Lapai Journal of Applied and Natural Sciences

LAJANS Vol 5(1): 177-182

RESEARCH PAPER

Detection and Analysis of DDoS Attacks in Internet Kiosk Voting Using Machine Learning Algorithms

A. J. Almustapha[1], O. M. Olaniyi[1], I. M. Abdullahi[2], J. Ndunagu[1] and M. Ibrahim[2*]

[1]*Department of Computer Science, Federal Polytechnic Bida, Niger-State, Nigeria*
[2]*Department of Computer Engineering, Federal University of Technology, Minna, Niger-State*

ABSTRACT

Classification algorithms recognize and differentiate class instances in a dataset to produce correctly classified output for better understanding of Distributed Denial-of-Service (DDoS) flood attacks captured on a workstation thereby improving availability of the e-voting server. Machine learning algorithms have been applied as detection mechanisms on DDoS attacks in securing network infrastructure by training the algorithms using datasets containing captured DDoS flood traffic on the network. In this paper, we compare and analyze the performance of Random Forest, Naïve Bayes and Multilayer Perceptron (MLP) machine learning classification algorithms on a sample of the Knowledge Discovery and Data Mining (KDD) Cup 99 Dataset containing four classes of DDoS attack using accuracy, precision and recall performance metrics. The training and testing of these classifiers on the dataset records was carried out in Waikato Environment for Knowledge Analysis 3.8.2 version tool using nine (9) best optimal attributes selected to produce confusion matrices at a reduced building model time. An accuracy of 98.65% in classifying DDoS flood attacks was achieved by MLP classifier. The study showed that the MLP classifier provides a better mechanism for DDoS detection for secure Internet voting system by increasing voting server's performance in terms of system's availability to voters during election process.

Keywords: Distributed Denial-of-Service (DDoS) Multilayer Perceptron (MLP) classifier, Normal Packets

INTRODUCTION

DDoS attacks are becoming more frequent especially on network services, such as Internet Kiosk Voting Systems. The goal of DDoS flood attack is to deny legitimate voters the ease and timeliness of casting ballots by making the e-voting server unavailable. Attack packets sent by an attacker are similar to legitimate packets and as such difficult to detect. With Internet kiosk voting, voters are allowed to cast their votes from computers placed in polling units and election officials are given the responsibility of authenticating voters as well as conducting the election process in a transparent manner. The security of Internet voting systems is paramount to the success of electronic democratic decision making and as such there is the need to address e-voting vulnerabilities that include Denial-of-Service (DoS) attacks. The goal of DDoS flood attack is to breach security and mar the availability as well as the reliability of voting service in e-voting systems during an election and thus questioning the required confidence and trust of the electronic democratic decision making, which could be carried out by overwhelming the vulnerable voting server with huge amount of packet requests thereby hijacking the server's resources. This form of attack could disrupt an election process by frustrating voters from casting their votes within a stipulated time frame. Detecting DDoS attacks early enough could minimize malicious flood traffic from reaching the voting server thereby increasing server's utilization and ensuring that voters do not get frustrated during voting. Machine learning algorithms have been applied as detection mechanisms on DDoS attacks in securing network infrastructure by training the algorithms using datasets containing captured DDoS flood traffic on the network.

Received 19 June, 2020
Accepted 05 October, 2020
Address Correspondence to:
almustapha@st.futminna.edu.ng

Researches based on Biometrics and Cryptographic schemes have been used to address security issues, such as voters' authentication, integrity and confidentiality during the voting process while more researches for detecting and preventing DDoS attacks have been conducted in physical network and cloud computing environment but less in e-voting as shown in Dhamdhere, *et al.* (2017) and Olaniyi, *et al.* (2015). Proposed schemes for DDoS attack detection include those based on Artificial Neural Network (ANN) explained in Perakovic, *et al.* (2017), Packet Sampling Threshold in Dominic, *et al.,* (2015), Multi-Filter Feature Selection Least Squares Support Vector Machine (LS-SVM) Aqeel, *et al.,* (2017) and Block-chain Wei, *et al.,* (2018).

Therefore, this paper focuses on detecting DDoS flood attacks in Internet voting by statistically analyzing both attack and legitimate traffic directed at the voting server by using KDD Cup 99 dataset. By training and testing the classifiers in Waikato Environment for Knowledge Analysis(WEKA) 3.8.2 version tool, DDoS flood attack packets were distinguished from normal packets using nine (9) optimal packet attributes from the KDD Cup 99' dataset.

A sample of the given dataset was analyzed using Random Forest, Naïve Bayes and MLP machine learning algorithms to classify DDoS flood attacks and these classifiers were evaluated based on accuracy, precision and recall performance evaluation metrics. The use of datasets for experimentation especially in machine learning has contributed to solving prediction and classification problems as this can be seen in where the dataset that included User Datagram Protocol (UDP), Smurf, Hyper-Text Transfer Protocol (HTTP) Flood and SQL Injection DDoS (SIDDoS) attacks by applying Naïve Bayes, Random Forest and MLP to classify DDoS attacks from normal traffic to achieve an accuracy of 98.63% for MLP. The paper of Thwe, *et al.,* (2013) is related to the research conducted here in that it attempts to identify DDoS flood attacks by using KNN classifier on UCLA dataset to produce an efficient and suitable system for recognizing DDoS flood attacks.

Specifically, this paper is an extension of the previous contribution in Brownlee, (2016) where MLP classification model was compared with Random Forest and Naïve Bayes classifiers in the process of detection of DDoS flood attacks from KDD Cup 99 dataset thereby improving the network performance and availability of the e-voting server in kiosk scenario

METHODOLOGY

Data Collection

This research made use of a sample of the KDD Cup 99 dataset, which contained 63,723 records, 27 attributes, and five (5) classes for training and testing in WEKA. An analysis of KDD Cup 99 dataset with to respect to Normal traffic class and four (4) DDoS flood attack traffic classes consisting of UDP Flood, Smurf, SIDDoS and HTTP Flood were used to train Multi-Layer Perceptron (MLP), Naïve Bayes and Random Forest classifiers. A 66% split method was used to determine accuracy (ACC), precision (PRC) and recall (RCL) of the classification models.

Sample collection and proceeding

Samples were collected on a monthly basis for six months (from March to September 2013). Microalgae samples were collected using a Van Dorn water sampler. Triplicate samples were collected, fixed with Lugol's iodine solution after which, were taken to the laboratory for taxonomic identification and enumeration.

Upon arrival in the laboratory, microalgae samples were allowed to settle for at least three days without disturbing. After which, the 500ml water sample were siphoned out until only 50ml of sample remained. The remaining samples were then transferred to labelled 100ml capacity opaque plastic bottle and kept in the dark until further analysis.

Attribute Selection

Attribute evaluator and search method are two important components of attribute selection. The entropy of each attribute from the KDD dataset is calculated and attributes with higher information gain value close to 1 are selected while attributes that contribute low information close to 0 are removed. The Ranker Search Method used with Information Gain Evaluator identifies the results of attribute selected by ranking from best to worst information gain as explained in Brownlee, (2016).

Investigation of the Classification Models

This section defines the various activities involved in selecting best attributes from the KDD dataset and the process of training the three classifiers to produce confusion matrix for each classification model.

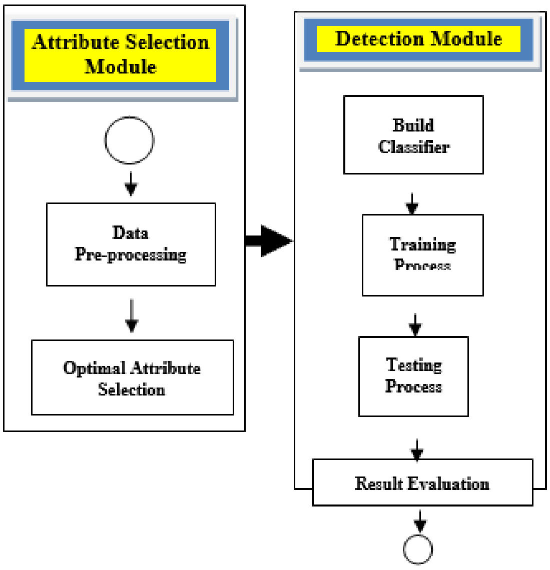

Figure 1: DDoS Flood Detection Model Investigation Process

Figure 1 shows an illustration of the various modules that are present in the proposed DDoS flood detection model and they are explained below.

- Data Pre-Processing Phase: This involves loading the dataset and selecting the number of attributes. The Class attribute PKT_CLASS (NOM) was used when filter was applied.

- Attributes Selection Phase: This involved applying the attribute evaluator called Information Gain (IG) and Ranker Search method on the full training set to determine the best attributes. Each attribute is ranked based on the information gain metrics and the best number of attributes that increases the classification model performance was chosen.

- Training Classifiers: Each classification model was trained to classify the data into Normal, UDP Flood, Smurf, SIDDoS and HTTP Flood. A 66% split option of the sample of KDD dataset was adopted since it is fast for training and testing dataset containing large records producing results by evaluating the performances of the models using accuracy, precision and recall metrics.

- Producing Classifier Evaluation Output: The outputs are the classification models on the training set, which can be viewed and these include confusion matrices for Random Forest, Naïve Bayes and MLP predictions as shown in Tables 2, 3 and 4 respectively.

Evaluation of Classifiers

The accuracy, precision and recall performance metrics derived from the four outcomes, namely true positive (TP), true negative (TN), false positive (FP) and false negative was used to evaluate the algorithms and get the best classifier for objective two. The classifier performance was evaluated based on the confusion matrices generated by Random Forest, Naïve Bayes and MLP algorithms shown in Figure 4.2, 4.3 and 4.4. The performance of the simulated model was also evaluated using the server utilization.

- Accuracy: This performance metrics measures how often the algorithm correctly classified the DDoS attacks, and it is denoted as

$$\text{Accuracy} = \frac{TP + TN}{\text{Total Classified Instances}} * 100\% \quad (1)$$

- Precision: This performance metrics measures how often the algorithm predicts yes when it is actually yes, and it is denoted as

$$\text{Precision} = \frac{TP}{FP + TP} * 100\% \quad (2)$$

- Recall: This performance metrics measures how often the algorithm predicts correctly when it predicts yes, and it is denoted as

$$\text{Recall} = \frac{TP}{FN + TP} * 100\% \quad (3)$$

RESULTS AND DISCUSSION

Packet Attribute Selection Results

The nine (9) best attributes chosen from applying Information Gain technique with Ranker search method are shown in Table 1.

In Table 2, the confusion matrix showed that correctly classified instances totalled 21370 while 296 were incorrectly classified for Random Forest algorithm. For Naïve Bayes algorithm, correctly classified instances totalled

20992 and 674 were incorrectly classified as shown in Table 3 while for MLP, correctly classified instances totalled 21374 and 292 were incorrectly classified shown in Table 4.

Table 5 is a summary of the results of all the classifiers. This result showed an improvement in the accuracy, precision and recall when 9 attributes are used for classification. The accuracy was 98.56%, 96.89% and 98.65% for Random Forest, Naïve Bayes and MLP respectively. Similarly, the precision and recall values indicate that MP is better with a value of 98.7 compared to 98.6 and 97.3 for naïve Bayes and random forest respectively. Based on the results obtained in Table 5, Figure 2 shows that MLP is the best classifier for detecting DDoS Flood attacks with promising performance results, hence, MLP was adopted for the mitigation model development.

Table 1: Selected Attributes with Highest Information Gain Value

Attribute Number	***Description***	***Type***
1	Byte_Rate	Continuous
2	Number_of_Byte	Continuous
3	Utilization	Continuous
4	Pkt_Rate	Continuous
5	Last_Pkt_Reserved	Continuous
6	Number_of_Pkt	Continuous
7	Pkt_Delay	Continuous
8	Pkt_Avg_Size	Continuous
9	Pkt_Size	Continuous

Table 2: Confusion Matrix for Random Forest

	Normal	UDP Flood	Smurf	SIDDoS	HTTP Flood
Normal	19435	0	0	3	0
UDP Flood	197	1795	0	0	0
Smurf	82	0	45	4	0
SIDDoS	4	0	0	55	0
HTTP Flood	1	1	4	0	40

Table 3: Confusion Matrix for Naïve Bayes

	Normal	UDP Flood	Smurf	SIDDoS	HTTP Flood
Normal	19095	95	0	159	89
UDP Flood	196	1796	0	0	0
Smurf	80	2	0	4	45
SIDDoS	4	0	0	55	0
HTTP Flood	0	0	0	0	46

Table 4: Confusion Matrix for MLP

	Normal	UDP Flood	Smurf	SIDDoS	HTTP Flood
Normal	19435	0	0	3	0
UDP Flood	197	1795	0	0	0
Smurf	82	0	45	4	0
SIDDoS	4	0	0	55	0
HTTP Flood	0	0	2	0	44

Table 5: Classifier Performance Evaluation Result

MACHINE LEARNING MODELS	PERFORMANCE METRICS		
	ACCURACY	PRECISION	RECALL
Random Forest	98.63	98.50	98.60
Naïve Bayes	96.89	97.30	96.90
MLP	98.65	98.70	98.70

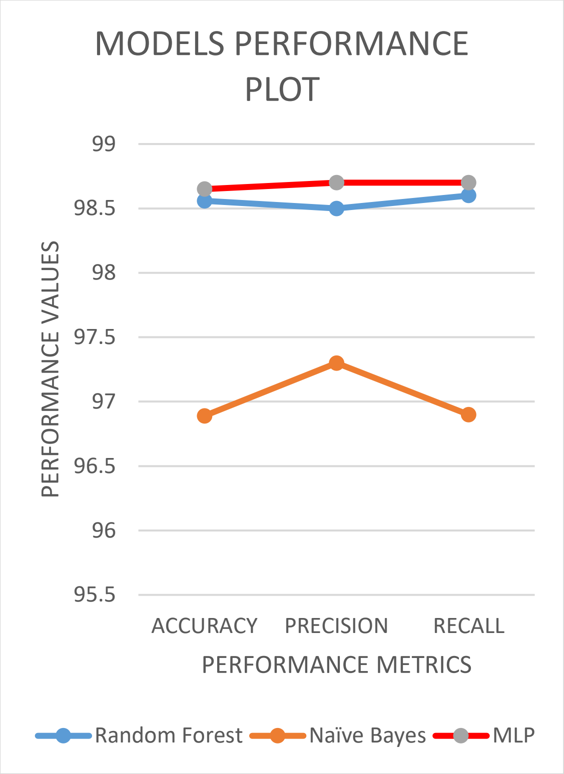

Figure 2: Classifier Performance Comparison

Conclusion

The findings of this research show that MLP is the most appropriate machine learning model for classifying DDoS flood attack traffic and normal traffic from the given KDD dataset comprising of the nine best packet attributes in ranking. KDD 99 dataset has been used as a benchmark dataset by many research conducted in the area of DDoS attack detection for network environment, such as in Idris, *et al.*, (2017). The study showed the Multilayer Perceptron (MLP) provides a better classification algorithm in the detection of DDoS attacks in Internet voting system by increasing voting server's performance and thus providing the required voting service availability for seamless electronic voting delivery to the populace. The MLP will provide a better secure Internet voting system by increasing voting server's performance in terms of system's availability to voters during election process. Furthermore, optimization techniques, such as Stochastic Gradient Descent and Adagrad could be investigated for DDoS attribute selection. Other machine learning models, such as Deep Learning and Support Vector Machine algorithms could also be investigated to compare performance of the proposed model.

REFERENCES

Al-Ameen, A. & Talab, S. The Technical Feasibility and Security of E-Voting. The International Arab Journal of Information Technology, Vol. 10, No. 4, pp. 398, July 2013.

Aljumah, A. (2017): Detection of Distributed Denial of Service Attacks Using Artificial Neural Networks. *International Journal of Advanced Computer Science and Applications (IJACSA),* 8(8), 310 – 311.

Alkasassbeh, M., Al-Naymat, G., Ahmad, H. & Almseidin, M. (2016): Detecting DDoS Attacks using Data Mining Techniques. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 7(1), 436-438.

Almustapha, A. J., Olaniyi, O. M., Abdullahi, I. M. & Ndunagu, O. (2019): Detection and Analysis of DDoS Attacks in Internet Kiosk Voting Using Machine Learning Algorithms. *Proceedings from the 3rd International Conference on Applied Information Technology (AIT)*, 133-140.

Aqeel, S., David, L., Yan, L. & Mohammed, D. (2017): An Efficient DDoS TCP Flood Attack Detection and Prevention System in a Cloud Environment. *IEEE Access*. DOI: 10.1109/ACCESS.2017.2688460

Bala, A. & Osais, Y. (2013): Modelling and Simulation of DDoS Attack using SimEvents. *International Journal of Scientific Research in Network Security and Communication* , 1(2), 5 – 6.

Brownlee, J. (2016): How to Use Ensemble Machine Learning Algorithms in Weka. Weka Machine Learning. Retrieved 5 August, 2019 from https://machinelearningmastery.com/use_ensemble-machine-learning-algorithms-weka/

Dhamdhere, P. B., Kasar, A., Satpute, N. & Lekurwale, P. (2017): A Secure E-voting System Using Biometrics Authentication Methods for Android. *International Journal of Advanced Research in Computer and Communication Engineering*. ISO 3297:2007 Certified 6(2).

Dominic, B., Inyiama, H. C., Ahmed, A., Abdullahi, M. B. & Olaniyi, O. M. (2015): A Packet Sampling Threshold Technique for Mitigating Distributed Denial of Service (DDoS) Attacks in a University Campus Network. Retrieved July 12, 2018 from https://www.researchgate.net/publication/283120300.

Idris, I., Obi, B. F., Abdulhamid, S. M., Olalere, M. & Meshach, B. (2017): Distributed Denial of Service Detection using MultiLayered Feed Forward Artificial Neural Network. *I. J. Computer Network and Information Security*. 12, 29-35. Published Online in MECS (http://www.mecs-press.org/). DOI: 10.5815/ijcnis.2017.12.04.

Jefferson, D. The Myth of 'Secure' Blockchain Voting. Available Online at https://www.verifiedvoting.org/jefferson_thyemythofsecure_blockchainvoting. Retrieved 8th July, 2019.

Olaniyi, O. M., Arulogun, O. T., Omidiora, E. O. & Okediran, O. O. (2015): Enhanced Stegano-Cryptographic Model for Secure Elecnnnnnnnbbtronic Voting. *Journal of Information Engineering and Applications,* 5(4), 3 – 7.

Perakovic, D., Perisa, M., Cvitic, I. & Husnjak, S. (2017): Model for Detection and Calssification of DDoS Traffic Based on Artificial Neural Network. *Telfor Journal*. 9(1).

Porup, J. M. (2018): Online Voting is Impossible to Secure. So Why Are Some Governments Using It. Availabile Online at https://www.google.com/amp/s/www.csoonline.com/so-are-some-governemnets-using-it.amp.html Retrieved August 14, 2019.

Thwe, T. & Phyu, T. (2013): Machine Learning Techniques used for the Detection and Analysis of Modern Types of DDoS Attacks. *International Journal of Advanced Research in Computer Engineering and Technology (IJARCET)*, 2(5).

Wei, C. C. & Wen, C. C (2018): Blockchain-Based Electronic Voting Protocol. *International Journal on Informatics Visualization,* 2(4), 2-5.