	95.4	98.69	92.15	96.17	55.98	100	99.75	100	100	99.94	99.99	92.45	94.96	100	6.11	29.93

TABLE V
SUMMARY OF AVERAGE METRICS VALUES WITH RANDOM FOREST MODEL ON GAS PIPELINE DATASETS

	Accuracy	Sensitivity								Specificity								Time (sec)	Memory
		CMRI	DoS	MFCI	MPCI	MSCI	NMRI	Normal	Recon	CMRI	DoS	MFCI	MPCI	MSCI	NMRI	Normal	Recon		
CFT	93.09	99.6	81.53	86.8	98.99	98.14	93.32	97.53	100	99.67	100	100	99.88	90.07	90.27	99.02	100	0.12	15.57
Info_Gain	95.76	99.9	84.39	96.2	98.75	93.67	94.73	98.31	100	99.79	98	100	99.93	99.99	97.85	99.4	100	0.15	18.32
Chi_Sqd	96.76	99.6	92.63	96.2	98.56	93.8	95.37	98.27	100	99.79	98	100	99.93	99.93	98.91	99.34	100	0.15	18.11
Gini_Index	95.82	99.7	86.39	95.5	99.19	92.69	94.91	97.77	100	99.7	98	100	99.96	100	98.13	99.36	100	0.16	18.35

TABLE VI
SUMMARY OF AVERAGE METRICS VALUES WITH KNN MODEL ON GAS PIPELINE DATASETS

	Accuracy	Sensitivity								Specificity								Time (sec)	Memory
		CMRI	DoS	MFCI	MPCI	MSCI	NMRI	Normal	Recon	CMRI	DoS	MFCI	MPCI	MSCI	NMRI	Normal	Recon		
CFT	88.10	98.23	76.67	100	99.51	92.62	63.05	78.57	100	99.90	97	99.30	99.81	99.85	96.98	93.68	100	5.96	20.86
Info_Gain	87.87	98.27	77.8	100	99.47	93.73	62.43	75.65	100	99.89	100	100	99.82	99.87	97.07	93.72	100	6.54	23.79
Chi_Sqd	87.87	98.27	77.8	100	99.47	93.73	62.43	75.65	100	99.89	97	99.1	99.82	99.87	97.06	93.72	100	7.37	23.71
Gini_Index	87.87	98.27	77.8	100	99.47	93.73	62.43	75.65	100	99.89	97	99.1	99.82	99.87	97.06	93.72	100	7.46	23.73

TABLE VII
SUMMARY OF AVERAGE METRICS VALUES WITH RANDOM FOREST MODEL ON WATER TANK DATASETS

	Accuracy	Sensitivity								Specificity								Time (sec)	Memory
		CMRI	DoS	MFCI	MPCI	MSCI	NMRI	Normal	Recon	CMRI	DoS	MFCI	MPCI	MSCI	NMRI	Normal	Recon		
CFT	91.32	52.95	94.58	100	99.94	96.85	98.13	99.97	100	99.29	90.56	100	100	99.99	99.98	98.76	100	0.38	44.30
Info_Gain	98.42	92.48	99.23	100	99.97	96.93	98.65	100	100	99.45	99.87	100	100	100	99.99	98.90	100	0.40	48.93
Chi_Sqd	98.30	90.83	99.34	100	99.95	97.10	99.05	100	100	99.69	99.60	100	100	100	99.98	98.83	100	0.45	49.10
Gini_Index	98.21	90.83	99.33	100	99.99	96.96	98.52	99.97	100	99.69	99.50	100	100	100	99.99	98.75	100	0.45	48.87

A. Measure of Significance - Hypothesis testing

The hypothesis is to use the Common Features Techniques (CFT) approach for a lightweight intrusion detection model. This is due to the lower computational cost of the approach while providing the same classification rate as traditional feature selection approaches. Because of the small size of our data sample (n = 20), we chose the two-tailed t-test to test our hypothesis in this study. To determine whether the null hypothesis should be rejected or accepted, a statistical test is used. The t-test is an inferential statistical test that is used to determine whether there is a statistically significant difference between the means of two variables. Using information gain, chi-squared, gini index, and CFT as variables with computed average memory usage and computation time, we went on to compute the p-values for average memory usage and average computation time as follows: Common Features Technique vs Information Gain - (CFT vs Inf-Gain), Common Features Technique vs Chi-Squared - (CFT vs Chi-Squared), and Common Features Technique vs Gini-index (CFT vs Gini-index).

The Two-Sample t-Test

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$

$$Two\ test\ value : t = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}} \tag{2}$$

The p-value indicates the likelihood that the statistical measure, which determines whether an observed outcome should be accepted or rejected based on a 90% or 95% confidence interval (CI), will be accepted or rejected. The smaller the p-value, the more evidence there is in a sample of data to reject the null hypothesis and accept the alternative hypothesis. Statistical significance is established and defined when the p-value is 0.05 or less. First, we calculated the p-values for the computation time and memory in Tables IV through VII.

TABLE VIII
P-VALUES FOR TIME AND MEMORY FOR TABLE III

	P-values		
	CFT-vs-Inf_Gn	CFT-vs-Chi_Sq	CFT-vs-Gini_In
Time	4.41E-09	5.78E-09	7.52E-09
Memory	2.20E-16	2.20E-16	2.20E-16

TABLE IX
P-VALUES FOR TIME AND MEMORY FOR TABLE IV

	P-values		
	CFT-vs-Inf_Gn	**CFT-vs-Chi_Sq**	**CFT-vs-Gini**
Time	2.23E-04	2.76E-04	3.03E-06
Memory	2.20E-16	2.20E-16	2.20E-16

TABLE X
P-VALUES FOR TIME AND MEMORY FOR TABLE V

	P-values		
	CFT-vs-Inf_Gn	**CFT-vs-Chi_Sq**	**CFT-vs-Gini**
Time	5.33E-11	5.17E-14	2.20E-16
Memory	6.81E-02	2.20E-16	2.20E-16

B. Discussion

Table IV shows that, while the SVM model with the three other feature selection approaches had an overall accuracy of 88%, the model with the CFT features selection technique had an overall accuracy of 87%. Except for the Normal traffic, the table also showed that the sensitivity, which are the True Positives of the classes, indicated high classification. More importantly, the p-values in Table VIII show that within a 95% confidence interval, there is a statistically significant difference in the computation time and memory used by the CFT feature selection technique and the other approaches. As a result, we can conclude that the CFT technique achieved the same level of classification as traditional feature selection approaches while consuming far less computational resources.

According to the results shown in Table V, the random forest model with the CFT features selection technique achieved an overall accuracy of 93%, with the p-values in Table IX indicating a statistically significant difference in computation time and memory used. When compared to the model's classification using the other feature selection approaches, which was 95%, we can conclude that the random forest model fitted on the CFT dataset achieved a high classification comparable to the other feature selection approach at a lower computation cost.

Similarly, when the KNN model was fitted on datasets from the four feature selection approaches (Table VI), the CFT approach achieved 88% classification. This rate is comparable to the classification rate of the model using the other feature selection techniques. Interestingly, while the computation time difference between the model's CFT approach and the other approaches is statistically significant at 95% confidence intervals, the p-value for the memory difference between the CFT vs Information gain approach indicates that it should be rejected at 95% confidence intervals but accepted within 90% confidence intervals (Table X). As a result, we can conclude that the CFT features model achieved the same classification rate as the other techniques while incurring a lower computation cost.

Finally, as shown in Table VII, the random forest model with the CFT technique had a 91% classification rate. This is a lower classification than the model's other approaches, which achieved 98%. This drop in classification could, interestingly, be attributed to the CMRI class's 52% sensitivity classification. The p-values in Table XI, on the other hand, show that the p-values for computation time and memory used are less than 0.05, indicating that the CFT technique can achieve high intrusion classification at a low computational cost.

TABLE XI
P-VALUES FOR TIME AND MEMORY FOR TABLE VI

	P-values		
	CFT-vs-Inf_Gn	**CFT-vs-Chi_Sq**	**CFT-vs-Gini**
Time	8.12E-03	2.55E-08	4.60E-09
Memory	2.20E-16	2.20E-16	2.20E-16

V. CONCLUSION

In this paper, we propose a lightweight intrusion detection technique based on the Common Features Technique (CFT). The technique involves ranking features in order of importance using feature selection techniques, and then selecting a subset of the features to form new datasets using saturated cumulative variance as the threshold. Following that, a subset of common features between the sub-datasets was generated. The fitting of RF, SVM, and KNN models on a multiclass ICS dataset was performed, and the model's outcome was evaluated using a two-tailed t-test statistical significance technique. The models using the CFT techniques achieved a high classification with fewer features as the models utilising Information Gain, Chi-Squared, and Gini-index feature selection techniques. More importantly, the CFT technique accomplished the classification of the data at a lower computation cost to the devices. In terms of future work, additional opportunities to address the low sensitivity classification observed in some of the classes should be explored. In addition, we also hope to explore how we can use deep learning models to fit on the dataset and then optimise the deep learning model using 8 and 16 bits quantisation. The essence is to further cause a reduction in the size of the data. Interestingly, we have applied this approach to IoT binary datasets and obtained comparable classification results when compared to traditional feature selection approaches. Finally, the feature selection approaches could be expanded beyond the three techniques that we used.

REFERENCES

[1] Stouffer, K., Falco, J., & Scarfone, K. (2011). Guide to industrial control systems (ICS) security. NIST special publication, 800(82), 16-16.

[2] Pliatsios, D., Sarigiannidis, P., Lagkas, T., & Sarigiannidis, A. G. (2020). A survey on SCADA systems: secure protocols, incidents, threats and tactics. IEEE Communications Surveys & Tutorials, 22(3), 1942-1976.

[3] Phillips, B., Gamess, E., & Krishnaprasad, S. (2020, April). An evaluation of machine learning-based anomaly detection in a SCADA system using the modbus protocol. In Proceedings of the 2020 ACM Southeast Conference (pp. 188-196).

[4] Pan, X., Wang, Z., & Sun, Y. (2020). Review of PLC security issues in industrial control system. Journal of Cybersecurity, 2(2), 69.

[5] N. Rachburee and W. Punlumjeak, "A comparison of feature selection approach between greedy, IG-ratio, Chi-square, and mRMR in educational mining," 2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE), 2015, pp. 420-424, doi: 10.1109/ICITEED.2015.7408983.

[6] Kuncheva, L. I. (2002). A theoretical study on six classifier fusion strategies. IEEE Transactions on pattern analysis and machine intelligence, 24(2), 281-286.

[7] Wang, J., Xu, J., Zhao, C., Peng, Y., & Wang, H. (2019). An ensemble feature selection method for high-dimensional data based on sort aggregation. Systems Science & Control Engineering, 7(2), 32-39.

[8] Zhang, H., Li, J. L., Liu, X. M., & Dong, C. (2021). Multi-dimensional feature fusion and stacking ensemble mechanism for network intrusion detection. Future Generation Computer Systems, 122, 130-143.

[9] Rodríguez, D., Ruiz, R., Cuadrado-Gallego, J., & Aguilar-Ruiz, J. (2007, August). Detecting fault modules applying feature selection to classifiers. In 2007 IEEE International Conference on Information Reuse and Integration (pp. 667-672). IEEE.

[10] McLaughlin, S. (2011). On dynamic malware payloads aimed at programmable logic controllers. In 6th USENIX Workshop on Hot Topics in Security (HotSec 11).

[11] D. Mayor, K. K. Mookhey, J. Cervini, and F. Roslan. Metasploit Tookit: For Penetration Testing, Exploit Devevlopment, and Vulnerability Research. Syngress, 2007

[12] Aljawarneh, S., Aldwairi, M., & Yassein, M. B. (2018). Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. Journal of Computational Science, 25, 152-160.

[13] Ayo, F. E., Folorunso, S. O., Abayomi-Alli, A. A., Adekunle, A. O., & Awotunde, J. B. (2020). Network intrusion detection based on deep learning model optimized with rule-based hybrid feature selection. Information Security Journal: A Global Perspective, 29(6), 267-283.

[14] Zhou, Y., Cheng, G., Jiang, S., & Dai, M. (2020). Building an efficient intrusion detection system based on feature selection and ensemble classifier. Computer networks, 174, 107247.

[15] Li, L., Yu, Y., Bai, S., Cheng, J., & Chen, X. (2018). Towards effective network intrusion detection: A hybrid model integrating gini index and GBDT with PSO. Journal of Sensors, 2018, 1-9.

[16] Morris, T., Gao, W., "Industrial Control System Network Traffic Data sets to Facilitate Intrusion Detection System Research," in Critical Infrastructure Protection VIII, Sujeet Shenoi and Johnathan Butts, Eds. ISBN: 978-3-662-45354-4. Due November 14, 2014

[17] Morris, T. Srivastava, A., Reaves, B., Gao, W., Pavurapu, K., Reddi, R. A Control System Testbed to Validate Critical Infrastructure Protection Concepts. International Journal of Critical Infrastructure Protection (2011). Elseiver. doi:10.1016/j.ijcip.2011.06.005

[18] Otokwala, U., Petrovski, A., & Kalutarage, H. (2021, December). Improving Intrusion Detection Through Training Data Augmentation. In 2021 14th International Conference on Security of Information and Networks (SIN) (Vol. 1, pp. 1-8). IEEE.