iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** II **Month of publication:** February 2024

DOI: https://doi.org/10.22214/ijraset.2024.58620

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

AI-Based Network Intrusion Detection System

Prof. Ms. S. P. Shinde[1], Dipali Chormale[2], Yash Bhise[3], Aditya Raj[4], Noanika Agarkar[5]
Department of Information Technology, Smt. Kashibai Navale College of Engineering (SPPU- Pune)

Abstract*: This report discusses the research done on the chosen topic, which is Developing an AI-based Network Intrusion Detection System using ML and DL algorithms. Recently we have seen so much progress in Internet and communication technologies it is not just connecting computer networks and people but it is also connecting devices involving Big Data. It has so many benefits in each field which are crucial in today's world like education, health, digital transactions, traveling, and anything we can think of. With so many benefits it comes with its negative effects like cyber-attacks which can happen to anybody who is connected to the Internet. So, Networks and Security become very desirable areas of research and work. To be saved from the attacks we have to detect the cyber-attack and stop the intruder from causing harm to our system. We use IDS (Intrusion Detection System), so we can identify incoming attacks. A Network Intrusion detection system provides security by constantly monitoring the network traffic for suspicious behavior. For building an Intrusion Detection System we need a good dataset with a huge amount of data which is of good quality and can be used for training the System so it can predict the output more accurately. In this paper we used NSL- KDD [6] data set a refined version of the KDD'99 dataset. We developed many models using machine learning and deep learning and compared them for detecting intrusion in networks. We used supervised machine learning and deep learning techniques to train and build many classification models that can differentiate between attacking traffic and normal traffic. We compared the accuracy of every model of different datasets so we can find the model that is performing best for network intrusion detection. After performing all the research and comparison we found that a fully connected Deep Learning model is giving better performance than a machine learning model. We used Autoencoder for feature selection for the best-performing Deep Learning Classification Model.

Keywords: IDS, Machine Learning, Deep Learning, Auto-Encoder, Network Security.

I. INTRODUCTION

Almost all people use the Internet to carry out essential activities such as bill payments, bank transfers, etc. But attacks on home networks are not uncommon nowadays, as everybody is connected through the internet, and the attacks have been growing more frequent and severe.

When an attack does occur, a comprehensive and organized analysis must be conducted to verify the causes of the attack and the damage caused by the attack. A comprehensive and fast analysis and reaction can help to reduce network downtime and keep essential business systems operational. The level of connectivity worldwide has provided opportunities for cybercriminals who earn a profession by getting into networks, as well as amateur hackers who have too much time on their hands. The determined hacker can find a way into your network either by establishing some type of connection and entering your virtual "front door" or by using social engineering tactics to obtain user ID and password information. Whatever technique is utilized, the fact is that an intruder can get into your network and cause damage to your organization.

A. Detection Method of IDS

1) *Signature-based Method:* Signature-based Approach With signature-based intrusion detection systems, attacks are identified based on predefined patterns in network traffic, such as the number of bytes, 1s, or 0s. Additionally, it detects malware based on the previously identified malicious instruction sequence that it employs. Signatures are the patterns that the IDS has discovered. Signature-based intrusion detection systems (IDS) find it easy to identify attacks whose pattern (signature) is already present in the system, but they have a hard time identifying new malware attacks because their pattern (signature) is not known.
2) *Anomaly-based Method:* Since malware is developing quickly, anomaly-based intrusion detection systems (IDS) were created to identify attacks involving unknown malware. A trustworthy activity model is created using machine learning in anomaly-based intrusion detection systems. Any new information is compared to this model and deemed suspicious if it does not match the model. With the ability to train models based on hardware configurations and applications, machine learning-based IDS has a more generalized property than signature-based IDS.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

B. Classification of Intrusion Detection System:

The classification of an Intrusion Detection System is as follows:

1) *Network Intrusion Detection System (NIDS):* Network intrusion detection systems (NIDS) are set up at a planned point within the network to research traffic from all devices on the network. It performs an observation of passing traffic on all subnets and matches the traffic that is passed on the subnets to the collection of known attacks. Once an attack is recognized or abnormal behavior is observed, an alert can be sent to the administrator. An example of an NIDS is installing it on the subnet where firewalls are located to see if someone is trying to crack the firewall.
2) *Host Intrusion Detection Systems (HIDS):* Host intrusion detection systems (HIDS) run on self-reliant hosts or devices on the network. A HIDS monitors or observes the incoming and outgoing packets from the device only and will alert the administrator if suspicious or malicious activity is detected. It takes a snapshot of existing system files and equates it with the previous snapshot. If the analytical system file 1 is edited or deleted, an alert is sent to the administrator to investigate. An example of HIDS usage can be seen on mission-critical machines, which are not expected to alter their layout.
3) *Protocol-based Intrusion Detection System (PIDS):* A protocol-based intrusion detection system (PIDS) comprises a system or agent that consistently resides at the front end of a server, controlling and interpreting the protocol between a user or device and the server. It is trying to secure the web server by regularly monitoring the HTTPS protocol stream and accepting the related HTTP protocol. As HTTPS is unencrypted, before instantly entering its web presentation layer, this system would need to reside in this interface to use HTTPS.
4) *Application Protocol-based Intrusion Detection Systems (APIDS):* Application Protocol-based Intrusion Detection Systems are systems or agents that commonly reside within a group of servers. It identifies the intrusions by observing and interpreting the communication on application-specific protocols. For example, this would observe the SQL protocol explicitly in the middleware as it transacts with the database on the web server.
5) *Hybrid Intrusion Detection Systems:* A hybrid intrusion detection system is created by combining two or more intrusion detection system methodologies. The host agent or system data is combined with network data in the hybrid intrusion detection system to provide a comprehensive picture of the network system. When compared to other intrusion detection systems, the hybrid system exhibits higher efficacy.

C. Purpose

In this research work, we wanted to compare various machine learning models with deep learning models and find which model gives the highest possible accuracy for Network Intrusion Detection Using ML and DL.

To attain a high level of threat visibility, organizations must ensure that intrusion detection technology is correctly installed and optimized.

D. Different classes of Attacks

1) *Denial of Service (DoS):* An attacker attempts to prevent authorized users from using a service. For example, SYN flood, smurf, and teardrop.
2) *User-to-Root (U2R):* An attacker has local access to the victim's computer and tries to gain super-user privilege. For example, buffer overflow attacks.
3) *Remote to Local (R2L):* An attacker tries to gain access to the victim's machine without having an account on it. For example, a password-guessing attack.
4) *Probe:* An attacker attempts to gain information about the intended host. For example, port-scan and ping-sweep.

E. Comparison of IDS with Firewalls

IDS and firewall are both related to network security, but an IDS differs from a firewall in that a firewall looks outwardly for intrusions to prevent them from occurring. Firewalls restrict access between networks to prevent intrusion, and if an attack comes from within the network, it is not detected. An intrusion detection system (IDS) characterizes a suspected intrusion after it has occurred and then signals it.

II. LITERATURE SURVEY

TABLE I. SUMMARY OF RELATED WORK/LITERATURE SURVEY

Title	Author	Year/Journal name	Summary
A Deep Learning Approach to Network Detection System	Nathan Shone, Tran Nguyen Ngoc, Vu Dinh Phai, and Qi Shi	IEEE-2018	This paper proposes a novel deep-learning model to enable NIDS operation within modern networks.
An Efficient Network Intrusion Detection and Classification System	Iftikhar Ahmad, Qazi Emad Ul Haq, Muhammad Imran, Madini O.Alassafi and Rayed A.AlGhamdi	ResearchGate-2022	In this paper, we learned the introduction of an AdaBoost-based network intrusion detection system utilizing feature selection and decision tree classification.
Intrusion Detection Systems using Supervised Machine Learning Techniques: A survey	Emad E. Abdallah, Wafa' Eleisah, Ahmed Fawzi Otoom	ScienceDirect-2022	In this paper, we investigate the subject of intrusion detection using supervised machine learning methods.
Network intrusion detection system: A systematic study of machine learning and deep learning approaches	Zeeshan Ahmad, Adnan Shahid Khan, Cheah Wai Shiang, Johari Abdullah, Farhan Ahmad	ScienceDirect-2022	In this paper, we discuss the challenges faced by traditional IDS in detecting novel attacks, explore the application
Network Intrusion Detection System using machine learning	Vinaya Bhalerao, Bhavan Shinde	JSRCSEIT-2022	In this paper, we came to know about all the Algorithm and comparison between them and we can select which algorithm we need to use.

III.METHODOLOGY

This study conducts a systematic literature review of the different ML- and DL-based NIDS and investigates the published journal articles between 2017 and to first quarter of 2020. A systematic literature review is a methodology followed to identify, examine, and extract needful information from the literature related to certain research topics.

A. AI Methods For NIDS

This section provides a general methodology of the AI-based NIDS along with the details of the most commonly used ML and DL algorithms used to design an efficient NIDS. Both ML and DL are broadly classified as supervised and unsupervised algorithms. In supervised algorithms, useful information is extracted from the labelled data. While unsupervised algorithms rely on unlabelled data to extract useful features and information.

1) A general AI-based NIDS Methodology

A NIDS developed using ML and DL methods usually involves the following three major steps, that is, (i) Data preprocessing phase, (ii) Training phase, and (iii) Testing phase. For all the proposed solutions, the dataset is first pre-processed to transform it into the format suitable to be used by the algorithm. This stage typically involves encoding and normalization. Sometimes, the dataset requires cleaning in terms of removing entries with missing data and duplicate entries, which is also performed during this phase. The pre-processed data is then divided randomly into two portions, the training dataset, and the testing dataset. Typically, the training dataset comprises almost 80% of the original dataset size and the remaining 20% forms the testing dataset. The ML or DL algorithm is then trained using the training dataset in the training phase. The time taken by the algorithm in learning depends upon the size of the dataset and the complexity of the proposed model. Normally, the training time for the DL models requires more training time due to its deep and complex structure. Once the model is trained, it is tested using the testing dataset and evaluated based on the predictions it made. In the case of NIDS models, the network traffic instance will be predicted to belong to either benign (normal) or attack class.

B. ML Algorithms

ML is a subset of AI that includes all the methods and algorithms that enable machines to learn automatically using mathematical models to extract useful information from large datasets. The most common ML (also called Shallow Learning) algorithms used for IDS are Decision Tree, K-Nearest Neighbour (KNN), Artificial Neural Network (ANN), Support Vector Machine (SVM), K-Mean Clustering, Fast Learning Networks, and Ensemble Methods.

1) Decision Tree

DT is one of the basic supervised ML algorithms that is used for both classification and regression of the given dataset by applying a series of decisions (rules). The model has a conventional tree structure with nodes, branches, and leaves. Each node represents an attribute or a feature. The branch represents a decision or a rule while each leaf represents a possible outcome or class label.57 The DT algorithm automatically selects the best features for building a tree and then performs pruning operation to remove irrelevant branches from the tree to avoid the over-fitting. The most common DT models are CART, C4.5, and ID3.58 Many advanced learning algorithms like Random Forest (RF)59 and XGBoost60 are made from multiple decision trees.

2) K-Nearest Neighbour

KNN is one of the simplest supervised ML algorithms that utilizes the idea of “feature similarity” to predict the class of a certain data sample. It identifies a sample based on its neighbours by calculating its distance from the neighbours. In the KNN algorithm, the parameter k affects the performance of the model. If the value of k is very small, the model may be susceptible to over-fitting. While a very large selection of k value may result in misclassification of the sample instance.61,62 Karatas et al63 compared the performance of different ML algorithms using an up-to-date benchmark dataset CSE-CIC-IDS2018. They addressed the dataset imbalance problem by reducing the imbalance ratio using Synthetic Minority Oversampling Technique (SMOTE),64 which resulted in detection rate improvement for minority class attacks.

3) Support Vector Machine

SVM is a supervised ML algorithm based on the idea of max-margin separation hyper-plane in n-dimensional feature space. It is used for the solution of both linear and nonlinear problems. For nonlinear problems, kernel functions are used. The idea is to first map a low-dimensional input vector into a high-dimensional feature space using the kernel function. Next, an optimal maximum marginal hyper-plane is obtained, which works as a decision boundary using the support vectors.65,66 For NIDS, the SVM algorithm can be used to enhance its efficiency and accuracy by correctly predicting the normal and malicious classes.

C. Deep Learning Algorithms

DL is the subset of the ML which includes many hidden layers to get the characteristics of the deep network. These techniques are more efficient than ML due to their deep structure and ability to learn the important features from the dataset on its own and generate an output. This section presents the DL approaches adopted to propose DL-based NIDS solutions in the reviewed articles.

1) Recurrent Neural Networks

Recurrent Neural Networks (RNN) extends the capabilities of the traditional feed-forward neural network and is designed to model the sequence data. RNN is made of input, hidden, and output units, where the hidden units are considered to be the memory elements. To make a decision, each RNN unit relies on its current input and the output of the previous input. RNN is widely used in different fields like speech processing, human activity recognition, handwriting prediction, and semantic understanding, to name a few. For an IDS, RNN can be used for the supervised classification and feature extraction. RNN normally can handle limited length sequences and will suffer from short-term memory if the sequence length is long. Different RNN variants like Long short-term memory (LSTM) and gated recurrent unit (GRU) are proposed to solve these issues. RNN-based IDS was proposed by Yin et al91 in the context of binary and multi class classification of the NSL-KDD dataset. The model was tested using a different number of hidden nodes and learning rates. Results showed that different learning rates and the number of hidden nodes affect the accuracy of the model. Best accuracy was obtained using 80 hidden nodes and a learning rate of 0.1 and 0.5 for binary and multi class scenarios. The proposed model performed well compared to ML algorithms and a reduced-sized RNN model proposed in Reference. The main shortcoming of this work is the increase in computational processing which results in high model training time and lower detection rate for the R2L and U2R classes. The article also lacks the performance comparison of the proposed model with different other DL methodologies.

In Reference 93, Xu et al proposed an IDS based on RNN using GRU as the main memory together with the multilayer perceptron and a softmax classifier. The proposed methodology was tested using KDD Cup'99 and NSL-KDD datasets. Experimental results showed good detection rates for comparing other methodologies. The major drawback of their model is lower detection rates for minority attack classes like U2R and R2L. Naseer et al94 performed a comparative analysis of IDS based on different DL and ML algorithms and implemented on a GPU-based testbed. NSL-KDD is considered as the benchmark dataset and the experimental results showed that LSTM and Deep CNN achieved higher accuracy results comparing other models.

2) AutoEncoder

AutoEncoder (AE) is a popular DL technique that belongs to the family of unsupervised neural networks.95 It works on the idea of matching the output as close to input as possible by learning the best features. It contains input and output layers of the same dimension, while the dimensions of the hidden layers are normally smaller than the input layer. AE is symmetric and works in Encoder-Decoder fashion. Different variants of AE are Stacked AE, Sparse AE, and Variational AE.96 Shone et al97 proposed an IDS based on deep AE and ML technique RF. To make the model efficient in terms of computational and time, only the encoder part of AE is utilized to make it work in a nonsymmetric fashion. Two non symmetric deep AEs, with three hidden layers each, are arranged in a stacked manner. RF was used for classification. Experiments were performed for multiclass classification scenarios using KDD Cup '99 and NSL-KDD datasets. The pro posed method showed their efficiency compare to Deep Belief Network (DBN) used in Reference 98 in terms of detection accuracy and reduced training time. But the model showed inefficiency for detecting R2L and U2R attacks due to lack of data for training the model. Yan et al67 proposed an IDS using stacked sparse autoencoder (SSAE) and SVM. The SSAE was used as the feature extraction method and SVM as a classifier. Binary-class and multi-class classification problem is considered for conducting experiments. The results showed the proposed model superiority in performance comparing different feature selection, ML, and DL methods using the NSL-KDD dataset. Although, the model achieves reasonable detection rates for U2R and R2L attacks but it is still less comparing the other classes of the dataset. A-Qatf et al99 also proposed a similar idea of self-taught learning based on sparse AE and SVM. To validate their performance, they performed experiments on the proposed model considering the NSL-KDD dataset. The results showed improved overall performance comparing other DL and ML models. But the proposed methodology performance in R2L and U2R class is not discussed. Papmartizivanous at all proposed an autonomous misuse detection system by combining the advantages of self-taught learning and MAPE-K frameworks. They used sparse AE for the unsupervised learning algorithm to learn useful features while performing the Plan activity within the MAPE-K Framework. Experiments performed using the KDD Cup'99 and NSL-KDD datasets. The main drawback is the lack of detection accuracy for U2R and R2L attack classes. Khan et al103 proposed an efficient two-stage model based on deep stacked AE. The initial stage classified the dataset into the attack and normal classes with probability values. These probability scores are then used as an additional feature and are input to the final decision stage for normal and multiclass attack classification. The performance of the proposed model was tested using KDD Cup'99 and UNSWNB15 datasets. To reduce the problems due to class imbalance of the datasets, a different methodology was adopted for both datasets. For KDD Cup'99, the downsampling was performed to remove repeated records. While, to balance the distribution of records in UNSWNB15, upsampling of the dataset was performed using SMOTE. This preprocessing of the dataset dramatically improves the DR efficiency of attack class with lower training instances. Malaiya et al104 proposed different IDS models based on fully connected networks, Variational AE, and Sequence-to-Sequence (Seq2Seq) structures, respectively. These models were examined for different datasets NSL-KDD, KyotoHoneypot, UNSW-NB15, IDS2017, and MAWILab traces.105 Results showed that the Seq2Seq model constructed using two RNNs performed the best comparing other models in terms of detection accuracy across all the datasets. Yang et al106 proposed a model for ID based on the supervised adversarial variational AE with regularization and DNN (SAVAER-DNN). The performance of the model was tested using benchmark data NSL-KDD and UNSW-NB15. Experimental results confirm the model's effectiveness in detecting low frequency and new attacks. Andresini et al107 incorporated the idea of AE to proposed a multistage model involving the ID convolution layer and two stacked fully connected layers. In the initial unsupervised stage, two AEs were trained separately using Normal and Attack flows to reconstruct the samples again. In the supervised stage, these new reconstructed samples are used to build a new augmented dataset that is used as input to a 1D-CNN. Then the output of this convolution layer is flattened and fed to fully connected layers, and lastly, a softmax layer classifies the dataset. Experiments were performed on the KDD Cup'99, UNSWNB15, and CICIDS2017 datasets and the proposed methodology achieves superior performance compar ing different DL models. They have not shown how the minority classes perform using this methodology. The second drawback is that it does not provide any information on the characteristics of the attack.

3) Deep Neural Network

DNN is a basic DL structure that allows the model to learn in multiple layers. It is composed of an input layer, an output layer, and many hidden layers. DNN is used to model complex nonlinear functions. Increased number of hidden layers enhances the abstraction level of the model to increase its capability.108 Jia et al109 proposed a network IDS based on DNN with four hidden layers to classify the datasets KDD cup'99 and NSL-KDD. The output layer included one fully connected layer and softmax classifier for classification purposes. For the hidden layer, a rectified linear unit was used as the activation function.110 Results showed the robustness of the proposed model as it achieved higher detection rates for almost all the attack classes except U2R due to presence of less number of records. According to the authors, increasing the number of nodes and layers leads to a complex structure that increases the computing time and consumes more resources. The solution to these issues is the optimization algorithm and automatic tuning. Wang et al111 studied the DNN-based IDS with adversaries and evaluated using the NSL-KDD dataset. They com prehensively studied the roles of individual features in generating adversarial examples. The adversarial samples were produced by FGSM,112 JSMA,113 DeepFool,114 and CW attacks.115 Results showed that the most commonly used attributes are more vulnerable to DL-based IDS and require more attention to safeguard the network from attacks. Vinayakumar et al116 proposed a hybrid scalable DNN framework called as scale-hybrid-IDS-AlertNet, for intrusion identification at both host and network level. Apache Spark cluster computing platform117 was used for implementing the scalable platform. For NIDS, the proposed model was tested using publically available datasets like KDDCup 99, NSL-KDD, Kyoto, UNSW-NB15, WSN-DS, and CICIDS 2017. Experiment results showed the superiority of the proposed model comparing different ML algorithms.

IV. CONCLUSIONS AND FUTURE WORK

This paper provides an extensive review of the network intrusion detection mechanisms based on the ML and DL methods. A systematic approach is adopted for the selection of the relevant articles in the field of AI-based NIDS. Firstly, the concept of IDS and its different classification schemes is elaborated extensively based on the reviewed articles. Then the methodology of each article is discussed and the strengths and weaknesses of each are highlighted in terms of the intrusion detection capability and complexity of the model. Based on this study, the recent trend reveals the usage of DL-based methodologies to improve the performance and effectiveness of NIDS in terms of detection accuracy and reduction in FAR. About 80% of the proposed solutions were based on the DL approaches with AE and DNN are the most frequently used algorithms. Although DL schemes have much superior performance than the ML-based methods in terms of their ability to learn features by itself and stronger model fitting abilities. But these schemes are quite complex and require extensive computing resources in terms of processing power and storage capabilities. These challenges need to be addressed to fulfill real-time requirements for NIDS and hence improves NIDS performance. The study also shows that 60% of the proposed methodologies were tested using KDD Cup'99 and NSL-KDD datasets mainly because of the availability of extensive results using these datasets. But these datasets are quite old to address modern network attacks, and hence limits the performance of the proposed methodologies in real-time environments. For AI-based NIDS methods, the model should be tested with the latest updated dataset like CSE-CIC-IDS2018 for better performance in terms of detection accuracy for intrusions. This article also highlights the research gaps in improving the model performance for low-frequency attacks in a real-world environment and to find efficient solutions to reduce complexity for the proposed models. Proposing an efficient NIDS framework using less complex DL algorithms and have an effective detection mechanism is a potential future scope of research in this area. For future research, we will use this knowledge to design a novel, lightweight, and efficient DL-based NIDS which will effectively detect the intruders within the network.

V. BENCHMARK DATSETS

This section provides detail about the popular datasets used by the researcher for testing the performance of their proposed methodology.

1) *KDD Cup'99:* It is one of the most popular and widely used dataset for IDS. It contains approximately five and two million records for training and testing respectively. Each record contains different features or attributes and is labeled as either normal or attack. The attacks are classified into four different types as Denial of Service (DoS), Probe, Remote to Local (R2L), and User to Root (U2R).
2) *Kyoto 2006+:* This dataset is created from the network traffic records, obtained by deploying honeypots, darknet sensors, email servers, web crawler, and other network security measures by Kyoto University. The most latest dataset includes the traffic record from 2006 to 2015. Each record has statistical features, of which are derived from KDD Cup'99 dataset while the remaining 10 are additional features.

3) *NSL-KDD:* This is the revised and refined version of the KDD Cup'99 dataset by removing several of its integral issues. This dataset is also a feature dataset with the attacks divided into four classes as discussed in KDD Cup'99.
4) *UNSW-NB15:* This dataset is created by the Australian Center for Cyber Security. It contains approximately two million records with a total of 49 features, that are extracted using Bro-IDS, Argus tools, and some newly developed algorithms. This dataset contains the types of attacks named as, Worms, Shellcode, Reconnaissance, Port Scans, Generic, Backdoor, DoS, Exploits, and Fuzzers.
5) *CIC-IDS2017:* This dataset is created by the Canadian Institute of Cyber Security (CIC) in 2017.133 It contains the normal flows and updated real-world attacks. The network traffic is analyzed by CICFlowMeter using the information based on timestamps, source, and destination IP addresses, protocols, and attacks.136 Moreover, CICIDS2017 includes common attack scenarios like Brute Force Attack, HeartBleed Attack, Botnet, Denial of Service (DoS) Attack, Distributed DoS (DDoS) Attack, Web Attack, and Infiltration Attack.
6) *CSE-CIC-IDS2018:* This dataset is jointly created by Communications Security Establishment (CSE) and CIC in 2018.63 The user profiles containing the abstract representation of the different events is created. For the generation of the dataset, all these profiles are combined with a unique set of features. It includes seven different attack scenarios: Brute-force, Heartbleed, Botnet, DoS, DDoS, Web attacks, and infiltration of the network from inside.

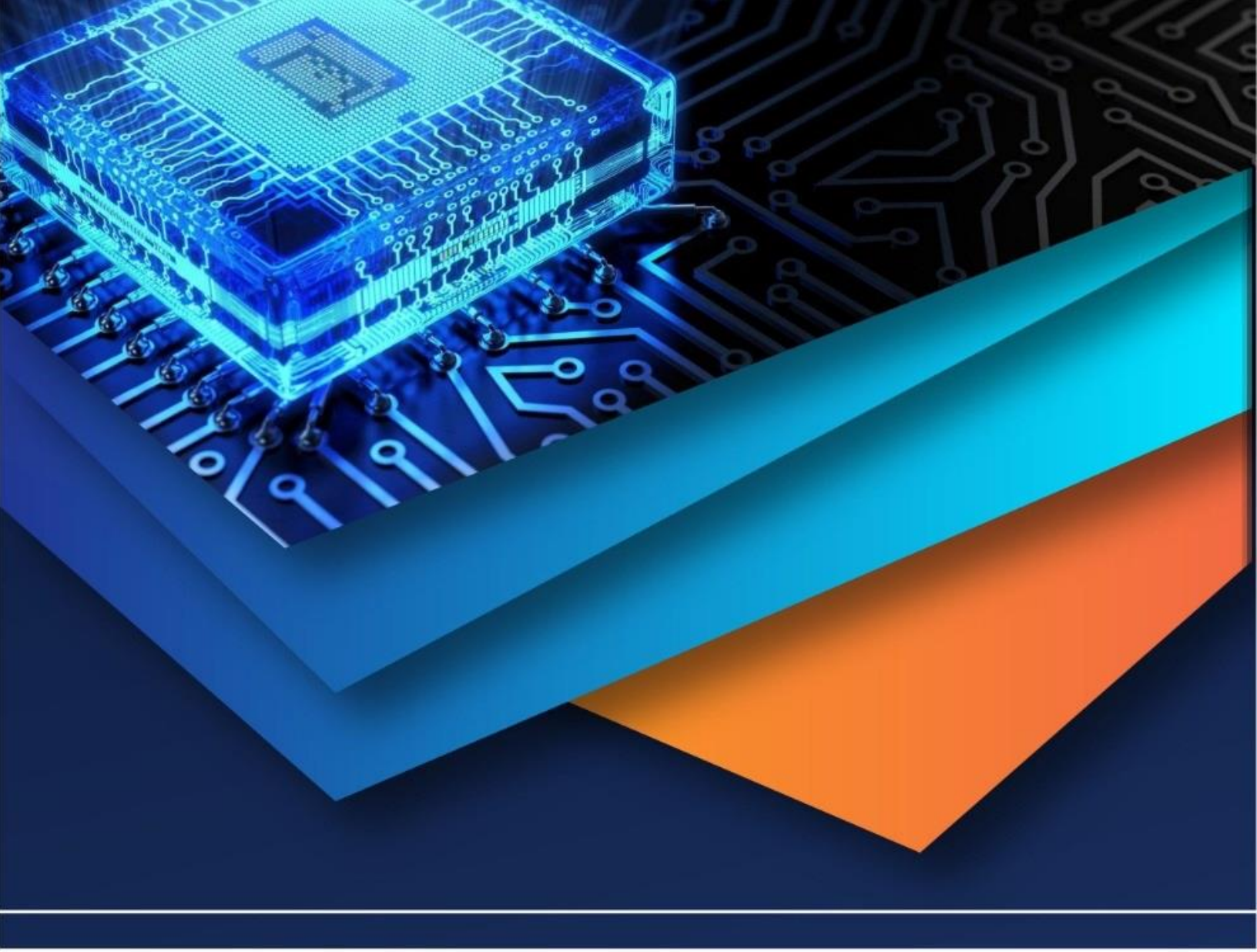
Dataset	Year	Attack types	Attacks
KDD Cup'99[129]	1998	4	DoS, Probe, R2L, U2R
Kyoto 2006+[130]	2006	2	Known Attacks, Unknown Attacks
NSL-KDD[131]	2009	4	DoS, Probe, R2L, U2R
UNSW-NB15[132]	2015	9	Backdoors, DoS, Exploits, Fuzzers, Generic, Port scans, Reconnaissance, Shellcode, worms
CIC-IDS2017[133]	2017	7	Brute Force, HeartBleed, Botnet, DoS, DDoS, Web , Infiltration
CSE-CIC-IDS2018[133]	2018	7	HeartBleed, DoS, Botnet, DDoS, Brute Force, Infiltration, Web.

TABLE 1 Summary of public benchmark datasets

REFERENCES

[1] Nathan Shone, Tran Nguyen Ngoc, Vu Dinh Phai, and Qi Shi, A Deep Learning Approach to Network Detection System, 1st February 2018.
[2] L. Xiao, Y. Chen, C.K. Chang, Bayesian model averaging of Bayesian network classifiers for intrusion detection, in: 2014 in IEEE 38th International Computer Software and Applications Conference Workshops on 35, 2014, pp. 1302–1310
[3] Khraisat A, Gondal I, Vamplew P, Kamruzzaman J. Survey of intrusion detection systems: techniques, datasets and challenges. Cybersecurity. 2019;2(1):20. https://doi.org/10.1186/s42400-019-0038-7.
[4] A REVIEW ON KDD CUP99 AND NSL NSL-KDD DATASET. International Journal of Advanced Research in Computer Science. Mar/Apr2019, Vol. 10 Issue 2, p64-67. 4p. Author(s): Bala, Ritu; Nagpal, Ritu
[5] Kumar and A. Yadav, "Increasing Performance of Intrusion Detection System Using Neural Network", 2014 IEEE International Conference on Advanced Communication Control and Technologies (ICACCCT), pp. 1935-1939
[6] Santosh Kumar Sahu Sauravranjan Sarangi Sanjaya Kumar Jena, "A Detail Analysis on Intrusion Detection Datasets", 2014 IEEE International Advance Computing Conference (IACC)
[7] http://nsl.cs.unb.ca/NSL-KDD/ - Dataset Used.
[8] Sapna S. Kaushik, Dr Prof. P. R. Deshmukh," Detection of Attacks in an Intrusion Detection System", International Journal of Computer Science and Information Technologies, Vol. 2 (3), 2011, 982-986
[9] Vipin Kumar, Himadri Chauhan, Dheeraj Panwar, "K-Means Clustering Approach to Analyze NSL-KDD Intrusion Detection Dataset", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-4, September 2013

[10] Y. Chang, W. Li, and Z. Yang, "Network intrusion detection based on random forest and support vector machine," in Proc. IEEE Int. Conf. Comput. Sci. Eng./IEEE Int. Conf. Embedded Ubiquitous Comput., Jul. 2017
[11] Zanero S, Serazzi G. Unsupervised learning algorithms for intrusion detection. Paper presented at: Proceedings of the IEEE Network Operations and Management Symposium. Salvador, Bahia, Brazil: IEEE; 2008:1043-1048.
[12] S.G. Kene, D.P. Theng, A review on intrusion detection techniques for cloud computing and security challenges, in: 2nd International Conference on Electronics and Communication Systems (ICECS), 2015, pp. 227–232, https://doi.org/ 10.1109/ECS.2015.7124898.
[13] J. Yan, D. Jin, C.W. Lee, P. Liu, A comparative study of off-line DL based network intrusion detection, in: Tenth International Conference on Ubiquitous and Future Networks, ICUFN), 2018, pp. 299–304, https://doi.org/10.1109/ ICUFN.2018.8436774.
[14] G. Poojitha, K.N. Kumar, P.J. Reddy, Intrusion detection using artificial neural network, in: 2010 in Second International Conference on Computing, Communication and Networking Technologies, 2010, pp. 1–7, https://doi.org/ 10.1109/ICCCNT.2010.5592568.
[15] Malhotra, H., & Sharma, P. (2019). Intrusion Detection using Machine Learning and Feature Selection. International Journal of Computer Network & Information Security, 11(4).



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY