INTERNATIONAL JOURNAL OF INFORMATION SECURITY SCIENCE
P.B. Dash et al., Vol.9, No.3, pp.164-171

Anomaly Detection in IoT Network by using Multi-class Adaptive Boosting Classifier

Pandit Byomakesha Dash[1],K. Srinivasa Rao[2]

[1,2]Department of Computer Science and Engineering, Sri Sivani College of Engineering, Srikakulam, Jawaharlal Nehru Technological University, Kakinada, Andhra Pradesh, India.
Corresponding Author: *byomakeshdash2000@gmail.com*

ORCID iD: 0000-0003-1643-9651, 0000-0002-1099-9669

Research Paper Received: 18.04.2020 Revised: 10.06.2020 Accepted: 24.06.2020

Abstract—Detection of anomaly and attack identification is some of the major concerns in IoT domain in recent days. With the exponential use of IoT based infrastructure in every domain, threats and anomalies are amplifying adequately. Attacks such as malicious operations, spying, service denial, etc. are the main cause for failure in IoT system. To solve such an important problem, it is highly desirable to develop some intelligent computing-based approaches with better security conventions for protecting the system. With the combination of several models, ensemble learning helps to enhance the performance of machine learning methods. As compared to any single method, the ensemble learning-based models are highly predictable for large dimensional data. In this paper, an adaptive boosting based model has been proposed to identify the anomaly in IoT based environment. The performance of the proposed method is compared with several other competitive machine learning-based methods and found to be superior with all the considered metrics.

Keywords—IoT Security, Anomaly Detection, Adaptive Boosting, Ensemble Learning

1. Introduction

Now-a-days, IoT is rising at a precariously quick pace with the broad use of Wi-Fi networks, and researchers guess that the amount of vigorous wireless linked devices will surpass huge usage by the upcoming generation. However, the huge number of linked devices signifies more attack vector and other probabilities for hackers to aim for us. So, security is becoming a major challenging task. Where IoT security is the expertise region concerned with protecting connected networks and devices in the IoT. Applying safety procedures is important to guarantee the security of networks with IoT devices associated with them. Attacks and threats are rising commensurately with the grown usage of IoT infrastructure in the daily field. Some of the anomalies which can cause an IoT system failure are: denial of service, malicious control, wrong setup, data type probing, spying, scan, and malicious operation, etc. [1]. The methods of anomaly detection are adherently used to preprocess the data with the removal of anomalous data, which seems like homogeneous with other data. Anomaly detection based methods are used to detect fraudulent operations, intrusion based activities and other important unusual ac-

tivities to protect the network. With the growing demand and rise of IoT automated network scheme, the IoT techniques are getting exigent day-by-day. Individuals are being familiar with data-driven communications, and this is providing importance to the research more on Machine Learning (ML) based applications alongside IoT [2] [3]. IoT and ML-based methods are used in every field of human life at present. ML has been considered as the main technology of independent smart network management and procedure. Particularly, the majority of IoT systems are turning into progressively more active, heterogeneous, and difficult. Therefore it is difficult to manage such IoT systems. Moreover, it necessitates enhancing such IoT system services in terms of efficiency and variety for attracting more users. Many studies have improved on concerning ML to IoT. Various applications of ML for IoT allows users to attain deep analytics and enlarge proficient intelligent applications of IoT. This is because ML can afford possible solutions to extract the hidden features and information in IoT data. Authentication, malware detection, access control, and anti-jamming are the significant network securities which are been improved with the usage of ML techniques [4]. Applications of ML create complicated ideas and visions which are presented to IoT systems for service alteration and elevation.

Due to imbalanced nature of IoT security data, the designing of model-based anomaly detection in IoT network poses a challenge for machine learning model as most of the machine learning model assumes an equal number of samples for each class, which results in poor predictive performance for identification of anomaly type. This is essentially a problem because of the fact that the anomaly type is more sensitive than the normal activity type. In order to address these issues, this work proposes ensemble learning-based methods with the Synthetic Minority Oversampling Technique for the prediction of an anomaly in IoT network. The remaining sections are segmented in the following way: Section 2 elaborates the literature about various ML-based methods in IoT applications with their limitations. Section 3 describes the Proposed Methodology. Section 4 discusses the details about Experimental Setup with Parameter Settings and Result Analysis. Section 5 concludes the work with some future directions.

2. Literature Study

Attack and anomaly detection IoT infrastructure is growing apprehension in the field of IoT. Hasan et.al [1] have compared many ML techniques such as SVM, RF, DT, LR, and ANN to identify the attacks in IoT sensors. Accuracy, f1 score, ROC curve, recall, and precision are the evolution metrics that are utilized by authors for the comparison of their performance. It was found that the RF approach is precise when contrasted to remaining approaches. They have attained a 99.4% rate accuracy for ANN, RF, and DT and majorly found RF outperforms better when compared to other techniques. Deng et.al [5] have analyzed the features of security problems and discussed the structure of internet security as well as few major security methodologies like key administration, access control, privacy safety, intrusion, validation and fault tolerance, etc. The authors have introduced a novel scheme for IoT, i.e., lightweight intrusion detection technique that is joined with the FCM and PCA algorithm. Comparison of the projected method with other applications like the multi-agent, Bayes, and game theory was made by the authors and found better simulation outcomes. Diro and Chilamkurti [6] have proposed a method of Distributed attack recognition with the DL technique for IoT. Detection Rate, False Alarm Rate, and accuracy are the evolution

metrics that are used by the authors for determining the efficiency of DL when compared to conventional ML in IoT. NSL-KDD data set has been utilized to evaluate the performance of distributed and parallel network recognition scheme. Authors have compared their deep models with the shallow models for attaining greater performance. Meidan et.al [7] have applied RF which is a supervised ML algorithm for removing characteristics from network traffic information. They gathered data from seventeen distinct IoT devices to train as well as test the multi-class classifiers. Authors have considered Synthetic data set along with accuracy as an evolution metric. They have examined the classification based on the moving window. Tama and Rhee [8] have addressed deep NN (DNN) for categorizing attacks in IoT. They have calculated the performance of the proposed method with the CIDDS-00, UNSW-NB15, and GPRS which are the 3 new standard data sets. A comparison of DNN with a grid search approach was used to attain the best parameter for an individual dataset. Their experimental analysis showed the efficiency of the proposed method with DNN in some evolutionary metrics of accuracy, recall, precision, and false alarm rate respectively. Pajouh et.al [9] have presented a new technique on 2-tier classification module and 2-layer dimension reduction module for identifying cruel activities such as Remote to Local (R2L) and User to Root (U2R) attacks. To recognize the mistrustful behavior, authors have used KNN as well as the naïve Bayes algorithm. It was found that SVM and 2-tier were compared with BIRCH, ESC-IDS, Association rule IDS, and HFR-MLR. The authors have proved that the Imbedded fault function decided the trained NSL-KDD dataset, to choose the most positive number of an element with the least error and information. Hodo et.al [10] have trained a multi-level Perceptron with internet packet outlines and charged on its capability to Distributed Denial of Service (DDoS) or Denial of Service (DoS) stabbings. They have authenticated the ANN process against a replicated IoT network. It was found using the KDD99 data set and examined data from many parts of IoT. Various advances of intrusion detection (ID) techniques on ID patterns were utilized and outperformed 99.4% accuracy. Authors have mentioned their future investigation is on recurrent and convolution NN advance. Veeramachaneni et.al [11] have presented AI2, an analyst-in-the-loop safety scheme where Analyst Intuition (AI) is combined with ML to construct a total end-to-end AI solution. They have presented four major characteristics such as an outlier recognition system, a big data behavioral analytics policy, a supervised learning section and a machine to get feedback from safety analysts. Recall and AUC are considered as evolution metrics for attaining better performance. D'Angelo et.al [12] have presented a novel inconsistency detection scheme based on ML and more accurately on U-BRAIN which is a batch relevance based fuzzified learning algorithm. They have utilized NSL-KDD data set in U-BRAIN by intending at understanding exact laws and policies governing regular or irregular network traffic to consistently form its operating dynamics. It was found that the proposed classification method has compared with J48, Naive-Bayes, MLP and SVM for outperforming accuracy as well as flexibility when handling with ambiguity in the recognition process. Ham et.al [13] have applied linear SVM to identify Android malware. The authors have contrasted the performance of SVM's malware detection with some other ML classifiers. They have introduced an ML method to remedy the drawback of the projected technique as well as to appropriately identify malware aiming at the Android platform. To raise efficiency, authors have chosen correlated features with malware. The proposed method was compared with Bayesian network, DT, RF, and naïve Bayesian

where SVM outperformed better results to deal with the problem of categorizing nonlinear information of the input characteristics.

3. Proposed Methodology

The proposed work has the objective to design a multi-class adaptive boosting based model for prediction of anomalies in IoT network traces data [14]. In this work, we have used IoT security dataset from kaggle [14] for the model evaluation. This dataset is produced in a virtual environment through Distributed Smart Space Orchestration System (DS2OS) which is composed of communications information among various IoT nodes in application layer. There are total of 357,952 samples each having 13 no. of attributes. This dataset covers eight types of anomalies those are 'data Probing' (dP), 'DoS attack' (DoS), 'malitious Control' (mC), 'malitious Operation' (MO), 'scan', 'spying' and 'wrong SetUp' (wSU). The proposed Multi-class Adaptive Boosting (AdaBoost) [15] model makes use of Decision tree (DT) as base classifier for prediction of anomaly type . The working schema of the proposed model and step by step computation is presented Algorithm 1 respectively.

$$W_i^0 = 1/n \tag{1}$$

In Eq.1, $W_i^0 = 1/n$ is the i^{th} weight for I_i at time $t = 0$ and n is the total of IoT profiles in I.

$$IGain\left(F_I, f_I^j\right) = IMeasure\left(F_{I'}\right) - \frac{F_I{}^L}{F_I} InfoMeasure\left(F_I{}^L\right) - \frac{F_I{}^R}{F_I} IMeasure\left(F_I{}^R\right) \tag{2}$$

$$IMeasure_{giniI[FIS]} = 1 - \Sigma_{ai \in a} P(a_i|I), S \in \{L, R\} \tag{3}$$

In Eq.2 and Eq.3, $f_I^j \in F_I$ is the selected feature from feature set $F_I = \{f_I^1, f_I^2 ... f_I^m\}$ of I for splitting. $F_I{}^L$ and $F_I{}^R$ are the feature sets at left and right part of the sub-tree of the DT^t .

$$a^{'} = DT^t(I) \tag{4}$$

In Eq.4, $a^{'}$ is the anomaly prediction vector and $DT^t(I)$ is the t^{th} Decision Tree.

$$e^t = Error(W^t \left[1_{a_i^{'} \neq a_i}\right]_{i=1}^n) \tag{5}$$

In Eq.5, e^t is the weighted anomaly prediction error vector and W^t is the t^{th} weight vector.

$$\delta^t = \frac{1}{2} \times \ l \ n\left(\frac{1 - e^t}{e^t}\right) \tag{6}$$

In Eq.6, δ^t is the weight parameter of t^{th} model.

$$W_{I_i}^{t+1} = \frac{W^t\left(I_{i,1}, I_{i,2} ... I_{i,m}, a_i\right) e^{\left(-\delta^t \times a_i \times D \ T^t(I_i)\right)}}{\theta} \tag{7}$$

In Eq.7, $W_{I_i}^{t+1}$ is the $(t+1)^{th}$ weight of I_i and θ is the normalization factor subjected to condition $\sum_{i=1}^n W_i^t = 1$

$$\rho_{AdaBoost}\left(I\right) = \sigma\left(\sum_{i=1}^{N} \delta^t DT^t\left(I\right)\right) \tag{8}$$

In Eq.9, $\sigma\left(.\right)$ is the sigmoid activation function and $\rho_{AdaBoost}\left(I\right)$ is the final prediction on I .

4. Experimental Setup and Results Analysis

In this section, the experimental set up and result analysis has been presented and discussed in details.

The proposed method has been implemented on a system having Intel(R) Core(TM) i7-6700 CPU

Algorithm 1: Multi-class Adaptive Boosting Model for Anomaly Prediction

Input: $I=\{I_1, I_2 .. I_n\}$ be the IoT network activity profiles, where $I_i=\{I_{i,1}, I_{i,2} .. I_{i,m}, a_i\}$ denotes i^{th} IoT network activity profile. Here, I_i denotes i^{th} IoT network activity profile, m is the number of features of communication profile in the dataset and $a_i \in a$, $a=\{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8\}$, where a_1, a_2, a_3, a_4, a_5, a_6, a_7 and a_8 symbolizes dP, DoS, mC, mO, scan, spying, wSU and *Normal* respectively.

1. Initialization of the weights of each $I_i \in I$ as presented in Eq.1.

2. Repeat for $t=0$ to N

i) Design the $DT^t(I)$ by using splitting along features by using information gain computation (Eq.2) using Gini index (Eq.3) and add the decision tree sequentially.

ii) Predict the anomalies by using trained model $DT^t(I)$ as presented in Eq.4.

iii) Pick the model with least weighted prediction error as presented in Eq.5.

iv) Find out the weight parameter of t^{th} model as in Eq.6.

v) Updating the weight of each I_i as in Eq.7.

vi) If $(e^t - e^{t+1} < \lambda)$ then Break

Else Continue

EndFor

@3.40 GHz, 4.00 GB RAM, 64 bit OS Windows 10 configurations. Simulation environment includes Python Anaconda open source distribution and Spyder IDE. The parameters of Decision Tree and AdaBoost are set as follows: Decision Tree (max_depth=5, random_state=1) AdaBoostClassifier(base_estimator = DecisionTree (max_depth= 5, random_state=1), n_estimators=50) with Training and Testing Spit: 70% - 30%. Various performance metrics such as Accuracy (Eq.9), Sensitivity (True Positive Rate (TPR)) (Eq.10), False Positive Rate (FPR) (Eq.11), Precision (Eq.12), Specificity (Eq.13), F1-Score (Eq.14) and ROC–AUC (Receiver Operating Characteristic - Area Under the ROC Curve) curve has been computed and compared (Table 1) to study the effectiveness of the proposed method.

$$Accuracy = \frac{tp+tn}{p+tn+fp+fn} \tag{9}$$

$$Sensitivity = \frac{tp}{tp+fn} \tag{10}$$

$$FPR = \frac{fp}{tn+fp} \tag{11}$$

$$Precision = \frac{tp}{tp+fp} \tag{12}$$

$$Specificity = \frac{tn}{tn+fp} \tag{13}$$

$$F1-Score = \frac{2 \times t\ p}{2 \times t\ p+fp+fn} \tag{14}$$

In Eq. 9 to Eq. 14, the tp ,fp ,tn and fn are symbolized for True Positive, False Positive, True Negative and False Negative respectively.

The ROC-AUC curve of DT, LDA, LR, MLP, NB, RF, and AdaBoost is shown in Fig. 1to Fig. 7 respectively. The comparisons of all these considered

models based on various performance metrics are shown in Table 1. The proposed ensemble-based methods are compared with some other competitive research in the literature. Table 1 depicts the performance comparison of some of the competent research on IoT anomaly detection with the other work. From the table, it may be conferred that, the proposed system is able to detect anomalies with higher accuracy as compared to other methods. The confusion metric of the proposed AdaBoost model has been demonstration in Table 2.

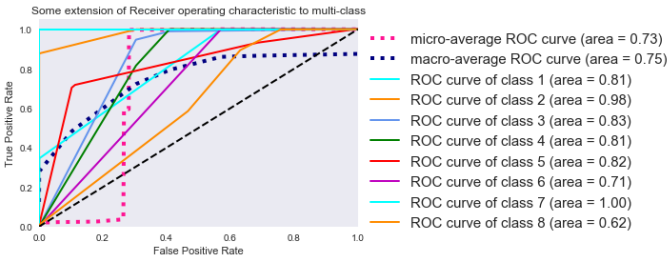

Figure 1. ROC analysis on performance of DT

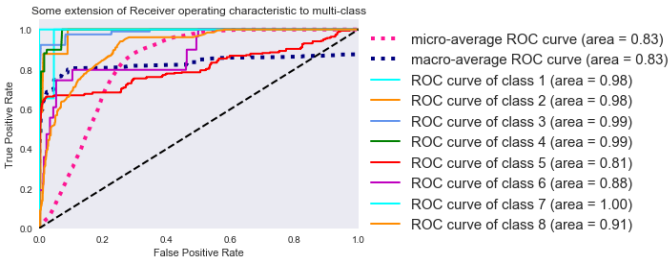

Figure 2. ROC analysis on performance of LDA

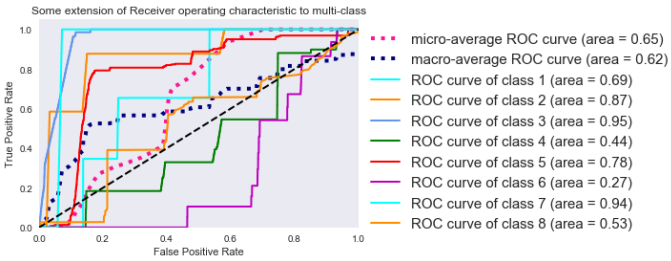

Figure 3. ROC analysis on performance of LR

The performance comparison of other existing model with proposed model has been presented

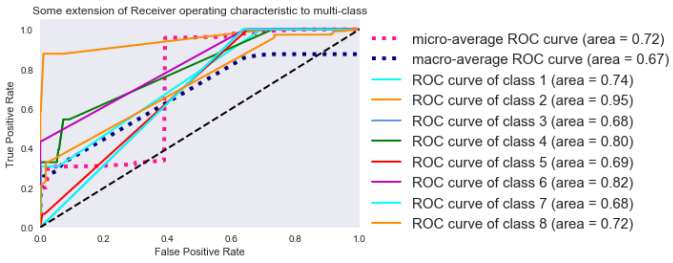

Figure 4. ROC analysis on performance of MLP

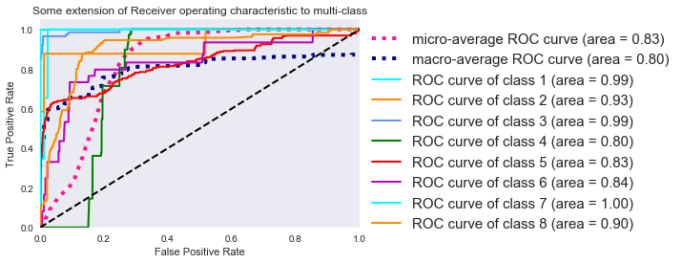

Figure 5. ROC analysis on performance of NB

in Fig. 8. Furthermore, the performance of the proposed ensemble model has been verified with increasing estimator for detecting the anomalies and shown in Fig. 9.

5. Conclusion

With the growing demand for IoT based automated systems, the complicacy of such models is becoming complex day by day. As the preference towards the data-driven system is increasing, so research on machine learning (especially advanced methods) is leading along with IoT. This work developed an AdaBoost based method for a smart and secure based IoT framework, which can identify its anomalies with a strong firewall system. Moreover, in abnormality cases, the proposed system can identify and protect the system from all types of threats and attacks. Based on the simulation results, it is evident that the proposed ensemble-based technique is efficient in handling such cyber-attacks an IoT framework. As compared to classical

Table 1
Evolution of Honeypot Against DDoS attacks

Prediction Models	Performance Metrics									
	TP	FP	TN	FN	Sensitivity (TPR)	FPR	Precision	Specificity	F1-Score	ROC-AUC
RF	11746041	0	5780	154691	0.98	0	1	1	0.99	0.99
NB	9135447	0	5780	251171	0.97	0	1	1	0.98	0.98
LR	4187764	5780	2000	576234	0.87	0.74	0.99	0.25	0.93	0.56
DT	9752627	18900	2000	2800	0.99	0.9	0.99	0.09	0.99	0.54
LDA	10330274	0	5780	139265	0.98	0	1	1	0.99	0.99
MLP	12306460	0	5780	24039	0.99	0	1	1	0.99	0.99
Proposed AdaBoost	16925026	14000	3780	0	1	0.78	0.99	0.21	0.99	0.6

Table 2
Confusion Metric of Proposed AdaBoost

	dP	DoS	mC	mO	scan	spying	wSU	Normal
dP	3780	0	0	0	0	0	0	2000
DoS	0	342	0	0	0	0	0	0
mC	0	0	889	0	0	0	0	0
mO	0	0	0	805	0	0	0	0
Scan	0	0	0	0	1416	0	0	131
Spying	0	0	0	0	0	420	0	112
wSU	0	0	0	0	0	0	122	0
Normal	0	0	0	149	18	12537	0	335231

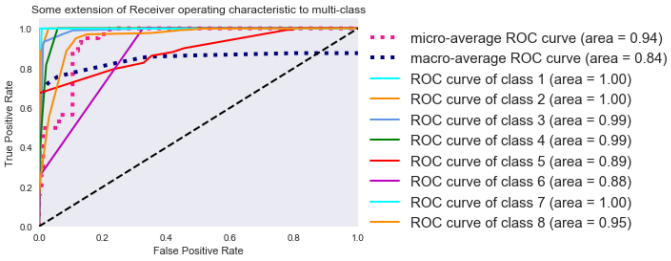

Figure 6. ROC analysis on performance of RF

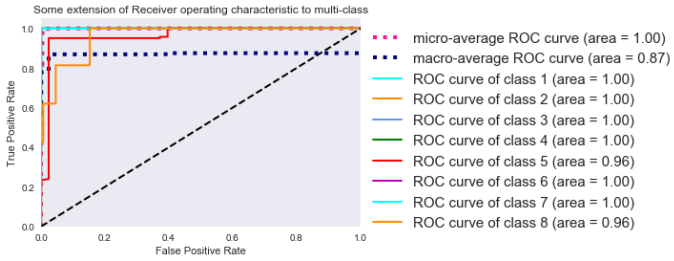

Figure 7. ROC analysis on performance of AdaBoost

machine learning-based methods(RF, NB, LR, DT, LDA, MLP), the proposed method is proved as a robust algorithm in handling such virtual environment data. However, future work may comprise of the simulation and development of efficient machine learning-based models that can handle real-time IoT data.

References

[1] M. Hasan, M.M. Islam, M.I.I. Zarif and M.M.A. Hashem, "Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches", Internet of Things, Vol.7, pp.1-14, 2019.

[2] O. Salman, I. Elhajj, A. Chehab and A. Kayssi., "IoT survey: An SDN and fog computing perspective", Computer Networks, Vol.143, pp.221-246, 2018.

[3] W.H. Hassan, "Current research on Internet of Things (IoT)

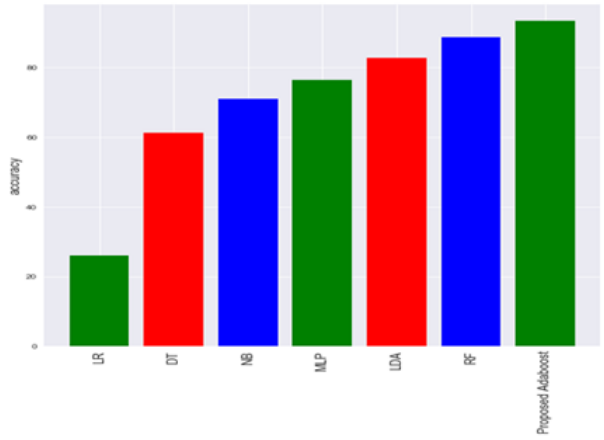

Figure 8. Comparison of accuracy

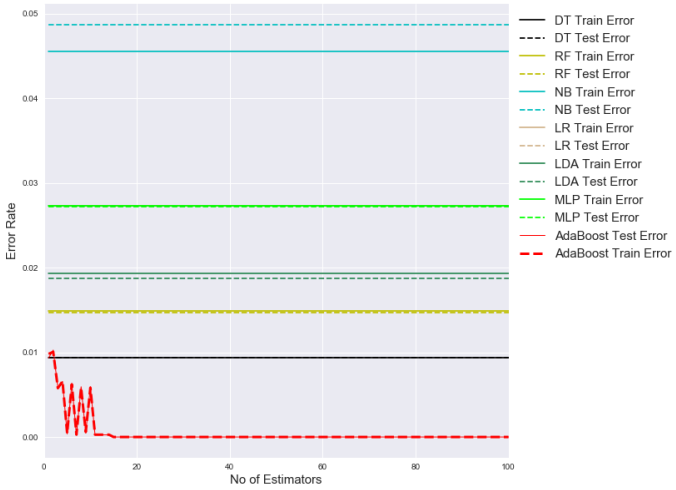

Figure 9. Performance of proposed AdaBoost with various numbers of estimators

security: A survey", Computer Networks, Vol.148, pp.283-294, 2019.

[4] L. Xiao, X. Wan, X. Lu, Y. Zhang and D. Wu, "IoT security techniques based on machine learning: How do IoT devices use AI to enhance security?", IEEE Signal Processing Magazine, Vol.35, No.5, pp.41-49, 2018.

[5] L. Deng, D. Li, X. Yao, D. Cox and H. Wang, "Mobile network intrusion detection for IoT system based on transfer learning algorithm", Cluster Computing, Vol.22, No.4, pp.9889-9904, 2019.

[6] A.A. Diro and N. Chilamkurti, "Distributed attack detection scheme using deep learning approach for Internet of Things", Future Generation Computer Systems, Vol.82, pp.761-768, 2018.

[7] Y. Meidan, M. Bohadana, A. Shabtai, M. Ochoa, N.O. Tippenhauer, J.D. Guarnizo, and Y. Elovici, "Detection of unauthorized IoT devices using machine learning techniques", arXiv preprint arXiv:1709.04647, pp.1-13, 2017.

[8] B.A. Tama and K. H. Rhee, "Attack classification analysis of IoT network via deep learning approach", Research Briefs on Information & Communication Technology Evolution (ReBICTE), Vol.3, pp.1-9, 2017.

[9] H.H. Pajouh, R. Javidan, R. Khayami, D. Ali and K.K.R. Choo, "A two-layer dimension reduction and two-tier classification model for anomaly-based intrusion detection in IoT backbone networks", IEEE Transactions on Emerging Topics in Computing, Vol.7, No.2, pp.314-323, 2019.

[10] E. Hodo, X. Bellekens, A. Hamilton, P.L. Dubouilh, E. Iorkyase, C. Tachtatzis and R. Atkinson, "Threat analysis of IoT networks using artificial neural network intrusion detection system" 2016 International Symposium on Networks, Computers and Communications (ISNCC), Yasmine Hammamet, Tunisia, pp.1-6, May 11-13, 2016.

[11] K. Veeramachaneni, I. Arnaldo, V. Korrapati, C. Bassias and K. Li, " AÎ2: training a big data machine to defend" 2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), New York, NY, USA, pp. 49-54,9-10 April 2016.

[12] G. D'angelo, F. Palmieri, M. Ficco and S. Rampone, "An uncertainty-managing batch relevance-based approach to network anomaly detection", Applied Soft Computing, Vol. 36, pp.408-418, 2015.

[13] H. S. Ham, H. H. Kim, M. S. Kim and M. J. Choi, "Linear SVM-based android malware detection for reliable IoT services", Journal of Applied Mathematics, Vol.2014, pp. 1-11, 2014.

[14] DS2OS traffic traces, IoT traffic traces gathered in a the DS2OS IoT environment, kaggle, 2018

[15] T. Hastie, S. Rosset, J. Zhu and H. Zou, "Multi-class adaboost", Statistics and its Interface, Vol.2, No.3, pp. 349-360, 2009.