Journal of trends in Computer Science and Smart technology (TCSST) (2020)
Vol.02/ No. 01
Pages: 15-25
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2020.1.002

Decision Tree Based Interference Recognition for Fog Enabled IOT Architecture

Dr. S. R. Mugunthan,
Associate Professor,
Department of Computer Science and Engineering,
Sriindu college of Engineering and Technology,
Sheriguda, Hyderabad, India.
Email: srmugunth@gmail.com

Abstract: The cyber-attacks nowadays are becoming more and more erudite causing challenges in distinguishing them and confining. These attacks affect the sensitized information's of the network by penetrating into the network and behaving normally. The paper devises a system for such interference recognition in the internet of things architecture that is aided by the FOG. The proposed system is a combination of variety of classifiers that are founded on the decision tree as well as the rule centered conceptions. The system put forth involves the JRip and the REP tree algorithm to utilize the features of the data set as input and distinguishes between the benign and the malicious traffic in the network and includes an decision forest that is improved with the penalizing attributes of the previous trees in the final stage to classify the traffic in the network utilizing the initial data set as well as the outputs of the classifiers that were engaged in the former stages. The proffered system was examined using the dataset such BOT-Internet of things and the CICIDS2017 to evince its competence in terms of rate of false alarm, detection, and accuracy. The attained results proved that the performance of the proposed system was better compared to the exiting methodologies to recognize the interference.

Keywords: Internet of Things, Network Security, Attacks Distinguishing System, Decision Tree and Rule Centered Conceptions, Fog Architecture

1. Introduction

The internet of things that has enabled the tangible things to communicate over internet, has become more and more vulnerable as it carries numerous of information's that are sensitized. Recently cyber-attacks that are highly erudite aims in attacking exclusively the systems that hold or compute information's that are very sensitive. Important infrastructures of the nations are the key targets of cyber-attacks, since vital information or services rely on their systems causing major problem for the both government administrations as well as the nation's security.

ISSN: 2582-4104

Journal of trends in Computer Science and Smart technology (TCSST) (2020)
Vol.02/ No. 01
Pages: 15-25
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2020.1.002

Bouts against these vital structures include penetrations into their network and the deployment of malicious software or programs that can expose sensitive data or alter the actions of specific physical equipment. To counter the progress in the interference, researchers and business professionals are working together to build new technologies and processes that can protect their technologies. In addition to other preventive protection measures, such as regulation on accessing and verification, the attack distinguisher system (ADS) are used as a second line of security. ADS is founded on particular rules or patterns of the system's usual behavior to distinguish the normal from the harmful attacks. Based on the distinguishing procedures followed the ADS are categorized as rule centered, hybrid and the ill use detector. The ADS are so far proffered under various taxonomies, and are capable of functioning in online by providing a constant monitoring in real time as well as in offline by operating over the data that are already gathered and reserved over a particular duration. The industrial regulation organization presently has devised ADS with new taxonomies that could be categorized as three different types, they are the methods based on the protocol examination, control process analysis and traffic mining.

The information that were collected from the previous attack were used as a fundamental sources to develop the identification system, better the identification process less would be the damages caused by the attacks in the network. But the network faces certain penalties due to the countermeasures taken than the attack itself. So to elude the consequences caused by the remedial measures and improve the performance of the attack detection system by finding even the attacks that are more similar to the network activities the proposed attack distinguishing system is devise, the ADS is developed is incorporated in the fog enabled internet of things architecture and validated using the BOT-internet of things dataset.

The proposed-ADS with the aim of distinguishing the common and the uncommon attacks with the higher rate of detection and less rate of false alarms is encompassed with the related works in section 2, proposed work in section 3, results analysis in section 4 and the conclusion in section 5.

2. Related Works

Maglaras et al [1] proposes the "Cyber security of critical infrastructures.", Smys, et al [2] presents the "DDOS Attack Detection In Telecommunication Network Using Machine Learning" Ferrag, et al [3] puts forth the "EPSA: an efficient and privacy-preserving scheme against wormhole attack on reactive routing for mobile ad hoc social networks." Suma, V et al [4] elaborates the "Security and Privacy Mechanism Using Block chain Leandros", Leandros et al [5] puts forward the "Teaching the process of building an Intrusion Detection System using data from a small-scale SCADA test bed." Anand, J. V. et al [6] proposes the "Design and Development of Secure and Sustainable Software Defined Networks."

ISSN: 2582-4104

Journal of trends in Computer Science and Smart technology (TCSST) (2020)
Vol.02/ No. 01
Pages: 15-25
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2020.1.002

Ahmim, et al [7] proposes the "A novel hierarchical intrusion detection system based on decision tree and rules-based models." Raj, Jennifer S. et al [8], puts forth the A Novel Classification via Clustering Method for Anomaly Based Network Intrusion Detection System Wang. et al [9] elaborates the designing of "A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering." Govindarajan, et al [10] proposes the "Intrusion detection using neural based hybrid classification methods." Chung. et al [11] puts forth the development of "A hybrid network intrusion detection system using simplified swarm optimization (SSO)."

Kim et al [12] proposes "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection." Kevric, et al [13] describes the designing of "An effective combining classifier approach using tree algorithms for network intrusion detection." Adnan et al [14] presents the "Forest PA: Constructing a decision forest by penalizing attributes used in previous trees." This is used as the third stage classifier in the proposed model to distinguish both the attack traffic and the specific type of attack Frank. et al [15] is the "Reduced-error pruning with significance Frank tests. Used in the parallel classification stage to distinguish between the normal network traffic and the attacked" Cohen, et al [16] puts forth the "Fast effective rule induction." This is used as second classifier employed in the parallel stage to distinguish between the specific types of the attacks. Koroniotis et al [17] develops a "the realistic botnet dataset in the internet of things for network forensic analytics." Folino et al [18] puts forth the "Evolving meta-ensemble of classifiers for handling incomplete and unbalanced datasets in the cyber security domain."

3. Proposed Frame Work

The inadequacy of the data set used in the machine learning is the major cause for the degradation in the performance of the classification algorithms. This was primarily due to two main reasons, one is the simple accuracy used as the key objective in the classification and the other one is due to the improper distribution of classes, leading to misjudging and inappropriate classification. Distinguishing the malicious as normal and the normal as the attacker.

To bring down the rate of false alarms and enhance the rate of identification as well as accuracy in distinguishing the network activities the paper devises an ADS constructed with three various classifiers that are founded on the decision tree and the rule centered conceptions to form a hierarchical architecture to recognize and distinguish the interferences. The block diagram below shows the proposed frame of ADS.

ISSN: 2582-4104

Journal of trends in Computer Science and Smart technology (TCSST) (2020)
Vol.02/ No. 01
Pages: 15-25
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2020.1.002

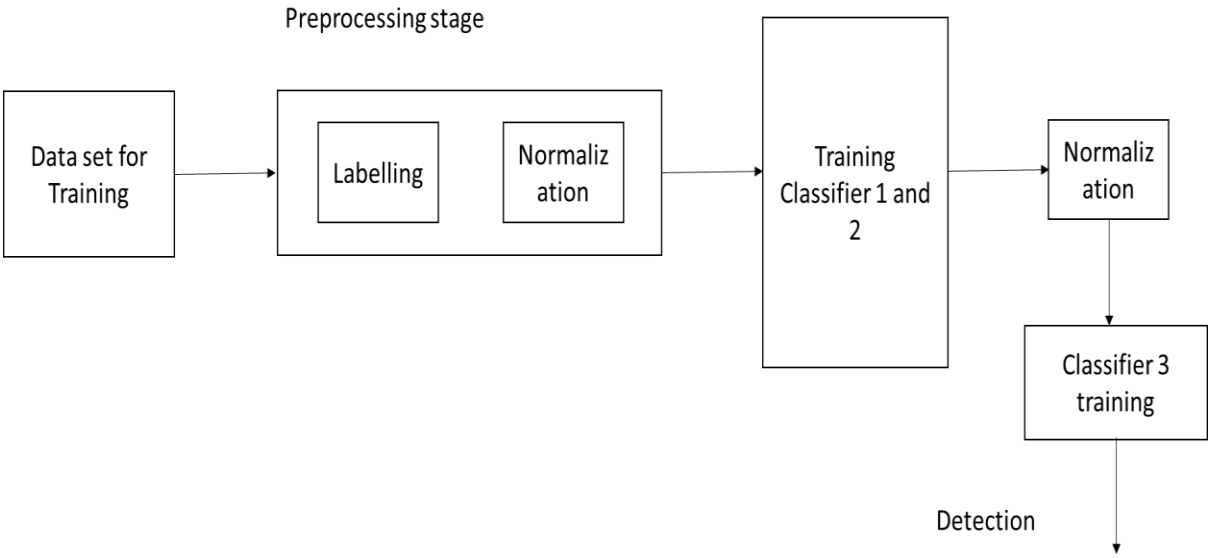

Figure.1 Proposed-ADS

The proposed ADS employs two classifier, one binary and one multiclass in parallel and engages another classifier in the final stage to reduce the consequences of the inadequacy in the data in the classification process. The hierarchical architecture put forth scopes to distinguish correctly among the normal and the malicious and to in return reduce the false alarms and maximize the rate of identification. The in parallel stage one classifier engaged distinguishes between the malicious and the normal traffic and the next one distinguishes the categories of bouts, using the various dataset features as input. The third classifier takes in the initial data set and the outputs of the parallel stage classifier as input and distinguishes the normal from the malicious as well the specific category of attack form the each row of dataset that is provided as input. The three classifiers used are reduced error pruning-REP tree [15], JRip [16], and the DF with penalizing attributes [14], the classifiers accurately classifies the datasets by framing rules according to the values of the features.

ISSN: 2582-4104

Journal of trends in Computer Science and Smart technology (TCSST) (2020)
Vol.02/ No. 01
Pages: 15-25
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2020.1.002

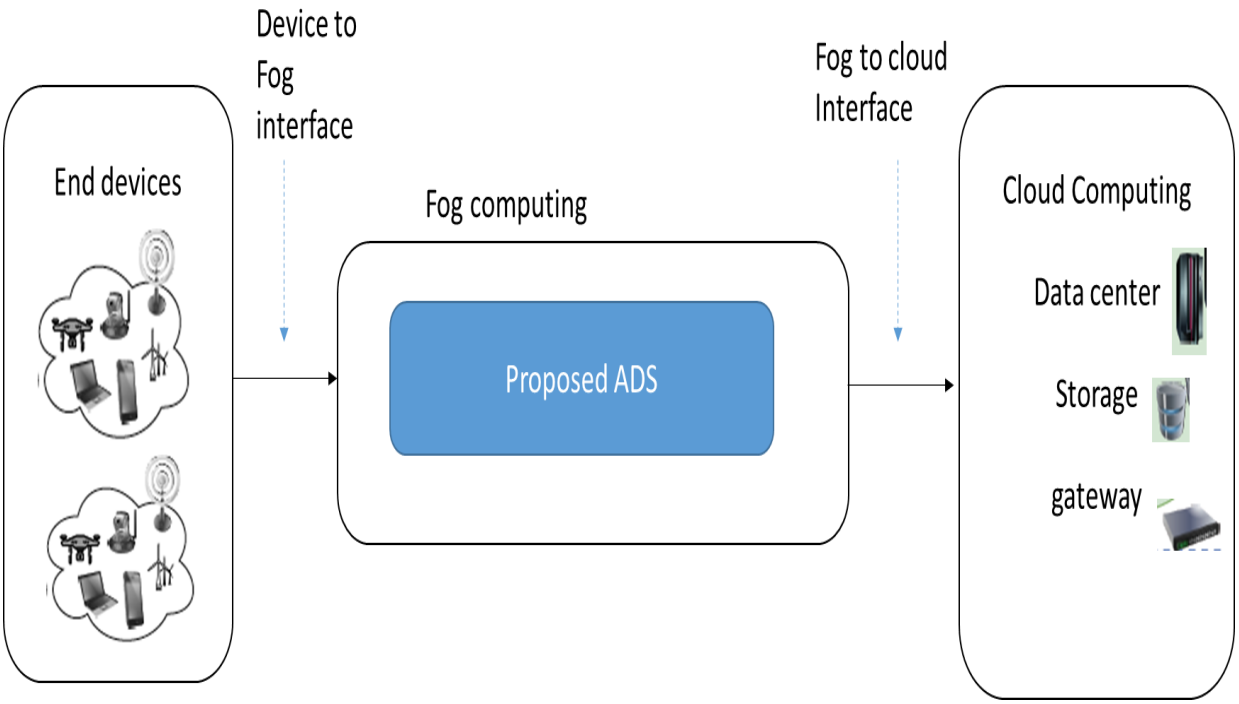

Figure.2 ADS in FOG Enabled IOT

The ADS is incorporated into the multi-tier FOG enabled IOT, and placed in the fog computing layer. The ADS distinguishes the malware from the normal by undergoing the training and the testing, the parallel stage classifiers are trained initially and the final stage classifiers are trained using the output of the parallel stage classifiers. In the parallel stage one distinguisher is trained with the attacks and the normal traffic and the other is trained using the normal and the specific category of attacks, this enables to identify both common and the uncommon attacks that affect the network. So we get two frame of data set, this is used in training the third classifier in the final stage, after the training of the each classifier is performed using the normalized features of the dataset. The preprocessing stages convert the symbolic values to numeric values and scales the numeric valued attributes in a greater numeric ranges dominating those in the smaller numeric ranges. The system proffered also enables to elude the numeric difficulties during the estimations. The normalization of the features (f) values (pi) is done using the equation 1

$$normalized_{value} = \frac{P_i(f) - \min(P(f))}{\max(P(f)) - \min(P(f))} \quad (1)$$

ISSN: 2582-4104

Journal of trends in Computer Science and Smart technology (TCSST) (2020)
Vol.02/ No. 01
Pages: 15-25
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2020.1.002

4. Result Analysis

In this portion, the data set used in conjunction with the data pre-processing method is discussed in detail. We also give measurements of the results used in our experiments. In addition, we're showing our concept structure. Finally, a comparative analysis is given between our model and that of different classifiers. The tests are carried out on a Windows 10, 64-bit PC with 8 GB of RAM and Intel(R) I5 2.7 GHz CPU. The proposed system is tested over the two data sets (i) CICIDS 2017 and the (ii) Bot-IoT dataset since both the datasets fulfills the eleven important features of a legitimate ADS dataset, namely Anonymity, Full Network Setup , Complete Collection, , Attack Diversity, Supported Protocols, Full Interaction, Set of Functions, , Complete Traffic Heterogeneity, Metadata and Labelling.

The Bot-IoT dataset encompasses, more than72, 000,000.001 records created on the 74 files with every row holding 46 features and the CICICDS contains 2, 30,743.01 rows segregated into 8 files and with every row holding 46 features. Rows with the identical feature are sorted out and eliminated after which the subsets that unique in each row are used for training, the rows are selected randomly after performing the process of suppression.

The table1 below s the total attacks of the data set used and the number of data used in training and testing.

Data set types	Attack Types	Training Data	Testing Data
CICIDS 2017	DoS slowloris	40,000.001	40,000.001
	DoS Goldeneye		
	Port scan		
	Web Attack-XSS		
	Benign		
BOT-IoT	Benign	5,877,647.001	1,468412
	DDoS TCP,uDP, Http		
	Key Logging		
	Data Theft		
	Service scanning		

Table .1 Dataset Utilized

ISSN: 2582-4104

Journal of trends in Computer Science and Smart technology (TCSST) (2020)
Vol.02/ No. 01
Pages: 15-25
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2020.1.002

The performance of the system is measured on the basis of accuracy, false alarm rate and the detection rate. The results obtained shows the improved detection rate achieved by the proposed method, the figure .3 depicts the detection rate of the proposed ADS.

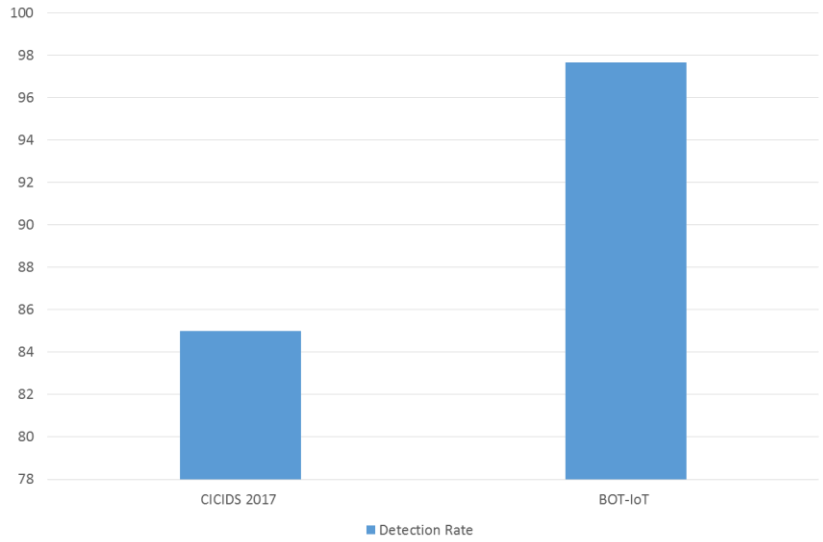

Figure.3 Detection Rate

The detection rate is estimated using the equation 2 shown below

$$Rate_{Detection} = \frac{attack\ type_{TP}}{attack\ type_{TP} + attack\ type_{FN}} \quad (2)$$

Where the 'TP 'and the 'FN' is the true positive and the false negative respectively.

The figure4 below shows the false alarm rate and the accuracy of the proffered system, for the two different types of dataset.

ISSN: 2582-4104

Journal of trends in Computer Science and Smart technology (TCSST) (2020)
Vol.02/ No. 01
Pages: 15-25
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2020.1.002

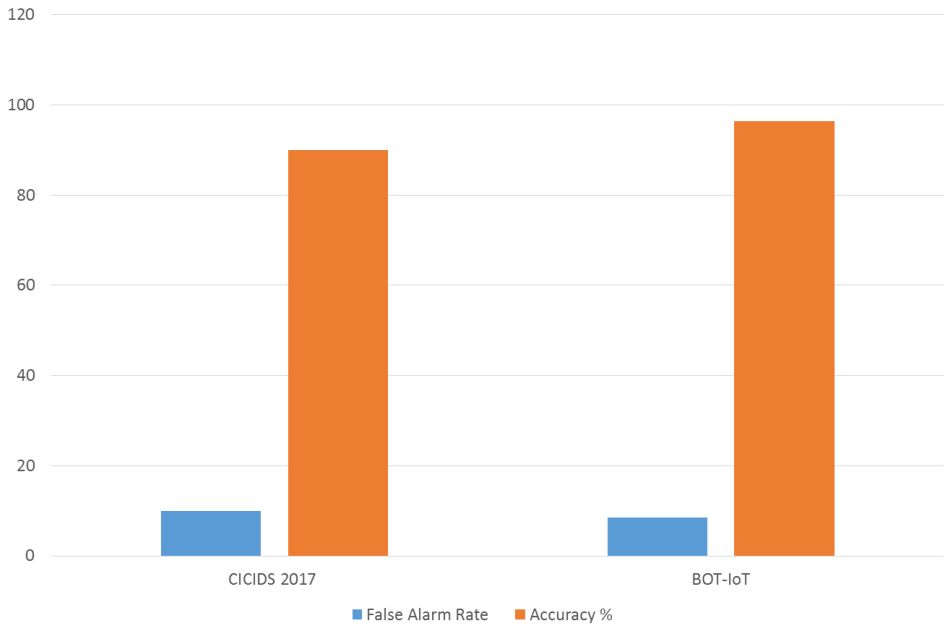

Figure. 4 False Alarm Rate and Accuracy

Where the rate of false alarm is estimated using the equation 3 and the accuracy is estimated using the equation 4.

$$Rate_{False\ Alarm} = \frac{normal_{TN}}{normal_{TN} + Normal_{FP}} \tag{3}$$

$$Accuracy\ \% = \frac{\sum specific\ type\ of\ attack_{TP} + normal_{TN}}{\sum specific\ type\ of\ attack_{TP} + \sum specific\ type\ of\ attack_{FN} + normal_{TN} + Normal_{FP}} \tag{4}$$

The Table.2 below provides the comparison of results obtained on the proposed method over the existing methods.

ISSN: 2582-4104

Journal of trends in Computer Science and Smart technology (TCSST) (2020)
Vol.02/ No. 01
Pages: 15-25
https://www.irojournals.com/tcsst/
DOI: https://doi.org/10.36548/jtcsst.2020.1.002

Methods	Dataset	Accuracy %	False Alarm Rate%	Detection Rate %
Conventional	CICIDS 2017	67.6	56	70
	BOT-IoT	65	45	62
Proposed	CICIDS 2017	89.9	10	85.6
	BOT-IoT	96.54	8.46	97.67

Table.2 Accuracy, Detection and False Alarm Rate Comparison

5. Conclusion

The interference recognize system or the ADS put forth in the paper is the combination of three various classifiers with two employed in parallel and one in the final stage where the output of the two parallel classifiers are fed. The proposed model was validated using two data set and was found outstanding compared to the other similar existing systems that were framed in this regard.

References

[1] Maglaras, Leandros A., Ki-Hyung Kim, Helge Janicke, Mohamed Amine Ferrag, Stylianos Rallis, Pavlina Fragkou, Athanasios Maglaras, and Tiago J. Cruz. "Cyber security of critical infrastructures." *Ict Express* 4, no. 1 (2018): 42-45.

[2] Smys, S. (2019). DDOS Attack Detection In Telecommunication Network Using Machine Learning. Journal of Ubiquitous Computing and Communication Technologies (UCCT), 1(01), 33-44.

[3] Ferrag, Mohamed Amine, Mehdi Nafa, and Salim Ghanemi. "EPSA: an efficient and privacy-preserving scheme against wormhole attack on reactive routing for mobile ad hoc social networks." *International Journal of Security and Networks* 11, no. 3 (2016): 107-125.

[4] Suma, V. (2019). Security and Privacy Mechanism Using Blockchain. Journal of Ubiquitous Computing and Communication Technologies (UCCT), 1(01), 45-54.

ISSN: 2582-4104

[5] Maglaras, Leandros, Tiago Cruz, Mohamed A. Ferrag, and Helge Janicke. "Teaching the process of building an Intrusion Detection System using data from a small-scale SCADA testbed." *Internet Technology Letters* 3, no. 1 (2020): e132.

[6] Anand, J. V. "Design and Development of Secure and Sustainable Software Defined Networks." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 1, no. 02 (2019): 110-120

[7] Ahmim, Ahmed, Leandros Maglaras, Mohamed Amine Ferrag, Makhlouf Derdour, and Helge Janicke. "A novel hierarchical intrusion detection system based on decision tree and rules-based models." In *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pp. 228-233. IEEE, 2019.

[8] Raj, Jennifer S., S. Smys, G. Josemin Bala, B. Praba, R. Sujath, V. Hilda Christy Gnanam, M. Abdul Rahiman et al. "Editorial Board v-vi A Novel Classification via Clustering Method for Anomaly Based Network Intrusion Detection System 1-6 Mrutyunjaya Panda and Manas Ranjan Patra Performance Improvement in MANETs: A Cross Layer Approach via TCP 7-11."

[9] Wang, Gang, Jinxing Hao, Jian Ma, and Lihua Huang. "A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering." *Expert systems with applications* 37, no. 9 (2010): 6225-6232.

[10] Govindarajan, Muthukumarasamy, and R. M. Chandrasekaran. "Intrusion detection using neural based hybrid classification methods." *Computer networks* 55, no. 8 (2011): 1662-1671.

[11] Chung, Yuk Ying, and Noorhaniza Wahid. "A hybrid network intrusion detection system using simplified swarm optimization (SSO)." *Applied soft computing* 12, no. 9 (2012): 3014-3022.

[12] Kim, Gisung, Seungmin Lee, and Sehun Kim. "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection." *Expert Systems with Applications* 41, no. 4 (2014): 1690-1700.

[13] Kevric, Jasmin, Samed Jukic, and Abdulhamit Subasi. "An effective combining classifier approach using tree algorithms for network intrusion detection." *Neural Computing and Applications* 28, no. 1 (2017): 1051-1058.

[14] Adnan, Md Nasim, and Md Zahidul Islam. "Forest PA: Constructing a decision forest by penalizing attributes used in previous trees." *Expert Systems with Applications* 89 (2017): 389-403.

[15] Frank, Eibe, and Ian H. Witten. "Reduced-error pruning with significance tests." (1999).

[16] Cohen, William W. "Fast effective rule induction." In *Machine learning proceedings 1995*, pp. 115-123. Morgan Kaufmann, 1995.

[17] Koroniotis, Nickolaos, Nour Moustafa, Elena Sitnikova, and Benjamin Turnbull. "Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset." *Future Generation Computer Systems* 100 (2019): 779-796.

ISSN: 2582-4104

[18] Folino, Gianluigi, and Francesco Sergio Pisani. "Evolving meta-ensemble of classifiers for handling incomplete and unbalanced datasets in the cyber security domain." *Applied Soft Computing* 47 (2016): 179-190.

ISSN: 2582-4104