2021 International Conference on Emerging Smart Computing and Informatics (ESCI)
AISSMS Institute of Information Technology, Pune, India. Mar 5-7, 2021

MLIDS: A Machine Learning Approach for Intrusion Detection for Real Time Network Dataset

Monika D.Rokade
Research Scholar, Department of Computer Science and Engineering
Shri J.J.T. University,
Rajasthan,, India
monikarokade4@gmail.com

Yogesh Kumar Sharma
Associate Professor (HOD/Research-Coordinator),
Department of Computer Science and Engineering ,
India
dr.sharmayogeshkumar@gmail.com

2021 International Conference on Emerging Smart Computing and Informatics (ESCI) | 978-1-7281-8519-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ESCI50559.2021.9396829

Abstract- **Computer network and virtual machine security is very essential in today's era. Various architectures have been proposed for network security or prevent malicious access of internal or external users. Various existing systems have already developed to detect malicious activity on victim machines; sometimes any external user creates some malicious behavior and gets unauthorized access of victim machines to such a behavior system considered as malicious activities or Intruder. Numerous machine learning and soft computing techniques design to detect the activities in real-time network log audit data. KKDDCUP99 and NLSKDD most utilized data set to detect the Intruder on benchmark data set. In this paper, we proposed the identification of intruders using machine learning algorithms. Two different techniques have been proposed like a signature with detection and anomaly-based detection. In the experimental analysis, demonstrates SVM, Naïve Bayes and ANN algorithm with various data sets and demonstrate system performance on the real-time network environment.**

Keywords- Intrusion Detection System, Network security, Naïve Bayes, SVM, Artificial Neural Network, KDDCUP99.

I. INTRODUCTION

The IDS is exclusively for detecting one type of attack, e.g. Sample or unknown attack, DoS attack or U2R attack or R2L attack. It then sequentially deploys a number of such subsystems, one by one. This serves a two-fold purpose: first, only a limited number of features that detect a particular type of attack can be trained in each sub-phase. Second, the sub-size device remains tiny and therefore functional. Similar to our scheme, a common downside is that it raises overhead communication between modules. However by having every sub-phase completely independent of each other layer, this can be easily avoided in our method. As a consequence, in more than one sub-phase, such characteristics may happen. Any sub-phase will simply block an attack if detected without a central decision maker, depending on the security policy of the network. As long as they are established during a certain layer, various sub phases primarily serve as filters blocking abnormal association, thus providing rapid reaction to the intrusion as well as reducing the analysis in subsequent stages.

The method defines the system for creating SVM rules in this research work, based on its role selection procedure, which works on both HIDS and NIDS. A genetic algorithm is an optimization algorithm used to find the best solution. For all types of master class sub-attacks, the ensemble approach with different classification algorithms may provide the best detection of NIDS. Our aim of the proposed study is to create clear rules and increase the detection rates of DOS, PROBE, U2R and R2L for NIDS and HIDS.

II. LITERATURE SURVEY

Bhosale, Karuna S. et al. [1] deep neural network (DNN), Creating scalable and efficient IDSs to recognize and recognize unintentional and unexpected cyber-attacks is studied as a form of the deep learning system. The constant change in network operation and the rapid creation of attacks involve reviewing frequent occurrences that are generated over the years through dynamic and static techniques. Such a study facilitates the identification of the correct algorithm that can work effectively to forecast future cyber-attacks. A thorough evaluation of the DNN experiments and other powerful machine learning classifiers is shown on numerous public information test malware databases. The optimized network parameters and modulation schemes for DNNs are selected using KDDCup 99 dataset number of hidden layers use around by the system.

Chamou et al. [2], It is because of that explanation that the science community has grown accustomed to the complexity and enhancement of the efficiency of intrusion detection systems that a large number of companies all over the world are being targeted and threatened by the constant emergence of new and emerging threats. This is a groundbreaking tool for evaluating suspicious activity in DDoS and malware cyber threats using deep learning models. According to most Internet users, cybersecurity efficiency, data protection, and safe communication are considered important due to the rapid growth of web applications and their use. Simultaneously, increased exposure to more sophisticated security threats has been ascertained over the computers and internet networks, in the digital world of academia and industry, especially in small and medium-sized enterprises (SMEs), with economic implications.

According to [3], a system has been developed to accurately detect potential attacks by using various decision-free, random forest, and KNN. A new technique is suggested to fix the shortcomings of the previous approach that could not detect IPV6 attacks. In detecting IPV4-based attacks, the established framework produces an impressive and efficient result, considering the possible reach—the efficiency of the different algorithms that have been measured. Detection consistency, accuracy, and recall percentage were measured.

According to[4], identifying a novel phenomenon called NEC, clustering, and KDD can be used effectively. An unsupervised anomaly is used to produce high detection rates and fewer false passive rates. It is an effective way to solve the issue and find an anomaly that does not involve collecting labeled data. Use the 2009 NSL-KDD dataset to test the system. The preprocessing model converts all characteristics into the actual number, and the standardized dataset will compare the predicate outcome with an accurate result at the end of the evaluation section.

A data mining and machine learning survey for cybersecurity intrusion detection is conducted to ensure cybersecurity concerning the CSID Data Mining and Machine Learning Survey[5]. To access packet header and net flow packet header are used to networks and kernel-level data for the intrusion detection process, the potential issue is that data mining and artificial intelligence cannot be achieved without actual databases, therefore, very time-consuming. The presence of various statistics and machine learning algorithms is discussed. The research paper contains a set of comparative criteria for the anomaly of methodologies for machine learning data mining — dependent intrusion detection helps discover, evaluate and recognize the unauthorized use, replication, alteration, and destruction of the information system.

Based on the [6] detection and monitoring of the intruder using machine learning, ranking, and Voronoi, security has improved. That is the size of the given dataset and the high precision of the detection. A series of data called ISOT was used with the processing delay in mind. The paper also uses network flow characteristics to predict botnet infiltration, provided the packet compliance feature, which helps in packet encryption.

According to ADS-B IDS[7], automated based surveillance-broadcast IDS techniques are proposed using ADS-B techniques. To enhance air traffic control performance, the HMAC data set is used. The approaches operate with minimum overhead. The future scope specifies that its distance from the corresponding one-time position is within the safe zone for the ADS-B location to be valid. GPS uses the cyber-physical environment validated by attack detection to determine radar specifications in aircraft signaling superior location precision, as ADS-B has emerged as an alternative to current radio. A protocol is suggested to swap the keys used for the HMAC algorithm safely. The ATC Centre initiates firm handshakes with ATCs that monitor another zone on the flight route to transfer the private key over public key infrastructure (PK1) networks.

According to [8], for a power device that supports data logs, the report stated that a combination IDS using data mining is created using safe process mining. The method is an integrated approach to designing the IDS prototype. Detection precision, which is up to 73%, has been one of the key advantages. But this approach is not adequate for collecting major issues such as data logs, although it is tricky. The framework leverages characteristic and specification-based IDS functionality. The data analysis methodology that consolidates audit logs to learn the typical route from different machine devices. In the automated system, and no need to manually analyse and code the sequence manually.

According to Al, Dr. Yogesh Kumar Sharma et.[10] defines a 6G Network Access system and Edge-Assisted Congestion Rule Mechanism using Software-Defined Networking To avoid buffering of traffic flow, the framework must take up large frequency data. The evaluation results of this proposed approach indicate that it may maximize the efficacy of the network. Another issue is the network's safety, as a centralized security framework is important for maintaining data reliability across the network. Protection is necessary as users transmit and connect so that the possibility of data malignancy or piracy is greater. Another form of slicing network virtualization to avoid sluggish wireless broadband is network optimization, and the broadband speed may not hinder individuals. The software suggests the System Slicing Edge Admission System, model.

III. Proposed system

The proposed research methodology carried out intrusion detection and prevention using machine learning techniques. Training, including packet choice for anomaly and remote monitoring, will be done by the packet environments characterized block. Then it will submit a function collection for a specific packet activity. If all is well, send it forward together. To detect individual attacks, misbehavior samples will be tested for feature selection for various qualities. There are two-phase in the proposed system; we have taken the network dataset for system training and testing purposes—components of the framework. Below, figure 1 demonstrates the entire execution of the system using defined algorithms. Various machine learning has been used to generate train modules as well as testing, respectively.

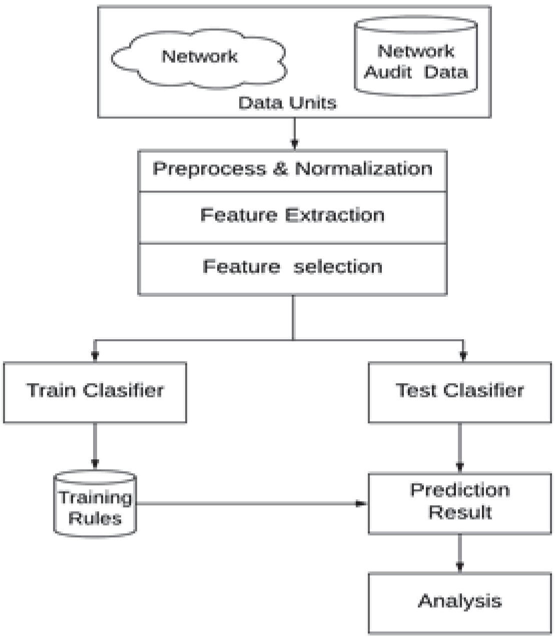

Fig. 1. System architecture

The NSL KDD CUP 1999 dataset is used for the experiments (KDD data set, 1999). The KDD CUP 1999 dataset is a version of the MIT Lincoln Laboratory's original 1998 DARPA intrusion detection assessment software, which is ready and controlled. Sampled awareness is stratified. It includes about 5 million association records as coaching knowledge and nearly 2 million association records as knowledge. In addition, the dataset contains a list of forty-one options extracted from each association and a mark defining the status of records of associations as a typical or particular form of attack. All kinds of continuous, discrete, and symbolic variables have these options, with wildly varying ranges falling into four categories: (1) the primary class consists of an association's intrinsic options, including the critical options for linking individual transmission control protocols. A variety of choices include the length of the affiliation, the type of protocol (TCP, UDP, etc.), and network access (HTTP, telnet, etc.) area unit. (2) Content choices within an association notified by the Domain Data Area Unit, such as the sum of failed login attempts, will not determine the first transmission control protocol packets' payload. (3) Constant host options analyze defined connections that have a continuous destination host within the last 2 seconds because of the current relationship and measure statistics related to protocol behavior, operation, etc. (4) Similar service options examine connections that have the same service as the present link within the last two seconds.

In the pre-processing point, the second block of Figure 1 shows that we use a packet sniffer designed with the winpcap library to extract network packet information from each packet, including the IP header, TCP header, UDP header, and ICMP header. After that, by considering links between each combination of science addresses (source science and destination IP), the packet data is partitioned and shaped into a record by aggregating each information. Every record consists of thought-about options details since musical notation options reflect the most features of network data and activities. In order to identify critical opportunities that outline the signatures of conventional vs. attack network traffic, we prefer to conduct extensive experiments. We like to use data acquisition to select thirty-five essential choices for our IDS strategy, according to the data type, the options, and the price of data benefit. The cost of each feature's data benefit reflects the relation of the component to the output group. Knowledge parameters obtain square measure X and Y, where X specifies individual choices such as the range of protocol packets for transmission control, the content of supply ports for transmission control protocol, and Y defines category teams that squarely measure conventional knowledge, probe attack, and DoS attack. Nevertheless, every aspect of the gift is essential to the DoS attack and the Probe attack. The data gain results show that we need to consider all 35 network data features for intrusion detection and classification.

IV. RESULTS AND DISCUSSION

After the successful implementation of the system, we calculate the confusion matrix for the system. Table 1 and Table 2 show the classification with SVM algorithms. Figure 2 demonstrates the classification performance of data collection by KDDCUP using the density-based approach of the machine learning algorithm program Figure 3 used to classify and predict the precision of the system using different methods like the RNN algorithm**.**

TABLE I. CONFUSION MATRIX CALCULATION USING SVM FOR CLASSIFICATION

Class	**Normal**	**Attack**
Normal	1760	19
Attack	9	1640
	1769	1659

TABLE II. CONFUSION MATRIX CALCULATION USING NB FOR CLASSIFICATION

Class	Normal	Attack
Normal	1830	227
Attack	169	1202
	1999	1429

TABLE III. PERFORMANCE EVALUATION WITH NB AND SVM

	NB	**SVM**
Accuracy	0.9892	0.9525
Precision	0.9867	0.9797
Recall	0.9933	0.9463
F-Score	0.9899	0.9529

According to both experiment analyses, SVM shows better classification accuracy than the NB algorithm, shown in figure 3. Conferring to the above experiment analysis, we can conclude the system produces better accuracy for trust computation in IoT in-service environment. The entire research follows some simulation environmental parameters as well as a combination of machine learning algorithms. Various computation parameters have been used cluster differentiation and id.mi.com using respect to machine learning algorithms.

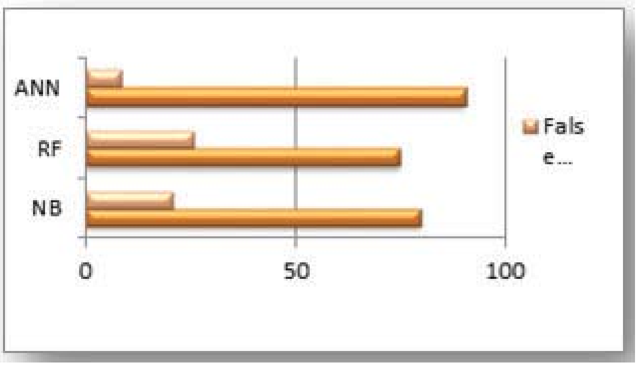

Fig. 2. Detection accuracy for KDD : CUP99 dataset using machine learning

The above figure 2 Shows accuracy of kddCup 99 results classification, with five different classes. Average software output is around the algorithm for the machine learning 88.50% for all classes.

A. Proposed Result

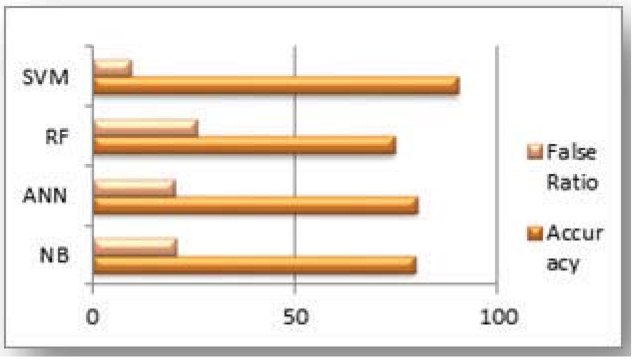

Fig. 3. Detection accuracy various network dataset using propsoed (SVM) vs exisitng

The above figure 3 Shows average efficiency of identification in various databases, of (n) different classes. The system's mean performance with the machine learning algorithm is around 95% for all (n) classes

V. CONCLUSION

This study proposed an SVM-IDS approach based on deep learning to suggest an efficient system of IDs. To test anomaly detection accuracy, we used the synthetic-based intrusion dataset - NSL-KDD. We plan to implement IDS into the cloud environment in the future using the deep learning technique. We also analyze and compare various methods of deep learning, namely. To detect intrusions in the network, NB ANN, RF, and SVM on the NSL-KDD dataset; The program essentially acts as artificial intelligence and conditioning algorithm to determine the unknown instances during the data check. The efficient rule structure makes for improved classification and high-class detection. Various experiments used experimental analysis to assess the algorithm's efficiency using several tests and conclude that we are achieving satisfactory results.

REFERENCES

[1] Bhosale, Karuna S., Maria Nenova, and Georgi Iliev. "Modified Naive Bayes Intrusion Detection System (MNBIDS)." 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS). IEEE, 2018.

[2] Chamou, Dimitra, et al. "Intrusion Detection System Based on Network Traffic Using Deep Neural Networks." 2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD). IEEE, 2019.

[3] Mohammed Anbar, Rosni Abdulah, Izan H. Hasbullah, Yung- Wey Chong; Omar E. Elejla, “Comparative Performance Analysis of classification algorithm for Internal Intrusion Detection “, 2016 14th Annual Conference on Privacy Security and Trust (PCT), Dec 12-14,2016, Penang, Malaysia.

[4] Weiwei Chen, Fangang Kong, Feng Mei, GuiginYuan, Bo Li, “a novel unsupervised Anamoly detection Approach for Intrusion Detection System”, 2017 IEEE 3rd International Conference on big data security on cloud, May 16-18,2017, Zhejiang, China.

[5] Anna L. Buczak, Erha n Guven, "A Survey of Data Mining and Machine Learning methods for cybersecurity intrusion detection", IEEE communication surveys and tutorials, vol. 18, Issue 2,2016.

[6] Manoj s. Koli, Manik K. Chavan, “An Advanced method for detection of botnet traffic using Interhnal Intrusion Detection”, 2017 International Conference on (ICICCT), March 10-11, 2017, Sangli, India.

[7] Thabet Kacem, Duminda Wijesekera, Paulo Costa, Alexander Barreto, “An ADS-B Intrusion Detection System”, 2016 IEEE on ISPA, 2016, Fairfax, Virginia.

[8] Shengyi Pan, Thomas Morris, Uttam Adhikari, “Developing a Hybrid Intrusion Detection System using Data Mining for power system”, IEEE Transactions on, vol. 6, issues. 6, Nov. 2015.

[9] Mehdi Ezzarii, Hamid Elghazi, Hassan El Ghazi, Tayeb Sadiki, "Epigenetic Algorithm for performing Intrusion Detection System", 2016 International Conference on ACOSIS, Oct17- 19,2016, Rabat, Morocco.

[10] BOROLE, Prajakta; SHARMA, Yogesh Kumar; NEMADE, Santosh. 6G Network Access and Edge-Assisted Congestion Rule Mechanism using Software-Defined Networking. International Journal of Future Generation Communication and Networking, 2020, 13.1s: 107-112.