INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

MITIGATION DISTRIBUTED DENIAL OF SERVICE ATTACKS IN MACHINE LEARNING

[1.] MS. K. ANUSHA, [2.] B.ARAVIND,[3.] B.KIRAN KUMAR [4.] K.SAI CHARAN

[1.] Assistant Professor Department of Computer Science and Engineering, Teegala Krishna Reddy Engineering College, Rangareddy (TS).India.

Email-:[1.] kommuanusha14@gmail.com

[2,3,4.]B.Tech StudentstDepartment of Computer Science and Engineering, Teegala Krishna Reddy Engineering College, Rangareddy (TS).India.

Email-:[2.] aravindb2301@gmail.com, [3.] rathodkiran1819@gmail.com,

[4.] saicharankatamouni@gmail.com.

Abstract- In imbalanced network traffic, malicious cyber-attacks can often hide in large amounts of normal data. It exhibits a high degree of stealth and obfuscation in cyberspace, making it difficult for Network Intrusion Mitigation System to ensure the accuracy and timeliness of detection . First, use the Nearest Neighbor algorithm to divide the imbalanced training set into the DOS set and the easy set. SDN networks are exposed to new security threats and attacks, especially Distributed Denial of Service (DDoS) attacks. For this aim, we have proposed a model able to detect and mitigate attacks automatically in SDN networks using Machine Learning (ML). Different than other approaches found in literature which use the native flow features only for attack detection, our model extends the native features.

KEYWORDS: Distributed Denial of Service, Cyber-Attacks, Machine Learning.

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

1. INTRODUCTION

Distributed Denial of Service (DDoS) attacks pose a significant threat to the availability and reliability of online services and systems. With the increasing reliance on machine learning in various domains, it becomes crucial to develop effective strategies to mitigate DDoS attacks specifically targeted at machine learning systems. DDoS attacks overwhelm a target system with a massive volume of malicious traffic, rendering it unable to respond to legitimate user requests. The dynamic and resource-intensive nature of machine learning systems makes them particularly vulnerable to such attacks.

Mitigating DDoS attacks in machine learning involves developing techniques and algorithms that can detect and respond to these attacks in real-time, allowing the system to continue functioning and providing services to legitimate users. Traditional DDoS mitigation methods, such as rate limiting or IP filtering, may not be effective for machine learning systems, as they can inadvertently block legitimate traffic and hinder the learning process.

The field of DDoS mitigation in machine learning is still evolving, and researchers are exploring various approaches to enhance the resilience of machine learning systems against these attacks. These approaches often leverage the unique characteristics of machine learning algorithms and architectures to develop proactive and adaptive defense mechanisms.

2. LITERATURE SURVEY

Securing computer and network information is important for organizations and individuals because compromised information can cause considerable damage. To avoid such circumstances, intrusion detection systems are important. Recently, different machine learning approaches have been proposed to improve the performance of intrusion detection systems.

Wang et al. proposed an intrusion detection framework based on SVM and validated

their method on the NSL–KDD dataset. They claimed that their method, which has 99.92% effectiveness rate, was superior to other approaches; however, they did not mention used dataset statistics, number of training, and testing samples. Furthermore, the SVM performance decreases when large data are involved, and it is not an ideal choice for analyzing huge network traffic for intrusion detection.

Machine Learning (ML) algorithms focus on the development of computer programs where they provide the systems with the ability to automatically learn and improve from experience without the intervention of humans and without being explicitly programmed. The feature selection (FS) process is one of the vital ML pre-processing phases where it removes unwanted and irrelevant features with the goal of improving prediction (i.e.; detection) accuracy and reducing computational complexity. Dash and Liu mentioned four basic procedures in a FS method. The procedures are generation, evaluation, stopping, and validation. Various support vector machine (SVM) models with NSL-KDD dataset, genetic-fuzzy rule mining approach, genetic algorithm approach, mutual information-based techniques, filter-based methods, etc. were used in feature selection process for intrusion detection systems. Several FS methods are also found in detecting DDoS attacks such as detecting DDoS in cloud computing, detecting robust backscatter DDoS, chi-square and information gain FS methods in detecting general DDoS attacks, etc. In addition, supervised and unsupervised ensemble frameworks were also used to detect DDoS attacks with better accuracy. In this research, we propose an ensemble framework for feature selection methods (EnFS) where all three types of methods are used and combined using a majority voting technique to extract a valid minimal subset of features that improves the performance of DDoS detection problem.

3. EXISTING SYSTEM:

In existing machine learning systems, several approaches can be employed to reduce the risk of DDoS attacks. Load balancing techniques can help distribute traffic across multiple servers, blacklisting can block requests from known attackers, captchas can prevent automated attacks. Additionally. cloud-based protection services can provide an additional layer of protection against DDoS attacks.

DISADVANTAGES OF EXISTING SYSTEM:

1. Training data: ML algorithms require

large amounts of high-quality training data to accurately detect and mitigate DDoS attacks. This can be challenging as DDoS attacks are relatively rare and can be difficult to replicate in a controlled environment.

2. Adaptability: DDoS attacks are constantly evolving and attackers can change their tactics frequently. ML models may struggle to adapt quickly enough to new attack methods and may require significant retraining to remain effective.

4. PROPOSED SYSTEM:

Our proposed ensemble framework, The detailed architectural diagram depicting the process flow is given in Fig. 1. It shows the processing phases, namely a) data preprocessing, b) feature selection, c) ensemble selection methods and d) model classification with performance analysis of the detection.

ADVANTAGES OF PROPOSED SYSTEM:

1. Maintaining Availability to analyze large amount of data in real-time.

2. Improving Performance and increase speed to respond to a DDoS event and optimize response resources.

5. MODULES:

1. DATASET

Dataset selection for experimentation is a significant task, because the performance of the system is based on the correctness of a dataset. The more accurate the data, the greater the effectiveness of the system. The dataset can be collected by numerous means, such as 1) sanitized dataset, 2) simulated dataset, 3) testbed dataset, and 4) standard dataset [9]. However, complications occur in the application of the first three methodologies. A real traffic method is expensive, whereas the sanitized method is unsafe. The development of a simulation system is also complex and challenging. Additionally, different types of traffic are required to model various network attacks, which is complex and costly. To overcome these difficulties, the NSL–KDD dataset is used to validate the proposed system for intrusion detection.

2. PRE-PROCESSING

The classifier is unable to process the raw dataset because of some of its symbolic features. Thus, pre-processing is essential, in which non-numeric or symbolic features are eliminated or replaced, because they do not indicate vital participation in intrusion detection. However, this process generates

overhead including more training time; the classifier's architecture becomes complex and wastes memory and computing resources. Therefore, the non-numeric features are excluded from the raw dataset for improved performance of intrusion detection systems.

3. CLASSIFICATION

Placing an activity into normal and intrusive categories is the core function of an intrusion detection system, which is known as an intrusive analysis engine. Thus, different classifiers have been applied as intrusive analysis engines in intrusion detection in the literature, such as multilayer perceptron, SVM, naive Bayes, self-organizing map, and DT. However, in this study, the three different classifiers of SVM, RF, and ELM are applied based on their proven ability in classification problems. Details of each classification approach are provided

4. SUPPORT VECTOR MACHINE

Implementation of the SVM model in the proposed system. The kernel function uses squared Euclidean distance between two numeric vectors and maps input data to a high dimensional space to optimally separate the given data into their respective attack classes. Therefore, kernel RBF is particularly effective in separating sets of data that share complex boundaries. In our study, all the simulations have been conducted using the freely available Lib SVM package.

5. RANDOM FOREST

RFs are ensemble classifiers, which are used for classification and regression analysis on the intrusion detection data. RF works by creating various decision trees in the training phase and output class labels those have the majority vote [13]. RF attains high classification accuracy and can handle outliers and noise in the data. RF is used in this work because it is less susceptible to over-fitting and it has previously shown good classification results.

6. EXTREME LEARNING MACHINE

ELM is another name for single or multiple hidden layer feed forward neural networks. ELM can be used to solve various classification, clustering, regression, and feature engineering problems. This learning algorithm involves input layer, one or multiple hidden layers and the output layer. In the traditional neural networks, the tasks of adjustment of the input and hidden layer weights are very computationally expensive and time-consuming because it requires multiple rounds to converge. To overcome this problem, Huang et al. proposed an SLFN by arbitrarily selecting input weights

and hidden layer biases to minimize the training time.

6. RESULTS:

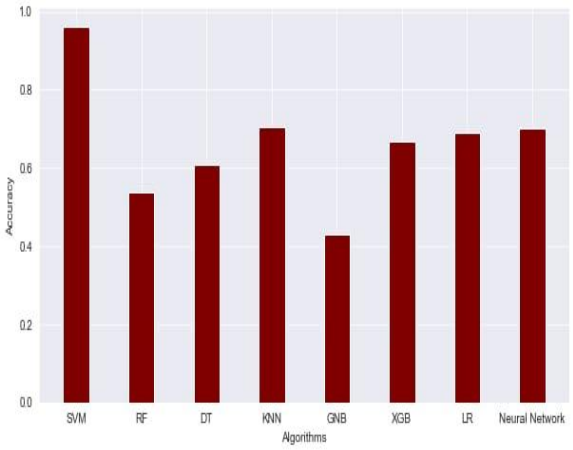

Fig 1 : DESCRIBING COLUMN INFORMATION

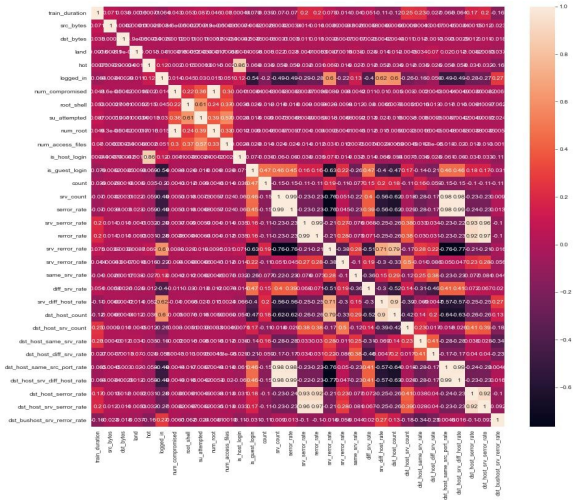

Fig 2 : CREATING THE BAR PLOT

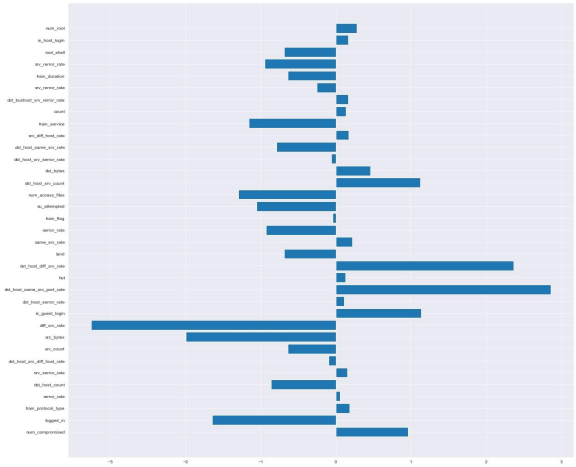

Fig 3 : PLOTING FEATURES

7. CONCLUSION

Intrusion detection and prevention are essential to current and future networks and information systems, because our daily activities are heavily dependent on them. Furthermore, future challenges will become more daunting because of the Internet of Things. In this respect, intrusion detection systems have been important in the last few decades. Several techniques have been used in intrusion detection systems, but machine learning techniques are common in recent literature. Additionally, different machine learning techniques have been used, but some techniques are more suitable for analyzing huge data for intrusion detection of network and information systems. To address this problem, different machine learning techniques, namely, SVM, RF, and ELM are investigated and compared in this

work. ELM outperforms other approaches in accuracy, precision, and recall on the full data samples that comprise 65,535 records of activities containing normal and intrusive activities. Furthermore, the SVM indicated better results than other datasets in half of the data samples and in 1/4 of the data samples. Therefore, ELM is a suitable technique for intrusion detection systems that are designed to analyze a huge amount of data. In future, ELM will be explored further to investigate its performance in feature selection and feature transformation techniques.

Feature selection is a vital part of any classification problem. In this research, we have proposed an ensemble framework for feature selection (EnFS) which combined seven well-known selection methods. The goal of combining these methods is to extract the most accurate set of features that produces better outcomes in detecting DDoS attacks. We have performed three experiments using the i) full feature set initially, then ii) seven feature sets obtained from the seven selection methods, and iii) finally the resultant feature set obtained from our EnFS that used the majority voting technique. The NSL-KDD dataset provided the basis for validating EnFS that reduced the number of features from 41 to 11. Subsequently, we performed an extensive set of experiments using our ensemble supervised ML framework [2] to evaluate the performance of the resulting feature set.

8. REFERENCES

[1] H. Wang, J. Gu, S. Wang, An effective intrusion detection framework based on SVM with feature augmentation, Knowledge-Based Systems, Volume 136, 2017, Pages 130-139, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2017.09.014.

[2] F. Kuang, X. Weihong , S. Zhang, A novel hybrid KPCA and SVM with GA model for intrusion detection, Applied Soft Computing, Volume 18,2014,Pages 178-184,ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2014.01.028.

[3] A. A. Aburomman, M.B. Reaz, A novel SVM-kNN-PSO ensemble method for intrusion detection system, Applied Soft Computing, Volume 38, 2016, Pages 360-372, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2015.10.011.

[4] M.R. Raman, N. Somu, K. Kirthivasan, R. Liscano, V.S. Sriram, An efficient intrusion detection system based on hypergraph – Genetic algorithm for parameter optimization and feature

selection in support vector machine, Knowledge-Based Systems, Volume 134, 2017, Pages 1-12, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2017.07.005.

[5] S. Teng, N. Wu, H. Zhu, L. Teng and W. Zhang, "SVM-DT-based adaptive and collaborative intrusion detection," in IEEE/CAA Journal of Automatica Sinica, vol. 5, no. 1, pp. 108-118, Jan. 2018. doi: 10.1109/JAS.2017.7510730.