A Survey of Learning Techniques for Detecting DDOS Assaults

K. Jeevan Pradeep and Pragnyaban Mishra

Abstract The distributed denial-of-service (DDOS) exploit is one of the most catastrophic assaults on the Internet, disrupting the performance of critical administrations offered by numerous organizations. These attacks have become increasingly complicated, and their number has been steadily increasing, making it harder to detect and respond to such assaults As a result, a sharp security system (IDS) is necessary to detect and control any unexpected system traffic behavior. In a DDOS Assaults, the intruder delivers a stream of packets to the server while exploiting known or unknown flaws and vulnerabilities.

Keywords DDOS assaults · Network security · Decision tree · Naïve Bayes · SVM · Neural network · Fuzzy logic · Learning techniques

1 Introduction

Communication infrastructure and information assurance play critical roles in both social and economic growth, as well as in our daily lives. Because of the fast growth of world wide web-related networking and communication networks, knowledge management are becoming increasingly exposed to a variety of cyberattacks. Communication network and system attacks are becoming a security concern. Cyber security assaults are on the rise in general. As a result, a comprehensive security system for detecting security breaches is essential to defend the network from all forms of assaults.

An intruder attempt is defined as an unlawful attempt or threat to (i) acquire information, (ii) modify or alter the content, or (iii) disable the system. As an example,

K. Jeevan Pradeep (✉) · P. Mishra
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
e-mail: jeevanpradeep33@gmail.com

P. Mishra
e-mail: pragnyaban@kluniversity.in

S. Fong et al. (eds.), *ICT Analysis and Applications*, Lecture Notes in Networks and Systems 517, https://doi.org/10.1007/978-981-19-5224-1_14

A. Malware exploits of other network nodes by interfering with routine operations by corrupting software with a fault (virus) or polluting the network with a bug. As a result, network traffic is slowed.
B. A denial-of-service (DoS) attack is one in which the attacker intends to make a system or a network resource inaccessible to legitimate users.

A distributed denial-of-service (DDoS) attack is one that uses several machines in a distributed method to target a victim. This research looked into the usage of machine learning algorithms to detect DDoS assaults. Every cyber-attack leaves a digital footprint. An IDS uses a signature, which is a collection of criteria, to detect harmful intrusive behavior in the network, such as DDoS attacks. Among the approaches for identifying signatures are:

- An attempt to connect from a reserved IP address was made. An IP header connection might be easily recognized by examining the source address information.
- In a packet, an illegal TCP flag combination was used. Positive and negative flag mixes can be detected by comparing the flag set in a TCP header to known positive and negative flag mixes.
- An email server has been infected with a virus. This may be discovered by inspecting the subject lines of every email sent to the subject of a virus-infected email.
- DOS attack on a POP3 server caused by sending the same command several times. One signature will be kept to keep track of how many times the same command has been issued and to alert if the number of times the same command has been issued exceeds a specific threshold.
- Attack the FTP server's file system by supplying directories and file commands without even logging in. A state tracking system may be extended to track FTP traffic to ensure proper sign-in. This might alert you if you issued any of the instructions before the user was authorized.

Several real intrusion detection systems rely on human research to distinguish between intrusive and non-invasive assaults. Because human intervention is required to build, debug, and deploy the extension on the studied datasets, finding and creating a new signature for an attack may take a long time or several hours.

Because of the catastrophic consequences of a compromised system on commercial and personal networks, intrusion systems have become a primary study area for researchers, cybersecurity administrators, and network administrators. An intrusion detection system (IDS) can identify several forms of harmful network activity and hostile network incursions, but a normal firewall cannot. In general, intrusion detection systems are divided into two types: Signature-based IDSs and anomaly-based IDSs.

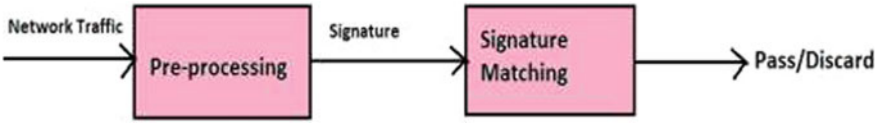

Fig. 1 Signature-based detection

1.1 Signature-Based Detection

Signature-based approaches rely on prior observed signatures to tally the signatures that are recorded in a database. This database contains a collection of signatures connected with prior assaults. Signature of IDs with some precise information in computation and preparation so that it does not check for every activity or network traffic on the environment that is being monitored. The signature-based technique is simple to use since it does not need learning about the environment; instead, known signatures are previously recorded in the database. The signature-based technique is effective against known attacks, but it cannot detect new attacks until it is activated with fresh signatures.

Signature-based intrusion detection systems are extremely difficult to evade since they are based on known attacks and require a new signature to be used before they can be recognized as new assaults. Signature-based approaches are easy to alter and enhance because their performance is determined by the signature or rules used.

Figure 1 depicts the design of the signature-based method. The network traffic is pre-processed in this architecture to obtain the desired properties and to find the signature. The signature is then matched to the action signature and validated against the signature database. If the signatures match, an alert is triggered; if no signatures match, nothing happens. Many businesses utilize this form of intrusion detection system to identify known threats financially and with few false positive failures.

1.2 Anomaly-Based Detection

Anomaly-based detection is also known as “Behavior based detection” that’s because the models are based on network behavior, and most computer systems issue an alert message when there is a normal detection in the behavior. This approach identifies undesirable traffic and is best suited for doing research on network hardware. Figure 2 depicts the overall architecture, which incorporates anomaly-based approaches. This design contains a pre-processing function that collects data and builds a pattern for the connection; if it deviates from regular behavior, an alarm message is generated.

This paper explored several machine learning techniques for detecting DDOS attacks. The following is the rest of the paper: Sect. 2 provides an overview of the literature review on machine learning approaches utilized in DDOS detection. Section 3 compares several ML algorithms for detecting DDOS assaults, as well

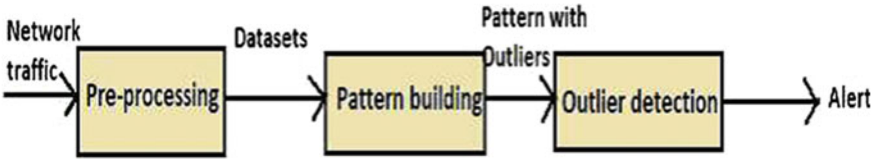

Fig. 2 Anomaly-based detection

as their benefits and drawbacks. Section 4 brings the study of machine learning techniques to a close.

2 Literature Survey

This section includes a literature overview of ML techniques used in the detection of DDOS assaults, such as decision tree, Naive Bayes, artificial neural networks, support vector machine, and fuzzy logic, as well as the research that has been conducted on them.

2.1 Decision Tree

Many researchers have been published on decision tree prediction models to identify DDOS attacks. Wu et al. [6] constructed a DDoS Detection system using a decision tree method, and the system uses a traffic-flow pattern-matching technique to trace back the attacker's location when an attack is detected. A C.45 classifier is used to detect DDOS assaults. The author of [8] discovers a method for efficiently detecting DDOS attacks. Many ML approaches take longer to identify an attack or yield lower accuracy. In [8], the C4.5 methodology is used, which has poorer accuracy and takes longer to generate the decision tree; however, the C5.0 method has been proved to be more efficient since it consumes less time and memory than the C4.5 technique. Bujlow, Khafhali, and Saadi [9] focused on network traffic categorization and found that the C5.0 algorithm outperformed the C4.5 method. Another research utilized ID3, C4.5, and C5.0 to construct a better decision tree with less error pruning and feature selection [10]. C5.0 outperformed in terms of accuracy and memory use, according to the results.

2.2 Support Vector Machines (SVM)

It is the most extensively used and popular strategy for machine learning assignments. In 2010, Das et al. [12] conducted an attempt to identify DDOS assaults using RTS and SVM. The RST preprocessed the network packet data that was first acquired. The SVM model is given the feature set chosen by the RST to train and test. When compared to principal component analysis (PCA), RST and SMV may reduce false positives and improve accuracy.

2.3 Naive Bayes

This machine learning algorithm is a basic probabilistic classifier [13]. Carl Livadas [16] employed Ml methods to identify commands and manage IRC-based botnet traffic. By comparing the performance of J48, Nave Bayes, and Bayesian networks, the author distinguishes between IRC and non-IRC traffic. The author of paper [16] discovers the qualities that provide greater accuracy. This classifier produces low false negative (2.49%) and false positive (15.04%) rates for real-life IRC/non-IRC flows, as well as low false negative (7.89%) rates for botnets tested on IRC flows, demonstrating naive Bayes to be an efficient classifier. In [14], Bains et al. suggested a layered hierarchical strategy for attack detection accuracy.

2.4 Neural Networks

Neural networks incorporate processing elements to turn a collection of inputs into a set of outputs that function in a manner comparable to organic nerve systems such as the human brain. Gavrilis et al. [15] employed a “RBF-NN detector” with nine packet parameters and computed the frequencies of these parameters. RBF-NN categorizes traffic as either normal or assault based on the expected frequencies. Distributed time delay neural network (DTDNN) has a high likelihood of detecting assaults with more precision. DTDNN performs data classification with quick conversion rates and great speed.

2.4.1 Fuzzy Logic

Anomaly detection mostly use fuzzy approaches. In IEEE 802.15.4, the author Vladimir [17] suggested a DDoS detection and prediction approach based on fuzzy logic. The fuzzy-based detection and prediction system (FBDPS) assisted in the identification of DDODS assaults by analyzing the energy usage of sensor nodes. The unusual energy usage of the node identifies it as a hostile attacker.

3 Comparision of Machine Learning Techniques

Despite the fact that machine learning-based algorithms are employed to identify intrusion in order to attain a high detection rate, they have their own set of advantages and drawbacks.

3.1 Fuzzy Logic

In [7], Yusof et al. reported that fuzzy c-means clustering outperforms other classifiers in categorization. Fuzzy c-means were shown to be faster than other machine learning methods. Fuzzy logic relies on reasoning that is approximate rather than exact. Suresh [4] used ML methods to construct and analyze fuzzy c-means clustering on DDOS assaults and obtained better categorization than previous solutions. Identification of reduced, relevant rule subsets is a tough process. In [5], Manjula and Anitha assessed ML approaches for detecting DDOS attacks using the CAIDA dataset, which is based on chi-square and information gaining ranking for the selected characteristics. The results demonstrate that fuzzy-c means better classification and is faster than other algorithms. Fuzzy logic works well against port scanning and probes.

3.2 Neural Networks

Jie-Hao et al. [2] utilized artificial neural networks to identify DDOS assaults and conducted a comparison study between ANN, decision tree, antropy, and Bayesian. It is capable of generalizing from limited and imperfect data. Neural networks require more time to train and are not ideal for real-time detection.

3.3 Support Vector Machine (SVM)

Li et al. [1] suggested a customizable intelligent module for network intrusion prevention systems by integrating SNORT and firewall. By combining an SVM classifier with SNORT, the false alarm rate is minimized, boosting the accuracy of the intrusion prevention system. SVM produces outcomes that are simple to comprehend and efficient. The only issue is that SVM only handles binary classification. For binary classifiers, it fails to provide further information about the identified attack type.

3.4 *Naïve Bayes*

Alkasassbeh et al. [13] compiled a new dataset of DDOS assaults at various network levels. He employed three algorithms to identify DDoS attacks: Multilayer perceptron (MLP), naive Bayes, and random forest. MLP demonstrated the best accuracy (98.63%) when compared to other approaches. The advantage of naive Bayes is that it is simple to implement. On greater data points, naive Bayes might be used, but it adds complexity. The disadvantage is the requirement for probability data and the assumption of conditional independence.

3.5 *Decision Tree*

C4.5 decision tree algorithm is more stable than k-nearest neighbor algorithm, according to Ismanto and Wardoyo [19]. Experiment on tree intrusion detection is carried out, analyzing multi-layer perceptron (MLP), C4.5, and SVM classifiers, with C4.5 demonstrating that it is better in detection (99.05%) and has the shortest training time. The authors of [20] examined the C4.5, naive Bayes, and C5.0 in identifying DDOS assaults. When compared to the other two algorithms, C5.0 has the highest accuracy. Many studies have been conducted to conclude that C4.5 has been working well with increased accuracy, but C5.0 has begun to perform even better than C4.5 [20]. The decision tree has the advantage of increasing accuracy as the quantity of the datasets increases. The disadvantage of decision trees is that they are unstable when dealing with complex numerical datasets.

4 Conclusion

Following a comprehensive investigation, it is determined that network attacks are destructive and that proper intrusion detection systems must be implemented. As previously stated, each machine learning approach worked well with one or more features that might be tallied in accordance with the identification of an intrusion in a system. Some machine learning-based detection algorithms improve in accuracy as the number of datasets increases, while others have a lower false alarm rate, and yet others are good for port scans and probes. As a result, it can be stated that machine learning methods have both advantages and downsides in identifying DDOS attacks. Appropriate techniques should be chosen based on whether the DDOS attack is originating at the port, network, or application layer, as well as the type of datasets accessible.

References

1. Li H, Liu D (2010) Research on intelligent intrusion prevention system based on snort. In: International conference on computer, mechatronics, control and electronic engineering (CMCE), vol 1. IEEE, pp 251–253
2. Li J, Liu Y, Gu L (2010) DDos attack detection based on neural network. In: 2nd international symposium on aware computing (ISAC). IEEE, pp 196–199
3. Suresh M, Anitha R (2011) Evaluating machine learning algorithms for detecting DDoS attacks. Commun Comput Inform Sci 441–452. https://doi.org/10.1007/978-3-642-22540-6_42
4. Livadas C, Walsh R, Lapsley DE, Strayer WT (2006) Using machine learning techniques to identify botnet traffic. In: Proceedings of 2006 31st IEEE conference on local computer networks, pp 967–974
5. Suresh M, Anitha R (2011) Evaluating machine learning algorithms for detecting DDoS attacks. In: Wyld DC, Wozniak M, Chaki N, Meghanathan N, Nagamalai D (eds) Advances in network security and applications. CNSA 2011. Communications in computer and information science, vol 196. Springer, Berlin, Heidelberg
6. Wu Y-C, Tseng H-R, Yang W, Jan R-H (2011) DDoS detection and traceback with decision tree and grey relational analysis. Int J Ad Hoc Ubiquitous Comput 7(2)
7. Yusof AR, Udzir NI, Selamat A (2016) An evaluation on KNN-SVM algorithm for detection and prediction of DDoS attack. In: Fujita H, Ali M, Selamat A, Sasaki J, Kurematsu M (eds) Trends in applied knowledge-based systems and data science. IEA/AIE 2016. Lecture notes in computer science, vol 9799. Springer, Cham
8. Zekri M, El Kafhali S, Aboutabit N, Saadi Y (2017, October) DDoS attack detection using machine learning techniques in cloud computing environments. In: 2017 3rd international conference of cloud computing technologies and applications (CloudTech). IEEE, pp 1–7
9. Bujlow T, Riaz T, Pedersen JM (2012, January) A method for classification of network traffic based on C5.0 machine learning algorithm. In: 2012 international conference on computing, networking and communications (ICNC). IEEE, pp 237–241
10. Pandya R, Pandya J (2015) Article: C5.0 algorithm to improved decision tree with feature selection and reduced errorpruning. Int J Comput Appl 117(16):18–21
11. Bhuyan MH, Bhattacharyya DK, Kalita JK (2011) Surveying port scans and their detection methodologies. Comput J 54:1565–1581
12. Das V, Pathak V, Sharma S, Sreevathsan, Srikanth MVVNS, Gireesh Kumar T (2010) Network intrusion detection system based on machine learning algorithms. Int J Comput Sci Inform Technol (IJCSIT) 2(6)
13. Alkasassbeh M, Al-Naymat G, Hassanat ABA, Almseidin M (2016) Detecting distributed denial of service attacks using data mining techniques. Int J Adv Comput Sci Appl (IJACSA) 7(1)
14. Bains JK, Kaki KK, Sharma K (2013) Intrusion detection system with multi layer using Bayesian networks. Int J Comput Appl 67(5). ISSN 0975-8887
15. Gavrilis D, Dermatas E (2005) Real-time detection of distributed denial-of-service attacks using RBF networks and statistical features. Comput Netw 48:235–245. https://doi.org/10.1016/j.comnet.2004.08.014
16. Sofi I, Mahajan A, Mansotra V (2017) Machine learning techniques used for the detection and analysis of modern types of DDoS attacks. IRJET 4(6)
17. Balsrengadurali C, Saraswathi S (2013) Fuzzy based detection and prediction of DDoS attacks in IEEE 802.15.4 low rate wireless personal area network. IJCSI Int J Comput Sci 10(6)(1)
18. Bains JK, Kaki KK, Sharma K (2013) Intrusion detection system with multi-layer using Bayesian networks. Int J Comput Appl 67(5). ISSN 0975-8887
19. Ismanto H, Wardoyo R (2016) Comparison of running time between c4.5 and k-nearest neighbor (k-nn) algorithm on deciding mainstay area clustering. Int J Adv Intell Inform 2(1):1–6
20. Hariharan M, Abhishek HK, Prasad BG (2019) DDoS attack detection using C5.0 machine learning algorithm. Int J Wirel Microw Technol (IJWMT) 9(1):52–59