IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, May 2022
Impact Factor: 6.252

Intrusion Detection for Real-Time Network Dataset Using Machine Learning: A Review Approach

Supriya S. Dahake[1], Dr. Monika D. Rokade[2]
ME Research Scholar, Department of Computer Engineering[1]
Research Guide & Assistant Professor, Department of Computer Engineering [2]
Sharadchandra Pawar College of Engineering, Otur, Pune, India

Abstract: *The current paper is related to the application of the machine learning approach for intrusion detection for real-time datasets. This is a review article. Some important papers are chosen to identify research gaps for intrusion detection for real-time datasets. In today's world, computer network and virtual machine security is critical. For network security or to restrict unwanted access by internal or external users, many designs have been proposed. Several techniques have previously been created to identify malicious activities on target computers. When an external user generates harmful behavior and gains illegal access to target machines, this is referred to as malicious actions or intruder behavior. A variety of machine learning and soft computing approaches have been developed to detect activities in real-time network log audit data. The most commonly used data sets to identify the Intruder on benchmark data sets are KKDDCUP99 and NLSKDD. In this study, we suggested employing machine learning methods to identify intrusions. Two distinct strategies, signature detection, and anomaly detection have been proposed. In the experimental study, SVM, Nave Bayes, and ANN algorithms are demonstrated with diverse data sets, and system performance in a real-time network environment is demonstrated.*

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

I. INTRODUCTION

The IDS is only capable of identifying one form of attacks, such as a sample or unknown assault, a DoS attack, a U2R attack, or an R2L attack. It then deploys a series of similar subsystems one at a time. This fulfills two functions: First, each sub-phase can only train a limited number of features that detect a certain sort of assault. Second, the sub-size gadget stays small and functional. A typical disadvantage of our method is that it increases the overhead communication between modules. This may be readily avoided in our system by having each sub-phase totally independent of the other layers. As a result, such traits may occur in more than one sub-phase. Depending on the network's security policy, any sub-phase will simply stop an attack if it is discovered without a central decision-maker. As long as they are created during a specific layer, various sub-phases primarily operate as filters inhibiting inappropriate association, giving immediate responsiveness to the incursion. Based on its role selection approach, the method describes the system for constructing SVM rules in this study effort, which operates on both HIDS and NIDS. A genetic algorithm is a solution-finding optimization method. The ensemble technique with several classification algorithms may give the best NIDS detection for all forms of master class sub-attacks. The planned study's goal is to provide clear guidelines and boost DOS, PROBE, U2R, and R2L detection rates for NIDS and HIDS.

II. LITERATURE SURVEY

2.1 Types of Attacks in Cloud Computing

The attacks that can be conducted must be discovered and understood in order to defend the Dataset from such threats and avoid any harm. The following are the most often mentioned attacks in intrusion detection systems: [3]

(a) Denial of Service (DoS) attack: The main motive of the attacker in this type of attack is to shut down a machine or network, making it inaccessible to its intended users.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, May 2022
Impact Factor: 6.252

(b) User to Root (U2R) Attack: In this approach, the attacker gains access to a normal user account first, then exploits system flaws to obtain access to the root account.
(c) Remote to Local (R2L):This exploit is commonly known to be carried out by an attacker in order to get unauthorized access to a victim's machine throughout the network.
(d) Probe: By seeing the physical silicon implementation of a device, probing attacks are an intrusive way for overcoming security mechanisms.

2.2 Application of Machine Learning Methods of Attack Detection for Intrusion Detection on Real-Time Dataset

In the publications we gathered, we noticed that ML is used in a number of approaches for identifying Dataset attacks. One of the most often used approaches is traditional detection, which identifies and alerts users when an assault occurs. Another approach of detection is to assess the Dataset security for any holes or flaws prior to the attack.

Table 1: List of Researchers Work on Similar Topics

Researchers	Year	Work Done by Researcher
1. Rokade, M. D., & Sharma, Y. K. [1]	2020	This work presented a deep learning-based SVM-IDS technique for recommending an efficient ID system. The synthetic-based intrusion dataset NSL-KDD was used to assess anomaly detection accuracy.
2. L. Alhenaki, A. Alwatban, B. Alamri, and N. Alari[5]	2019	The benchmarking dataset NSL-KDD is utilized for training and testing in this work. In order to improve and optimize the algorithm's performance for accurate prediction, as well as to support a smooth training process with minimal time and resources, feature normalization, feature selection, and data pretreatment techniques are utilized.
3. Bhosale, K. S., Nenova, M., & Iliev, G. [3]	2018	The Modified Nave Bayes Intrusion Detection System (MNBIDS) was shown in this study to improve the detection of DDoS attacks. For this, we use the Real Time Captured Network Packets dataset and the KDD Cup 99 dataset.
4. Ezzarii, M., Elghazi, H., El Ghazi, H., & Sadiki, T. [4]	2016	We looked into biological inspiration and tried to apply it to the intrusion detection industry, looking for epigenetic characteristics that would lead to a successful solution and efficient security.
5. Hoque, M. S., Mukit, M., Bikas, M., & Naser, A. [5]	2012	This paper describes and constructs an Intrusion Detection System that uses a genetic algorithm to identify various forms of network intrusions efficiently. He utilized the standard KDD99 benchmark dataset to develop and test the performance of our system, and He were able to get a decent detection rate. He utilized the standard deviation equation with distance to determine the fitness of a chromosome.
6. Kanimozhi, V., & Jacob, T. P. [6]	2019	This article examines the artificial intelligence framework utilized in this study. The Scikit learn framework optimization is based on the CPU (Central Processing Unit).
7. Kok, S. H., Abdullah, A., Jhanjhi, N. Z., & Supramaniam, M. [7]	2019	This study focuses on KDDCup99 and its variants. Although they are over 20 years old, the NSL-KDD datasets are the two most extensively utilized datasets. While intrusion threats continue to change in tandem with new technology and user behaviors, this tendency might result in stagnation in IDS.
8. Kumar, A., Maurya, H. C., & Misra, R. [8]	2013	The goal of this study is to integrate two classification approaches. As a result, a growing variety of methods for achieving this goal have been developed, including KNN classification, Nave Bayes classification, support vector machines (SVM), decision tree (DT), neural network (NN), and maximum entropy.
9. Kumar, G., Kumar,		This research looked at several intrusion detection systems (IDS) and how

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, May 2022
Impact Factor: 6.252

K., & Sachdeva, M. [9]	2010	they were classified based on different modules. A detailed overview of several AI-based intrusion detection (ID) algorithms is offered. The use of a multi-classifier-based strategy (hybrid/ensemble approach) to identify known and new threats with high accuracy is presented.
10. Liu, H., & Lang, B. [10]	2019	The study first presents an IDS taxonomy that takes data sources as the major thread. We then study and discuss IDSs applied to various data sources, such as logs, packets, flow, and sessions, using this taxonomy. Because IDSs are designed to identify attacks, it's critical to choose the right data source based on the attack's characteristics.
11. Mohit, T., Raj, K., Akash, B., & Jai, K. [11]	2017	This work presents IDS are becoming the main part for many organizations after deploying firewall technology at the network perimeter. IDS can offer protection from external users and internal attackers, where traffic doesn't go past the firewall at all.
12. Panda, M., Abraham, A., Das, S., & Patra, M. R.[12]	2011	The results of this study's empirical analysis show that our suggested technique, which employs a variety of machine learning approaches and the NSL-KDD dataset, performs admirably in all categories of majority attacks. However, owing of the substantial bias present in the sample, a low detection rate for U2R and R2L minority assaults is unavoidable.
13. Repalle, S. A., & Kolluru, V. R. [13]	2017	This thesis research provided an overview of machine learning methods and demonstrated their use in an intrusion detection system. K-Nearest Neighbors was shown to be the most effective.

III. RESULTS AND DISCUSSIONS

The primary goal of this research is to calculate the confusion matrix for the system. The classification with SVM techniques is shown in Tables 2 and 3. Figure 2 shows the classification performance of data collected by KDDCUP using the density-based technique of the machine learning algorithm programmer Figure 3 shows how multiple approaches, such as the RNN algorithm, were used to classify and forecast the precision of the system.

Table 2. Confusion Matrix Calculation Using SVM for Classification

Class	Normal	Attack
Normal	1760	19
Attack	9	1640
	1769	1659

Table 3. Confusion Matrix Calculation Using NB for Classification

Class	Normal	Attack
Normal	1830	227
Attack	169	1202
	1999	1429

Table 4. Evaluation of Performance using NB and SVM

Class	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 outperforms the NB method in terms of classification accuracy (see Figure 3). According to the results of the fore mentioned experiment, the system generates improved accuracy for trust calculation in the IoT in-service context. The entire study is guided by a set of simulated environmental settings and a mix

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, May 2022
Impact Factor: 6.252

of machine learning methods. With regards to machine learning techniques, many calculation parameters have been employed cluster differentiation and id.mi.com.

Figure 1: Machine learning detection accuracy for the KDD: CUP99 dataset.

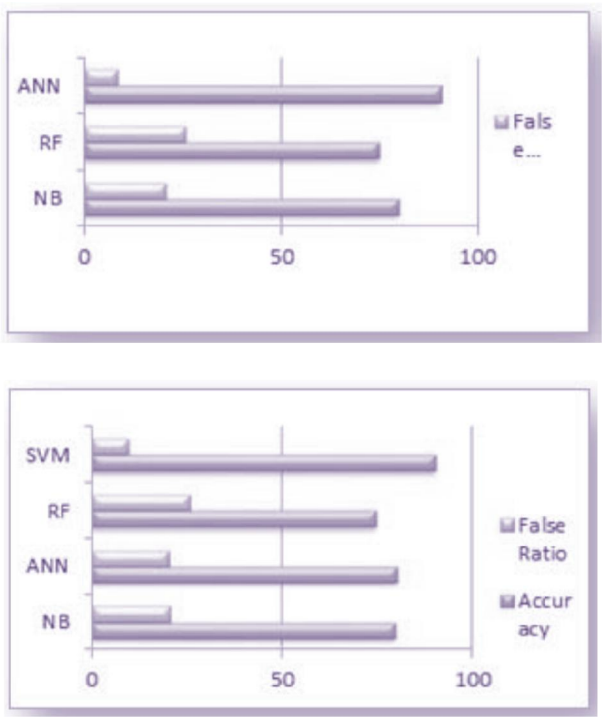

Figure 2: Detection accuracy of diverse network datasets using proposed (SVM) vs. existing datasets

The average effectiveness of identification in several databases, of (n) distinct classes, is shown in Figure 2. For all (n) classes, the system's average performance with the machine learning method is about 95%.

IV. CONCLUSION

This work presented a deep learning-based SVM-IDS technique for recommending an efficient ID system. The synthetic-based intrusion dataset NSL-KDD was used to assess anomaly detection accuracy. We want to use deep learning to build IDS in the cloud environment in the future. We also examine and compare several deep learning approaches, such as.The software mainly functions as artificial intelligence and conditioning algorithm to find the unknown occurrences during the data check by using NB ANN, RF, and SVM on the NSL-KDD dataset to detect network intrusions. Improved classification and high-class detection are possible because to the efficient rule structure. Several trials employed experimental analysis to evaluate the algorithm's effectiveness using a variety of tests, resulting in the conclusion that we are obtaining adequate results.

ACKNOWLEDGMENT

The main author is wishing thank to the research guide Assistant Professor, Dr. Monika Rokade from Sharad Chandra Pawar College of Engineering, Otur for her cooperation and technical help during the survey phase of my research work.

REFERENCES

[1]. Rokade, M. D., & Sharma, Y. K. (2021, March). MLIDS: A Machine Learning Approach for Intrusion Detection for Real Time Network Dataset. In 2021 International Conference on Emerging Smart Computing and Informatics (ESCI) (pp. 533-536). IEEE."

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, May 2022
Impact Factor: 6.252

[2]. Alzahrani, A. O., & Alenazi, M. J. (2021). Designing a network intrusion detection system based on machine learning for software defined networks. Future Internet, 13(5), 111.

[3]. Bhosale, K. S., Nenova, M., & Iliev, G. (2018, December). Modified naive bayes intrusion detection system (mnbids). In 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS) (pp. 291-296). IEEE.

[4]. Ezzarii, M., Elghazi, H., El Ghazi, H., & Sadiki, T. (2016, October). Epigenetic algorithm for performing intrusion detection system. In 2016 International Conference on Advanced Communication Systems and Information Security (ACOSIS) (pp. 1-6). IEEE.

Hoque, M. S., Mukit, M., Bikas, M., & Naser, A. (2012). An implementation of intrusion detection system using genetic algorithm. arXiv preprint arXiv:1204.1336.

[5]. Kanimozhi, V., & Jacob, T. P. (2019, April). Artificial intelligence based network intrusion detection with hyper-parameter optimization tuning on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing. In 2019 international conference on communication and signal processing (ICCSP) (pp. 0033-0036). IEEE

[6]. Kok, S. H., Abdullah, A., Jhanjhi, N. Z., & Supramaniam, M. (2019). A review of intrusion detection system using machine learning approach. International Journal of Engineering Research and Technology, 12(1), 8-15.

[7]. Kumar, A., Maurya, H. C., & Misra, R. (2013). A research paper on hybrid intrusion detection system. International Journal of Engineering and Advanced Technology (IJEAT) Vol, 2.

[8]. Kumar, G., Kumar, K., & Sachdeva, M. (2010). The use of artificial intelligence based techniques for intrusion detection: a review. Artificial Intelligence Review, 34(4), 369-387.

[9]. Liu, H., & Lang, B. (2019). Machine learning and deep learning methods for intrusion detection systems: A survey. applied sciences, 9(20), 4396.

[10]. Mohit, T., Raj, K., Akash, B., & Jai, K. (2017). Intrusion detection system. International Journal of Technology Research and Applications, 5(2), 2320-8163.

[11]. Panda, M., Abraham, A., Das, S., & Patra, M. R. (2011). Network intrusion detection system: A machine learning approach. Intelligent Decision Technologies, 5(4), 347-356.

[12]. Repalle, S. A., & Kolluru, V. R. (2017). Intrusion detection system using ai and machine learning algorithm. International Research Journal of Engineering and Technology (IRJET), 4(12), 1709-1715.