© 2023 IJNRD | Volume 8, Issue 5 May 2023 | ISSN: 2456-4184 | IJNRD.ORG

Detection and Elimination of Network Anomaly in SDN using Trust Analysis

M.Shanmugam,
Assistant Professor,
Sri Manakula Vinayagar Engineering College,

S.Santhosh Rajan,
Student,
Sri Manakula Vinayagar Engineering College,

***Abstract*— The necessity for networking and data exchange has dramatically increased in the modern world. Network security is necessary given the rapid development and globalization of information technology. Although they may offer some amount of security, firewalls never warn administrators of impending assaults. A trustworthy detection system is required to locate such aberrant network packet behavior in order to increase efficiency and accuracy. As a result of how quickly today's network environment is evolving, the network is constantly at risk from new sorts of attacks. Therefore, regular updates to the network administration system are required for upgrading the security level. Intrusion Detection Systems is one of the network packet monitoring systems (IDS). The proposed model was created using a machine learning approach to identify malicious network packet activity. KDD-99 dataset is utilized for that. The dataset is first normalized to reduce calculation complexity, and then further features are reduced using a Deep Neural Network technique. Only effective features can be employed for harmful behavior identification, according to the reduced features. According to the results analysis, DNN works best when choosing more than 15 features, whereas co-relation performs best when choosing less than 15. The k-mean clustering algorithm is used to accomplish data clustering after feature reduction. Deep Neural Network, are designed for classification of dataset into five attack categories i.e. DOS, U2R, R2L, Probe and Normal. As compared to some other multilevel classifier work the proposed algorithm proves its efficiency in terms of high accuracy, high detection rate and False Alarm Rate (FAR).**

Keywords—SDN, Random Forest Classifier, Intrusion Detection System, KNN Clustering

I. INTRODUCTION

The Internet has significantly changed our social and professional lives since the advent of mobile and wireless gadgets like Internet of Things (IOT) and Radio-frequency identification (RFID) devices, among others. For many other types of transactions, including social, economic, financial, healthcare, industrial, and government-related ones, the Internet has emerged as a standard medium. Meanwhile, malevolent individuals or groups known as cybercriminals have taken advantage of numerous internet vulnerabilities to launch a variety of cyberattacks that have resulted in significant loss and damage on a social, individual, industrial, and national level. In general, there are four basic sorts of cyberattacks that have been detected and documented: probe, denial of service (DoS), user to root (U2R), and remote to local (R2L). Securing the internet from these vulnerabilities and cyber-attacks have been a long standing researched problem.

Different security solutions have been suggested to defend against different types of cyber-attacks. In order to monitor the network for malicious activity or policy breaches, we present in this paper a method known as Random Forest Classifier for intrusion detection in software defined networking (SDN) [11–13]. The SDN controller will be notified by the IDS wherever an intrusion activity or policy violation is found. The controller will then perform one of two actions depending on when the detection was made: if it was made during a PACKET_IN event, it will log the packet information; if it was made during an analysis of flow statistics, it will install a rule in the concerned switch to drop packets of the anomalous flow. Both signature-based detection and abnormality-based detection are methods the intrusion recognition system uses to find attacks. In order to determine attack complements, signature-based analysis is employed to compare an individual to a database of acknowledged attack autographs. Contrarily, anomaly-based detection compares monitored data to a baseline of expected behavior and can send alerts in response to anomalous behavior.

picking of a dataset for training and creating the model is one of the key elements for any Intrusion Detection System's (IDS) performance. In this study, we examine the performance of the suggested strategy using the common NSL-KDD [1,2] dataset. The NSL-KDD dataset was offered as a solution to some of the KDD'99 dataset's intrinsic issues, which are listed in [1]. The NSL-KDD dataset offers a manageable number of train and test data sets (not too few for the analysis to be unsuccessful nor too many for it to be computationally expensive), allowing experiments to be run on the entire set without the requirement to arbitrarily choose a portion of the main dataset. The context and foundation for the huge data problems that Intrusion Diagnosis encounters are thoroughly outlined in this section.

Various intrusion detection and mitigation systems have been proposed till date. These intrusion detection schemes do, however, face a number of obstacles and problems. The issue of incorrect negatives (misclassified as non-anomalous packets) is one of these difficulties and false positives (misclassified as attacks) in all attack detection approaches. A mechanism that can accurately pinpoint the source of the attack is required to ensure accountability for these attacks when detection and mitigation fail. By enabling the ability to identify the true source of the faked packet(s),

II. RELATED WORK

Numerous study findings on intrusion detection systems utilising data-driven machine learning algorithms that have been published in literature reviews are focused on finding ways to increase the effectiveness of these systems using machine learning techniques. Table 1 lists a few reviews of recent studies on various machine learning techniques used for big data in intrusion detection systems.

Table 1 Summary Of Related Work

Authors	Research statement	Methodology	Advantage	Disadvantage
Zhang Lin, Du Hong et.al [5]	**Research on SDN intrusion detection based on online ensemble learning algorithm**	Experimental	The experimental results show that the algorithm can improve the detection accuracy, especially the recognition rate of unknown intrusion behavior.	But from the experimental results, we can see that the fluctuation of the algorithm is large, and the difference between the best and the worst is 20.63%.
Saifudin Usman, Idris Winarno et.al [4]	**Implementation Of SDN-based IDS To Protect Virtualization Server Against HTTP Dos Attacks**	Experimental	Network intrusion detection system (IDS) which is used to perform network traffic analysis on SDN networks to protect virtualization servers from HTTP DOS attacks	But adding more rules to IDS will degrade the performance of the SDN
Anderson Clayton Alves, Admilson De Ribamar Lima Ribeiro et.al [3]	**Intrusion Detection In SDN Environment Using Continuous Data Stream**	Experimental	From experiments, the solution obtains 97.83% accuracy, 99% recall, 80% precision and 2.3% FPR for 10% DDoS attacks on the normal traffic.	Though the Outlier Den Stream algorithm reached 97.83% accuracy in detection of DDOS attacks, but it suffers from overfitting of machine learning model
Josy Elsa Varghese & Balachandra Muniyal et.al [2]	**An Efficient IDS Framework for DDoS Attacks in SDN Environment**	Experimental	Detection of DDoS attacks by incorporating intelligence in the data layer using Data Plane Development Kit (DPDK) in the SDN architecture	Due to the limitation of the experimental environment, the scaling of the framework is not performed
Abdelouahid Derhab & Mohamed Guerroumi et.al [1]	**Blockchain and Random Subspace Learning-Based IDS for SDN-**	Experimental	Blockchain can be used to tamper Open Flow rules of the SDN-enabled industrial IoT systems.	It would also be intriguing to use Blockchain technology to stop the introduction of fraudulent flow rules into the flow tables rather than just detecting them.
Rachana Sharma et.al [6]	**Classified BIG data Intrusion Detection system**	Experimental	Map Reduce performance with KNN models for FPR and detection rate	It gives less accuracy to attacks

III. Data set NSL-KDD

The NSL_KDD data is employed in order to test the IDS proposed model. The data set created by NSL-KDD is not the first of its sort. An international competition for discovering information and data mining tools was called the KDD Cup. This contest was launched in 1999 with the intention of gathering traffic statistics. Building a network intrusion detector—a prediction model that can tell "good" connections from "bad" connections, such as invasions or attacks—was the goal for the competition. Due to this competition, a sizable number of internet traffic records were gathered and combined into a data set known as the KDD'99. From this, the University of New Brunswick's NSL-KDD data set was created as a revised, cleaned-up version of the KDD'99.

The data set contains information on four main types of attacks: DoS, Probe, User to Root (U2R), and Remote to Local (R2L). Below is a quick explanation of each attack:

A. DOS Attack:

The goal of a DoS attack is to obstruct traffic going to and from the target system. Because of the unusually high volume of traffic, the IDS must shut down in order to defend itself. This stops regular traffic from accessing a network. When an online store receives a large volume of orders on a day when there is a huge discount, the network may become overwhelmed and shut down, preventing paying consumers from making any purchases. In the data set, this attack occurs the most frequently.

B. Probe Attack:

An attack that tries to gather information from a network is called a probe or surveillance. Whether it's financial information or customer personal information, the objective is to pose as a thief and steal vital information.

C. U2R Attack:

U2R is an attack that starts out using an ordinary user account and attempts to log in as root to the system or network. To get access or root privileges, the attacker tries to use a system's deficiencies.

D. R2L Attack:

An assault known as R2L seeks to acquire physical access to a distant machine. An attacker attempts to "hack" their way into the network even if they do not have local access to the system or network.

IV. METHODOLOGY

The current work's main objective is to successfully analyse massive data in attack detection systems. The data collection from an attack detection system, which contains both common and odd packets, is used to undertake a full experimental investigation. The data is preprocessed using the information entropy approach. The information entropy method is used to construct faster and more accurate devices. The Random Forest Classifier technique is frequently used to divide the input into constructive and destructive packets. The results of the classifier are evaluated using the efficiency measures.

Finally, an examination of the comparison between the suggested and several existing classifiers is offered. The following subsection typically contains a detailed discussion of each step used in the particular suggested model.

V. PREPROCESSING

Preprocessing is a critical stage in the organization of real-world datasets into a comprehensible manner. Undoubtedly, the real-world datasets have been noisy and sparse in several behaviors. For the purpose of analyzing large-scale data trends, preprocessing is essential. In order to enhance the machine learning method for pattern categorization in massive data intrusion detection systems, preprocessing techniques are consequently necessary. In order to further increase the accuracy and effectiveness of the resultant machine learning work. The information gain method is employed in the current research study to extract important features from the data collection. The next subsections provide a full description of the information gain method.

The values in the dataset should be accurate and not null in order to apply the machine learning method. For some machine learning models, data in a specific format is required because the Random Forest Classifier approach does not allow null values.

The data set should be structured to allow for the simultaneous usage of many neural networks and Deep Learning algorithms, with the optimal approach being selected.

VI. INFORAMTION GAIN METHOD

The gain of information, which is used to determine the worth of the features, includes a measurement of entropy with regard to class. The assumption behind the entropy theory is that a higher feature's entropy implies a greater information richness. To choose the subset of characteristics from the data set with the highest information rank, information gain uses the idea of entropy. This feature offers it more power to classify the data. Information gain is frequently described as a joint set of features using the decrease in entropy that occurs from learning a joint feature set F.

$$Entropy(s) = Info(G) = -\sum m\ pi\ log\ (p_{\ i=1i})$$

Where G represents the likelihood that the class label will appear inside the entire set of class label data. where pi is a random probability that belongs to the class lab CI.

The most important features from the dataset were chosen for this work using the information gain method. Out of 41 features, 13 have been determined to be the most important by using the gain approach.

Figure 1 displays the technique of generating the subset features:

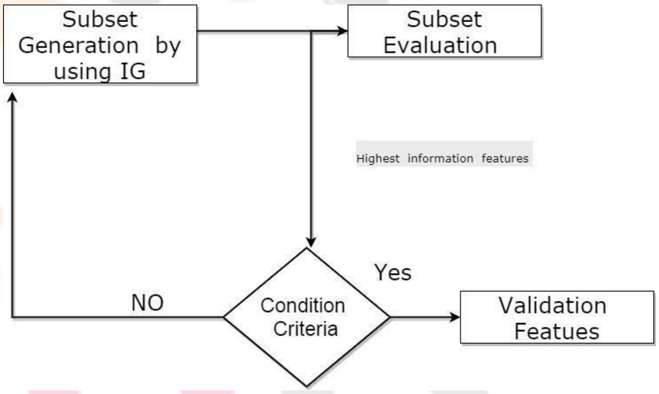

Figure 1 Phases involved in generating subset

VII. RANDOM FOREST CLASSIFIER ALGORITHM

An incredibly common supervised machine learning technique used for Classification and Regression issues in machine learning is called the Random Forest technique. A forest is made up of many different types of trees, and the more trees there are, the more robust the forest will be. Similar to this, the accuracy and problem-solving capacity of a Random Forest Algorithm increase with the number of trees in the algorithm.

In order to increase the dataset's predictive accuracy, a classifier called Random Forest uses many decision trees on different subsets of the input data. It is based on the idea of ensemble learning, which is the practice of integrating various classifiers to solve a challenging problem and enhance the model's performance.

The Random Forest Classifier method combines multiple different trees to create a powerful learner tree. The Random Forest Classifier method produces several classification trees in order to produce one strange, leaner tree. The Random Forest Classifier then makes an effort to construct each tree using a unique bootstrap sample from the initial data set.

VIII. Experimental setup

Several performance assessment methodologies have been used to evaluate the proposed model. In this experiment, a hybrid Random Forest Classifier model and information acquisition methodology are both used. The experiment selects 31 noteworthy assaults. The data only included 185560 attacks, and usual cases were applied.

These attacks match the value 185560 over the whole data set. The original data collection contains 25 MB of data. A hybrid Random Forest Classifier and information gain method is employed in Matlab. Table 2 shows the performance of the proposed model . It investigated that the correct classification of instance is 184331 out 185560 instances. Furthermore, 1229 instances is misclassification out of 185560 instances.

Table 2: Performance evaluation of the suggested model

Performance	Proposed model
Time	16.89 seconds
Correctly Categorized Occurrences	184332
Incorrectly categorized Occurrences	1229
Total Number of Occurrences	185560

It is advised to use the information acquisition strategy to enhance the Random Forest Classifier. The feature selection method improves classifier accuracy while expediting model development. It was difficult to select the ideal number of pertinent subsets from the original data set after collecting the goodness features. In order to improve classification accuracy, the traits with the greatest information rank have been chosen.

Table 3 compares the performance of various existing classifiers with the suggested model utilising the feature selection method. The proposed model has been found to perform better than all techniques currently in use.

Table 3: Results of proposed model with different existing algorithms

Classifiers	FP	TP	Accuracy	Precision
Naïve Bayes	0.005	0.948	94.9232	0.949
REP Tree	0.004	0.987	98.762	0.721
SVM	0.006	0.955	95.42	0.987
KNN	0.007	0.932	93.12	0.975
Proposed model	**0.002**	**0.995**	**99.35**	**0.996**

.

IX. COMPARISON AND PERFORMANCE OF PROPOSED MODEL

Using a comparison criterion that took into account the intrusion detection system's classification accuracy, the proposed IDS model was assessed and looked at. Using the FP, TP, accuracy, and precision performance measures, the suggested technique is compared to all the other current methods. Table 2 provides a summary of the outcomes obtained by the current and proposed algorithms when using the features selection technique. It has been discovered that the suggested model outperforms any current algorithm in terms of results. The accuracy performance of the proposed model is compared to numerous other methods in Figure 3. It illustrates that the proposed model is more accurate and easier to build. Figure 2 illustrates how the proposed model has the highest TP and accuracy measures when compared to other current classifiers. Finally, it is found that the recommended model can accurately identify different types of attacks when compared to other current systems.

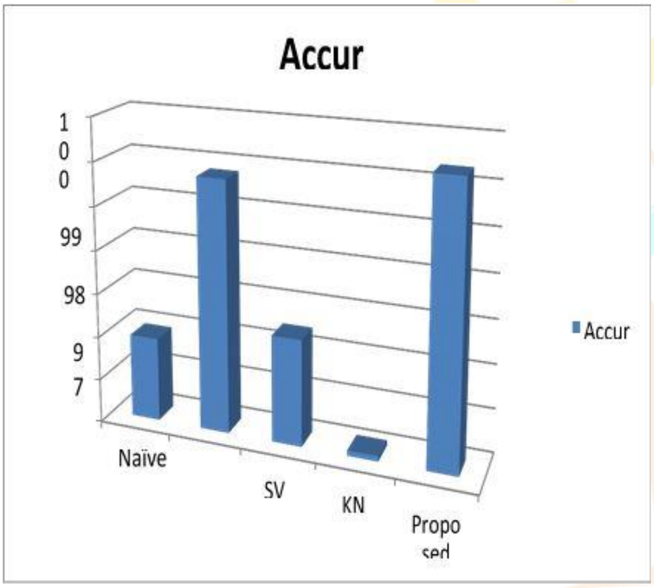

Figure 2 correctness of proposed model's performance compared to current models

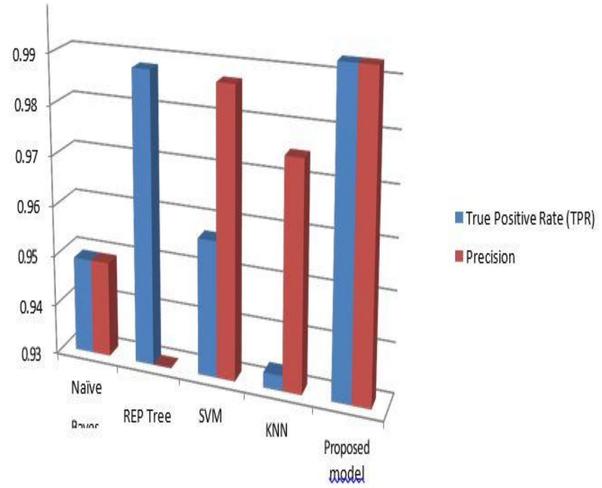

Figure 3 TP and precision of proposed and existing models

X. CONCLUSION

The purpose of the project is to improve the current Intrusion detection system construction methods. The main goal was achieved using the provided model. Different attack types have been used in this investigation. The creation of IDS systems now takes more time due to big data. However, the features selection strategy is employed to resolve the problem. When information gain is used, the 41 initial data features are reduced to their 13 most crucial subset features.

The most crucial elements of the suggested model are also used in application. It has been noted that the accuracy has increased and the model's construction time has decreased. The proposed model was tested using the various performance measures. The proposed model's accuracy is 99.35%. It has been noted that the suggested model performs better than the current classifiers. The researcher will attempt to utilize soft computing in the future while utilizing various datasets.

XI. REFERENCES

[1] Abdelouahid Derhab & Mohamed Guerroumi, " **Blockchain and Random Subspace Learning-Based IDS for SDN Enabled Industrial Iot Security", IEEE 2020 paper**

[2] Josy Elsa Varghese & Balachandra Muniyal,"**An Efficient IDS Framework for DDoS Attacks in SDN Environment**" IEEE 2021 paper

[3] Anderson Clayton Alves, Admilson De Ribamar Lima Ribeiro,"**Intrusion Detection In SDN Environment Using Continuous Data Stream Machine Learning Algorithms" IEEE paper 2021**

[4] Saifudin Usman, Idris Winarno, "**Implementation Of SDN-based IDS To Protect Virtualization Server Against HTTP Dos Attacks**", International conference 2021 IEEE

[5] Zhang Lin, Du Hong ,"**Research on SDN intrusion detection based on online ensemble learning algorithm**" International conference 2020 IEEE

[6] Rachana Sharma & Priyanka Sharma, Preeti Mishra & Emmanuel S. Pilli "Towards MapReduce Based Classification approaches for Intrusion Detection".

International conference 2016 IEEE PP-361-366

[7] R . Chitrakar, and C. Huang, "Anomaly based Intrusion Detection using Hybrid Learning Approach of combining k-Medoids Clustering and Naïve Bayes Classification," In Wireless Communications, Networking and Mobile Computing (WiCOM), 8th International Conference on, pp. 1-5, IEEE, 2012.

[8] M. Dhakar, and A. Tiwari, " A novel data mining based hybrid intrusion detection framework," Journal of Information and Computing Science, vol 9, no. 1, pp. 037-048, 2014.

[9] W. Huai-bin, Y. Hong-liang, X. U. Zhi-Jian, and Y. Zheng, "A clustering algorithm use SOM and K- means in intrusiondetection," In E-Business and E- Government (ICEE), 2010 International Conference on, pp. 1281-1284, IEEE, 2010.

[10] S. Warnars, "Mining Patterns with Attribute Oriented Induction," In Proceeding of The International Conference on Database, Data Warehouse, Data Mining and Big Data (DDDMBD2015), pp. 11-21, 2015.

[11] V. Kachitvichyanukul, "Comparison of three evolutionary algorithms: GA, PSO, and DE," Industrial Engineering and Management Systems,11(3), pp. 215-223, 2012.

[12] R. Chitrakar, and C. Huang, "Anomaly based Intrusion Detection using Hybrid Learning Approach of

combining k-MedoidsClustering and Naïve Bayes Classification," In Wireless Communications, Networking and Mobile Computing (WiCOM),8th International Conference on, pp. 1-5, IEEE, 2012.

[13] Shi, X., Manduchi, R., 2003. A study on Bayes feature fusion for image classification. In: Conference on Computer Vision and Pattern Recognition Workshop, CVPRW, Madison, Wisconsin, USA, pp. 95–95.

[14] http://www.kdd.ics.uci.edu/databases/kddcup99/task.

html 7

[15] Nassar M, al Bouna B, Malluhi Q (2013) Secure

outsourcing of network flow data analysis. In: Big Data (BigData Congress), 2013 IEEE International Congress On. IEEE, Santa Clara, CA, USA. pp 431– 432.

[16] Kezunovic M, Xie L, Grijalva S (2013) The role of big data in improving power system operation and protection. In: Bulk Power System Dynamics and Control - IX Optimization, Security and Control of the Emerging Power Grid (IREP), 2013 IREP Symposium. IEEE, Rethymno, Greece. pp 1–9

[17] Denning DE (1987) An intrusion-detection model. Softw Eng IEEE Trans SE-13(2):222–232. doi:10.1109/TSE.1987.232894

[18] Suthaharan S, Panchagnula T (2012) Relevance feature selection with data cleaning for intrusion detection system. In: Southeastcon, 2012 Proceedings of IEEE. IEEE, Orlando, FL, USA. pp 1–6

[19] Marcelo D. Holtz, Bernardo M. David and Rafael Timeote "Building Scalable Distribute Intrusion Detection System Based on the Map Reduce Framework. 2011, Intrenation journal of Revista Telecommucation pp 23-31

[20] Lidong Wang*, Randy Jones "Big Data Analytics for Network IntrusionDetection: A Survey. International Journal of Networks and Communications 2017, 7(1): 24-31 DOI: 10.5923/j.ijnc.20170701.03

[21] Jingwei Huang, Zbigniew Kalbarczyk, and David M. Nicol. "Knowledge Discovery from Big Data for Intrusion Detection Using LDA. 2014 IEEE International Congress on Big Dat pp760-762