International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 13, Issue 3, March 2024)

DDoS Attacks Detection Using Machine Learning Algorithms

Shivani, Prof. Saurabh Sharma, Prof. Vishal Paranjape

Global Nature Care Sangathan Group of Institutions, Jabalpur(M.P)

***Abstract*— Entropy-based characteristics have been widely exploited in DDoS attack detection in recent years. However, existing approaches mostly use entropy-based features to discriminate between regular and attack data. The challenge with this strategy is reducing DDoS attacks that can be detected. We have suggested additional entropy-based functions to help with attack detection in an effort to enhance the overall and precise identification of low- and high-intensity DDoS attacks. Additionally, we have unveiled a brand-new multi-classification system built on the suggested collection of numerous classed devices and entropy-based characteristics. We carried out analysis and measurements using various intensity data sets. The results of the investigations show that our method was more accurate and efficient than other state-of-the-art techniques. For different assault intensities, our technique consistently produces four different data sets, including the best or second-best classification.**

Keywords — prediction models, data models, logistics, machine learning, mathematical model, techniques for machine learning and analysis, DDoS, entropy. Introduction

I. INTRODUCTION

An assault includes a coordinated DDoS attack from several sources[1]. DDoS attack sources are often dispersed throughout the world and can number in the hundreds or even thousands. Conventional traffic measurement methods like MULTOPS[2] and LADS[3] can be used to identify DDoS assaults. This technique detects a DDoS assault if the volume of traffic surpasses a preset threshold. DDoS assaults are usually divided into two categories: high intensity attacks and moderate intensity attacks. The main differentiator between these attacks is the trans-packet mission rate. Application-level attacks with moderate intensity, such as Slowloris[5], transmit data at a slow speed; on the other hand, high-intensity attacks, such as Smurf[4], transfer data at a high speed, significantly increasing the volume of traffic. Other low intensity attacks try to create longer-term economic loss by slowing down services instead of completely stopping them[6] or low intensities since the scattered nature of the attack makes it impossible to inspect every packet involved. Entropy has been widely employed by DDoS researchers as a workaround for the volumetric approach's limiting application in recent years[7]. Entropy quantifies the unpredictability of information. It can show patterns of a network attack and provides an overview of the distribution of network communication[8]. Any abrupt shift in entropy levels could be a sign of a DDoS attack. However, the majority of contemporary[9]–[11] detectors only employ a small number of particular subtypes of entropy-based attack detection. Traffic features at regular intervals that exhibit entropy are known as entropical characteristics, and they include things like the originating IP address. Different entropy-based criteria suggest that the detection accuracy may be limited by different types of DDoS attacks. It can be difficult to choose the ideal detection threshold for entropy-based methods[9]. For a certain amount of time, traffic with an entropy value greater than the threshold will be regarded as containing attack traffic. An suitable detection threshold needs to be supplied in order to distinguish between DDoS attacks of varying intensity. It can be difficult to choose a fair threshold, though, so that different DDoS attack types can be identified and are less likely to cause mistakes.
One way to address this problem is to use machine learning classifiers (ML). A more sophisticated and intricate way of classifying traffic is through the use of machine learning classificators, which employ historical data to look for trends and build classification models[12]. These classificators may automatically identify recurring trends in assault and traffic data in order to generate a categorization model. The need to manually adjust the detection settings is removed by the learning process. The challenge with machine learning classifiers, however, is that there isn't a single, universal classifier that can accurately classify all attack flows. For example, DDoS attacks of moderate and high intensities are essentially distinct from one another. As such, it is not possible to classify both attacks correctly using a single ML classifier.

In response to worries about the accuracy and scope of DDoS attacks, we have created a novel method for detecting DDoS attacks called Machine Learning Classifier Determination (E3ML).

In order to correctly identify high and low-intensity DDoS attacks, our main contribution is the development of a multi-classifier system with majority voting, an arbiter, and new entropy-based characteristics that integrate common entropy-driven characteristics and novel entropy-based characteristics. We used many intensity data sets to examine and evaluate E3ML. Our testing results clearly demonstrate that E3ML performs best and second best when compared to other models.

II. Literature Review

A. Classification of traffic

A classification of traffic can be used to determine between DDoS and typical network traffic attacks. Significant steps to classify traffic are packet titles including source and destination IP addresses, source and destination port numbers, protocol, time to live (TTL) and flag.
The goal of this research is to construct a model of classification that classifies traffic classes that use machinery-learning techniques based on entropy values.

B. Entropical

Entropy has been commonly used in contemporary DDoS assault detection processes[9],[13] and[14], often used to produce important traffic grading features. Entropy is a statistical approach that evaluates the uncertainty of knowledge. Entropy detects the distribution of network traffic modifications using a single value measurement[15]. Sufficient observation of these changes showed anomalies in the network[9],[13],[14].

Recent detection investigation shows higher accuracy of detection than existing entropy-based detection methods[16]. The benefits of this approach are: quick estimation, high susceptibility, low FPR, without traffic or devices and without the usage of the network[9], more than classic volumetric approaches[2][3].

Entrropy measurements are typically employed for raw traffic functions including source and destination IP addresses, port and entropy-based protocol numbers. The high entropy value, for example, indicates a considerable fluctuation and the low entropy value means that traffic packets have a smaller origin variation. This is helpful for detecting attacks, since prominent DDoS attacks with many target attack sources are often very different from ordinary traffic at source and destination.

Methods of so-called entropy-based variants can also produce entropical functions.

[9] In these functions, you can measure the difference between two distinct entropy-based properties, such as the difference between the source and the target.

Address IP. IP address. IP address. IP address. IP address. Address IP. The source IP and destination IP address for legitimate traffic usually shows a similar entropy value[9], whereas the majority of DDoS assault traffic has substantial differences between the two entropy-based properties as multiple distributed sources of attack send assault traffic to a certain goal.

C. Multiple Classification System.

The Multiple Classification System (MCS) is the technique for developing a broader and more accurate classification system employing distinct ML classifiers and merging one or more models.

In the past, neural network combinations and decision tree classifiers have shown the ability to generalise and detect accuracy[17]. Robinson et al. [18] analysed and graded many machine classifications based on their success in detecting DDoS attacks. They have also shown that classification methods such as AdaBoost and Random Forest are better combined than classification approaches. Chand et al.[19] created MCS approaches using attack stacking methods and showed the maximum accuracy in Vector Machine Support (SVM) stacking with other ML classificators, rather than using the SVM classification alone.

However, MCS approaches only focus on the detection of a specific sort of DDoS assault using standard entropy-based calculations. No further entropy functions were considered in any of the existing works. It is desirable to have a broad and reliable method of detection. We are presenting a system for identifying various forms of DDoS attacks with wide knowledge integrated in numerous entropy-based functions and a classification based on MCS.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 13, Issue 3, March 2024)

III. ENTROPY AND MACHINE LEARNING BASED CLASSIFIER SYSTEM FOR DDOS DETECTION

We offer a new classification model that uses entropy based features and MCS approaches to detect various forms of DDoS attacks. Our approach, E3ML, combines the comprehensive knowledge of various entropy features and intensity to differentiate between attack and legitime traffic in several classification models. We are using both old entropy-based characteristics and adding new entropy-based features to reveal various forms of network DDoS assault. In three classifications of the machine, we apply various forces to increase generality and precision in the detection of various DDoS attack traffics utilising novel entropy features to classify models.

A. System overview

Our strategy includes building the features displayed in Fig. 1 and detecting attacks as indicated in Fig. 2.

B. Entropical Building Feature

We construct two types of entropy function; typical entropy-based functions that depend on raw entropy

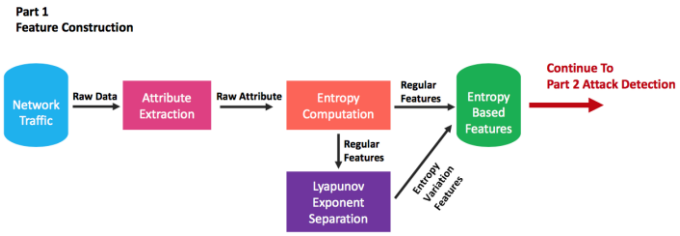

Fig. 1. Construction function

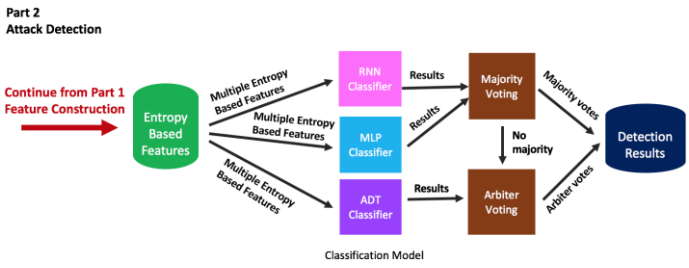

Fig. 2. Detection of Attack

Characteristics and entropy variance based on two conventional entropy characteristics. We have selected two separate features to increase the efficiency of traffic classification. While standard entropy (source and destination IP addresses, address and protocol, source and destination port) features commonly used in the classification of traffic, they are restricted to certain types of DDoS (common DDoS attacks) and cannot operate in other types of DDoS, such as low-intensity DDoS attacks. The variation in IP in Entropy showed promising improvements in continuous, low-intensity attacks[20]. We can boost the overall detection of various sorts of DDoS attacks by providing more useful traffic classification.
The design of the function takes place in three phases.

• Extraction of raw features: we extract raw features from every packet header. Features include source IP, target IP, source port address, destination port and protocol commonly used for DDoS detection[21]. We also examine the extraction of other functionality not usually used for entry-based function formatting, such as delta time, packet length, TCP sequence, and TCP window duration (cf: Table I.)
• Entropy Calculation: each entropy shall be computed by a specific time interval based on the Shannon entropy, with W = 60 seconds in all our past tests. The literature considers the time interval of 60 seconds as an adequate classification and standard practice[9]. The complex time interval change remains a projected target of the investigation. Shannon entropy is used since it is a standard traffic entropy measure. We have also tried different techniques to entropy, such Tsallis[22].

TABLE I
REGULAR ENTROPY-BASED FEATURES GENERATED FROM RAW TRAFFIC FEATURES

No	Features	Definition
1	Delta time *c*	Time since the previous packet was captu
2	Source IP Address *c*	Source IP address of the packet
3	Destination IP Address *c*	Destination IP address of the packet
4	Source Port Address *c*	Source port number of the packet
5	Destination Port Address *c*	Destination port address of the packet
6	Source MAC Address	Hardware address of the previous netwo router/host the packet is coming from
7	Destination MAC Address	Hardware address of the next-hop netwo router/host the packet is going to
8	Source Network Address	Source Network address of the packet
9	Destination Network Address	Destination Network address of the pack
10	Protocol *c*	Type of protocol (HTTP, TELNET, DN
11	Packet Length	Size of packets in bits
12	IP DSCP Value	Differentiated Services Code Point value each packet
13	TCP Sequence Number	TCP sequence number relative to the first segment in a TCP session

c - commonly used features in traffic classification

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 13, Issue 3, March 2024)

[21] TABLE II
ENTROPY VARIATION FEATURES GENERATED USING A VARIATION O F
LYAPUNOV EXPONENT SEPARATIO N

No	Features	Definition
1	Separation IP	Rate of separation between source and destination IP Address
2	Separation Port *new*	Rate of separation between source and destination Port Address
3	Separation MAC *new*	Rate of separation between source and destination MAC Address
4	Separation Network *new*	Rate of separation between of source and destination Network Address
5	Separation TCP *new*	Rate of separation between TCP window size and destination TCP Length

new - new entropy-based features

Reinyi [23] and produce results similar to the entropy of Shannon. Features that are calculated at this stage are referred to in Table I as normal entropy-based features.

• Build entropy variation characteristics: We create entropy variation features using the method given by Ma et al.[9] by combining two normal entropy-based features obtained by pre-violent stages. Ma et al. recommended to compute the rate of separation between the source and the destination IP address by the Lyapunov Exponent Separation method. They used the separation rate to highlight the disparities between two separate characteristics. This IP separation capability can successfully separate DDoS attacks from a single type of DDoS attack. We have introduced new entropy modification characteristics that can be used for detecting different kinds of DDoS attacks, such as Lyapunov separation technique, separation mac, separation network and separation TCP, as described in Table
II.

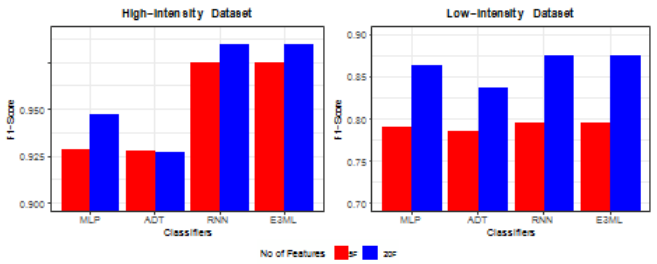

Fig. 3. Effects of conventional 5-fold (5F) vs 20 F1 scoring (20F) compared with traditional 5-fold data set. This applies to all four classifiers, with the exception of the high-intensity ADT classifier. However, the difference in F-measure performance in both sets of characteristics can be minimal as it is only 0.06 percent. In other classifications, 20 features have a 5-fold output of at least 1%. We therefore employ all 20 features as input for our Attack Detection classifier model.

A. Attack detection

Attack detection (Fig. 2) involves (1) the categorization of traffic through many ML classificers (RNN, Multilayer Perceptron, ADT) and (2) the use of a simple majority vote method based on the classification results to determine the occurence of DDoS. The vote takes place between two standard ML classifications (i.e. MLP and RNN). Each classifier provides votes on the basis of its categorization result (attack traffic or normal traffic). If voting is equal for attack and normal, we propose a third classification, i.e. ADT for arbitrators. We only compare the ADT detection result to the RNN detection results because our preliminary test suggests that RNN is more accurate. We compare the outcomes of ADT and RNN detection with simple majority voting to decide the ultimate outcome by comparing RNN and MLP in the former phase. However, we identify traffic as an assault if ADT voting and RNN voting are different.

IV. PERFORMANCE EVALUATION

We validated our technique by using a relatively fresh ISCX 2012 (ISCX'12)[24] and 1998 DARPA Intrusion Detection (DARPA'98) data set [25]. ISCX'12 is a newer dataset containing more recent attacks, such as high intensity attacks generated with IRC botnets and low intensity attacks made by the Slowloris tool. ISCX'12 is a new dataset (Low-Intensity). The dataset DARPA'98 features more typical attacks, such as SMURF, Neptune and Land that have a mixing intensity due to many forms of attacks in a single dataset. In our assessment we use data from Week 1 (Mix Intensity 1) and Week 2 (Mix Intensity 2). The dataset is still used to evaluate DDoS detection and defence approaches[26] for comparison in the literature, however the DARPA'98 has various problems[27]. In the literature, these datasets were widely utilised to evaluate detection of DDoS and protection approaches[9],[28].

IV. Conclusion & Future Work

In this study we proposed a DDoS detection approach using several entropy-based capabilities and E3ML classificators. E3ML comprises of function builds which provide two types of entropy functions (i.e. conventional entropy-based and entropy) and an attack detector to recognise traffic in the network in an attack or normal using our detection algorithm. E3ML is a voting method which compares the results of MLP and RNN classification as arbitrators with ADT.

Results from performance evaluations demonstrated that our technology can efficiently detect DDoS attacks with varied intensities via data sets. Our approach surpasses other previous approaches (EMD-Li and ESDA) and has shown that data sets with various types of DDoS attacks can consistently produce high-precision results. We have shown that a combination of features and ML approaches on entropy produces promising results for the identification of DDoS.

References

[1]Neural network. https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11 /report. html.
[2] Peer to peer lending and alternative investing. https://www.lendingclub.com/.
[3] (2017). Variable selection with elastic net. https://www.r-bloggers.com/ variable- selection-with-elastic-net/.
[4] Addo, P. M., Guegan, D., and Hassani, B. (2018). Credit risk analysis using machine and deep learning models. Risks, 6(2):38.
[5] Antonakis, A. and Sfakianakis, M. (2009). Assessing naive bayes as a method for screening credit applicants. Journal of applied Statistics, 36(5):537–545.
[6] Attigeri, G. V., Pai, M., and Pai, R. M. (2017). Credit risk assessment using machine learning algorithms. Advanced Science Letters, 23(4):3649–3653.
[7] Batista, G. E. and Monard, M. C. (2003). An analysis of four missing data treatment methods for supervised learning. Applied artificial intelligence, 17(5-6):519–533.
[8] Bekhet, H. A. and Eletter, S. F. K. (2014). Credit risk assessment model for jordanian commercial banks: neural scoring approach. Review of Development Finance, Universiti Tenaga Nasional (UNITEN), 43000 Kajang, Selangor, Malaysia, 4(1):20–28.
[9] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16:321–357.
[10] Chen, T., He, T., Benesty, M., et al. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, pages 1–4.
[11] DAVIS, R. H., Edelman, D., and Gammerman, A. (1992). Machine-learning algorithms for credit-card applications. IMA Journal of Management Mathematics, 4(1):43–51.
[12] Duan, K.-B. and Keerthi, S. S. (2005). Which is the best multiclass svm method? an empirical study. In International workshop on multiple classifier systems, Nanyang Technological University, Nanyang Avenue, Singapore, pages 278–285. Springer.
[13] Faggella, D. (2017). The rise of neural networks and deep learning in our everyday lives - a conversation with yoshua bengio -.
[14] Faggella, D. (2018). What is machine learning? - an informed definition. https://www.techemergence.com/what-is-machine-learning/.
[15] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, Stanford University, USA, pages 1189–1232.
[16] Friedman, J. H. (2002). Stochastic gradient boosting. Computational Statistics & Data Analysis, Stanford University, Stanford, CA 94305, USA, 38(4):367–378.
[17]Hamid, A. J. and Ahmed, T. M. (2016). Developing prediction model of loan risk in banks using data mining. Machine Learning and Applications: An International Journal,University Khartoum, Sudan, 3(1):1–9.
[18] Hsu, C.-W., Chang, C.-C., Lin, C.-J., et al. (2003). A practical guide to support vector classification. National Taiwan University, Taipei 106, Taiwan.
[19] Islam, S. (2017). Bad loans cripple the banking sector.
[20] Khandani, A. E., Kim, A. J., and Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. Journal of Banking & Finance, 34(11):2767–2787.
[21] Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. Expert Systems with Applications, Lefkosa, Mersin 10, Turkey, 37(9):6233–6239.
[22] Kohavi, R. et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Ijcai, Stanford University Stanford, CA., volume 14, pages 1137–1145. Montreal, Canada.
[23] Liaw, A., Wiener, M., et al. (2002). Classification and regression by randomforest. R news, 2(3):18–22.

[24] Lopes, R. G., Carvalho, R. N., Ladeira, M., and Carvalho, R. S. (2016). Predicting recovery of credit operations on a brazilian bank. In Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on, pages 780–784. IEEE.

[25] Mackiewicz, A. and Ratajczak, W. (1993). Principal components analysis (pca). Com- puters and Geosciences, Department of Mathematics, Technical University of Poznan´ , Piotrowo 3a, Poznan´ Poland, 19:303–342.

[26] Mowla, G. (2018). Default loans plague banking sector.

[27] Nova, A. (2018). More than 1 million people default on their student loans each year. [28] of India, P. T. (2018). 9 million loan defaulters blacklisted in china; 27 billion dollar frozen.

.

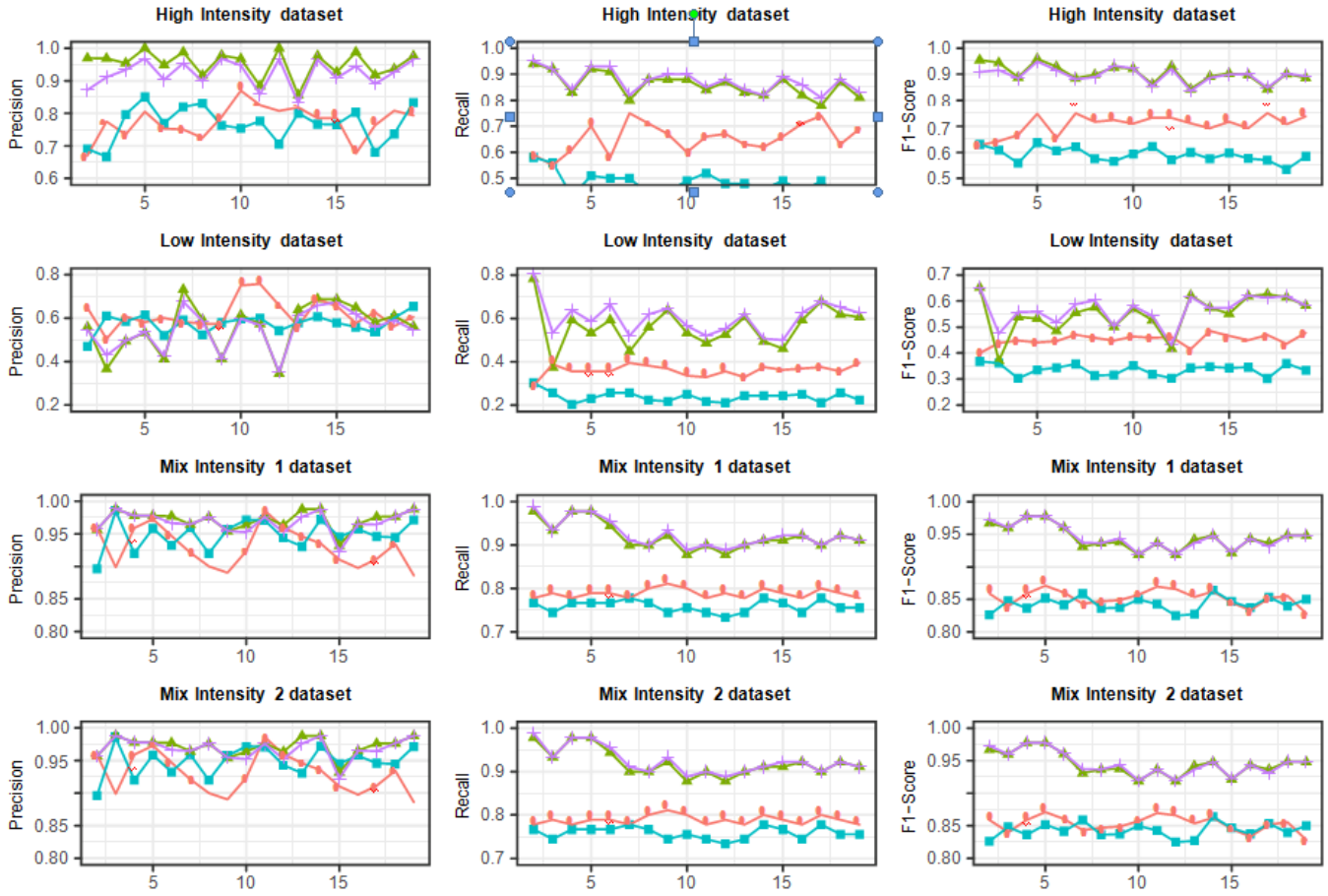

Fig. 4. Performance Evaluation