[1]Rituparna Borah
[2]Satyajit Sarmah

Detection of Various Botnet Attacks Using Machine Learning Techniques

Journal of Electrical Systems

Abstract: - With the rapid growth in the quantity of Internet of Things (IoT) devices linked with the network, there exists a concurrent rise in network attacks, including overwhelming and service disruption incidents. The increasing prevalence of network attacks, such as overwhelming and service denial, poses a threat to IoT devices, leading to network disruptions and service disruption. Detecting these attacks is challenging due to the diverse array of heterogeneous devices in the IoT environment, making traditional rule-based security solutions less effective. Developing optimal security models for diverse device types is challenging. Machine learning (ML) offers an alternative approach, enabling the creation of effective security models by leveraging empirical data specific to each device. We utilize machine learning techniques for the detection of Internet of Things (IoT) attacks. Our focus is on botnet attacks directed at variety of IoT devices. We undertake the development of machine learning-based models tailored to each specific category of device for enhanced security. We utilize the N-BaIoT dataset, which incorporates injected botnet attacks (specifically Gafgyt and Mirai) across diverse IoT device types, including Doorbell, Baby Monitor, Security Camera, and Webcam. We develop models for detecting botnets for each IoT device by utilizing diverse machine learning algorithms. Following model development, we assess the utility of the models with a strong detection F1-score through classification analysis. The novelty of this work lies in crafting a Machine Learning-based framework designed to identify IoT botnet attacks, followed by successful detection of such attacks across diverse IoT devices utilizing this framework. Among the most widely used machine learning algorithms on the NBaIoT dataset, Decision Trees, Random Forests, and K-Nearest Neighbors (KNN) demonstrate superior performance.

Keywords: IoT; N-BaIoT dataset; Machine Learning

I. INTRODUCTION

The Internet of Things (IoT) is a network that links billions of smart devices, enabling communication over the internet [1]. With the escalating interconnectivity, there is a heightened risk of more devices becoming potential botnet targets. The primary emphasis of this paper is to employ machine learning techniques for identifying botnet attacks in this evolving landscape.

A botnet operates by deploying bots on multiple internet-connected devices, forming a network controlled by a central command and control (C&C) system [2]. Botnets have the capability to execute distributed denial-of-service attacks (DDoS), engage in data theft, or gain unauthorized access to devices. A botnet attack is a malicious form of attack that leverages a network of connected computers to target and compromise a network, network device, website, or IT environment. The primary objective of such attacks is to disrupt normal operations or degrade the overall service of the targeted system. Therefore, the successful detection and prevention of botnets hold significant importance in the realm of computer security.

With an increasing number of devices susceptible to becoming botnet targets, the identification and differentiation of these devices can be achieved through the application of various machine learning techniques. The objective of this work is to recognize botnets or malicious traffic activity by leveraging advancing machine learning methods.

[1] *Satyajit Sarmah:

[2] Assistant Professor

The paper is structured as follows: Section II provides a review of the literature on related work. Section III elaborates on the proposed approach. Section IV discusses the outcomes derived from the obtained results. Section V outlines the conclusion and outlines prospective areas for further exploration.

II. LITERATURE REVIEW

In recent years, numerous studies have demonstrated the efficacy of employing Machine and Deep Learning techniques in the identification of escalating botnet attacks.In [3] the authors provide a comprehensive examination of the use of machine learning techniques in identifying and combating botnet activities. It explores various methods and strategies in the field of machine learning for effectively detecting and mitigating botnet threats. The objective is to present a broad perspective on the current landscape of botnet detection, shedding light on the advancements, challenges, and potential future directions in the field. In [4] the authors proposed a model that combines feature engineering techniques with different machine learning algorithms for the detection of DDoS attacks. The study benchmarks the performance of k-nearest neighbors, Naive Bayes, support vector machine, Random Forest, and artificial neural network to determine their effectiveness in detecting DDoS attacks. Feature engineering is employed to enhance the quality of input attributes for the machine learning model. The authors used feature selection methods, specifically chi2 and information gain scores, to fine-tune and optimize the selected features. The author emphasizes the possibility of improving the detection model's performance through feature reduction, demonstrating that a reduced feature set can still maintain high accuracy. This is particularly useful for minimizing processing overhead in the system. The outcome the suggested model is validated using ROC AUC analyses and cross-validation techniques. This ensures a robust validation of the model's ability to generalize and perform well on new, unseen data. Among the evaluated machine learning algorithms, the authors discover that the k-nearest neighbors (KNN) algorithm achieves the optimal outcome in detecting DDoS attacks. The authors conclude that feature reduction has a minimal impact on accuracy, highlighting the potential to streamline the system's processing overhead without significantly sacrificing detection performance. In [5] the author introduces a detection system for botnets utilizing artificial neural networks, enhanced by the use of the SMOTE data resampling technique to handle class imbalance. The authors employ the BoT-IoT dataset alongside artificial neural networks (ANN) to construct their detection system, incorporating both ANN and the Synthetic Minority Over-sampling Technique (SMOTE) for modeling purposes. The proposed system is shown to be effective in detecting DDoS attacks, even with a basic configuration of the ANN, demonstrating its potential for practical applications. In [6] the authors employed machine learning as a tool to detect and mitigate botnet attacks, emphasizing the need for dedicated research in this area. The study utilizes two datasets, Bot-IoT and University of New South Wales (UNSW), to facilitate the development and evaluation of machine learning models. Four machine learning models are built, each based on a different classifier: K-Nearest Neighbor, Naïve Bayes, Decision Trees and Support Vector Machine. The performance of the models is assessed using the UNSW-NB15 dataset, consisting of 82,000 records. The Decision Trees model demonstrates superior performance, achieving a testing accuracy of 99.89%, 100% recall, 100% F1-score and 100% precision In identifying botnet assaults. The author’s presents the outcomes of the experiments, with a specific focus on the Decision Trees model, which outperforms the other classifiers in terms of accuracy and effectiveness in detecting botnet attacks. In [7] the authors proposed an approach that utilizes a hybrid deep learning model, combining a Convolutional Neural Network (CNN) with a Long Short-Term Memory (LSTM) algorithm (CNN-LSTM), to strengthen botnet discovery. Empirical research is conducted using an authentic N-BaIoT dataset derived from an actual system, encompassing both non-malicious and malicious patterns to simulate real-world conditions. The study evaluates the effectiveness of the CNN-LSTM model on different commercial Internet of Things (IoT) devices, specifically doorbells from the Danmini and Ennio brands and thermostats. The model achieves accuracies of 88.61% and 90.88% for identifying and thwarting botnet attacks originating from doorbell devices and 88.53% for thermostats. The suggested framework demonstrates accuracies of 89.23%, 87.76%, 89.64% and 87.19% in identifying botnet intrusions through security cameras, emphasizing the versatility of the CNN-LSTM model across various IoT devices. The CNN-LSTM model proves effective in detecting and mitigating botnet attacks originating from a variety of connected devices, achieving optimal accuracy in the experimental evaluations. In [8] the author addresses the proliferation of intelligent systems and energy-efficient sensing devices in the Internet of Things (IoT), which contributed to a surge within IoT-based botnet attacks due to limitations in processing, memory, and

connectivity capacities. In response, the author proposes ELBA-IoT, An amalgamation learning framework for identifying botnet attacks in IoT networks. ELBA-IoT leverages behavioral attributes of IoT networks and employs utilizing ensemble learning for the detection of abnormal network activity stemming from compromised IoT devices. The evaluation includes three decision tree-based machine learning techniques (AdaBoosted, RUSBoosted, and bagged) within the ELBA-IoT framework. The N-BaIoT-2021 dataset, containing both normal and traffic indicative of botnet attacks records from infected IoT devices, is used for experimental validation. Findings indicate that ELBA-IoT achieves exceptional detection accuracy (99.6%) while maintaining a minimal inference overhead of 40 µ-seconds. The author compares ELBA-IoT against state-of-the-art methods, affirming its superior performance in IoT-based botnet attack detection. In [9] the authors employed a multi-window convolutional neural network in conjunction with clustering techniques to identify over 350 IoT botnets within darknet traffic. In [10] the authors introduced an unsupervised intelligent system utilizing a combination of SVM (Support Vector Machine) and Grey Wolf Optimization to detect IoT botnets. The model demonstrated efficient performance with low detection times and a reduction in the number of features required for detection. In [11] the authors conducted a study focused on identifying Android malware, specifically emphasizing anomaly-based mobile botnet detection. They employed five machine learning classifiers, including K-NN, Multi-Layer Perceptron (MLP), Decision Trees (DT), and Support Vector Machine (SVM). The study utilized malware data samples from the Android Malware Genome Project and assessed the classifiers based on three selected network features: connection duration, TCP size, and the number of GET/POST parameters. The findings revealed that the most effective classifier was K-NN, achieving a true positive rate of 99.94% and a false positive rate of 0.06%. In [12] the author highlights the rapid growth of the Internet of Things (IoT) and the multitude of devices integrated into smart systems. However, this development attracts attackers who target these devices, turning them into controlled bots. These compromised devices pose a significant threat to organizations, making it essential to have robust security mechanisms. While rule-based systems, such as public Intrusion Detection Systems (IDS), exist, they can be circumvented by attackers with knowledge of the rules. To address this, the author proposes machine learning-based detection architecture, utilizing the CART algorithm and the N-BaIoT dataset. Experimental results demonstrate that the detection accuracy of the CART classifier surpasses that of the Naïve Bayes classifier, achieving an overall detection rate of up to 99%. This suggests that machine learning-based approaches, particularly the CART algorithm, offer an effective solution to bolster the security of IoT devices against malware attacks. In [13] the authors constructed a botnet detection model utilizing supervised machine learning classifiers, specifically combining Support Vector Machine (SVM) with the artificial fish swarm algorithm (AFSA). They gathered a dataset through a Local Area Network designed to collect packet data, simulating both botnet attack and normal traffic. The outcomes, determined through 5-fold cross-validation, demonstrated that the SVM and AFSA combination outperformed other classifiers, achieving an impressive 99% average accuracy rate.

These investigations collectively reveal that machine learning is exceptionally applicable and successful in the realm of botnet detection. The primary contribution of this study centers on the creation of a model for detecting botnets through the application of a machine learning classification approach.

III. METHODOLOGY

We have established a comprehensive approach for crafting IoT device-controlled botnet identification model, encompassing the complete process starting from delineating the botnet dataset to the current detection of botnets. This section outlines the application of the N-BaIoT dataset within our approach and delineates the design of the suggested approach.

A. NBaIot Dataset

The N-BaIoT dataset, described in [14], comprises data instances with 115 attributes. Collected through replication of IoT port traffic devices, the non-malicious information was recorded promptly after network setup to ensure their benign nature. The dataset includes features related to packet dimensions (solely outgoing/both outgoing and incoming), packet quantities, packet jitters, and intervals among packet arrivals for statistical values across five temporal windows (1 m, 10 s, 1.5 s,500 ms,100 ms). A cumulative of 23 attributes was derived for each time frame, summing up to 115 features in total. Our framework utilizes all 115 features from

the dataset. Elaborate characteristics regarding the dataset are presented in Table 1. The datasets were generated by introducing a pair of attack types into diverse IoT devices, as detailed in Table 2.

Every dataset was created by introducing different instances of Bashlite and Mirai attacks. Bashlite, alias gafgyt, is a C-written botnet developed by Lizard Squad, primarily used for infiltrating IoT devices running on Linux to execute Distributed Denial of Service (DDoS) attacks.The attacks orchestrated by Bashlite include various flooding techniques including UDP and TCP attacks. On the other hand, Mirai, developed by Paras and Unearthed in August 2016, is employed for orchestrating extensive attacks utilizing IoT devices. Over the years, Mirai has evolved significantly, and it is currently accessible as open-source. In [16] the authors recommended an approach utilizing four heuristic algorithms to identify Mirai and Gafgyt C&C servers. Table 3 details the ten particular attack categories associated with Gafgyt and Mirai.

B. Proposed Framework

Our approach includes a dataset of botnets, models trained on botnet data, and. models for detecting botnets. The botnet dataset comprises four subdatasets from N-BaIoT. We specifically choose devices, such as a doorbell (Damini), baby monitor (Philips B120N/10), security camera (Provision PT-737E), and webcam (Samsung SNH 1011 N), that include all 10 attack samples outlined in Table 3 of N-BaIoT. The distribution of samples across the four datasets, categorized by device type, is presented in Table 4.

For botnet training models, we utilize a selection of widely-used machine learning (ML) algorithms. Our approach incorporates five ML models (naïve Bayes (NB), K-nearest neighbors (KNN), logistic regression (LR), decision tree (DT), and random forest (RF)). The classification categorizes the N-BaIoT dataset into three groups, distinguishing between Mirai and Bashlite and benign instances without considering different protocols. Figure 1 illustrates our framework for developing ML based IoT botnet detection models.

Table 1: In depth characteristics of the NBaIot dataset [15]

Aggregated by	**Value**	**Statistics**	**Total No. of features**
Source IP	Packet size (only outbound)	Mean, Variance	**3**
	Packet count	Integer	
Source MAC_IP	Packet size (only outbound)	Mean, Variance	**3**
	Packet count	Integer	
Channel	Packet size (only outbound)	Mean, Variance	**10**
	Packet count	Integer	
	Amount of time between packet arrivals	Mean, Variance, Integer	

	Packet size (both inbound and outbound)	Magnitude radius, covariance, correlation, coefficient	
Socket	Packet size (only outbound)	Mean, Variance	**7**
	Packet count, Packet size (both inbound and outbound)	Integer	
Total			**23**

Table 2: Device category and model designation specifics within the NBaIot dataset

Device Type	Device Model Name
Dorbell	Damini
	Emnio
Thermostat	Ecobee
Baby Monitor	Philips B120 N/10
Security Camera	Provision PT-737E
	Provision PT-838
	Simple Home XCS7-1002-WHT
	Simple Home XCS7-1003-WHT
Webcam	Samsung SNH1011N

Table 3: Botnet and attack categories employed.

Major Attacks	Subattacks	Description
Bashlite	Junk	By sending spam data
	TCP Flood	Sends flood of request
	UDP Flood	Sends flood of request
	Scan	Scan the network for victim devices
	Combo	Opens connection IP address and network port by sending spam data
Mirai	ACK	Sends flood of Acknowledgement
	SYN	Sends synchronize packet flood
	Plain UDP	UDP flood by optimizing seeding packet per second
	UDP flood	Scans the network for victim devices
	Scan	

Table 4: The quantity of samples utilized in this paper

Botnet	**Attack type**	**Dorbell**	**Baby Monitor**	**Security Camera**	**Webcam**
Benign		3278	43810	15538	4882
Bashlite	Combo	13253	14538	15345	14850
	Junk	15156	14174	15449	13706
	Scan	13747	13930	14648	14286
	TCP	14253	13887	15676	14711
	UDP	15719	15867	15602	15447
Mirai	Ack	28321	22781	15138	26797
	Scan	6479	15543	14517	6551
	Syn	29202	29532	16436	30620
	UDP	15148	21703	15625	15708
	UDP Plain	23589	21818	15304	22798

IV. EXPERIMENTAL EVALUATION

In this segment, we determine the most efficient model for detecting IoT botnets through a comprehensive analysis of performance variations depending on the category of IoT devices and the ML models employed. We initially construct a model for detecting IoT botnets according to the suggested approach. For the N-BaIoT dataset samples, we randomly partition the training and testing sets using a 70-30 split ratio through Scikit-learn, an open-source ML library. This ensures the independence of the training and testing sets. To mitigate overfitting, 20% of the training set is designated as a validation set. Throughout training, we compute the validation loss to monitor its behavior relative to the training loss. Additionally, this section includes the classifications which discerns not only benign but also mirai and bashlite attack by learning them.

4.1 Classification

The categorization model consolidates 5 distinct detailed Bashlite and 5 distinct detailed Mirai attacks introduced into IoT devices as a two attack i.e. Mirai and Bashlite. It differentiates between Mirai and Bashlite and benign states, where the latter indicates the absence of an injected attack. The model is trained using datasets collected from each device, leveraging machine learning (ML) models. The design of these models is implemented using Keras and Scikit-learn. Table 5 provides a description of the model designs.

Subsequently, we evaluate the models' performance using F1-score measurements. The F1-score, serving as a comprehensive metric incorporating both precision and recall, is represented as a unified value. The F1-score

assigns a weighted beta score of 1 to precision in the F-score calculation and can be formulated using the following equation:

F1-Score= (2*Precision*Recall)/ (Precision+Recall)

where,

Precision= (TP)/ (TP+FP) and

Recall= (TP)/ (FN+TP)

True positive (TP) denotes the count of samples correctly classified as benign. False negative (FN) represents the instances where benign data is falsely identified as a botnet. False positive (FP) corresponds to the samples incorrectly predicting an actual botnet as benign. True negative (TN) indicates the number of samples accurately detected as a botnet. The comprehensive detection results, encompassing precision, recall, and F1-score for each ML model, are detailed in Table 6.

Through experimental evaluation, it was determined that the decision tree and random forest are the most effective machine learning (ML) models for detecting Bashlite and Mirai botnets.

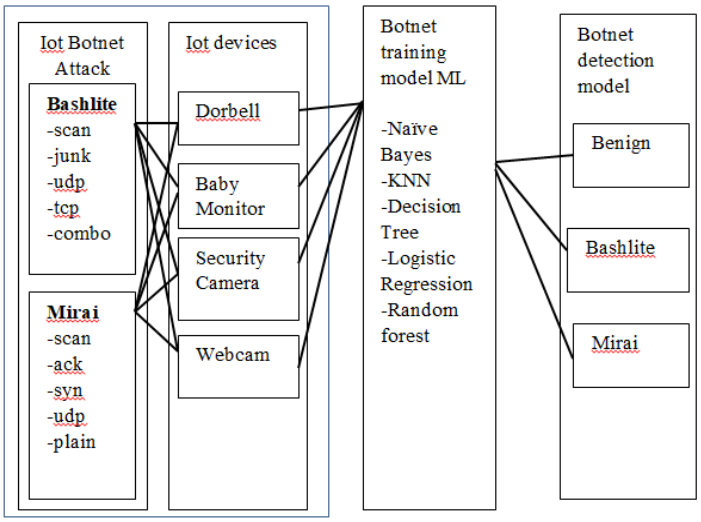

Fig 1: The outlined architecture for identifying IoT botnets in our proposal

Table5: Details of our Machine Learning (ML)

Model	**Description**
Naïve Bayes (NB)	Portion of the largest variance of all featueres 10^{-9}
K -nearest neighbors (KNN)	Number of neighbors:5 Weight function:uniform
Logistic Regression (LR)	Penalt:12 Tolerance for stopping creiteria:10^{-4}
Decision Tree (DT)	Impurity measure:entropy Splitter:best
Random Forest (RF)	Number of trees:100 Impurity measure:Gini

Table 6: Outcomes of the detection process for the five machine learning (ML) models

ML Models	INDEX	Dorbell			Baby Monitor			Security Camera			Webcam		
		Benign	Gafgyt	Mirai	Benign	Gafgyt	Mirai	Ben	Gaf	Mir	Ben	Gaf	Mir
Decision Tree	Precision	1	1	1	1	1	1	1	1	1	1	1	1
	Recall	1	1	1	1	1	1	1	1	1	1	1	1
	F1 Score	1	1	1	1	1	1	1	1	1	1	1	1
Logistic Regression	Precision	0.99	0.87	0.94	1	0.98	0.99	0.99	0.83	0.94	0.99	0.85	0.95
	Recall	0.92	0.90	0.92	0.89	0.92	0.90	0.59	0.93	0.92	0.96	0.92	0.90
	F1 Score	0.96	0.89	0.93	0.96	0.95	0.95	0.74	0.88	0.93	0.98	0.88	0.92
Random Forest	Precision	1	1	1	1	1	1	1	1	1	1	1	1
	Recall	1	1	1	1	1	1	1	1	1	1	1	1
	F1 Score	1	1	1	1	1	1	1	1	1	1	1	1
KNN	Precision	1	1	1	1	1	1	1	1	1	1	1	1
	Recall	1	1	1	1	1	1	1	1	1	1	1	1
	F1 Score	1	1	1	1	1	1	1	1	1	1	1	1
Naive Bayes	Precision	1	0.66	1	0.99	0.31	0.48	1	0.66	0.98	1	0.66	0.99
	Recall	1	0.99	0.68	0.48	0.99	0.64	0.86	0.99	0.63	0.98	0.99	0.68
	F1 Score	1	0.79	0.91	0.99	0.64	0.78	0.92	0.79	0.77	0.99	0.79	0.71
KNN	Precision	1	1	1	1	1	1	1	1	1	1	1	1
	Recall	1	1	1	1	1	1	1	1	1	1	1	1
	F1 Score	1	1	1	1	1	1	1	1	1	1	1	1

V. CONCLUSION

We designed a framework utilizing machine learning (ML) to identify IoT botnet attacks. Subsequently, we employed this approach to identify botnet assault aimed at various connected devices. The components of our framework encompass a dataset of botnets, a model trained on botnet data, and a model for botnet detection.

Utilizing the N-BaIoT dataset, we injected incidents involving the Gafgyt and Mirai botnet assaults into four distinct types of IoT devices: doorbell, baby monitor, security camera, and webcam. The Gafgyt and Mirai attacks encompass five attack types each, involving TCP, UDP, and ACK. For constructing the botnet training model, we utilized five machine learning models, namely naïve Bayes, K-nearest Neighbors, logistic regression, decision tree, and random forest. Building upon this training model, we formulated a botnet detection model capable of identifying pertinent botnet attacks. The model for detecting botnets comprises a classification model, which treats the 5 Bashlite and 5 Mirai sub-attacks as a two distinct attack (distinguishing them from benign data). In the results of the experimental phase for classification using machine learning (ML), the F1-score for the ML models, with the exception of Logistic Regression (LR), exhibited high values, mostly reaching 1.

Our research adds value by presenting an approach that facilitates the straightforward comparison of various machine learning (ML) models in identifying botnets within the context of IoT (Internet of Things). In future we can apply the same framework for various deep learning models.

References:

[1] K. Chopra, K. Gupta, and A. Lambora, “Future Internet: The Internet of Things-A Literature Review,” in Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Compute Trends, Prespectives Prospect. Com. 2019,pp. 135–139, Institute of Electrical and Electronics Engineers Inc., feb 2019.

[2] Thingbots: The Future of Botnets in the Internet of Things. Available online: https://securityintelligence.com/thingbots-the-future-of-botnets-in-the-internet-of-things.

[3] 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.

[4] M. Aamir and S. M. A. Zaidi, “DDoS attack detection with feature engineering and machine learning: the framework and performance evaluation,” Int. J. Inf. Secur., vol. 18, no. 6, pp. 761–785, 2019.

[5] Y. N. Soe, P. I. Santosa, and R. Hartanto, “DDoS Attack Detection Based on Simple ANN with SMOTE for IoT Environment,” Proc. 2019 4th Int. Conf. Informatics Comput. ICIC 2019, pp. 0–4, 2019.

[6] Mustafa Alshamkhany, Wisam Alshamkhany, Mohamed Mansour, Mueez Khan, Salam Dhou, Fadi Aloul,” Botnet Attack Detection using Machine Learning”, 14th IEEE International Conference on Innovations in Information Technology (IIT), November 2020.

[7] Hasan Alkahtani1 and Theyazn H. H. Aldhyani,” Botnet Attack Detection by Using CNN-LSTM Model for Internet of Things Applications”, Security and Communication Networks Volume 2021, Article ID 3806459, 23 pages https://doi.org/10.1155/2021/3806459.

[8] Abu Al-Haija, Qasem, and Mu’awya Al-Dala’ien. "ELBA-IoT: an ensemble learning model for botnet attack detection in IoT networks." Journal of Sensor and Actuator Networks 11.1 (2022): 18.

[9] Pour, Morteza Safaei, et al. "On data-driven curation, learning, and analysis for inferring evolving internet-of-Things (IoT) botnets in the wild." Computers & Security 91 (2020): 101707.

[10] Al Shorman, Amaal, Hossam Faris, and Ibrahim 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 11 (2020): 2809-2825.

[11] Feizollah, Ali, et al. "A study of machine learning classifiers for anomaly-based mobile botnet detection." Malaysian Journal of Computer Science 26.4 (2013): 251-265.

[12] Htwe, Chaw Su, Yee Mon Thant, and Mie Mie Su Thwin. "Botnets attack detection using machine learning approach for iot environment." Journal of Physics: Conference Series. Vol. 1646. No. 1. IOP Publishing, 2020.

[13] 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.

[14] Al-Garadi, Mohammed Ali, et al. "A survey of machine and deep learning methods for internet of things (IoT) security." IEEE Communications Surveys & Tutorials 22.3 (2020): 1646-1685.

[15] Kim, Jiyeon, et al. "Intelligent detection of iot botnets using machine learning and deep learning." Applied Sciences 10.19 (2020): 7009.

[16] Bastos, Gabriel, et al. "Identifying and Characterizing bashlite and mirai C&C servers." 2019 IEEE Symposium on Computers and Communications (ISCC). IEEE, 2019.