Performance Analysis of Classification Algorithms for Distributed Denial of Service Attacks Detection in a Distributed Network Environment

Olawale S. Adebayo
Cyber Security Science Department,
Federal University of Technology
Minna, Nigeria
waleadebayo@futminna.edu.ng

Abdulmutalib Abdullahi
Cyber Security Science Department,
Federal University of Technology
Minna, Nigeria
abdulmutaleeb97@gmail.com

Moses Noel
Cyber Security Science Department,
Federal University of Technology
Minna, Nigeria
moses.noel@futminna.edu.ng

Shafi'i Muhammad Abdulhamid
Cyber Security Science Department,
Federal University of Technology
Minna, Nigeria
,shafii.abdulhamid@fut.edu.ng

Abstract— *Organization network and its infrastructures persistently face challenges of Distributed Denial of Service (DDoS) attacks [19]. Mostly the attacks are targeted at the crucial network infrastructures such as the database server, cloud computing server, web server and other computing devices. The occurrence of such attacks causes a serious negative impact to the organization and its vital infrastructures. In this paper, six well-known classification algorithms (Random Forest, Decision Stump, NNge, OneR, RART and Naïve Bayes algorithms) were applied on NSL-KDD dataset to examine the performance of individual algorithm in terms of accuracy and false detection rate. The dataset was streamlined for optimum performance of the selected algorithms. The experimental result shows that Random Forest algorithm has 98.7% Detection accuracy and false detection rate of 0.022%.*

Keywords*: Denial-of-Service (DoS) Attacks; Distributed Denial of Service (DDoS) Attacks; Intrusion Detection Systems (IDS); Infrastructures; Classification Algorithms*

I. Introduction

The increase in dependency rate of military, commercial and government organizations on computer and its applications have no doubt increased the computing system and data is at continual risk [4] This is due to the increase in the rate at which computer and its application are being used and have rapidly grown in the past decades. As a result of frequent security bridge in our computer and increase in rate at which tools and devices are available to trick or bypass the security of the systems and its infrastructure by attacking or intruding networks. As a result of these challenges there is a need to combat the external attacks. Therefore, these attacks are external to these systems and its infrastructures and the aims of attackers are to steal, destroy, degrade, damage, disrupt or deny access to vital network resources [6]. In order to secure the systems and its infrastructures against unknown intrusions, a great deal of research has been focused on the development of intrusion detection system (IDS) and intrusion protection systems (IPS), which normally attempt to filter out such attacks from the network traffics. Intrusion detection systems are software tools that are normally use to harden the security of communication and information systems [4]. An IDS also monitor network traffic and logs, intrusions are identified in a network via applying detection algorithm. The assumption of intrusion detection is that it notifies intrusive activity different from other normal activities, therefore intrusion detection systems are not used to replace prevention–based techniques such as access control and authentication rather it is used to complement the existing security measures. Hence the intrusion detection is considered as a second guard and defense for computer network system and they normally notice and detect those actions that bypass the control component of the system and security monitoring [4]. Normally false alarms are largely produce by signature-based IDS than it is expected. Due to the large number of false positives in the log makes the process of taking necessary action for the true positive, i.e. successful attacks are delayed and labor intensive [6]. This DDoS attack can be in form of malware which is a malicious program that aim to disrupt the normal execution of computer program [1]. The intruders do not usually attack in order to steal but to reduce the performance of the network. [11]

The aim of this research work is to carry out a performance analysis on different classification algorithms for DDoS attack detection in a distributed computing environment. The other sections of this paper are organized as follows: Section II examined the related works in this domain of the research. Section III presents the methodology used in the experiment and also the analysis. Section IV chronicles the experimental results and discussion, and finally the conclusion in presented in Section V.

II. Related work

The research by Jie-Hao et al. [5] proposed an approach that used artificial neural network (ANN) to detect DDoS attacks. The result of this research was compared with other results using different classification algorithms like decision tree, Bayesian and entropy. The authors were able to identify users' access and requests to a specific resource to their communicatory data. In this process sample of those requests are sent to detection mechanism which test for abnormalities in the request.

Prakasha *et al* [7] proposed an algorithm which make use of three layer to authenticate traffic and users in the network. It normally takes short time to identify authorized and unauthorized users. The algorithm tends to permit the authorized user to have access to the server and deny unauthorized user access. These three layers make used of puzzle, cryptography based and MAC filtration.

Akilandeswari *et al* [2] in their work proposed a probabilistic neural network based attack traffic classification to detect DDoS attacks. The research concentrates on differentiating DDoS attacks from events. Bayes decision rules as Bayes inferences together with radical basis function neural network (RBFNN) were used. The algorithm was able to differentiate DDoS normal traffic.

Gupta et al [3] in another work were able to detect a number of slaves/zombies that are using neural network to attack the system. The aim of the study was to ascertain the relationship that exist between zombies in the network using entropy variation. In this method, it was assumed that the system work load depend on prediction using a feed – forward neural network.

Li et al [8] make use of learning vector quantization (LVQ) neural network to detect attacks. These LVQ is a supervised version of quantization that recognizes and carryout multi- class classification of data and compression of data. The datasets that was used in the work was converted to numerical form and send to the network as input.

Andrew et al [9] In this research the authors proposed a technique for intrusion mechanism for detecting DDoS attack in the cloud. The system work by investigating exploited and compromised virtual system to execute large amount of Distributed Denial of Service attacks.

Kanchan et al [10]. in this paper the authors introduced a new scheme for early detection of Distributed Denial of Service attacks in Wireless Sensor Networks. Which normally detect attacks on it early stages so that the data loss is prevented and large amount of energy is reserved after the prevention attacks.

III. Methodology

This research makes use of data mining approach to evaluate the performance of some classification algorithms in detecting DDoS attack in an organization network. This research is carried out in three (3) phases: Dataset acquisition, Data classification and Performance evaluation. The dataset was partitioned into sixty and fourty percent for training and testing set respectively.

Six different classification algorithms (Random Forest, Decision Stump, NNge, OneR, PART and Naïve Bayes algorithm) were used for classification experiment. The experimental results were compared to obtain the classification algorithm with better performance.

PHASE ONE **Dataset acquisition**	• Dataset description
PHASE TWO **Data classification**	• **Classification** Random forest Decision Stump NNge OneR RART Naïve Bayes
PHASE THREE **Performance Evaluation**	• Experiment setup • Performance Testing

Fig 1. Research Framework

A. Dataset acquisition

The NSL-KDD is an improved version of KDDcup99, which is obtained from machine learning repository of University of California, Irvine (UCI) library. The NSL-KDD dataset consist of 42 attributes with 75236 instances. Table 1 shows the detail samples of the dataset attributes

TABLE 1. NSL-KDD Dataset attributes

NO.	Features	Types
1	Duration	Continuous
2	protocol_type	Discrete
3	Service	Discrete
4	Flag	Discrete
5	src_byte	Continuous
6	dst_byte	Continuous
7	Land	Discrete
8	wrong_fragment	Continuous
9	Urgent	Continuous
10	Hot	Continuous
11	num_failed_login	Continuous
12	logged_in	Discrete
13	num_compromised	Continuous
14	root_shell	Continuous
15	su_attempted	Continuous
16	num_root	Continuous
17	num_file_created	Continuous
18	num_shells	Continuous
19	num_access_files	Continuous
20	num_outbound_cmds	Continuous
21	is_host_login	Discrete
22	is_guest_login	Discrete
23	Count	Continuous
24	svr_count	Continuous
25	serror_rate	Continuous
26	srv_serror_rate	Continuous
27	rerror_rate	Continuous
28	srv_rerror_rate	Continuous
29	same_srv_rate	Continuous
30	diff_srv_rate	Continuous
31	srv_diff_host_count	Continuous
32	dst_host_count	Continuous
33	dst_host_srv_count	Continuous
34	dst_host_same_srv_count	Continuous
35	dst_host_diff_srv_count	Continuous
36	dst_host_same_srv_port_port	Continuous
37	dst_host_srv_diff_host_rate	Continuous
38	dst_host_serror_rate	Continuous
39	dst_host_srv_serror_rate	Continuous
40	dst_host_rerror_rate	Continuous
41	dst_host_srv_rerror_rate	Continuous
42	Class	Continuous

B. Experimental setup

The input file which is used for the experiment is NSL-KDD dataset, which was obtain from University of California data library and it contain about 75236 instances with 42 attributes. The duration, protocol_type, services, flag, source_byte, destination_byte among others are the features of the dataset which are of the type continuous or discrete in nature. Hence the dataset is divided into 12 different portion for the experiment. In order to effectively evaluate the performance of this work, the NSL-KDD dataset was implemented with six machine learning algorithms namely Random forest, Decision Stump, NNge, OneR, PART and Naïve Bayes. Hold-out technique and 10 fold cross- validation were used to carry out the evaluation.

C. Performance evaluation

Statistical parameters were used to evaluate the performance of the technique and compare the results with contemporary techniques. The parameters used include False positive rate (specificity measure), F-measure, precision, Recall, and Receiver Operating Characteristic (ROC)

True Positive Rate

TP is defined as DDos attack that was actually classified as DDos attack i.e. **TPR** is the proportion of positive instances classified correctly.

True Negative Rate

TN is defined as non-DDos attack that was classified as non-DDos attack i.e. **TNR** is the proportion of negative instances classified correctly.

False Positive Rate

FP is defined as non-DDos attack that was classified as DDos attack i.e. **FPR** is the proportion of negative instances classified wrongly as positive (DDos attack).

False Negative Rate

FN represents DDos attack that was classified as non-DDos attack i.e. **FNR** is the proportion of positive instances wrongly classified as negative (non-malware).

Precision: this is the fraction of correctly classified instances.

Recall is used to determine how many anomaly are been detected correctly.

F –measure: this is used to determine the accuracy of the proposed model

Therefore:

$$FPR = \frac{FP}{FP+TN}$$ (1)

$$Recall\ (R) = \frac{TP}{TP+FN}$$ (2)

$$Precision\ (P) = \frac{TP}{TP+FP}$$ (3)

$$F - measure\ (F) = \frac{2PR}{P+R}$$ (4)

IV. Results and discussion

Results of the experiment were analyzed based on the statistical tests and parameters. The NSL-KDD dataset was partitioned into twelve portions, in which each iteration is incremented with 100 instance of the dataset the machine learning classification algorithm were applied on each partition of the dataset. The results experiment is presented in Table 2.

TABLE 2. Performance measure of the classification algorithms

Algorithm	FDR	Precision	Recall	f-measure	ROC
Random forest	0.025	0.987	0.729	0.987	0.985
Decision stump	0.23	0.797	0.979	0.729	0.800
NNge	0.036	0.979	0.979	0.979	0.971
OneR	0.23	0.797	0.729	0.729	0.749
RART	0.039	0.972	0.972	0.972	0.982
Naïve Bayes	0.305	0.836	0.842	0.842	0.915

Table depicts the result of six (6) selected machine learning algorithm. The Decision Stump and NNge algorithms according to the result perform better in Recall at 97.9 %. While other algorithms like OneR, RART, and Naïve Bayes have partial high performance in the experiment. Random Forest algorithm shows better performance in false detection rate (FDR) at 0.025%, precision at 98.7%, F-Measure has 98.7% and Receiver Operating Characteristic (ROC) 98.9% detecting accuracy. Hence the result clearly shows that Random Forest algorithm is the best for classification of DDoS attacks.

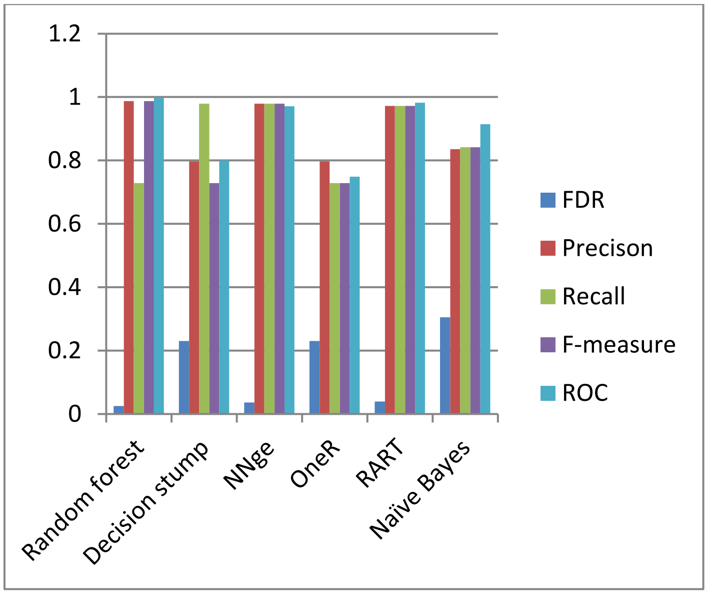

Fig.2. Performance measure of the classification algorithms' with NSL-KDD

Analysis from Table 2 and Figure 2 show that Random forest algorithm performed better in terms of FDR, precision, *F*-measure and ROC.

V. Conclusion

The DDoS attack is one of the commonest threat that normally has a devastating effect on organization's network and its infrastructures. As a result of that, a research was conducted to measure and analyze performance rate of classification algorithms on DDoS attacks based on the characteristics and parameters used to classify the normal traffic from the DDoS attacks. Some of the DDos attack parameter used are source address, destination address, and protocol bit service count. The dataset used is NSL-KDD datasets. The experimental results show Random Forest algorithm has best performance with Precision of 98.7%, F-measure of 98.7%, ROC of 98.9% and lowest false detection rate (FDR) of 0.025%. Hence based on the experiment conducted Random Forest algorithm perform better than other algorithms in the classification of DDoS attacks.

REFERENCE

[1] Adebayo O.S and Azziz N.A (2016). Static code Analysis of Permission based Features for Android Malware Classification Using Aprion Algorithm with Particle Swarm Optimization.. J*ournal of Information Assuarance and Security* 10(4), 2015.

[2] Akilandeswari V., Shalinie S.M. (2012). Probabilistic neural network based attack traffic classification *In: Proceedings of the Fourth International Conference on Advanced Computing (ICoAC*), Chennai: (pp. 1-8).

[3] Gupta B.B., and Misra M. (2011). ANN based scheme to predict number of zombies in a DDoS attack. *International Journal on Network Security vol.*13(3): (pp.216–225).

[4] Inadyuti D. and Samarjeet B. (2015). Some Studies in Intrusion Detection using Data Mining Techniques. *In the proceeding of International Journal of Innovative Research in Science,Engineering and Tchnology*.ISSN : 2319-8753

[5] Jie-Hao C.; and Feng-Jiao C. Z. (2012). DDoS defense system with test and neural network. *In: Proceedings of the IEEE International Conference on Granular Computing (GrC.)* Hangzhou, China: (pp38 - 43)

[6] Subbulakshmi, T., Mercy, S.S., Suneel, R.C., and Ramamoorthi, A. (2010). Detection and Classification of DDoS attacks using Fuzzy Inference System Communications. *In the proceedings of Computer and Information Science Journal*. DOI: 10.1007/978-3-642-14478-3_25

[7] Prakasha, A. Sri M. S., T .Sai Bhargava and N. Bhalajia (2016). Detection and Mitigation of Denial of Service Attacks Using Stratified Architecture. 4th *International Conference on Recent Trends in Computer Science & Engineering* 87: 275 – 280.

[8] Li J.; Liu Y.; Gu L. (2010). DDoS attack detection based on neural network. *In: Proceedings of the 2nd International Symposium on Maware Computing(ISAC*), Tainan. (pp. 196–199).

[9] Andrew C., Mohammad H. & Omar A. (2015) Defence for Distributed Denial of Service Attacks in Cloud Computing. *In the proceeding of International Conference on Advanced Wireless, Information, and Communication Technologies*. 73 (2015) 490 – 497

[10] Kanchan K. and Varsha S. (*2016).* Early Detection of DDoS Attack in Wireless Sensor Networks. *In the proceeding of International Journal of Computer Applications (0975 – 8887))* Volume 134 – No.13, January (2016)

[11] Abdullahi A. and Tariq A. (2016). A Novel Approach for Detecting DDoS using Artificial Neural Networks. *In the proceeding of International Journal of Computer Science and Network Security,VOL.16 No.12 December 2016*