International Journal of Industrial Engineering
2022;6(1):1-6.
ISSN: 2456-8449

A Network Security Application for Detecting DDOS Attacks Using Data Mining Approach

R. Mohandas[1*], N. Sivapriya[2], A. Sanyasi Rao[1], K. Radhakrishna[1]

[1]*Department of Electronics and Communication Engineering, Balaji Institute of Technology & Science, Warangal, Telangana, India.*
[2]*Department of Computer Applications, Cauvery College of Women, Trichy, Tamil Nadu, India.*

[*]*Corresponding author: mohandasbe@gmail.com*

Abstract. Distributed denial of service (DDoS) attacks is difficult for individuals and organisations to cope with on a consistent basis. Keeping a service running at all times is the job of the security engineer. Unusual activity can be detected and classified using an intrusion detection system (IDS). Attack packets will contain more packets than normal, but the inter arrival rate will be too short for attackers to deplete resources quickly. To keep the service's confidentiality, integrity, and availability, such IDS have to be regularly rationalized over newest prowler outbreak. Due to a lack of common data sets for recent DDoS attacks across multiple network levels, a new dataset was created in this article (SIDDoS, HTTP Flood). Multilayer Perceptron (MLP), Naive Bayes, and Random Forest are three well-known classification methods used in this study. Also, a new dataset was created for this study. A 97.63% accuracy rate was attained by the MLP in the experiments.

Keywords: DDoS Attacks, Perceptron, Data Mining, Classification, Security.

INTRODUCTION

A wide range of industries, including banking, email, social media, and university e-services have all evolved to place a high value on network security. Intruders have recently targeted online and network services. New varieties of DDoS are constantly being created by hackers to attack the application logic and also the network layer. Access to online services and network resources might be denied and slowed down because of the problems in the above stated areas [1]. Cloud hosting and computer system infrastructure can be protected from DDoS attacks with the help of an IDS system. Detecting and categorising new forms of DDoS attacks using machine learning techniques is a challenge, but the IDS system will remove intrusions without requiring the System Security Officer (SSO) to be notified of the intrusion. Naïve One sort of DDoS attack is the Smurf attack, which delivers massive quantities of Internet-controlled message protocol messages to the desired targets [2].

Numerous varieties of DDoS attacks are previously known. There is another sort of DDoS known as RUDY, which essentially eats up all of the web applications' sessions, preventing them from ever ending. In other words, new requests for the web application will be denied. HTTP POST/GET is a more recent form of DDoS attack in which an attacker sends genuine posted messages to a web server operating a web application at an extremely slow rate, such as 1 byte/240 seconds [3]. It's possible that a DDoS attack of this nature might force a web service to go down or perhaps go down completely. One of the most common current DDoS attacks is a SQL Injection Dos, or SIDDoS, when an attacker inserts a malevolent query as a text that would make corrupt a file, and then unlawfully allows contact to possessions or kept information on systems [4].

Detecting and classifying a DDoS is impossible because to the abundance of duplicate and redundant data in most commonly available open data sets. According to KDD 99, new DDoS varieties, such as HTTP floods and SIDDOS are not included [5]. A new dataset was created for this study, which comprises four forms of damaging attacks: UDP flood, Smurfing, HTTP Flooding, and SIDDOS. Detecting and classifying network based on some variables, such as mean incoming packets, inter arrival time, and other features such as the size of the packets and the pace at which they arrive, may be done using machine learning [6].

Received: 09.10.2021 Revised: 04.11.2021 Accepted: 24.11.2021

The average packet size of a DDoS attack is rather consistent. Attack packets will contain more packets than normal, but the inter arrival rate will be too short for attackers to deplete resources quickly. For network layer attacks, DDoS payloads often have a significant bit rate. Characteristics that let them drain resources and disrupt service to end users are prioritised by attackers [7]. Figure 1 shows the overview of attack detector.

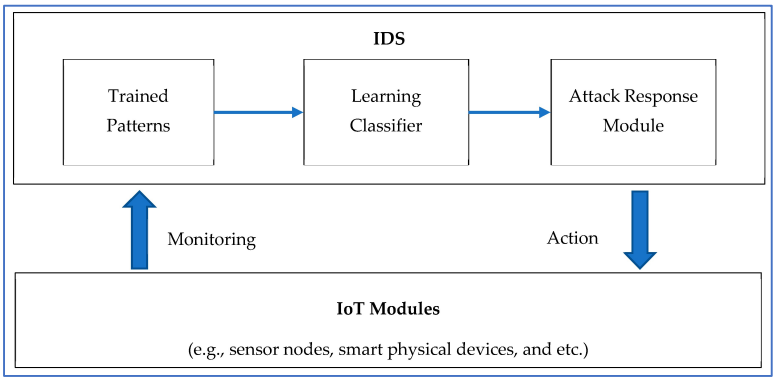

FIGURE 1. Overview of Attack Detector

EXISTING WORKS

SYN Flood was identified in 1996, smurf attacks began in January 1998, and the current DOS debuted in 2004 with an HTTP flood. A mix of anomaly- and signature-based detection technologies, as well as full security architecture, is recommended by the area's experts. DoS detection is increasingly relying on classifications also including machine learning approaches because of their ability to automatically classify [8]. This defence system either intercepts or discards the incoming packets, or it disables them. Some researchers have found that the most reliable indicators of a denial-of-service attack include the various range of attributes [9]. Using OPNET simulation, Guiomar et al. developed a Network Intrusion Detection System (NIDS). The network traffic was imported into OPNET through an ACE module based on abuse detection [10].

For this test, an NMAP port analyzer was utilised to mimic an attack, and the suggested IDS was evaluated with an accuracy rate of roughly 93%. For IDS, a variety of approaches have been employed. For an individual attack, Chandrika Palagiri demonstrated that a modelling network may produce realistic results. With MLP, it employed a machine learning classifier to create a uniform or group, highly helpful system [11]. Neural Network which can make judgments rapidly and identify things in real time is typically the subject of research. Layer 7 DDoS flooding attacks have been studying considering the growing severity and frequency of these attacks. The intrusion detection system made use of a variety of machine learning approaches. A good detection rate was reached by some of these approaches, whereas a low detection rate was attained by others. It also has a higher rate of success in classifying and detecting a layer seven attack using the Nave Bayes (NB) algorithm than other machine learning methods [12].

DDoS attacks may be detected, and attack packets can be quickly identified. The framework's goal is to take advantage of DDoS attacks' geographical and temporal correlations. By using these methods, routers may properly detect DDoS attacks and identify incoming packets without having to change their existing IP forwarding algorithms. The suggested methodology helped this study reach a detection probability of 97%. Wei Pan and Weihua Li employed a hybrid Neural Network approach to identify and categorise DDoS attacks [13]. Detecting and categorising them with the suggested method was successful with a high rate of accuracy. One of the most successful methods for identifying and categorising attacks into normal and threat categories was described by Norouzian et al. It's built on a Deep Convolutional Neural Network that can recognise and categorise input into six distinct categories. In their implementation, they used two hidden levels of axons and attained a 90% performance of the MLP architecture [14].

A NIDS used a 2-layered, feed-forward computational model. The suggested system was able to distinguish between legitimate connections and those that had been compromised. A variety of attack types were identified, with an emphasis on utilising training functions, information authentication, and a pre-process collection to use

less memory, utilise fewer resources, and speed up training. The suggested approach performed admirably on the KDD cup 99 dataset once it was put into practise [15].

Preprocessing of statistical variables and a Neural Network are used to identify attacks in Intrusion Detection (HIDE) system. As a part of their research, they examined Intuitive and Risk assessment involves considering Function. For identifying and categorising network threats, BP and PBH were found to be the most accurate, according to them. Using a risk assessment involves considering function (RBF) Neural Network, they were able to identify a DDoS attack with 96% accuracy using a UCLA dataset. Fuzzy clustering, evolutionary algorithms, and artificial neural networks (ANNs) have also been used to prevent attacks [17]. ANN has the greatest accuracy rate. In addition to the support vector machine and back propagation, they suggested a multilayer perceptron system with several units, like packet collecting and information preparation. It has a detection rate of 88.65 percentages.

PROPOSED SYSTEM

A DDoS attack can be prevented by using an intrusion detection system on a network that has been secured. Excellent IDS may identify a new DDoS in a brief span of time without the involvement of a person. There are two main types of IDS systems: centralised and distributed. An intrusion detection system could be installed on system plans and terminals. These methods could be used to protect a specific device against a DDoS attack, but they are unable to monitor a whole network [18-19]. Intrusion Detection System (IDS) is a security approach that categorise completely system circulation coming after all policies. Network intrusion detection systems (NIDS) will be used in this study.

IDS may be used to identify and categorise network traffic using either anomaly detection or signature detection. Analyzing network traffic based on anomalies is a complex process that requires a large amount of training data. Using a signature-based approach, each individual packet is matched to a recorded characteristic or known intruder attack and the results are compared [16]. Anomaly-based detection takes longer than signature-based detection because training data is required for anomaly-based detection, whereas a stored signature is required for signature-based detection. By depleting the servers' resources to the point that the service is no longer available to any users, the attackers utilise DDoS to overwhelm the targets.

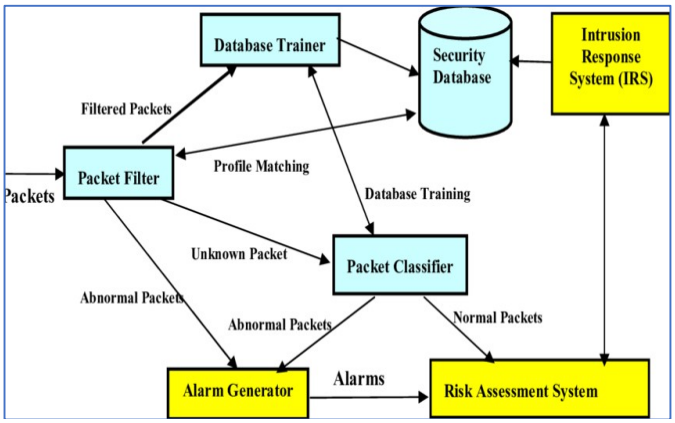

FIGURE 2. Proposed System Architecture

Using malicious (TCP/UDP) network traffic, an attacker can prevent a server service from being accessed. Smurf attacks and UDP flood attacks are examples of this type of attack. On order to transmit a high number of ICMP floods to the target system, attackers spoof IP addresses in a system and use them to send malicious data. Denial of service occurs when the server receives a significant number of ICMPs. A UDP flood is another form of network-layer attack that utilises the connectionless nature of UDP. The network workstation is used in an attack by sending a command towards the slave system. UDP traffic from all workstations is flooding the affected server. The HTTP protocol is the greatest often supported procedures by the network system. Due to the fact that HTTP-based web applications may be accessed from anywhere, it is difficult to identify application

layer attacks and prevent them. When an intruder attempts to access a server's resources, they do so in a way that appears to be a normal user's request for assistance.

DDoS attacks on the application layer have been increasingly common in recent years, with the most common kind being SQL Hacking Distributed Denial of Service (DDoS), where attackers begin by injecting malicious code into a web form and then sending it to the server-side input mailbox data. Unless the malware is subsequently passed to the server's execution forever, the attack drains the server's resources. When it comes to stealing personal information and making a service unavailable to customers, an attack may be used to do both.

RESULTS AND DISCUSSIONS

Free-Flowing Woods Tuning Options Random Forest (RF) detectors may be tested and trained by dividing a dataset into training and testing sections, e.g. 40% and 60%, respectively (60 percent). The error rate may be calculated using a model test (40 percent) can be assessed by associating properly categorised examples with erroneously categorised ones. The error rate can also be calculated using the "out of bag" (OOB) method. And no need to separate the dataset in this method because the computation takes place during training. For the best accuracy and lowest error rate, these settings need to be adjusted precisely.

Linux 13.10 framework, IBM, Diamond (R) Processor E5-2680 @ 5.5 GHZ x 4, and 8 GB RAM were used for the studies in this paper. WEKA version 3.7.12, a machine learning software platform, was utilised to implement classification strategies. Every technique was given training on our sample using 76% of the gathered data, and the remaining 34% was utilised as a test data set for each method. We attempted the ten-fold validation, but our earlier partitioning performed better. Based on a confusion matrix, we employ key performance indicators to examine the classification performance in this study.

TABLE 1. Attack Detection Analysis

	Normal	Flooding	Smurf	HttP_Flooding
Normal	3524	0	70	54
Flooding	5241	654	0	46
Smurf	524	5247	20	325
HttP_Flooding	0	0	0	86

Using this matrix, you can see both the actual and anticipated classifications that the classification models made. The information in the matrix is widely used to evaluate the performance indicators of such systems. Based on the data acquired in this study, we apply MLP, Random Forest, and Naive Bayes classifiers. Confusing matrixes were used to calculate model's accuracies, precisions, and recalls. The total accuracy for MLP, Random Forest, and Nave Bayes was 98.63 percent, 98.02 percent, and 96.91 percent, respectively.

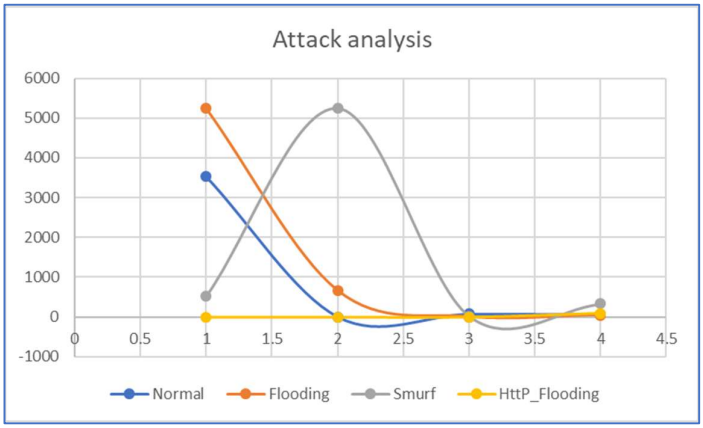

FIGURE 3. Attack Analysis Model

As in our situation, with a significantly greater number of cases in the classification step than in any of the other classes, simply looking at the accuracy rate is insufficient. To account for these differences, we computed the accuracy and recall separately for each kind of data. Percentage of applicable outbreaks compared to the entire amount of both unrelated and related outbreaks.

TABLE 2. Performance Analysis

	Accuracy	Precision	Recall	F1-Value
Normal	89.56	88.52	79.32	81.02
Flooding	88.25	80.12	90.21	88.32
Smurf	78.95	79.63	81.54	74.32
HttP_Flooding	80.35	82	84.36	87

There are a lot of ICMP echo message requests in the Smurf attack, which makes it difficult to distinguish between regular and anomalous traffic. Note that IP4 lacks flow control and traffic management, which ICMP provides. For the Smurf class, Random Forest and Nave Bayes did not perform well, however MLP performed exceptionally well. It can be argued that MLP is the popular algorithm for predicting DDoS with excellent innovation performance the created models' accuracy and recall values, broken down by class. That all classifiers obtained good accuracy and recall rates for normal class can be shown from Statistics In terms of the other four kinds of attacks, their performance differs.

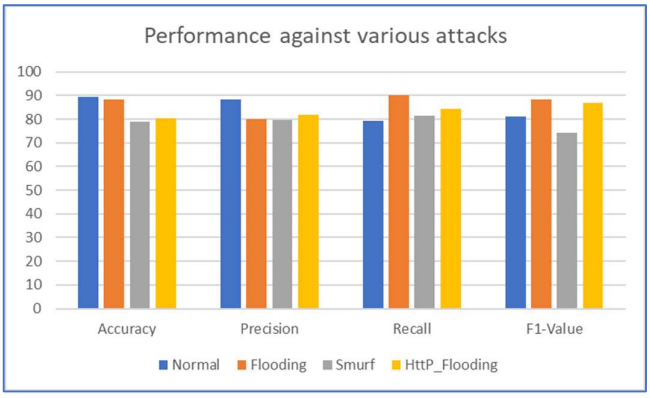

FIGURE 4. Algorithm Performance Analysis

CONCLUSIONS

New varieties of DDoS are constantly being created by hackers to attack the application logic and the network layer. Access to online services and network resources might be denied and slowed down because of the problems in the above stated areas. There are new forms of attacks that haven't been included in earlier studies in this study. There are 25 features in the collection, and there are five classifications. In this study, a network simulator (NS2) was utilised because of the high degree of confidence in its ability to produce meaningful results that accurately represent real-world conditions. To account for these differences, we computed the accuracy and recall separately for each kind of data. Application and network-layer attacks are among the many examples of intrusions reported in the datasets. Smurf, UDP-Flood, HTTP-Flood, and SIDDOS were all classified using three machine learning methods such as MLP, Random Forest, and Nave Bayes. In terms of accuracy, the MLP classifier had the best results.

REFERENCES

[1]. G. Y. Chan, C. S. Lee and S. H. Heng, 2013, “Discovering fuzzy association rule patterns and increasing sensitivity analysis of XML-related attacks,” *J. of Network and Computer Applications*, **36(2)**, pp. 829-842.

[2]. F. Haddadi, S. Khanchi, M. Shetabi and V. Derhami, 2010, "Intrusion detection and attack classification using feed-forward neural network," *In 2010 Second Int. Conf. On Computer and Network Tech.*, pp. 262-266.

[3]. J. Han and M. Kamber, 2006 "Data mining: concepts and techniques," *2nd. University of Illinois at Urbana Champaign: Morgan Kaufmann*.

[4]. Z. Zhang, J. Li, C. N. Manikopoulos, J. Jorgenson and J. Ucles, 2001, "HIDE: a hierarchical network intrusion detection system using statistical preprocessing and neural network classification," *In Proc. IEEE Workshop on Information Assurance and Security*, **85**, pp. 85- 90.

[5]. R. Karimazad and A. Faraahi, 2011, "An anomaly-based method for DDoS attacks detection using RBF neural networks," *Proc. of the Int. Conf. on Network and Electronics Eng.*, **11**, pp. 44-48.

[6]. K. Lu, D. Wu, J. Fan, S. Todorovic and A. Nucci, 2007, "Robust and efficient detection of DDoS attacks for large-scale internet," *Computer Networks*, **51(18)**, pp. 5036-5056.

[7]. J. H. Nord and G. D. Nord, 1995, "MIS research: journal status assessment and analysis," *Information & Management,* **29(1)**, pp. 29-42.

[8]. M. R. Norouzian and S. Merati, 2011, "Classifying attacks in a network intrusion detection system based on artificial neural networks," *13th Int. Conf. on Advanced Communication Tech. (ICACT2011)*, pp. 868-873.

[9]. W. Pan and W. Li, 2005, "A hybrid neural network approach to the classification of novel attacks for intrusion detection," *Int. Symposium on Parallel and Distributed Processing and Applications*, pp. 564-575. *Springer, Berlin, Heidelberg*.

[10]. R. Sadoddin and A. A. Ghorbani, 2009, "An incremental frequent structure mining framework for real-time alert correlation," *Computers & Security*. **28(3-4)**, pp. 153-173.

[11]. M. Y. Su, G. J. Yu and C. Y. Lin, 2009, "A real-time network intrusion detection system for large-scale attacks based on an incremental mining approach," *Computers & security*, **28(5)**, pp. 301-309.

[12]. Z. Tan, A. Jamdagni, X. He, P. Nanda and R. P. Liu, 2013, "A system for denial-of-service attack detection based on multivariate correlation analysis," *IEEE Trans. on Parallel and Distributed Systems*, **25(2)**, pp. 447-456.

[13]. B. A. Tama and K. H. Rhee, 2015, "Data mining techniques in DoS/DDoS attack detection: A literature review. International Information Institute (Tokyo). *Information*. **18(8)**: pp. 3739-3748

[14]. Yu J, Lee H, Kim MS and Park D. Traffic flooding attack detection with SNMP MIB using SVM. *Computer Communications*. 2008 **31(17)**, pp. 4212-4219.

[15]. T. A. Tang, L. Mhamdi, D. McLernon, S. A. Zaidi and M. Ghogho, 2016, "Deep learning approach for network intrusion detection in software defined networking," *2016 Int. Conf. on Wireless Networks and Mobile Communications (WINCOM)* pp. 258-263.

[16]. S Murugan, A. Bhardwaj, and T. R. Ganeshbabu, 2015, "Object recognition based on empirical wavelet transform," *Int. J. of MC Square Scientific Res.* **7(1)**, pp. 74-80.

[17]. MM Ismail, M Subbiah and S Chelliah, 2018, "Design of pipelined radix-2, 4 and 8 based multipath delay commutator (MDC) FFT, *Indian J. of Public Health Res. and Development*, **9(3)**, pp. 765–768.

[18]. V Jaiganesh and S. Murugan, 2005, "PC based heart rate monitor implemented in xilinx fpga and analysing the heart rate,"*Proc. of the Third IASTED Int. Conf. on Circuits, Signals, and Systems, CSS 2005*, pp. 319–323.

[19]. A. Unnikrishnan and V. Das, 2022, "Cooperative Routing for Improving the Lifetime of Wireless Ad-Hoc Networks, *Int. J. Adv. Sig. Img. Sci*, **8(1)**, pp. 17–24.