Proceedings of the 38th Chinese Control Conference
July 27-30, 2019, Guangzhou, China

A Novel Intrusion Detection System based on Extreme Machine Learning and Multi-Voting Technology

Jianlei Gao, Senchun Chai, Chen Zhang, Baihai Zhang, Lingguo Cui

School of Automation, Beijing Institute of Technology, Beijing 100081, P. R. China

E-mail: chaisc97@bit.edu.cn

Abstract: With the fast development of networking technology, billions of devices have been developed with network function. When the scale of a network traffic grows by an order of magnitude, traditional intrusion detection system (IDS) are no longer effective to detect malicious network intrusions. In our work, we propose a novel network intrusion detection framework based on extreme learning machine (ELM) and multi-voting technology (MVT). Due to the real time feature of ELM, several independent ELM networks can be trained simultaneously. The final results are obtained by MVT strategy. The standard UNSW-NB15 data set has been used to evaluate the performance of the proposed method. The experimental result illustrated that the high accuracy can be achieved by using the proposed method.

Key Words: Intrusion Detection System (IDS), Extreme Learning Machine (ELM), Multi-Voting Technology (MVT), UNSW-NB15

1 Introduction

With the development of modern technology, Internet technology has developed rapidly. The emergence of internet technology has been brought us into a new world of interconnection, which makes network become a very import part of our modern life, and provides us with convenience and promotes the progress of our world.

But, it also brings us lots of security problems. According to the report from Kaspersky Laboratory in the second quarter of 2018, more than $962,947,023$ malicious intrusions are launched in 187 countries, which is significantly higher than the number of previous quarters. Actually, with the rapid development of mobile networking, attacks against mobile devices are also showing a trend of explosion, which aggravates the severity of the situation and takes malicious mobile software as an example shown in Fig. 1.

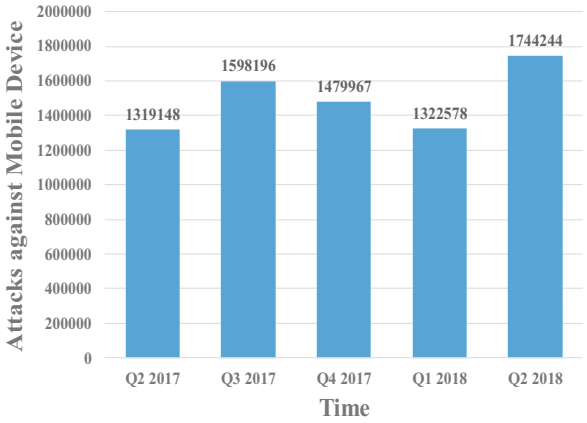

Fig. 1: Global Mobile Threat Situation

Due to the advantages of internet technology, more and more industrial control system (ICS) has been widely used and many industries have adopted internet technology to communicate with each other to exchange various data. Especially, most ICSs are applied in critical infrastructures that are related to national economy, public livelihood and national security. In addition, a variety of attack tools emerged with the feature of randomness of downloading and extremely simple way to use, thus making cyber attacks occur more and more easily and frequently. Once these facilities are subjected to cyber attacks, there will be no end of trouble for the future.

Therefore, protecting devices and systems against malicious attacks becomes an important and urgent task, since this intrusion can result in great loss. In order to solve this knotty problem, network intrusion detection system was created to defend against various network intrusions. In our paper, we propose a method combined ELM with MVT to detect aggressive behaviors of network, which get a relative better performance in efficiency and time.

The major contributions of our paper are summarized as follows: (1) A new method combined ELM and MVT is proposed in IDS, which not only provides better performance but also reduces the consuming time. (2) This method provides more computation capacity , which is suitable for dealing with massive data. (3) This method provides an new idea for the application of IDS with less training samples.

This work is supported by National Nature Science Foundation of China (No.61573061).

The remainder of our paper is as follows. Some related works are introduced in section II. In section III, we give an introduction about ELM, multi-voting technology and its work flow. A standard data set named UNSW-NB15 and some evaluation criterions are presented in section IV. What's more, an experiment is done to test the efficiency of method proposed by us in section V. Finally, this article ends with a conclusion.

2 Intrusion Detection System

The concept of IDS is firstly proposed by Anderson, which is a kind of network device and is designed to detect malicious cyber actions on communication networks[1]. IDS is a process that protects the system from malicious invasion by collecting,analyzing and distinguishing some effective information which can be data size of packets, the characteristic information of packets, the behavior model of attacker, the rule of access.etc. from networks or computers. It is the most widely used active security defense strategy in information security field.

Many people are devoted to studying and designing an IDS to detect anomaly. The traditional IDS is aimed to distinguish abnormal action of network environment, which usually is based on the statistics technology or blacklist[2]. However, traditional IDS is unable to detect more deliberate and powerful attacks. Therefore, more intelligent algorithms are proposed to design IDS, for example: data mining [3], artificial immune system [4], decision tree deep[5], support vector machine (SVM) [6], extreme learning machine [7] and artificial neural network [8].

Anomaly detection is a two-class problem in fact. That is to say, it can separate normal data from abnormal data by its function. SVM is a new machine learning algorithm developed from statistical learning theory [9], which is widely applied in anomaly detection. In order to overcome the problems of long training time and impracticability in large-scale data set, various kinds of deep learning algorithms are improved by researches. For example: particle swarm optimization (PSO) is proposed to optimize the model of SVM to find the optimal parameters, which is called PSO-SVM and can improve the detection efficiency of IDS [10]; paper [11] chooses convolutional neural network to construct a massive network to deal with the intrusion detection problem, and solve imbalance problem of NSL-KDD data set through setting the cost function weight coefficient of attack classes. A method combined preprocess of SOM network with BP neural network to build a model of IDS is proposed by paper [12]. But, above methods have their shortcomings. SVM is not suitable for dealing with multi-class and massive data. The existence of weight parameter's backward iteration makes Back Propagation (BP) and Convolutional Neural Networks (CNN) spend too much time on the training process of network. These problems are not conductive to anomaly detection and solution of the over-fitting problem.

In the view of this, a machine learning algorithm named ELM and its improved methods are utilized in IDS to check abnormal network behaviors. Paper [13] chooses ELM as the core learning algorithm to build a novel multiple kernel learning framework to improve the efficacy of IDS. A novel dual adaptive regularized online sequential ELM is introduced to IDS to distinguish network intrusions which selects ridge regression factor based on Tikhonov regularization to solve the over-fitting problem [14]. Huang *et. al.* [15] proposes an ELM based on principal component analysis (PCA) to enhance detection rate, which firstly use PCA to reduce the dimension of data set and then use ELM to detect attacks.

However, the continuous progress of technology and research of attackers on targets and increasing number of cyber devices produces massive data with non-linearity and high dimension. Those improved ELMs become more and more inadequate to deal with the ever-increasing volume and repetitive samples in data set and can not meet the rapid detection requirements of IDS. So, in this paper, in order to accelerate the detection speed, improve the detection accuracy and reduce degree of over-fitting, we use ELM as our key algorithm of IDS and train the network construct with samples which is choosed randomly from data set in a fixed proportion. We repeat this process for certain times and get the results handled by IDS. Finally, we got the final outcome according to MVT.

3 Proposed Method

3.1 Extreme Learning Machine

In this subsection, we present some algorithm principles of ELM, the critical detection algorithm of IDS. ELM is a new kind of machine learning algorithm proposed by [16, 17], which actually is a feedforward neural network and different from feedback neural network algorithm, for example: back propagation neural network. ELM is a novel learning algorithm and a single-hidden-layer neural network as shown in Fig.2.

Actually, this machine learning algorithm has a stronger learning capacity, more computation ability, faster conver-

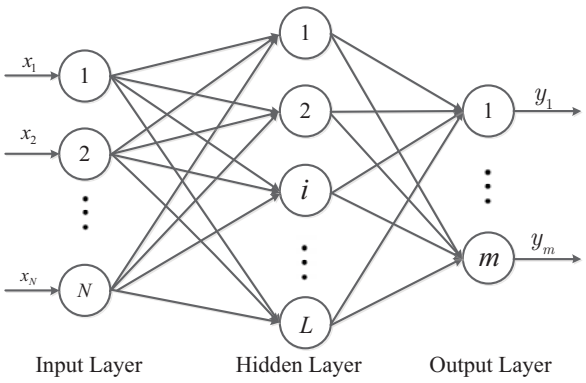

Fig. 2: The Structure of ELM

gence, faster training speed than other machine learning methods due to no existing feedback error iteration calculation. It is assumed that there exist N arbitrary and different samples $N = \{(x_i, y_i), i = 1, \cdots, N\}$, where $x_i = [[x_{i1}, x_{i2}, \cdots, x_{in}]^T \in \mathbf{R}^n$, $y_i = [[y_{i1}, y_{i2}, \cdots, y_{im}]^T \in \mathbf{R}^m$ and L additive hidden nodes, so ELM can be showed:

$$o_L(x) = \sum_{i=1}^{L} \beta_i G_i(\alpha_i \bullet x_i + b_i), x_i \in \mathbf{R}^n, \alpha_i \in \mathbf{R}^n, \beta_i, b_i \in \mathbf{R} \tag{1}$$

where $\boldsymbol{\alpha_i} = [x_{i1}, x_{i2}, \cdots, x_{in}]^T$ is the input weight of ith hidden node; b_i is the bias of ith hidden node; $\boldsymbol{\beta_i} = [\beta_{i1}, \beta_{i2}, \cdots, \beta_{in}]^T$ is the output weight of ith hidden node; G_i is the output activation function; o_i is the output of ith hidden node; and $\bullet$ represents the inner product in mathematical sense. What excitation functions commonly used can be selected from $Sigmoid$, $Sine$ and RBF . The purpose of training process using this machine learning is to find some appropriate parameters that can match the training data set without error. So, the formula (1) can be implemented:

$$o_L(x) = \sum_{i=1}^{L} \beta_i G_i(\alpha_i \bullet x_i + b_i) = y_i \tag{2}$$

The above formula (2) can be simplified:

$$H\boldsymbol{\beta} = Y \tag{3}$$

where,

$$H = \begin{bmatrix} G(\alpha_1 \cdot x_1 + b_1) & \cdots & G(\alpha_L \cdot x_1 + b_1) \\ G(\alpha_1 \cdot x_2 + b_2) & \cdots & G(\alpha_L \cdot x_2 + b_2) \\ \vdots & \cdots & \vdots \\ G(\alpha_1 \cdot x_N + b_N) & \cdots & G(\alpha_L \cdot x_{jN} + b_L) \end{bmatrix}$$

$$\boldsymbol{\beta} = [\beta_1^T, \beta_2^T, \cdots, \beta_L^T]^T$$

$$\boldsymbol{Y} = [y_1^T, y_2^T, \cdots, y_L^T]^T$$

where H is the output matrix of hidden layer. We can finish the training of network by finding the solution of $H\boldsymbol{\beta} = Y$ after initializing a random input weight and bias of ELM.

$$\hat{\boldsymbol{\beta}} = H^{\dagger} Y = (H^Y H)^{-1} H^Y Y \tag{4}$$

Where $H^{\dagger}$is the Moorse-Penrose generalized inverse matrix.

3.2 Multi-Voting Technology

MVT is a basic decision method and is usually applied in electoral activities. The principle of multi-voting is that it can get theoretical optimal result probability according to the voting outcomes, which is assumed that each voting is independent and identically distributed. That is to say, the probability of each event is the same.

The classical multi-voting technology is defined the probability of single event as the same p. When we take N votes, the probability of theoretical voting can be expressed as:

$$P = \sum_{j=(N+1)/2}^{N} C_N^j p^j (1-p)^{(N-j)} \tag{5}$$

where, C_N^j indicates that j events occurred in N votes. Assuming that events are independent of each other and that the probability of occurrence of the first event is pI, the theoretical probability of occurrence of events can be obtained by voting method.

It is assumed that every event is independent of each other in observed N times, and the probability of occurrence of ith event is p_i, therefore we can obtain the theoretical probability of occurrence of the event by MVT:

$$P = \sum_{j=(N+1)/2}^{N} C_N^j \prod_{i=1}^{j} p_i^j (1 - p_i^{N-j}) \tag{6}$$

Fig. 3[18] shows the relation between theoretical probability of voting decision and the probability of single event. From this picture, it shows obviously that if the $p_i \geq 0.5$ the result of multi-voting is better than the result without voting; the $p_i < 0.5$ the result is opposite.

3.3 Workflow of Method

In this subsection, the ELM algorithm is selected as our IDS's kernel, and the MVT is used for decision-making, whose work flow is displayed as followed in Fig. 4.

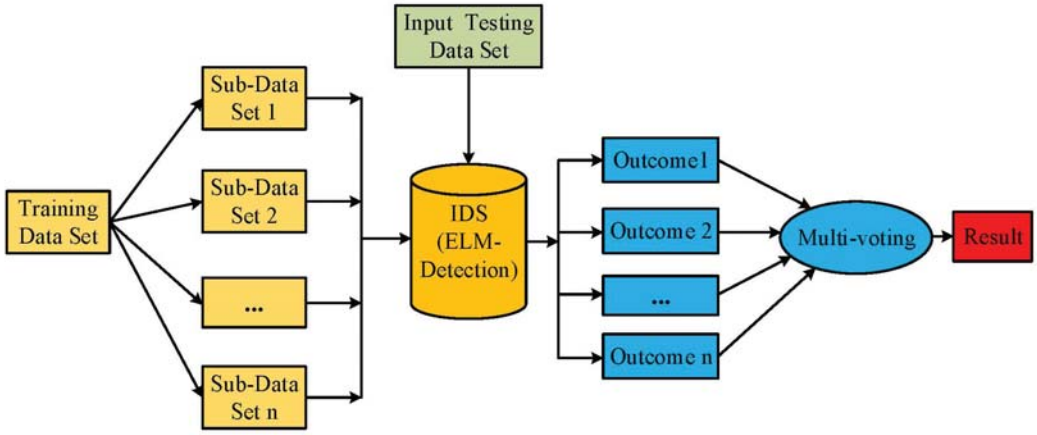

Fig. 4: The workflow of IDS

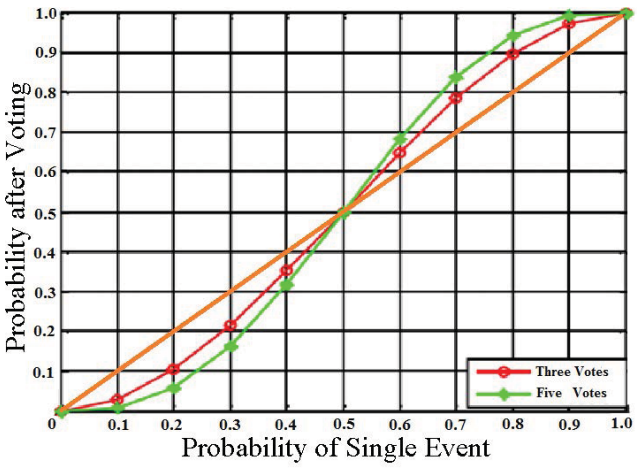

Fig. 3: The relation between probability and single event

Table 1: The specific detail of UNSW-NB15 Data Set

Index	Attack Types	Training Set	Testing Set
1	Analysis	2,000	677
2	DoS	12,264	4,089
3	Fuzzers	18,184	6,062
4	Generic	40,000	18,871
5	Shellcode	1,133	378
6	Worms	130	44
7	Exploits	33,393	11,132
8	Backdoor	1,746	583
9	Normal	56,000	37,000

4 Data Set and Evaluation Criterion

4.1 UNSW-NB15 Data Set

The UNSW-NB15 data set is created by Australian Center for Cyber Security (ACCS) in 2015[19, 20]. It is a new data set in research about IDS. This data set is to solve the inherent problems of classical KDD99 and improved NSL-KDD data set. UNSW-NB15 data set have nine different modern attack types with 49 features, which has 5 more attacks types than NSL-KDD.

This data set consists 2,540,044 samples, and include 9 attack types, such as Fuzzers, DoS, Analysis, Reconnaissance, Exploit, Shellcode, Worm, Backdoor, Generic, whose specific amount is shown in Table 1. For easy use, it is divided into 2 parts: a training data set (82,332 samples) and a testing data set (175,341 samples) .

4.2 Evaluation Criterion

In this subsection, some criterions are given to evaluate our method's function. As is known to all, there are many quotas provided to evaluate the IDS's performance, such as the detection rate, the false alarm rate, the detection accuracy, the detection time and so on. In order to simplify the performance evaluation, we select the detection accuracy Acc and the detection time T as our criterions. Let us show some concepts before we give the specific definitions of evaluation criteria.

- **TP** :(true positive) which indicates that the normal samples are detected as normal samples.
- **FP** :(false positive) which indicates that the abnormal samples are detected as normal samples.
- **FN** :(false negative) which indicates that the normal samples are detected as abnormal samples.

- **TN** :(true negative) which indicates that the abnormal samples are detected as abnormal samples.
- **T** : (detection time) It is the cost time spent by IDS on detecting dataset, which is the interval between starting of IDS and ending of IDS.

Therefore, it can be obtained:

$$Acc = \frac{TP + TN}{TP + TN + FN + FP} \tag{7}$$

Obviously, the more Acc and the less T, the better IDS is.

5 Experiment

5.1 Preprocessing of Data Set

In this paper, we adopt min-max normalization method to normalize data samples, which transforms the original value to make sure that it is mapped between $[0, 1]$. The function is as follows:

$$x' = \frac{(x - min(x))}{max(x) - min(x)} \tag{8}$$

Where, x' is the new value after normalization; $min(x)$ and $max(x)$ are the minimum value and maximum value of the samples of $x-$fitting feature respectively.

5.2 Results Analysis

In our paper, we extract some samples from training data set randomly to construct a new sub-data set. Then train this sub-data set, test our testing data set, and repeat it N times to get the original results. In order to simplify the analysis, we take an example of 1%,5%,10% here, and repeat 100 times .

Table 2: Accuracy of different ratio of Data Set

Index	1%	5%	10%	100%
Average Acc	83.823%	88.317%	88.937%	89.281%

It can be seen from above Table. 2, the detection accuracy Acc is 89.281% when we use all samples of training data set. And whether 1%, 5% or 10% of the training data set, the detection accuracy Acc obtained is greater than 80%. According to Fig. 5, due to detection accuracy $Acc > 0.5$, the performance with MVT will be improved when we just use $N = 3$..

In this section, the parameter of MVT $N = 3$. Fig. 6 shows the average detection time T and average detection accuracy Acc of IDS. Obviously, the detection accuracies Acc with MVT 89.29% and 89.71% when we select 5% and

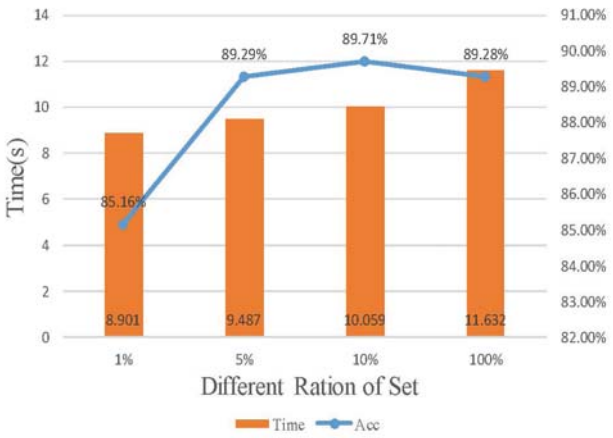

Fig. 5: The Acc and T of Detection with MVT

10% of data set are bigger than the accuracy we use 100% of data set. What's more, their average detection time T 9.487s and 10.059s are also smaller than the result of using entire data set.

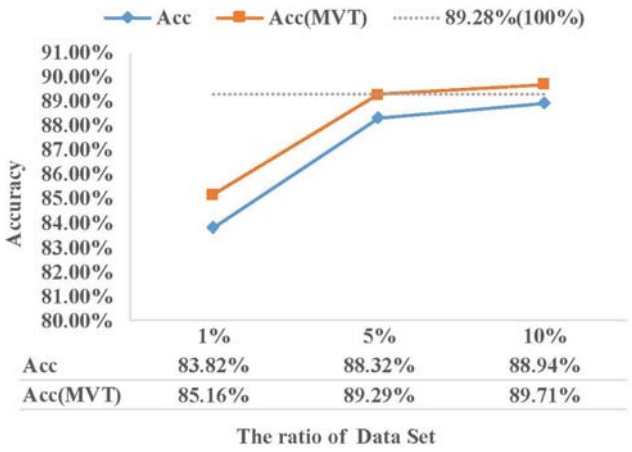

Fig. 6: The Acc and T with and without MVT

Fig. 6 displays some average detection accuracies Acc of IDS with MVT and without MVT when we randomly select 1%, 5% and 10% samples of data set, and they all have been improved by MVT. It is important to note that when only 5% of the data is taken we can achieve the accuracy that we used to get by using full data set.

6 Conclusion

The rapid development of internet technology has brought endless security issues. Especially, an increasing number of samples of data bring great challenges to IDS. In order to improve the detection accuracy of IDS and reduce the time cost of detection process, we firstly choose ELM as our IDS's critical detection algorithm, and train the network construction by selecting 1%, 5%, 10% of data set, and then

repeat this three times. Finally, we get the best result through using MVT. Experimental results show 1): the accuracy of IDS with MVT is better than the accuracy without MVT; 2) the accuracies using 5%,10% data set are better than the accuracy using 100% data set, and the time using 5%,10% data set is also lower than the accuracy using 100% data set. Thus, the proposed method in this paper can obtain the detection accuracy using whole data set and greatly reduce the consumed time.

References

[1] J. Anderson, Computer security threat monitoring and surveillance (1980), *http://csrc.nist.gov/publications/ history/ande80.pdf*

[2] H. Javitz, A. Valdes, The SRI IDES statistical anomaly detector, in *proceedings of the IEEE Computer Society Symposium on Research in Security & Privacy*, 1991.

[3] G. Nadiammai, M. Hemalatha, Effective approach toward Intrusion Detection System using data mining techniques, *Egyptian Informatics Journal*, 15(1): 37-50, 2014.

[4] S. Powers, J. He, A hybrid artificial immune system and Self Organising Map for network intrusion detection, *Information Sciences*, 178(15): 3024-42, 2012.

[5] T. Vuong, G. Loukas, D. Gan, A. Bezemskij, Decision Tree-based Detection of Denial of Service and Command Injection attacks on Robotic Vehicles, in *proceedings of the IEEE International Workshop on Information Forensics & Security*, 2016.

[6] L. Hui, X. Guan, X. Zan, H. Zhao, Network Intrusion Detection Based on Support Vector Machine, in *proceedings of the International Conference on Management & Service Science*, 2009.

[7] C. cheng, W. Tay, G. Huang, Extreme learning machines for intrusion detection, in *proceedings of the International Joint Conference on Neural Networks*, 2012.

[8] A. Zewairi, S. Almajali, A. Awajan, Experimental Evaluation of a Multi-Layer Feed-Forward Artificial Neural Network Classifier for Network Intrusion Detection System, in *proceedings of the International Conference on New Trends in Computing Sciences*, pp 167-172, 2018.

[9] V. Vapink, The Nature of Statistical Learning Theory, NewYork, 1995.

[10] W. Shang, S. Zhang, M. Wan, P. Zeng, Modbus/TCP Communication Anomaly Detection Algorithm Based on PSO-SVM, *Acta Electronica Sinica*, 490-491(11): 1745-53, 2013.

[11] D. Wu, Z. Chen, W. Li, A Novel Intrusion Detection Model for a Massive Network Using Convolutional Neural Networks, *IEEE Access*, vol. 6, pp. 50850-50859, 2018.

[12] M. Ramadas, S. Ostermann, B. Tjaden, Detecting Anomalous Network traffic with self-organizing Maps, *Lecture Notes in Computer Science*, pp. 36-54, 2003.

[13] J. Fossaceca, T. Mazzuchi, S. Sarkani, MARK-ELM: Application of a novel Multiple Kernel Learning framework for improving the robustness of Network Intrusion Detection, *Expert Systems with Applications*, 42(8):4062-4080, 2015.

[14] Y. Yu, S. Kang, H. Qiu, A new network intrusion detection algorithm: DA-ROS-ELM, *IEEJ Transactions on Electrical and Electronic Engineering*, 2018.

[15] S. Huang, W. Chen, J. Li, Network Intrusion Detection Based on Extreme Learning Machine and Principal Component Analysis, *Journal of Jilin University*, 2017.

[16] G. Huang, Q. Zhu, C. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, in *proceedings of the IEEE International Joint Conference on Neural Networks*, 2005.

[17] G. Huang, D. Wang, Y. Lan, Extreme learning machines: a survey, *International Journal of Machine Learning & Cybernetics*, 2(2): 107-22, 2011.

[18] J. Zou, Implementation of Narrowband Radar Air Target Classification Method Based on PowerPC, 2017.

[19] N. Moustafa, J. Slay, The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set, *Information Systems Security*, 25(1-3): 18-31, 2016.

[20] N. Moustafa, J. Slay, UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set), in *proceedings of the Military Communications & Information Systems Conference*, 2015.