IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11623

NETWORK INTRUSION DETECTION SYSTEM USING MACHINE LEARNING

Dr Maria Manuel Vianny[1], Meghana Prasanna[2], Aakanksha D[3], Shalini Menon[4]

[1]Project guide, Assistant Professor, Department of Computer Science and Engineering (Data Science),

Jain (Deemed-to-be University), Bangalore, India

[2-4]B.Tech, Computer Science and Engineering, Jain(Deemed-to-be university) Bangalore, India.

Abstract: Home and network intrusion detection system (HNIDS) is an intriguing problem that is being addressed in more recent research studies.this review is focused on recognizing intrusion in homes and computer networks based on image and numerical datasets respectively.diverse approaches have been tried over the past years.profuse methods are offered to acknowledge varied types of attacks such as botnet,DoS,DDos with respect to networks and breaking and entering,jumping a wall as actions in home intrusion to list a few. For each methodology,multiple algorithms and datasets are used.data is obtained in several ways such as cctv footage, images,sensors,computer logs,attack signatures.the expected outcomes by each approach and dataset are then compared.

Keywords: Home and network intrusion detection system,HNIDS,ML,GAN,LSTM

I. INTRODUCTION

Intrusion Detection System(IDS) is a comprehensive system that monitors traffic for fraudulent activity and warrants an alarm when such activity is discovered. It may refer to a broad class of systems whose input is a traffic source and decide to classify whether an instance is malicious. Appropriate design of intrusion detection systems is extremely important to safeguard all critical infrastructure and personal data from vandalism, theft and misuse[1]. Lack of intrusion prevention can degrade the credibility of security services, namely data confidentiality, integrity and availability.

The rapid development of network technology has dramatically improved people's daily lives, but it has also brought many threats.[2].Intrusion Detection Systems (IDSs) have been widely adopted as an effective method to detect and defend against network attacks in response to the growing network threats. Although the sparsity of attack data, the training set for this type of approach is unbalanced, affecting analysis performance. Conventional IDSs tend to get low detection rates and high false positive rates due to their reliance on patterns of known attacks. Intrusion detection systems (IDS) are able to collect samples of network traffic using fiber optical taps at various points in a network .

There are two basic types of intrusion detection based on how the malicious attacker is being recognised. Firstly,

Any deviations from the behavior characteristics are regarded as intrusion behaviors. This is a process referred to as anomaly detection. Misuse detection can record the behavioral characteristics of known cyber attacks, with a low false alarm rate, but it lacks learning ability. Therefore, the matching database must be constantly updated to adapt to the changing environment. In addition, the misuse detection usually has a poor detection effect for new attacks. On the other hand, anomaly detection can effectively detect unknown attacks, but its false alarm rate is high.

Nowadays the internet is ubiquitous and an enabler of the concept known as the internet of things (IoT)[5]. This concept allows billions of devices to be accessible through the internet, revolutionizing our day-to-day lives including houses, the appliances we use, the lighting systems, cars. The internet affects every major and minor aspect of our life. The interconnected nature of the devices requires the security-by-design approach, which brings pros and cons to building security protection. current protection may not be adequate as new exploitation techniques arise. refers to an emerging technology that connects physical devices to the internet. The revolution of smart home devices has made it very complicated and challenging to detect any anomalous behavior, whether it comes from the network or user [11].

The majority of these devices and sensors currently available in the market have failed to comply with modern security standards which are Confidentiality, Integrity, Authorization, Authenticity, and Availability; as vulnerabilities for IoT devices are discovered and exploited routinely despite them. In environments involving multiple devices of varying levels of trustworthiness and likely inter-dependencies between them, such as those found in smart homes, it makes sense to try to detect security breaches when they occur.

As more platforms and applications are being connected to networks, data becomes increasingly vulnerable to malicious attacks. Use of CCTV is a popular option adopted by many people in order to ensure the safety of their home front and this has currently been developed from Analog system to IP system. However, at present, the system still lacks the ability to alert users timely and this drawback results in property damages as well as risks to human life.There is a huge need to automate the systems and successfully identify the intruders, minimize the false alarms and minimize the human supervision of such systems[8].

Home intrusion plays a role and requires a system to notify and activate homeowners to be alert and protect their home assets in time[12].Over the past decade, the smart home technology usage along with concerning the security and safety has been increased according to a variety of crimes such as intrusion and stealing. Therefore, an automated system to monitor and notify attackers to a homeowner is required in order to prevent the security and safety issues.

Researchers have applied artificial intelligence (AI) algorithms in the designing part of IDSs to provide better performances[3]. Machine-learning (ML) is becoming a prevalent way of detecting advanced attacks with unexpected patterns [4]. ML is based on statistical and mathematical algorithms rather than rule-based algorithms. ML techniques contribute to improving performance of intrusion detection systems (IDS). The samples can be used as inputs to a trained ML model to classify a data sample as benign or malicious in the simplest case of binary classification [6].Predictions could also be made on potential attacks using such models.

Various machine learning (ML) or deep learning (DL) algorithms have been proposed for implementing anomaly-based IDS (AIDS)[7].. These supervised ML algorithms include the artificial neural network (ANN), decision tree (DT), k-nearest neighbor (k-NN), naive Bayes (NB), random forest (RF), support vector machine (SVM), and convolutional neural network (CNN) algorithms, whereas the unsupervised ML algorithms include the expectation-maximization (EM), k-means, and self-organizing maps (SOM) algorithms. The most basic concept on which the IDS works is outlier detection.outlier detection refers to the problem of finding patterns in the data that do not meet the required behavior. These anomalous patterns are often referred to as anomalies, inconsistent observations, exceptions, glitches, defects, noise, errors, or contaminants in various application domains. Outlier detection is a widely researched problem and finds great use in a wide range of application domains such as credit cards, insurance, tax fraud detection, cyber security intrusion detection, critical security system flaw detection, military surveillance for enemy activities and many more.The importance of Outlier detection system from the fact that anomalies in the data translate into meaningful information across a wide range of application domains.

An Intrusion Detection System (IDS), a significant research achievement in the information security field, can identify an invasion, which could be an ongoing invasion or an intrusion that has already occurred.

the massive amount of data being generated today and the rate of it being generated is alarming. It has brought huge difficulties and challenges to network security, a research topic which has attracted more and more attention. Intrusion Detection (ID)[9].ML-based approaches have been extensively used in detecting several types of malicious attacks and ML techniques can assist the network administrator with taking the appropriate actions to prevent these malicious attacks in the network[10].Ensemble learning (EL) also helps us to solve the problem at hand.

The rest of this survey paper is as follows: the related works section provides information on existing systems,their flaws and a comparison between the systems. The methodology section is split into two parts and has a description of every machine learning and deep learning algorithm and compares the results achieved by them. This section also includes the various datasets used in the existing work and which ones are the best.

II.RELATED WORK:

Intrusion detection systems use image sensors, cctv cameras, online site traffic and network logs. image sensors and cctv cameras capture movement only through fixed cameras thus not providing an overall view. and using site traffic and network logs give high false positives. the development of an intrusion detection system uses algorithms from machine learning and deep learning to increase the efficiency. many researchers have developed various systems using the same. some of the systems and frameworks are listed below:

Ziadoon Kamil Maseer et al (2021) discusses research on how an intrusion detection system (IDS) is an important protection instrument for detecting complex network attacks. They achieve results by Using 10 supervised and unsupervised ML algorithms for identifying effective and efficient Machine Learning Anomaly Intrusion Detection System (ML–AIDS) of networks and computers. They use several machine learning models such as Artificial Neural Network(ANN), Decision Tree, K-Nearest Neighbors, Naive Bayes, Random Forest, Support Vector Machine, Convolution Neural Network, Expectation maximization, K means, and Self-organizing maps. Supervised learning performs classification based on data

instances that are marked in the training phase and unsupervised machine learning algorithms have been evaluated for performing in Anomaly Intrusion Detection (AIDS). The dataset used is the CICIDS2017.[1]

Gustavo De Carvalho Bertoli (2021) suggests how the increase of connected devices and the constantly evolving methods and techniques by attackers pose a challenge for network intrusion detection systems from conception to operation. The proposed framework Describes the AB-TRAP framework that enables the use of updated network traffic and considers operational concerns to enable the complete deployment of the solution.[2]

Muhammad Ashfaq Khan et al (2021) used a Convolutional Recurrent Neural Network (CRNN) which is used to create a Deep Learning DL-based hybrid Intrusion Detection framework that predicts and classifies malicious cyberattacks in the network. To assess the efficacy of the Hybrid Convolutional Recurrent Neural Network Intrusion Detection System(HCRNNIDS), experiments were done on publicly available Intrusion Detection data, specifically the modern and realistic CSE-CIC-DS2018 data.[3]

Lirim Ashikuand Cihan (2021) discuss the use of deep learning architecture to develop adaptive and resilient Network Intrusion Detection Systems (NIDS) to detect and classify network attacks. Their emphasis is on how Deep Learning(DL) or Deep Neural Networks (DNNs) can facilitate a flexible Intrusion Detection System with learning capability to detect recognized and new or zero-day network behavioral features, consequently ejecting the systems intruder and reducing the risk of compromise. To demonstrate the model's effectiveness, they used the UNSW-NB15 dataset. [4]

Zhendong Wang et al.(2021) wants to improve the searchability and optimization ability of the traditional gray wolf optimizer algorithm; a novel optimization strategy combining the inner and outer hunting is introduced. Experiments on KDDCup99, NSL-KDD, UNSW-NB15, and CICIDS2017 datasets show that the proposed DBN-EGWO-KELM algorithm has greater advantages in terms of its accuracy, precision, true positive rate, false-positive rate, and other evaluation indices compared with Radial Bias Function, Support Vector Machine, Kernel Extreme Learning Machine, LIBrary of Support Vector Machines, Convolution Neural Networks, Deep Belief Network-KELM, and other intrusion detection models. [5]

Sara Al-Emadi Et Al (2020) developed an intelligent detection system that can detect different network intrusions. Additionally, NLS evaluates the performance of the proposed solution. Use of ensemble learning and k fold cross-validation in the preprocessing of the data and feature extraction. They have worked with deep learning techniques, namely, Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to design an intelligent detection system that can detect different network intrusions.[6]

Liqun Yang et al. developed an effective wireless intrusion detection for real-time data mechanism. Using various algorithms such as Conditional Deep Belief Network (CDBN)that is composed of the Conditional Gaussian-Bernoulli RBM (CGBRBM), SamSelect, a Stacked Contractive Auto-Encoder (SCAE) they intended to provide a better detection performance when compared to deep learning and other shallow learning methodologies. [7]

Kaiyuan Jiang et al. discuss on how to remove noise in measurements,generic build up of dataset,difficult to extract features, Traditional, ML algorithms belong to shallow learning.the aim is to detecting nids combined hybrid sampling with deep hierarchical networks on an imbalanced datasets.the algorithms used are one-side selection (OSS), Synthetic Minority Over-sampling Technique (SMOTE), neural networks and Bi-directional long short-term memory (BiLSTM).the algorithm provides an accuracy of 83.58 on nsl-kdd and 77.16 on unsw-n datasets.[9]

Table 1 shows a comparison of the papers under study :

Year	Title	Author(s)	Objective	Feature selection and validation technique	Algorithm	Accuracy	Drawbacks
2021	An End-to-End Framework for Machine Learning-Based Network Intrusion Detection System	GUSTAVO DE CARVALHO BERTOLI et al	To develop an IDS framework that solves the problem of real-time updated network traffic	Embedded feature selection. Stratified K-fold technique is implemented	AB-TRAP FRAMEWORK	FI score of .95 on MAWILab	As ab trap is complex it can lead to overfitting when used on smaller datasets like the mawilab when compared to cicids2018

2021	Benchmarking of Machine Learning for Anomaly Based Intrusion Detection Systems	IADOON KAMIL MASEER, etal	Using 10 supervised and unsupervised ML algorithms for efficient ML–AIDS of networks and computers	Self-organizing maps for visualization, and expectation-maximization is used to find maximum likelihood as the dataset has missing values	ANN, DT, K-NN, NB, RF, SVM, CNN, EM, K means, Self-organizing maps	98% on CICIDS2017	The results obtained are varied and incomparable as the models are trained on different datasets.
2020	Using Deep Learning Techniques for Network Intrusion Detection	Sara Al-Emadi et al	Designing an intelligent detection system which can detect different network intrusions.	Use ensemble learning and k fold cross-validation in the preprocessing of the data and feature extraction.	Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN)	F1-score cnn-98.4% Rnn & lstm-89.54% on NLS kdd	investigate the performance of different deep learning techniques by implementing various combinations.
2021	Network Intrusion Detection System using Deep Learning	Lirim Ashiku, Cihan Dagli	Proposing the use of dl architecture, to develop adaptive and resilient nids	The data set used is highly imbalance and has a high false-positive rate	DNN,LSTM GRU	95.4% on UNSW-NB-15	The data set used is highly imbalance and has a high false-positive rate
2021	Deep Belief Network Integrating Improved Kernel-Based Extreme Learning Machine for	ZHENDONG WANG et al	Improve the searchability of the traditional grey wolf optimizer algorithm,	Improve the searchability and optimization ability of the traditional grey wolf optimizer algorithm	Deep belief network-EGWO KELM(Kernel-based Extreme Learning Machine)	Around 92 % on KDDCup99, NSL-KDD, UNSW-NB15, and CICIDS2017	Algorithms are very complex for the dataset being used. This leads to overfitting and high false-positive rates
2021	HCRNNIDS: Hybrid Convolutional Recurrent Neural Network-Based Network Intrusion Detection System	Muhammad Ashfaq Khan.	To develop a hybrid convolutional recurrent neural network-based IDS using a real-world dataset.	The algorithm is the best classification model and provides significant accuracy	Hybrid convolutional recurrent neural network (HCRNN)	up to 97.75%	The dataset used is larger in volume compared to the other datasets and has a lesser imbalance and false-positive rate

DOI: 10.17148/IJARCCE.2022.11623

2020	Real Time Intrusion Detection in Wireless Network: a Deep Learning Based Intelligent Mechanism	LIQUN YANG, JIANQIANG LI	To develop an effective wireless intrusion detection for real time data mechanism.	the dataset used is a fairly balanced data set	Conditional Deep Belief Network (CDBN)that is composed of the Conditional Gaussian-Bernoulli RBM (CGBRBM), SamSelect, a Stacked Contractive Auto-Encoder (SCAE)	up to 75%	the work using the models is limited to the specific dataset and the model is a fairly complex approximation on the used data set
2020	Network intrusion detection combined hybrid sampling with deep hierarchical network	KAIYUAN JIANG	Approaches to remove noise in measurements,generic build up of dataset,difficult to extract features, Traditional, ml algorithms belong to shallow learning	the dataset is very unbalanced so sampling methods are used to rectify this problem	one-side selection (OSS), Synthetic Minority Over-sampling Technique (SMOTE), neural networks and Bi-directional long short-term memory (BiLSTM)	83.58 on nsl-kdd and 77.16 on unsw-nb15	the dataset used is highly imbalanced and leads to higher faslse positives

III. METHODOLOGY:

We start by looking at the database. In this process, we get unique and empty values. When done we combine the labels into one "malicious" label. By using numerical values, we easily determine the probability of our values. Turn Benign labels into zeros and Malicious labels into one.
Now that the refined site has been split into two parts.The training site contains 70% of the site and the test data set contains 30% .The train test separation process is used to measure the performance of machine learning algorithms. when used to predict data that can be used to model a model. It is a quick and easy process, the results that allow you to compare the performance of machine learning algorithms with your model guessing problem.

In order to continue producing the most efficient model we use a 5-point verification grid search. By using counter-verification and grid search we were able to have a much better impact compared to our actual split / test split and minimal tuning. Brightness verification is the most important method used to create models that are best suited for training and testing across all parts of the training site. We are now conducting a comparative study of machine learning with in-depth learning models such as decision tree, systematic retrieval, and long-term memory.

Unbalanced data refers to those types of data sets where the target class has an unequal distribution of visuals, i.e. one class label has the highest number of observations and the other has the lowest number of observations. We can better understand it by example.

The big problem with unequal data forecasting is how accurately we predict both the majority and sub-category? Let us explain an example of a diagnosis. Let us assume that we will predict the disease in an existing database where only 100 records of 5 patients are diagnosed with the disease. Therefore, the majority of the population is 95% free of the disease and the minority only 5% are infected. Now, imagine that our model predicts that every 100 out of 100 patients are free of the disease.
In a class of unequal class F1 school is a very appropriate matrix.

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

Category data refers to variables made with label values, for example, variables of "color" may have values of "red", "blue," and "green". Think of values that are similar to the categories that sometimes have a natural order in them.
Some machine learning algorithms can work directly with category data depending on the application, such as the decision tree, but most require any input or output output to be a number, or a number of numbers. This means that any category data must be mapped into numbers.

One hot code coding is one way to convert data to optimize the algorithm and get a better prediction. With one heat, we convert each category value into a new category column and assign a binary value of 1 or 0 to those columns. The value of each number is represented as a binary vector. All values are zero, and the index is marked 1.
One hot code coding makes our training data very useful and clear, and can be easily modified. By using numerical values, we easily determine the probability of our values. In particular, a single hot code text is used for our output values, as it provides predictions that are significantly different than a single label.

Decision Tree: The Decision Tree is a powerful and popular tool for classification and prediction. The decision tree is a flow chart similar to a tree structure, where each internal node refers to a test in the attribute, each branch represents the test result, and each leaf node (end node) holds a class label.

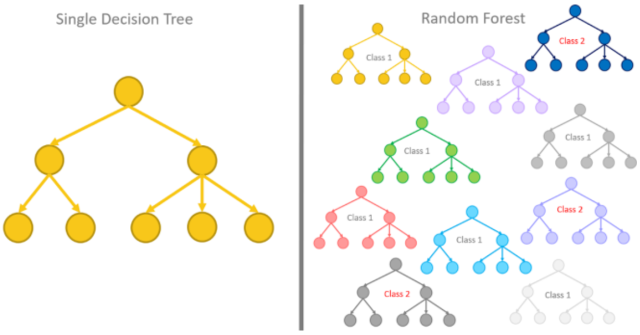

Figure 1: Decision Tree

Decision Tree is a supervised learning method that can be used for both planning and retrospective problems, but is often preferred in solving planning problems. It is a tree-shaped divider, where the internal nodes represent the elements of the database, the branches represent the rules of decision and each leaf node represents the result.in the Decision Tree, there are two areas, namely Resolution Node and Leaf Node. Decision Nodes are used to make any decision and have many branches, while Leaf nodes are the result of those decisions and have no other branches.

Decisions or tests are made on the basis of the data provided. It is a pictorial representation of all possible solutions / problem-based solutions. It is called a deciding tree because, it is similar to a tree. , begins with a root node, extends to additional branches and builds a tree-like structure.

Random Forest is a popular machine learning algorithm that is part of a supervised learning strategy. Can be used for both Scheduling and retrieve problems in ML. It is based on the concept of integrated learning, which is the process of integrating multiple dividers to solve complex problems and improve model performance.

As the name suggests, "The Random Forest is a subdivision that contains a number of decision trees for the various datasets set and takes measurement to improve the prediction accuracy of that database." Instead of relying on a single decision tree, the random forest takes a prediction from each tree and is based on multiple predictable votes, and predicts the final outcome.

The large number of trees in the forest leads to high accuracy and prevents the problem of overcrowding.

DOI: 10.17148/IJARCCE.2022.11623

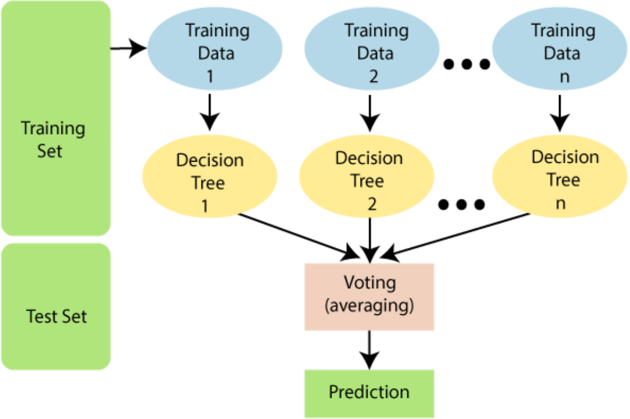

Figure 2: Random Forest

why use the Random Forest algorithm: -It takes less training time compared to other algorithms.Decreased performance is one of the most popular methods of machine learning, which comes under the supervision of a supervised learning strategy. It is used to predict phase-dependent variables using a given set of independent variables.Reversion predicts phase-out volatility. Therefore the result should be phase or separate value. Either Yes or No, 0 or 1, True or False, etc. but instead of giving a fixed value such as 0 and 1, it provides possible values between 0 and 1.

Logistic regression is very similar to Linear. Reversing regardless of how they are used. Linear Regression is used for troubleshooting problems, and Logistic regression is used for troubleshooting problems.

In Logistic regression, instead of inserting a regression line, we are equal to the "S" shaped editing function, which predicts two higher values (0 or 1).

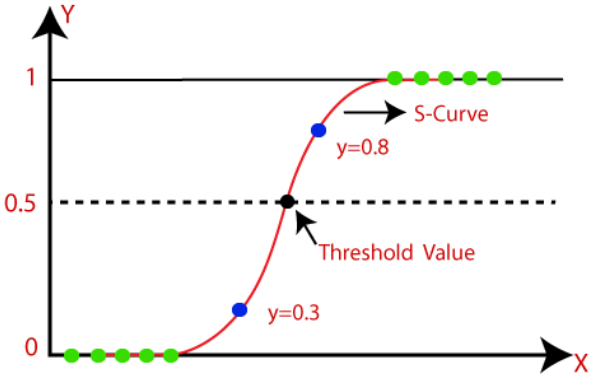

Figure 3: Logistic Regression Curve

The curve from the entry function indicates the possibility of something like cancer cells or not, the mouse is fat or not based on its weight, etc. Logistic Regression is an important machine learning algorithm because it has the potential to provide opportunities. and classify new data using continuous and differentiated data sets.Logistic retransmission can be used to visualize the use of different types of data and can easily determine the most effective variables used for classification.

K-Nearest Neighbor is one of the simple Machine Learning algorithms based on the Guided Reading method. The K-NN algorithm captures the similarities between new cases / data and available cases and places the new case in the most similar category to the available categories. .

The K-NN algorithm stores all available data and classifies new data points based on similarities. This means that where new data comes from it can be easily categorized into a well suite component using the K-NN algorithm.The K-NN algorithm may be used for Depression and Editing but mainly for Partition problems.K- I -NN is a non-parameter algorithm, which means it does not make any assumptions about basic data.

DOI: 10.17148/IJARCCE.2022.11623

It is also called the lazy student algorithm because it does not learn from the training set faster than it keeps the database and during the breakdown, performs the action on the database.

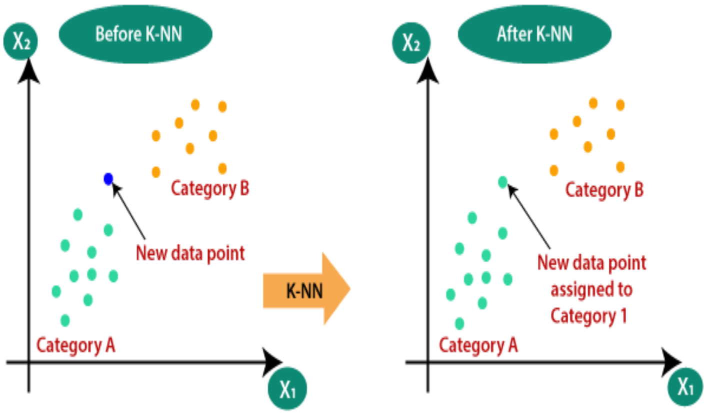

Figure 4 : K Nearest Neighbours

LSTM networks increase the continuous neural network (RNNs) especially designed to deal with situations where RNNs do not work. Speaking of RNN, an algorithm that processes current input by considering the output of previous events (response) and storing it in its users' memory for a short time (temporary memory). Of the many applications, the most popular are those in the non-Markovian language control areas and music formats. However, there are some obstacles to RNNs. The first one fails to store information for a long time. Sometimes a database of data stored in the past is needed to determine the current result. However, RNNs can never manage this "long-term dependence." The second issue is that there is no better control over which part of the context is needed to proceed and which part of the past should be forgotten. Other problems related to RNNs are cracks or disappearances of slopes (described later) that occur in RNN retreat training. Therefore, Long-Term Memory (LSTM) is presented in the image. It is designed in such a way that the problem of gradient disappearance is eliminated almost completely as the training model is not affected. Long-term delays within certain issues are resolved using LSTMs, which also deal with the effects of noise, distributed presentations, or endless numbers. For LSTMs, they do not meet the requirement to keep the same number of regions ahead of the time required by the Markov HireM (HMM) model. LSTMs provide us with a wide range of parameters such as reading and output levels and input bias. Therefore, there is no need for minor adjustments. Attempt to revise each weight was reduced to O (1) by using LSTMs similar to those used in Back Propagation Through Time (BPTT), which is a huge benefit.

DATASET:

The CICIDS2017 data set contains the most common and most recent attacks, similar to real-world data (PCAs). From the beginning of the CICIDS2017 database, the data set began to attract researchers to analyze and develop new models and algorithms. CICIDS-2017 database from ISCX Consortium. Machine Learning CSV contains eight (8) traffic monitoring times, each in the form of a comma-separated file (CSV). This file contains normal traffic which is defined as “Benign” traffic and unidentified traffic called “Attacks” traffic. In addition to normal traffic and optimal traffic, there are 14 types of attacks on this database. There is a more updated dataset that exists and it is the CSE-CIC-IDS2018 dataset. The CSE-CIC-IDS2018 dataset, we use the perception of profiles to generate datasets in a systematic way, a good way to comprise exact descriptions of intrusions and summary distribution models for programs, protocols, or lower-level network entities. Those profiles may be used by agents or human operators to generate occasions in the community. Due to the abstract nature of the generated profiles, we will follow them to a diverse range of network protocols with distinctive topologies. Datasets along with CSE-CIC-IDS2018 have been created to train predictive fashions on anomaly-primarily based intrusion detection for community site visitors. CSE-CIC-IDS2018 is not a completely new venture, but part of an existing venture that produces present day, sensible datasets in a scalable way. There are eighty functions within the dataset, presenting statistical information of the flows from each uplink and downlink. as compared to the trustworthy TCP/IP visitors header portions of data provided by way of the previous datasets, like DARPA and KDD, it's miles extensively believed that the data based on go with the flow may want to offer extra beneficial facts for intrusion detection .that allows you to get a well known concept of the dataset, we plot a bar chart of each function in opposition to the label. As redundant features will increase computation fee, set off chaos and decrease accuracy, we delete the features that do not display any distinction among benign traffic and a malicious one.

DOI: 10.17148/IJARCCE.2022.11623

IV.CONCLUSION:

Today's computer network is a dangerous place, full of people with millions of human hours available to use them against the strongest security strategies. The only way to defeat them is to know when they are trying to attack and oppose their efforts. Improving the accuracy of NIDS diagnosis is an important issue in the field of network security.

In this paper, the methods of obtaining network access based on machine learning and in-depth learning methods provide new researchers with updated information, the latest trends, and advances in the field. There are two broad categories of entry-level strategies; one is based on a clear set of rules, technically known as signature acquisition and the other is based on morality, technically known as an anomaly-based acquisition. The test was performed on the CICIDS2018 database to determine the intervention. Since data is limited we have selected a specific set of dates for data testing and training. The models have shown that our proposed method has achieved 99% complete accuracy.

To make the system output more usable and accurate, Artificial Intelligence features can be installed. This can be achieved by creating an alarm system for home network login systems and connecting them to CCTV images for real-time analysis. As detection of smart home access remains a major challenge, especially with regard to testing and forecasting. A warning email is sent to the owner with the latest photo taken using IoT.

V.REFERENCES :

1. Bertoli, Gustavo De Carvalho, et al. "An end-to-end framework for machine learning-based network intrusion detection system." IEEE Access 9 (2021): 106790-106805.
2. Maseer, Ziadoon Kamil, et al. "Benchmarking of machine learning for anomaly based intrusion detection systems in the CICIDS2017 dataset." IEEE Access 9 (2021): 22351-22370.
3. Peng, Ye, et al. "Detecting Adversarial Examples for Network Intrusion Detection System with GAN." 2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS). IEEE, 2020.
4. Wang, Zhendong, et al. "Deep belief network integrating improved kernel-based extreme learning machine for network intrusion detection." IEEE Access 9 (2021): 16062-16091.
5. Al-Emadi, Sara, Aisha Al-Mohannadi, and Felwa AlSenaid. "Using deep learning techniques for network intrusion detection." 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT). IEEE, 2020.
6. Ahmad, Zeeshan, et al. "Network intrusion detection system: A systematic study of machine learning and deep learning approaches." Transactions on Emerging Telecommunications Technologies 32.1 (2021): e4150.
7. Liao, Dashun & Huang, Sunpei & Tan, Yuyu & Bai, Guoqing. (2020). Network Intrusion Detection Method Based on GAN Model. 10.1109/CCNS50731.2020.00041.
8. Jiang, Kaiyuan, et al. "Network intrusion detection combined hybrid sampling with deep hierarchical network." IEEE Access 8 (2020): 32464-32476.
9. Zhang, Guoling, et al. "Network intrusion detection based on conditional Wasserstein generative adversarial network and cost-sensitive stacked autoencoder." IEEE Access 8 (2020): 190431-190447.
10. Khan, Muhammad Ashfaq. "HCRNNIDS: Hybrid
11. Convolutional Recurrent Neural Network-Based Network Intrusion Detection System." Processes 9.5 (2021): 834.
12. J. Guo, J. Cheng and J. Cleland-Huang, "Semantically Enhanced Software Traceability Using Deep Learning Techniques," 2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE), 2017, pp. 3-14, doi: 10.1109/ICSE.2017.9
13. Khan, Muhammad Ashfaq, Md Karim, and Yangwoo Kim. "A scalable and hybrid intrusion detection system based on the convolutional-LSTM network." Symmetry 11.4 (2019): 583.
14. Viegas, Eduardo, Altair Olivo Santin, and Vilmar Abreu Jr. "Machine Learning Intrusion Detection in Big Data Era: A Multi-Objective Approach for Longer Model Lifespans." IEEE Transactions on Network Science and Engineering 8.1 (2020): 366-376
15. Kilincer, Ilhan Firat, Fatih Ertam, and Abdulkadir Sengur. "Machine learning methods for cyber security intrusion detection: Datasets and comparative study." Computer Networks 188 (2021): 107840.
16. Laghrissi, FatimaEzzahra, et al. "Intrusion detection systems using long short-term memory (LSTM)." Journal of Big Data 8.1 (2021): 1-16.
17. Salim, A. Alrahman Abdulrahman, and Abdullahi Abdu Ibrahim. "Intrusion Detection System In Iot Network Using Machine Learning." 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). IEEE, 2020.
18. Krishna, Akhil, et al. "Intrusion Detection and Prevention System Using Deep Learning." 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE, 2020.
19. Patel, Ahmed, Qais Qassim, and Christopher Wills. "A survey of intrusion detection and prevention systems."

DOI: 10.17148/IJARCCE.2022.11623

Information Management & Computer Security (2010).

20. Hao, Xingran, et al. "Producing More with Less: A GAN-based Network Attack Detection Approach for Imbalanced Data." 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, 2021.
21. Li, Xinghua, et al. "Detection of low-frequency and multi-stage attacks in industrial Internet of things." IEEE Transactions on Vehicular Technology 69.8 (2020): 8820-8831.