International Journal on Applications in Engineering and Technology
Volume 9 : Issue 1 : February 2023, pp 22 – 27 www.aetsjournal.com ISSN (Online) : 2455 - 0523

BOTNET ATTACK DETECTION USING MACHINE LEARNING

Mr.K.Giri[1], Mrs.N.Jayanthi[2], K.Sabari[3]

[1][2] Associate Professor, Dept. of Electronics and Communication Engineering, Mahendra Engineering College, Namakkal, TamilNadu, India.

[3]Student, Dept. of Electronics and Communication Engineering, Mahendra Engineering College, Namakkal, TamilNadu, India.

ABSTRACT

Security dangers are emerging at a rapid rate as computers and technology progress. Botnets are one such security problem that demands much investigation and commitment to eradicate. In this study, we apply machine learning to detect Botnet assaults. Using the Bot-IoT and University of New South Wales (UNSW) datasets, four machine learning models based on four classifiers are built: Nave Bayes, K-Nearest Neighbor, Support Vector Machine, and Decision Trees. Using 82,000 records from the UNSW-NB15 dataset, the decision trees model produced the best overall results, with 99.89% testing accuracy, 100% precision, 100% recall, and 100% F-score in identifying botnet assaults.

Keywords: IoT, botnet, machine learning, computer security, DDoS, cyber-attack, classification

I. INTRODUCTION

As the Internet of Things (IoT) evolves, more everyday home products are becoming internet-connected [1]. This opens the door for additional devices to possibly become botnet devices. The goal of this work is to detect botnet assaults using Machine Learning techniques. A botnet is a collection of internet-connected devices that have been infected with malware on purpose by cyber hackers. Botnets can be used to launch distributed denial-of-service (DDoS) attacks, steal data, or gain access to devices. A botnet assault is a sort of malicious attack that use a network of linked computers to attack or bring down a network, network device, website, or IT environment. Various machine learning techniques may be used to identify and separate these botnet devices when additional devices become candidates to be botnet devices. This research intends to increase the accuracy of previous related work by detecting botnets or illicit traffic behaviour using new machine learning techniques.

The structure of the essay is as follows. The literature evaluation of relevant work is included in Section II. The proposed technique is presented in Section III. A description of the experimental findings is included in Section IV. The conclusion and further work are presented in Section V.

II.REVIEW OF LITERATURE

Many studies have been conducted in recent years that demonstrate the efficiency of employing Machine and Deep Learning in identifying botnet assaults, which have been on the rise.Some studies also concentrate on identifying the essential elements or characteristics of a botnet that can aid in distinguishing between an attack and regular traffic. The application and efficacy of machine learning in botnet identification. They examined the structure of botnets in order to identify critical criteria that can distinguish botnet traffic from regular traffic. These characteristics may then be used to pick features for our machine learning model. A honeypot was used to entice attackers and generate data from an IoT network in one such way. This new data was then utilised to examine various aspects of an attack, such as IP addresses, MAC addresses, packet size, and so on. The concept of applying hybrid feature selection models to lower the size of the feature set in order to obtain reliable results.The dataset utilised has 115 characteristics, which is quite huge for any dataset. To limit the amount of features, feature selection was performed using the filter, wrapper, and hybrid models. These characteristics were then loaded into a K-Nearest Neighbor (K-NN) and Random Forest model, with both models achieving 99% accuracy.

The Decision Tree (DT) Classifier is one of the most promising classifiers in P2P botnet identification. For botnet identification, they employed a variety of classifiers as well as clustering.Over 38,000 records of network traffic were utilised in the sample, which included both attack and routine traffic. The DT classifier has the highest accuracy of 90.2723% in this article, followed by the Decision Tree classifier with an accuracy of 87.7853%. Similarly, P2P botnets were difficult to identify due to their usual centralization and dispersion characteristics. In 2013, a two-stage detection approach for P2P botnets was developed. To filter non-P2P traffic, the first stage included port judgement, DNS query, and data flow count. The second step employed session characteristics to decrease the amount of packets evaluated.To categorise and identify the traffic, machine learning methods were also deployed. The experiment

was carried out using the CTU-dataset, which comprises 13 distinct botnet samples.

To detect P2P botnet traffic, three primary ML algorithms were developed based on session characteristics. Nave Bayes (NB), DT classification, and ANN were the methods employed.

The detection rate using NB and ANN was 75.5% and 93.8%, respectively, however the DT method had an accuracy of 94.4%. This demonstrated that the two-stage approach of P2P traffic filtering and the DT classifier based on session characteristics could detect P2P botnet traffic successfully.Random forest and decision tree classifiers are two other excellent classifiers.Stevanovic and Pedersen [7] investigated how supervised machine learning may be used to detect botnets with high accuracy. They first suggested a botnet detection method that use flow-based traffic analysis and supervised machine learning to identify botnets. They then put eight of the most significant machine learning algorithms (MLAs) for categorising botnet traffic to the test. Finally, they investigated how much traffic must be seen for categorization to be successful. Traffic analysis was performed using either "batch" analysis, which monitors from the beginning to the conclusion of the trace, or "limited" analysis, which limits time intervals and packet counts.

The trials were carried out utilising the ISOP dataset, which contains both malicious and non-malicious entries. While the random forest classifier had the greatest accuracy in botnet detection, the random tree classifier was deemed optimum because it had the best balance of accuracy and detection time. Hoang et al. [8] offered an assessment of botnet detection models utilising machine learning techniques in contrast to anomaly-based botnet detection approaches in a recent study published in 2018. For their machine learning model, the authors employed K-NN, C4.5, random forests (RF), and NB classifiers.They opted to employ Domain Name Service's superior and the classifier with the greatest performance for the machine learning model's success. The results indicated that random forest has an overall accuracy of 90% in detecting botnets.

Six distinct characteristics of botnet domain traffic were identified using DNS data. The chosen features were all name-based, such as the relevant length ratio. Then came message-based features including the number of source IPs, kinds, and A, AAAA, NS, and MX inquiries.

Finally, there are quantity-based features, such as the total number of inquiries per day and the amount of querying done per hour. Following the selection of characteristics, three prominent ML classifiers were employed to identify malicious domains in DNS traffic.Adaboost, Bagging, and NB were the classifiers employed. The findings indicated that all classifiers performed well, with accuracy rates over 90% and just slight discrepancies between them. Such findings clearly indicate the effectiveness of detecting and halting harmful botnet activity when their domain names occur in traffic. Jin et al. propose that in future investigations, this technology be used with bigger DNS logs.

Three machine learning algorithms were tested for their capacity to distinguish between botnet and regular traffic. This was accomplished by picking critical elements of network traffic and then arranging them in various combinations to generate many test cases for the algorithms.The NB, Nearest Neighbour (IBk), and J48 algorithms were examined. The detection accuracy of the J48 and IBk algorithms was greater than that of NB after testing each algorithm with all of the examples independently, although with the constraints that J48 required a long training period and IBk required a long testing time. Despite these restrictions, the detection of P2P botnets was possible due to the high detection rates of more than 99%.

The SVM classifier was also shown to be particularly successful in detecting P2P botnets utilising network activity analysis and machine learning. In their experiment, they employed two datasets totaling over 370,000 that exclusively included fraudulent activity.They utilised a labelled dataset with a million packets of typical traffic received from Ericsson Research Traffic Lab for normal traffic. They employed K-NN, Linear Support Vector Machine (SVM), Artificial Neural Network, Gaussian Based Classifier, and NB Classifier as classifiers.

The results of their investigation revealed that SVM had the longest training and classification times. SVM has the highest detection rate of almost 98% and the lowest error rate of roughly 6%. The Gaussian Classifier placed in second place with 96% accuracy, however it had the greatest error rate of 20% of all classifiers. NB had the lowest accuracy (89%), as well as the second-highest mistake rate (approximately 12%).In 2014, I developed a model for botnet identification using supervised machine learning classifiers, specifically a mix of SVM and the artificial fish swarm technique (AFSA). The data set they utilised was obtained using a Local Area Network, which was designed to collect network packet data and use it as a prototype simulation of botnet assaults traffic or regular traffic. After 5-fold cross-validation, their results demonstrate that the combination of SVM and AFSA performed better than other classifiers, with an average accuracy rate of 99%.The KNN classifier is another popular classifier in this sector. In 2013, a research on Android malware detection, especially anomaly-based mobile botnet detection, was undertaken. The study included five machine learning classifiers, including K-NN, MLP, DT, and SVM. They tested the classifiers on malware data samples from the Android Malware Genome Project. This study focused on three network characteristics: connection time, TCP size, and the amount of GET/POST arguments. The KNN was shown to be the best classifier, with a true positive rate of 99.94% and a false positive rate of 0.06%.

A subsequent article published in 2018 presented a network to identify HTTP botnets using machine learning classifiers such as DT, KNN, NB, and Random Forest (RF).The dataset they utilised for their study was retrieved from network traffic using the TCP packet characteristic. According to the data, the best classifier for detecting HTTP botnet assaults in network traffic is the KNN classifier, which has an average accuracy of 92.93% for each botnet family.

Deep and unsupervised learning have also demonstrated promising results in identifying botnet assaults. Detecting over 350 IoT botnets in darknet data using a multi-window convolution neural network paired with clustering. developed an unsupervised intelligent system for identifying IoT botnets based on SVM and Grey Wolf optimization. This model was able to achieve a low detection time while also reducing the amount of characteristics needed for detection.According to these findings, machine learning is very useful and successful in botnet identification. The primary contribution of this work is the development of a botnet detection model based on a machine learning classification technique.

II. RELATED WORK

The usefulness of utilising machine learning and deep learning to identify the growing number of botnet assaults has been demonstrated in several research in recent years. Finding the main traits or characteristics of a botnet that can aid distinguish between an attack and regular traffic is another area of study. The use and efficacy of machine learning in botnet detection were explored by Dong et al. in [2]. In order to identify botnet traffic from regular traffic, researchers looked at the structure of botnets. When creating our machine learning model, these features may then be utilised to choose other features. One such technique was utilised by Vishwakarma et al. [3], who produced data from an IoT network and deployed a honeypot to entice attackers. The various aspects of an assault, such as IP addresses, MAC addresses, packet sizes, etc., were then examined using this new data.

Additionally, Guerra-Manzanares et al proposal's [4] to use hybrid feature selection models to condense the size of the feature set in order to obtain reliable findings. 115 features are present in the data utilised, which is a very big number for any dataset. To cut down on the amount of features, feature selection was carried out utilising the filter, wrapper, and hybrid models. Following the ingestion of these characteristics into a K-Nearest Neighbor (K-NN) and Random Forest model, both models demonstrated high accuracy of 99%.

The Decision Tree (DT) Classifier is one of the most promising classifiers for P2P botnet identification. For botnet identification, Haq and Singh [5] employed a variety of classifiers in addition to clustering. Over 38,000 records of network traffic, including both attack and regular traffic, were included in the collection. The DT classifier in this study had the highest accuracy (90.2723%), followed by the Decision Tree classifier (87.7853%). Similar to this, Khan et al. [6] found that P2P botnets were challenging to identify since they possessed conventional characteristics of centralization and spread. A two-stage detection approach for P2P botnets was proposed by Khan et al. in 2013. Data flow count, port judgement, and DNS query made up the initial step of the filtering process for non-P2P traffic.The experiment was conducted using the CTU-dataset, which comprises 13 distinct botnet samples.To identify P2P botnet traffic, three primary ML algorithms based on session characteristics were used. DT classification, ANN, and Naive Bayes (NB) were the methods employed.

According to the findings, the NB and ANN algorithms had detection rates of 75.5% and 93.8%, respectively, while the DT method had a detection rate of 94.4%. This demonstrated that P2P botnet traffic could be successfully detected using a two-stage method that included P2P traffic filtering and DT classifier based on session characteristics.

The decision tree and random forest classifiers are further efficient classifiers. Stevanovic and Pedersen [7] investigated how supervised machine learning may be used to detect botnets with high accuracy. To start, they suggested a technique for detecting botnets that makes use of supervised machine learning and flow-based traffic analysis. The performance of eight of the most significant machine learning algorithms (MLAs) for categorising botnet traffic is then tested. Finally, they looked at how much traffic need be shown for a categorization to be effective. Either "limited" analysis, where time intervals and packet counts are restricted, or "batch" analysis, which monitors from the beginning to the conclusion of the trace, were used for traffic analysis.

The ISOP dataset, which contains malicious and non-malicious records, was used for the trials. The findings demonstrated that while the random forest classifier had the greatest accuracy of botnet detection, the random tree classifier was deemed the best since it had the best balance of accuracy and detection time. In a more recent work from 2018, Hoang et al. [8] suggested comparing machine learning techniques for botnet identification to anomaly-based botnet detection approaches. For their machine learning model, the authors of the research employed K-NN, C4.5, random forests (RF), and NB classifiers. They decided to utilise Domain Name Service's superior and the classifier with the greatest performance for the machine learning model to succeed.

The results indicated that employing random forest, botnet detection has an overall accuracy of 90%. DNS has also been used in studies by Jin et al. [9] to identify botnets. Using DNS data, six unique characteristics of botnet domain traffic were chosen. Name-based characteristics, such the meaningful length ratio, made up the chosen features. Message-based features such as the quantity of source IPs, kinds, and A, AAAA, NS, and MX inquiries came next.

Finally, aspects that depend on quantity, such the number of queries executed daily and per hour. Three well-known ML classifiers were used to determine the malicious domains from the DNS traffic once characteristics were chosen. The classifiers utilised were NB, Adaboost, and Bagging. Three machine learning techniques were examined by Garg et al. [10] for their capacity to distinguish between botnet and regular traffic. This was accomplished by picking important aspects of network traffic, combining them into various combinations, and then creating a variety of test cases for the algorithms. The NB, Nearest Neighbor (IBk), and J48 were the three algorithms that were put to the test. The detection accuracy of the J48 and IBk algorithms was greater than that of NB after testing each algorithm with

each instance independently, although with the constraints that J48 required a high training time and IBk required a high testing time. Despite these drawbacks, P2P botnets might be found thanks to the high detection rates of more than 99%.A extremely successful classifier was the SVM algorithm. A research on the detection of P2P botnets using network behaviour analysis and machine learning was carried out by Saad et al. [11]. They employed two datasets totaling over 370,000, all of which represented fraudulent traffic, in their experiment. They utilised a labelled dataset with a million observations for typical traffic the Ericsson Research Traffic Lab provided them with packets of typical traffic. K-NN, Linear Support Vector Machine (SVM), Artificial Neural Network, Gaussian Based Classifier, and NB Classifier were the classifiers they employed.

The results of their experiment indicated that SVM required the most time for both training and classification. SVM has the lowest error rate of around 6% and the maximum accuracy of about 98% for the detection rate. The second-best classifier overall, the Gaussian Classifier, with a 96% accuracy rating but the highest error rate of 20%. NB had the second-highest mistake rate of almost 12% and the lowest accuracy of 89%. Using supervised machine learning classifiers, Lin et al. [12] developed a model based on a botnet detection in 2014.The data set was gathered using a Local Area Network that was designed to gather network packet data and utilise it as a prototype simulation of botnet assaults or regular traffic. Their findings indicate that the SVM and AFSA combination outperformed other classifiers with an average accuracy rate of 99% after fivefold cross-validation.The KNN classifier is another popular classifier in this field. A research on Android malware detection, especially anomaly-based mobile botnet detection, was carried out in 2013 by Feizollah et al. [13]. Five machine learning classifiers were utilised in the study: K-NN, Multi-Layer Perceptron (MLP), DT, and SVM. They tested the classifiers using samples of malware data from the Android Malware Genome Project for their work. Three network characteristics were used for this study: connection time, TCP size, and quantity of GET/POST parameters. The outcome shown that the KNN classifier is the best one available, with a true positive rate of up to 99.94% and a false positive rate of just 0.06%.

A more recent article in 2018Another well-liked classifier in this area is the KNN classifier. Feizollah et al. conducted research on Android malware detection in 2013 [13], focusing on anomaly-based mobile botnet detection. The study used the K-NN, Multi-Layer Perceptron (MLP), DT, and SVM machine learning classifiers. They used samples of malware data from the Android Malware Genome Project to test the classifiers. This study examined three network characteristics: connection time, TCP size, and number of GET/POST arguments. The results shown that, with a true positive rate of up to 99.94% and a false positive rate of just 0.06%, the KNN classifier is the best one currently available.

The ability to recognise botnet assaults has also been demonstrated through deep and unsupervised learning. Pour et almethod .'s [15] of detecting over 350 IoT botnets in darknet data used a multi-window convolution neural network with clustering.In order to identify IoT botnets, Al Shorman et al. [16] suggested an unsupervised intelligent system built on SVM and Grey Wolf B. Dimensionality reduction and feature selection

We used feature selection and dimensionality reduction for our model to lower the dimensionality of the data while maintaining its variance. Principal Component Analysis is one technique for dimensionality reduction (PCA). PCA is a data transformation technique that places the data into a new feature space where the first coordinate of the new space (known as the first principal component) represents the majority of the variance in the data, the second most variance on the second principle component, etc.of characteristics utilised for detection and obtain a low detection time.These research have shown that machine learning is quite useful and successful in detecting botnets. The primary contribution of this work is the development of a machine learning-based classification strategy for botnet identification.

III. PROPOSED SYSTEM

A. Dataset Overview

For this effort, a number of datasets are available, including the UNSW-NB15 and Bot-IoT datasets. Over 72 million records with 42 attributes (27 integer, 13 float, and 2 string kinds) make up the Bot-IoT dataset, which was produced by establishing a botnet network in a safe place and watching network traffic to catch any packets being transmitted. The dataset includes assaults like DDoS, DoS, OS Scan, and others that are categorised as malicious and regular traffic, respectively. The other dataset, UNSW-NB15, has 2.5 million records and 43 characteristics (14 Float, 6 Strings, and 23 Integer types) that are divided into two categories: attack traffic and regular traffic. The attack category is further divided into subcategories. The collection also includes records for fuzzers, backdoors, reconnaissance, and worm assaults in addition to DDoS and DoS attacks [17].To construct the dataset, these records were first gathered in pcap files and converted to CSV. UNSW Canberra has assembled both datasets and made them accessible to the public for research purposes [18]. Additionally, it is discovered that the UNSW-NB15 dataset is a more refined dataset with characteristics comparable to those of the Bot-IoT dataset but a wider variety of harmful entries. The UNSW-NB15 dataset was utilised in this study because it is more thorough and has been deemed the best dataset for training by UNSW. In this study, 82,000 records were chosen at random and used. To categorise the training and testing models, the data was cleaned and processed. Numbers were coded with categorical information such as the "proto", kind of "service," "state," "sptks," "sload," and "attack cat." The data was divided randomly into two sets: a training set with 80% of the data and a testing set with 20% of the data.so on. In this study, the

characteristics with the highest variance are chosen to provide the classifiers' input.

By following this procedure, we make sure that only pertinent features are chosen, enhancing the computational effectiveness and simplicity of the machine learning models. Additionally, by utilising all the characteristics, this technique has the potential to minimise any overfitting that could develop.

C. ALGORITHMS FOR CLASSIFICATION

In this work, a number of classifiers were employed and assessed. This study made use of the Python Scikit-learn module. Using the real and predicted labels from the model, the confusion matrix was constructed to determine precision, recall, and F-Measure for the evaluation. These classifiers were employed in this study:

1) Naive Bayes (NB) using Gaussian probabilities: This probabilistic classifier, which applies the Bayes theorem, assumes conditional independence across the various dataset characteristics. On the basis of the training set, NB calculates the class probability.

2) The non-parametric method K-Nearest Neighbor (k-NN), which is used for regression and classification. The majority of the test sample's k closest neighbours are used by the model to determine the test sample's class in order to predict it.

3) Support Vector Machine (SVM) with Radial Basis Function (RBF), a nonlinear kernel, generates a decision boundary based on samples from several classes.The decision boundary's shape is determined by the kernel function that is employed as well as important hyperparameters like C, which regulates the tradeoff between the decision boundary's smoothness and the classification's accuracy, and gamma, which specifies how much the distribution of the data points affects the decision boundary's shape.

4) Decision Tree (DT): A classification model that resembles a tree in which each node indicates a test on a single feature and each branch descending from that node represents one of the potential values for that feature.

The dataset was randomly divided into training and testing datasets before these classifiers were used. The classifiers were trained using the training data. The testing dataset was then used to evaluate the classifiers by predicting the labels. According to what was covered in Section IV, each classifier was assessed and compared.

RESULT AND DISCUSSION

Figure 1 – IOT not attack

Figure 2-IOT attack

IV.CONCLUSION

The identification of botnet or malicious traffic behaviour using new machine learning algorithms was proposed in this research. This paper used four classifiers: Nave Bayes, K-Nearest Neighbor, Support Vector Machine, and Decision Trees. The experimental findings showed that the decision tree model

outperformed the other classifier models and was a minor improvement over the methods previously indicated in the studied literature. In theory, this approach can identify many botnet assaults and other types of harmful network behaviour.

To validate the results, the complete UNSW-NB15 dataset must be evaluated using the same model in the future.This experiment may be expanded to incorporate more recent datasets, such as the Bot-IoT dataset and the CTU-13, to evaluate the performance of the algorithms with different types of botnet traffic. Additional classifiers, such as logistic regression and neural networks, can be evaluated.

Unsupervised learning approaches, such as clustering, can also be examined and compared to the supervised learning methods employed in this study. Furthermore, various feature selection approaches might be investigated to further enhance these results. Finally, the machine learning model may be evaluated in a real-time controlled environment to properly quantify its performance and how it handles various sorts of threats, such as zero-day attacks.

REFERENCES

[1] S. Ranger, "What is the IoT? Everything you need to know about theInternet of Things right now | ZDNet," ZDNet, 2020. [Online].

[2] X. Dong, J. Hu and Y. Cui, "Overview of Botnet Detection Based on Machine Learning," International Conference on Mechanical, Control and Computer Engineering, Huhhot, pp. 476-479, 2018.

[3] R. Vishwakarma and A. Jain, "A Honeypot with Machine Learning based Detection Framework for defending IoT based Botnet DDoS Attacks," International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, pp. 1019-1024, 2019.

[4] A. Guerra-Manzanares, H. Bahsi and S. Nõmm, "Hybrid Feature Selection Models for Machine Learning Based Botnet Detection in IoT Networks," International Conference on Cyberworlds (CW), Kyoto, Japan, pp. 324-327, 2019.

[5] S. Haq and Y. Singh, "Botnet Detection using Machine Learning," International Conference on Parallel, Distributed and Grid Computing (PDGC), India, pp. 240-245, 2018.

[6] R. Khan, R. Kumar, M. Alazab and X. Zhang, "A Hybrid Technique To Detect Botnets, Based on P2P Traffic Similarity," Cybersecurity and Cyberforensics Conference (CCC), Melbourne, Australia, pp. 136-142, 2019.

[7] M. Stevanovic and J. Pedersen, "An efficient flow-based botnet detection using supervised machine learning," International Conference on Computing, Networking and Communications, HI, pp. 797-801, 2014.

[8] X. Hoang and Q. Nguyen, "Botnet Detection Based On Machine Learning Techniques Using DNS Query Data," Future Internet, vol. 10, no. 5, p. 43, May 2018.

[9] J. Jin, Z. Yan, G. Geng and B. Yan, "Botnet Domain Name Detection based on machine learning," International Conference on Wireless, Mobile and Multi-Media (ICWMMN), Beijing, China, pp. 273-276, 2015.

[10] S. Garg, A. Singh, A. Sarje and S. Peddoju, "Behaviour analysis of machine learning algorithms for detecting P2P botnets," International Conference on Advanced Computing Technologies, pp. 1-4, 2013.

[11] S. Saad et al., "Detecting P2P botnets through network behaviour analysis and machine learning," International Conference on Privacy, Security and Trust, Montreal, QC, pp. 174-180, 2011.

[12] K.-C. Lin, S.-Y. Chen, and J. C. Hung, "Botnet Detection Using Support Vector Machines with Artificial Fish Swarm Algorithm," Journal of Applied Mathematics, vol. 2014, Article ID 986428, 2014.

[13] A. Feizollah, N. B. Anuar, R. Salleh, F. Amalina, R. Ma'arof, and S. Shamshirband, "A Study Of Machine Learning Classifiers for Anomaly-Based Mobile Botnet Detection," Malaysian Journal of Computer Science, 26(4), 2013.

[14] R. Dollah, Faizal. A., F. Arif, M. Mas'ud and L. Xin, "Machine Learning for HTTP Botnet Detection Using Classifier Algorithms," Journal of Telecommunication, Electronic and Computer Engineering, 10(1-7), pp. 27- 30, 2018.

[15] M. Pour, A. Mangino, K. Friday, M. Rathbun, E. Bou-Harb, F. Iqbal, S. Samtani, J. Crichigno, and N. Ghani, "On Data-driven Curation, Learning, and Analysis for Inferring Evolving Internet-of-Things (IoT) Botnets in the Wild," Computers & Security, vol. 91, April 2020.

[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," Journal of Ambient Intelligence and Humanized Computing, vol. 11, pp. 2809–2825, July 2020.

[17] Moustafa, Nour, and Jill Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," Military Communications and Information Systems Conference (MilCIS), Australia, 2015.

[18] N. Koroniotis, N. Moustafa, E. Sitnikova and B. Turnbull, "Towards the Development of Realistic Botnet Dataset in the Internet of Things for Network Forensic Analytics: Bot-IoT Dataset," Future Generation Computer Systems, vol. 100, p. 779-796, November 2019.