Detecting IoT Botnet Attacks Using Machine Learning Methods

Celil OKUR
Gazi University
Graduate School of Natural and Applied Sciences
Information Security Engineering
Ankara, Turkey
celil.okur@gazi.edu.tr

Murat DENER
Gazi University
Graduate School of Natural and Applied Sciences
Information Security Engineering
Ankara, Turkey
muratdener@gazi.edu.tr

***Abstract*--Today, with the technological developments, the use of internet connected devices is increasing. It is a fact that life has become easier with the "Internet of Things (IoT), which contributes to the simultaneous operation of these devices with each other. IoT is a technology that designs and does the things people need to do - within a program - and increases the comfort of the user. All the advantages of IoT devices are valid as long as they work correctly and securely. However, when these devices do not work properly and securely or are abused by someone, their advantages as well as disadvantages emerge. The best example of this is the IoT-based Botnet attacks in 2016. Machine learning methods are used to prevent IoT-based attacks and planned attacks. The aim of this study is to detect the normal network traffic and attack traffic with high accuracy by using machine learning methods. The data set used is the N-BaIoT Provision 737E security camera data set, which includes normal network traffic and attack network traffic, and has been used in the literature. Machine learning has been carried out using this data set. The study was carried out in two ways, with and without supervision. EM (Expectation Maximization) algorithm was used while performing unsupervised learning and 76.73% success was achieved. In the application performed with supervised learning, the decision tree (J48) algorithm was used and 99.95% success was achieved. The application was carried out with the Weka 3.8 program.**

Keywords—IoT Botnets, N-BaIoT DDOS Attacks, Machine Learning, Cyber Security

I. INTRODUCTION

With the increasing use of IoT devices in health, military, industry, commercial and daily fields, it is observed that the number of devices connected to the internet is increasing day by day. According to researches, it is thought that the number of devices connected to the internet will be 24 billion in 2020 [1]. Undoubtedly, IoT devices [have a large share in this increase. The fact that most of these devices are small does not usually give the user the impression of a computer. However, in the cyber world, every device with an IP address is viewed as a computer. IoT devices have computer features in terms of generating packages, having IP addresses, and having an operating system. Users forget that these devices are connected to the internet and see them as devices that serve them, are harmless and make life easier. However, each of them is a computer that watches and follows the lives of its users. From this point of view, it can be easily seen that the devices in question are not so innocent about security and privacy. Each of the IoT devices are produced for different purposes. While performance and cost are in the first place in the production of IoT devices, security is in the next place. In addition, IoT devices are sometimes produced with limited hardware and software according to their intended use. The limited resources on them prevent the allocation of resources for security. Since the security standard changes according to countries and brands, a common security concept cannot be developed worldwide. This situation leads to deficiencies and weaknesses in the security point of IoT devices and makes these devices attractive against cyber attacks. In attacks, IoT devices are both directly targeted and used as a tool for other attacks. Botnet attacks are the leading attacks using IoT devices.

In this study, detection of IoT Botnet attacks with machine learning methods is explained. Botnet attacks are presented in the second chapter, and learning models are presented in the third chapter. In the fourth section, the development of the application is given. While the comparison results are given in the fifth section, the results of the study are given in the last section.

II. BOTNET ATTACKS

While IoT devices make our lives easier, they work simultaneously with other tools over the internet. This increases the number of devices connected to the internet day by day and attracts the attention of cyber attackers. The synchronous operation feature of the devices is used for different purposes with some changes made by the attackers. This leads to Botnet attacks. Botnet attacks have increased in popularity with the IoT. The best example of this is the Mirai Botnet attack in 2016 [2]. Cyber attacks on IoT devices can be listed as inaccessibleto the device, stealing the user's information and capturing the device for other activities. After the IoT devices are captured, these devices are assigned tasks in DDOS attacks. The number of devices used in DDOS attacks is important. The number of devices used in the attack determines the effect of the DDOS attack and the result to be obtained. That's why

978-1-6654-1863-8/20/$31.00 ©2020 IEEE

attackers want to reach more Internet-connected devices and join their team. These devices are mostly devices of unconscious users or IoT devices that are used incorrectly. Attackers, while taking over the management of IoT devices, usually either use the factory settings of the devices or take advantage of the weak, breakable structure of the passwords given to the devices even if the settings are changed. In short, the attackers capture the target device with factory settings information, user vulnerabilities or brute force attacks. Captured devices are made members of the organization established by the attacker, the Botnet. The member device is moved simultaneously with other devices on the instruction of the administrator during the attack. Botnet life cycle; It consists of four stages: the occurrence stage, the command and control stage, the attack stage and the post attack stage [3]. If the network behaviors of botnet member IoT devices are detected and shaped, it is possible to protect the network from these devices. Thus, normal network traffic continues to be transmitted without being affected by the situation. Some approaches have been developed in the literature to prevent such attacks using machine learning. These approaches can be listed as signature-based, anomaly-based and hybrid-based where both are used in different combinations [4]. The first of these methods is to detect malicious traffic by comparing the malicious traffic signatures in the database with the signature of the incoming traffic. The signature here; It is obtained by passing the information of incoming traffic through the hash algorithm. The second approach is the one developed on the basis of abnormal events in network traffic. The third approach is to use two approaches together according to the needs of the network. The working logic of firewalls and programs used in information systems is based on these approaches. With the above approaches, the way firewalls work is divided into two as intrusion detector (IPS) and intrusion prevention (IDS) [4]. The model developed in this study is in the field of intrusion detection.

III. MACHINE LEARNING MODELS

The increase of IoT devices day by day causes an increase in network traffic. These traffics can sometimes contain a large number and variety of attacks. This disadvantageous situation can be turned into an advantageous situation by analyzing the network traffic data sets and performing a learning. The realized learning can be used in applications and devices with an algorithm. Machine Learning takes place in 4 steps. Network data is collected first. Collected data cannot be used directly in machine learning algorithms. Network traffic packets are made available to learning after necessary analysis, transformation and simplification processes. Machine Learning model is determined for the adapted data. The Machine Learning model can be chosen from either supervised learning models or unsupervised learning models. Training is carried out with the determined model. Training is tested in the next stage. Data used in education or appropriate new data can be used in the test process. In the last stage, the success of the learning is evaluated. Evaluation of success is made according to the accuracy matrix. After learning, the model training is completed and it is ready to use with other data. Choosing the appropriate model is important in machine learning. As mentioned above, machine learning models are divided into two as supervised and unsupervised.

Supervised learning is the distribution of a data set among different classes using pre-labeled data. The algorithm used in classification learns this classification from the tagged training set. The data in the training set is labeled and applied. Thus, it is determined in advance which data will be in which class. The machine interprets and classifies the next, unclassified data as it learns. In other words, the machine is expected to carry out the training with the data whose classes are certain, then give this data to the trained model and classify them. This type of learning model is called a supervised learning model. Logistic Regression, Decision Trees, Linear Regression, Support Vector Machines, Nearest Neighborhood are examples of supervised learning models. In this study, decision trees algorithm is used for supervised learning. Decision trees algorithm, named C4.5 in the first version, was named J48 with some changes made on it [5]. Decision tree algorithm works with classification logic by dividing the data from upper level to lower level. In decision tree learning, class labels at the level of the leaves of the tree by creating a tree structure and the manipulations on the features with the branches leading to these leaves from the beginning are expressed [6]. In the unsupervised learning model, data sets are divided into clusters. The main basis of this clustering is that the data in the same cluster has more similar features, but the similarity ratio between the clusters is the least. So this method is unattended, unlike classification. The clustering process is completely left to the machine. The machine applies a clustering process to the data using various algorithms and in doing so, the above mentioned path is followed. Clustering, EM, PCA are examples of unsupervised learning models. Expectation-Maximization (EM) method, one of the unsupervised learning methods, was used in this study. EM is an iterative search method used to find the greatest likelihood or the largest aftershock estimates of the parameters of statistical models that are dependent on variables not observed in statistics [7].

IV. IMPLEMENTATION OF THE APPLICATION

In this study, network data sets formed by Provision 737E model security cameras, one of the N-BaIoT data sets, were used [5]. Since the data sets were also used in previous studies and were pre-processed, in this study, no pretreatment was applied to the data set except for feature size reduction. This data set was obtained by simulating normal and abnormal network traffic behaviors in the laboratory environment [8]. The data set consists of malicious network traffic and attack traffic by IoT devices.

With the normal network traffic data generated by the Provision 737E security camera, network traffic data (in packet type such as udp, tcp) during the DDOS attack was obtained and supervised and unsupervised learning was performed through the Weka program. Due to its Weka feature, it can preprocess the data. In addition, Weka offers different algorithm options ready to use under these two learning models. In addition to these, Weka is a program that has options such as feature extraction and sorting by feature weight. The size of the data is 876 MB and consists of 828260 lines. The data set consisting of different types of network traffic packets (tcp, udp) has been reduced to a single file using the Python programming language. Combined data is saved in csv file type so that it can be processed in Weka program. The size of the data set examined in this study includes 115 features. Using all of these 115 features causes both software and hardware problems [8]. This situation causes the data set to not be evaluated properly with the algorithms in Weka. The feature reduction process was carried out by using the One -R Attribute Evaluator feature in Weka and sorting was made according to the weight of the features. 10 features were extracted according to their importance and studies were made with these features. Related features are given in Table I. These features are MI (L1, L3, L5, L0.1, L0.01) and H (L1, L3, L5, L0.1, L0.01).

TABLE I. FEATURES USED IN THE STUDY

Number	10 Features Selected from 115 Features
1	MI_dir_L0.1_Mean
2	H_L0.1_Mean
3	MI_dir_L0.01_mean
4	MI_dir_L1_mean
5	H_L1_Mean
6	H_L0.01_Mean
7	H_L5_Mean
8	MI_dir_L3_Mean
9	H_L3_Mean
10	MI_dir_L5_Mean

In this study, after extracting the features in order of importance, the J48 algorithm, one of the decision tree classification algorithms, was used while the supervised learning was performed in the first part of the learning. Cross-validation technique is used while learning is performed in this algorithm. Cross-validation determines how far the training dataset is split and how the segregated parts will be modified in each training cycle. In this study, training was carried out by setting the cross-validation value to 10. If the cross-validation value is 10, it means that the data set to be processed will be divided into 10 equal parts and the training will be carried out by progressing one piece in each training. This enabled each of the data set divided into 10 parts to be used both in education and testing [9]. In the study, Receiver Operator Characteristics Curve (ROC) curves of these classes were examined after the classification process was successfully performed with this method and the accuracy rate was calculated as 99.95%. ROC is a probability curve used in machine learning to evaluate performance after learning occurs. The area under the ROC curve is called AUC. Area Under the Curve (AUC) in machine learning provides information about the performance of the developed model. The size of the field is directly proportional to the success of learning [10]. According to the classification process types of data sets (Benign, Bashlite, Mirai), ROC graphics are as in Figure 1, Figure 2, Figure 3, respectively.

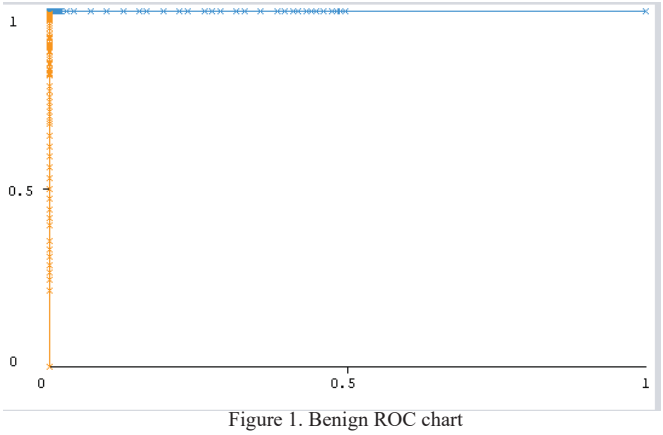

Figure 1. Benign ROC chart

Figure 1 shows the ROC traffic of normal network traffic. The size of the area under the curve shows the success of the classification.

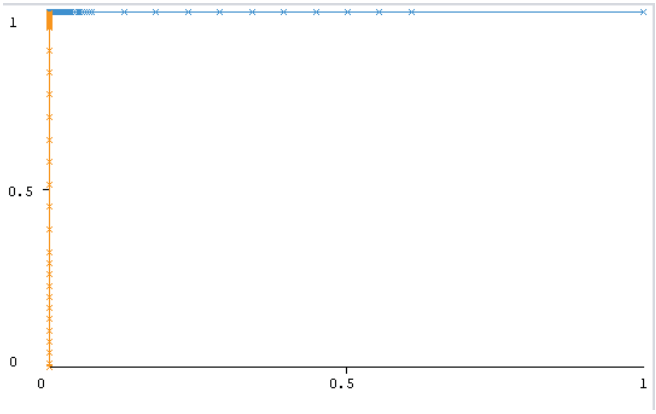

Figure 2. Bashlite ROC chart

ROC graphics of the attack traffic are given in Figure 2 and Figure 3. The size of the field shows the success of learning. When the areas of the curves in the ROC graphs given in the figures are examined, it is seen that the separation process from all data according to their classes has been successfully done.

In the continuation of the study, unsupervised learning was carried out. In this model, EM algorithm is used for learning. As expected, 3 clusters were formed as a result of the process. The accuracy percentage is calculated as 76.73%. Cluster formations have been visualized as seen in Figure 4. While doing this, the graphical feature of the Weka program was used.

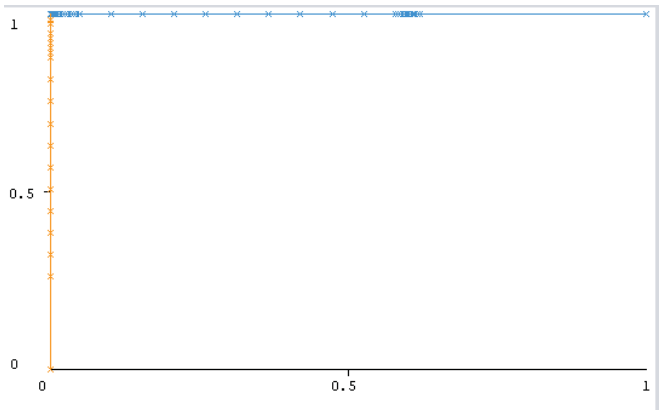

Figure 3. Mirai ROC chart

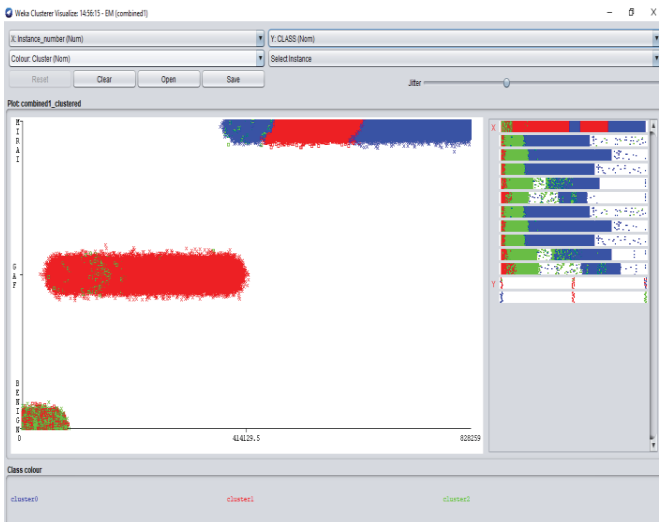

Figure 4. Benign, Bashlite, Mirai data scatter plot

When Figure 4 is examined, it is seen that the traffic is gathered in three groups. It is seen that some colors are mixed with each other in the graphic. This shows the deviations, that is, incorrect clustering resulting from learning, even if a little.

TABLE II. DISTRIBUTION OF DATA SET

Benign	Bashlite	Mirai
61928	219	7
139	329948	9
6	23	435981

In addition, the numerical distribution of the packages according to the groups is shown in Table II after learning.

V. COMPARISON RESULTS

Ten studies found as a result of the literature review were examined and the success rates of these studies were compared. Looking at Table 3, it is striking that the accuracy percentages are quite high, in this case it shows that the learning was successful. Similarly, in this study, the accuracy percentages are 99.95% in supervised learning; it was found to provide an accuracy of 76.73% in unsupervised learning. As in this study, feature reduction was performed in only one of the reviewed studies. The reason why decision tree algorithm is used is that it gives higher accuracy rate when compared with the other learning supervised models. Besides, why EM algorithm is used is that there are no other studies (at least to our knowledge) used EM algorithm directly for NBaIoT data set.

When the studies in the literature are examined, in addition to the data set examined in this study, attack traffics obtained both in the laboratory environment and in the real environment are also examined in the data sets. Most of the studies reviewed in this part used the N-BaIoT data set, which was also used in this study. In some of the studies only supervised learning models were used, whilst in some of the studies both supervised and unsupervised models were used similar to this study. The Weka program, which includes ready-made tools, was used in the studies based on the researcher's preference. In terms of capacity maintenance, Weka can give proper results with limited size data sets. As in some of the studies, feature size reduction was also performed. In all of the studies, the success percentage was specified. Accuracy, F1 score, precision, recall values were counted as success criteria. In this study, the decision tree model was used. In the literature, another study was also conducted using the same data set as thi model, but the accuracy percentage result was found to be less than this study. The N-BaIoT data set was examined for the first time with EM algorithm.

In [11], the hierarchical clustering, X-Means clustering, and rule-based classification were used for achieving fast and accurate recognition of bootnet attacks. While X-Means algorithm led to the highest cohesion inside the clusters and the maximum distance between clusters by choosing optimal K, each cluster with the similar flow is placed in a bot cluster, a semi-bot cluster or a normal cluster with the help of rule-based classification. Through network traffic flow analysis with the help of proposed method, sets of botnets have been evaluated and the results indicated that more than 95% accuracy in detection. In another s study [12], a framework especially for IoT devices identification and malicious traffic detection were proposed. The framework extracts feature per network flow to identify the source, the type of the generated traffic, and to detect network attacks by pushing the intelligence to the network edge. Consequently, various machine learning algorithms were compared with random forest, which gives the best results: up to 94.5% accuracy for device-type identification, up to 93.5% accuracy for traffic-type classification, and up to 97% accuracy for abnormal traffic detection. [13] used three sets of experiments with the purpose of to reveal the effectiveness of classification methods applied to the problem of network-based botnet detection, and to offer opportunities for improvement based on careful selection of a small subset of attributes. While the first set of experiments demonstrated very high accuracy in classifying network

TABLE III. LITERATURE STUDIES

Unique ID	Supervised/ Unsupervised	Method (Algorithm)	Purpose (Prediction, Classification	Data Set Name	Cross Validation/ Seperated Data	Success Rate
[11]	Unsupervised	The Hierarchica l Clustering, Xmeans Clustering, And Rule-Based Classification.	Accuracy	Virut, Agobot, Rbot, Zeus, And Njrat 2013	NA	%95
[12]	Supervised	Decision Tree, Random Forest, Neural Network	Accuracy	Experime ntal Set Up (Wifi Devices)	Cross- Validat ion 4fold	%97
[13]	Supervised	ZeroR, OneR	Accuracy	N-BaIoT	Cross Validat ion 10	%95
[14]	Supervised	SVM	Predict ion	N-BaIoT	NA	%90
[15]	Supervised	KNN, SVM, DT, Random Forest and Artificial Neural Networks	Classification	MTC	NA	%99
[5]	Supervised	Decision Tree	Classification	N-BaIoT	Cross Validat ion 10 Fold	%99 .87
[16]	Supervised	SVM	Classification	N-BaIoT	NA	%99
[17]	Supervised/Unsup ervised	RandomForest, k-NN, Gaussian Naive Bayes	F1 Score	Experimental setup	NA	%96
[18]	Supervised	Ensamble-KNN, DT, MLP	Recall, Precision, Accuracy, F1-Score	N-BaIoT	NA	%99 %99 %99 %49

activity into the 11 classes as well as significant redundancy in the 155 attributes; the second set of experiments focused on identifying the attributes which are the most beneficial for network-based botnet detection and on quantifying this benefit. In contrast to the first set of experiments, the third sets of experiments chose attributes in order of merit scores. The results demonstrate the benefit of this method by exhibiting high accuracy with 20 to 30 attributes. Consequently, in a deployment using fewer attributes is likely to ease the load on the monitoring infrastructure suggesting that such trimming need not incur a penalty in accuracy. The focus of [14] is that feature selection procedure can reduce the required number of features in an unsupervised learning model providing anomaly-based detection function in IoT networks. Reduced feature (reduction from 115 to 10) set helps less consumption of computational resources and more interpretable results. To contribute to the development of the timely DDoS traffic detection generated in the environments like smart home environment, [15] seeks to establish the diversity of traffic generated by IoT devices in such environments with respect to the traffic generated through human type communication. The results of the study are expected to represent base for the future development of new models aimed at detecting this specific DDoS traffic type. In [5], the researchers propose and empirically evaluate a novel network-based anomaly detection method which extracts behavior snapshots of the network and uses deep autoencoders to detect anomalous network traffic emanating from compromised IoT devices. To evaluate their method, they let their nine IoT devices get defected with Mirai and BASHLITE. The results revealed that their method is capable of detecting the attacks accurately and instantly. Last but not the least, [16] proposed another model to identify IoT botnet attacks from compromised IoT devices by exploiting the efficiency of a recent swarm intelligence algorithm called Grey Wolf Optimization algorithm (GWO) and to optimize the hyperparameters of the OCSVM and as well as finding the features that best describe the IoT botnet problem. With the aim of showing the efficiency of the method, its performance is evaluated using typical anomaly detection evaluation measures over a new version of a real benchmark dataset. The results showed that the method proposed in [16] performed better that all other algorithms regarding true positive rate, false positive rate, and G-mean for all IoT device types. Furthermore, the method was capable of having the lowest detection time and reducing the number of selected features. In [17], the network traffic classification was carried out using the data of the Mirai botnet attack in 2016 with the technique developed in the form of EDIMA. The classification was done with the machine learning models Naive Base KNN and Random Forest Learning models. Information about the results are presented in *Table3.* It was carried out in an experimental environment. With the aim of having the correct classification of normal and harmful POST and get traffic packets, three different machine learning models were used. At the end, it was seen that the highest score was obtained with 96% with the KNN algorithm. With the aim of highlighting limitations and particularities of individual algorithms for network traffic classification, [20] presents a comparative analysis among meta-learning approaches and individual classifiers to classify network traffic. Investigating and evaluating a range of meta-learning techniques like Voting, Stacking, Bagging and Boosting, the study proposes a new experimental analysis of different meta-learning techniques and compare them with their own base classifiers when used individually. Then, regarding the emerging popularity of Neutral Networks, the study analyzed this scenario using the Multi-layer Perceptron classifier. Data provided by the UCI Machine Learning Repository were used in the experiments. As a result, the best performance was obtained by an ensemble technique (Bagging), which obtained accuracy of 99.972% and false positive rate of 0.00018%.

As can be seen in Table III, some of the studies were conducted with supervision and some without supervision. In addition, cross-validation was used in different folds in the studies. Some of the studies used the data set used in this study. Percentage of accuracy is shown as the evaluation matrix for all studies.

VI. CONCLUSION

In this study, machine learning has been carried out by using the network traffic data of Provision 737E security cameras, one of the N-BaIoT data sets. The aim of the study is to distinguish between normal traffic and attack traffic in a network with high accuracy through machine learning. Training was carried out with the machine learning methods applied in the study, and then, by using this training, the machine was expected to distinguish between normal traffic and attack traffic. The result of the work carried out was evaluated according to the accuracy percentage as in other machine learning. Since smooth and accurate feature extraction is important in machine learning, the dimension with 115 features was reduced to 10 dimensions in this study. Feature extraction application of Weka program was used while performing feature reduction. In this way, computational complexity has been reduced. In the next step, the data set was trained with two different learning methods. Training was carried out using the Decision Tree (J48) algorithm in the supervised learning model. During the training, 10fold option was preferred in the cross-validation section. The accuracy percentage in supervised learning was calculated as 99.95%. EM algorithm is used in unsupervised learning method. The accuracy rate was calculated as 76.73%. Working with two different learning models has been successfully completed.

REFERENCES

[1] Ş. Kılınç, "2020 Yılında İnternete Bağlı 24 Milyar Cihaz Olacak!", Webtekno,2020.[Online].Available:https://www.webtekno.com/2020-yilinda-internete-bagli-24-milyar-cihaz-kullanimda-olacakh30300. html. [Accessed: 09- Nov- 2020].

[2] C. Kolias, G. Kambourakis, A. Stavrou and J. Voas, "DDoS in the IoT: Mirai and Other Botnets," Comp, vol. 50, no. 7, pp. 80-84, 2017, doi: 10.1109/MC.2017.201.

[3] J. Leonard, S. Xu and R. Sandhu, "A Framework for Understanding Botnets," 2009 International Conference on Availability, Reliability and Security, Fukuoka, 2009, pp. 917-922, doi: 10.1109/ARES.2009.65.

[4] S. Dua, X. Du, Data Mining and Machine Learning in Cybersecurity Book," New York: CRC Press, 2011.

[5] Y. Meidan et al., "N-BaIoT—Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders," IEEE Perva Comp, vol. 17, no. 3, pp. 12-22, Jul.-Sep. 2018, doi: 10.1109/MPRV.2018.03367731.

[6] S. Patil and U. Kulkarni, "Accuracy Prediction for Distributed Decision Tree using Machine Learning approach", 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 1365-1371, doi: 10.1109/ICOEI.2019.8862580.

[7] B. Karaçalı, "Improved quasi-supervised learning by expectation-maximization," 2013 21st Signal Processing and Communications Applications Conference (SIU), Haspolat, 2013, pp. 1-4, doi: 10.1109/SIU.2013.6531366.

[8] H. Bahşi, S. Nõmm and F. B. La Torre, "Dimensionality Reduction for Machine Learning Based IoT Botnet Detection," 15th International Conference on Control, Automation, Robotics and Vision (ICARCV), Singapore, 2018, pp. 1857-1862, doi: 10.1109/ICARCV.2018.8581205.

[9] M.B. Durna, "Cross validation nedir? Nasıl çalışır?", 2020. [Online]. Available: https://medium.com/bilişim-hareketi/cross-validation-nedir-nasıl-çalışır-4ec4736e5142. [Accessed: 09- Nov- 2020].

[10] D. K. Mclish, "Analysing a Portion of the ROC Curve", Soc of Medical Decision Making, vol 9, pp. 190-195,1989.

[11] P. Amini, R. Azmi and M.A. Araghizadeh, "Analysis of network traffic flows for centralized botnet detection" Jour of Telecomm., Elec. And Elec. Engineering, vol 11, pp. 7-19, 2019.

[12] O. Salman, I. Elhajj, A. Chehab and A. Kayssi, "A machine learning based framework for IoT device identification and abnormal traffic detection", Trans on Emerging Telecomm. Techn, 2019. Available: 10.1002/ett.3743.

[13] S. S. Chawathe, "Monitoring IoT Networks for Botnet Activity," IEEE 17th International Symposium on Network Computing and Applications (NCA), Cambridge, MA, pp. 1-8, 2018. doi: 10.1109/NCA.2018.8548330

[14] S. Nõmm and H. Bahşi, "Unsupervised Anomaly Based Botnet Detection in IoT Networks," 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, 2018, pp. 1048-1053.

[15] I. Cvitic, D. Prekovic, M. Perisa and M. Botica, "Smart home IoT traffic characteristics as a basis for DDoS traffic detection", 3rd EAI International Conference on Management of Manufacturing Systems, Dubrovnik November 06-08, 2018.

[16] A.Al Shorman, H.Faris, and I. Aljarah, "Unsupervised intelligent system based on one class support vector machine and Grey Wolf optimization for IoT botnet detection", J Ambient Intell Human Comput, 11, pp. 2809– 2825, 2020. Available at: https://doi.org/10.1007/s12652-019-01387-y.

[17] A. Kumar, T. J. Lim, 2019, "EDIMA: Early Detection of IoTMalware Network Activity Using Machine Learning Techniques",IEEE 5th World Forum on Internet of Things.

[18] I.P.Possebon et all, 2019 "Improved Network Traffic Classification Using Ensemble, IEEE Symposium on Computers and Communications (ISCC).