IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DETECTING CYBER ATTACKS AND SECURE 5G WIRELESS NETWORK USING REINFORCEMENT LEARNING TECHNIQUES

[1]R.S.SubhasreeVignani, [2]V.Mounika, [3]Y.Kamakshi, [4]K.Kondaiah, [5]N.Johnson
[1]Associate Professor, [2-6]UG Scholar
[1-6] Department of Computer Science and Engineering,
[1-6] Siddharth Institute of Engineering & Technology, Puttur, India

Abstract: The security of a 5G wireless network will be challenged by a wide range of sophisticated hacking methods. We propose and develop a novel cooperative attack detection method that makes use of a hierarchical Reinforcement Learning (RL) mechanism in order to better identify network assaults and focus on protecting 5G wireless networks from the most severe and cutting-edge forms of network assault, such as DDoS and jamming. Distributed detection systems are operated at the several important nodes of the 5G network (AP, BTS, and servers) to accomplish the cooperative detection. Results from our trials show that the suggested RL detection system improves the ability to identify novel malicious actions and assaults.
All network activity is analysed for malicious behaviour using IDS powered by machine learning. Enhancing the intrusion detection system's detection rate was the primary objective of the system design, with a focus on false negative and false positive performance measures. Different machine learning models, including SVM and KNN classifiers, are built and compared.

Index Terms - 5G Wireless Network, Reinforcement Learning, Network Attack Detection, Cyber Threats

I. INTRODUCTION

The global internet population is expanding at an exponential rate. Now more than ever, it is essential to the survival of every industry, from academia to the IT industry to banking to the stock market to healthcare. DDoS attacks are the most effective method currently available for penetrating internet defences. It is safe to assume that a Distributed Denial of Service (DDoS) assault is to blame whenever there is talk of a website being down or users being unable to access the site's web pages.

In a denial of service (DoS) attack, malevolent users block legitimate programmes from communicating with the network. Without stealing sensitive data or compromising credentials, they generate enormous amounts of traffic in the vicinity of the victim machine in an effort to slow down network access and deplete system resources. This enormous data generation makes it difficult for individual packets to be sent. In order to establish a connection across a network, a "three-way handshake" is often required. The server confirms the connection the client established in the wake of getting the solicitation and apportioning limit inside the repository association. After the receiver verifies the connection, the procedure is complete. The offender makes the necessary connections and sends the photons as planned, but they miss the provider's last warning.
As a result, the host seldom benefits financially from the connection and often ends up in the reservoir for uncompleted deals. If all of the necessary conditions are satisfied, the server will be unable to launch a denial-of-service attack because its buffer will be full with unresolved attachments and there will be no place for legitimate TCP connections. An attack known as "Denial of Service" (DoS) aims to prevent authorized users from gaining access to a specific computer system's network and data resources. DDoS

assaults are the most challenging security concerns of today to detect, mitigate, and track due to the constraints of ordinary networking equipment, the variety of assault systems and software engineers' mistiness toward have locales. Distributed denial of service (DDoS) attacks can be stopped using firewalls and vendor-specific workarounds.

Attacks may be classified as either fixed sourceIP assaults (FSIA) or random source IP attacks (RSIPA), respectively (RSIA). Spoofing an IP address is a common tactic used by attackers to conceal their true locations before launching malicious assaults. The faked IP address or addresses may be static or dynamic. Furthermore, attackers may initiate assaults from a botnet using non-spoofed IP addresses, allowing them to circumvent any anti-spoofing techniques that may be in place (fixed IP addresses). However, many IP address-based DDoS detection systems may not provide protection against port scan assaults (FSIA and RSIA). There has been a lot of focus on DDoS assaults as a study topic in recent years. Christological dualisms. Offers insight into the nature of DDoS assaults by laying out the issue they pose and organising various countermeasures into several categories. The most prevalent type of distributed attack on a network is known as distributed denial of service (DDoS). While many contemporary methods for detecting DDoS assaults are effective, they often fall short when it comes to tracking out the origin of the attacks themselves. This renders the victims completely helpless in terms of taking the initiative. The framework of the authorised organization's website, for example, might be vulnerable to these assaults because of general restrictions that apply to all arrangement assets. Distributed denial of service attacks (DDoS) include the simultaneous transmission of a large number of requests to a website over a large number of IP addresses, with the goal of overwhelming the site's resources and making it inaccessible to normal users.

Figure 1 depicts a DDOS attack scenario, as described by Singh KJ et al., in which an attacker infects a trusted system using malicious software to gain access to its Command and Control (C & C) network (2015). In this context, "BOTS" means infected computers under the attacker's command. McGregor S. et al. have given the network of these robots the name BOTNETS (2013). Sheng L et al. (2012) provide a thorough explanation of how attackers use C&C to control BOTNETs that employ protocols like HTTP and P2P. If BOTNETs are in place, an attacker who sends precise orders to each bot may take full control of the devices. The delay that results from the botnet's delivery of requests to the targeted machines once it has identified the end-IP user's address is inevitable. Either creating a detection algorithm or launching a phishing attack to gather malware components are the most common methods for locating botnets. In order to identify any suspicious activity, network intrusion detection systems perform a log check on both the system and the network. Nameservers, biometrics, and suspicious behaviour are all potential entry points for hackers. A botnet may be used to launch a DDoS assault for a variety of reasons, including the capacity to conceal the identity of the primary perpetrator and the availability of a large number of zombie connections, and the freedom to abuse recommendations and countermeasures. The line between actual and virtual time blurs. In Fig. 1, hostile user systems represented by bots are hacked by attackers, resulting in a decrease in server processing time. BOTNETS are the most viable recommendation for dealing with this problem.

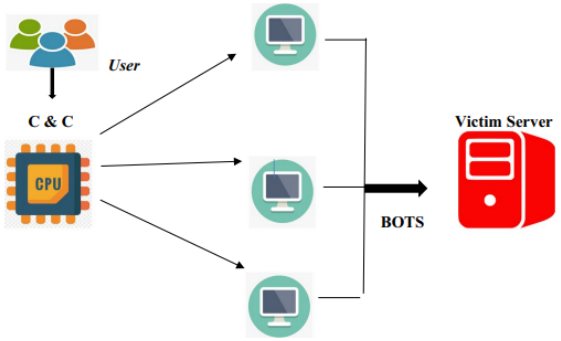

Figure1. DOS Attack Scenario on Victim Server

CYBER DEFENSE ARCHITECTURE

Figure 2 depicts the use of dispersed IDS agents to keep tabs on and safeguard the 5G network's wireless communications. First-Level (FL)-IDS and Second-Level (SL)-IDS are the two types of IDS agents available to us. Network intrusion and anomaly detection solutions rely heavily on machine learning methods. When enabled, the FL-IDS can identify abnormalities using a single-class Support Vector Machine (SVM) method across all access points. We chose a lightweight strategy because the FL-IDS only has a partial picture of the network and the access point doesn't have enough memory to handle a lot of

data. When an anomaly is found, an Anomaly alert is sent from the FL-IDS to the SL-IDS for confirmation and further investigation. The suspected user device's id and other relevant information, such as the pace at which packets are sent, the length of time communication has taken, and the intensity of the signal, are included in this warning. Each robust appliance, base station, and server in the CloudRadio Access Network is equipped with the SL-IDS (C-RAN). To effectively identify threats, the SL-IDS should use a multiclasses machine learning method, thanks to its capacity for handling massive amounts of data. Due to its shorter training and detection durations, a multi-classes SVM algorithm is used in SL-IDS, a component of this cyber protection system. To protect wireless networks from intrusion, SL-IDS monitors potential points of entry, including client devices, APs, BTSs, and C-RAN server nodes.

When the SL-IDS receives an Attack report, it passes the information to the SOC for further detection and feature updates. This report contains information regarding the suspected characteristics and the suspected target (client hardware, passageway, base station, or C-RAN servers). The SoC is furnished with a bunch of recognition and forecast calculations to recognize and expect the new attack designs, and a human (security master) is integrated into the dynamic interaction to bring down the misleading positive rate [8]. It is vital to bring up that the SoC sends new assault examples to the FL-IDS and SL-IDS remotely over time so that they can update their assaults databases. All critical communications (such as Anomaly alerts and Attack reports) are assumed to be encrypted and digitally signed in this architecture. As a result, the IDSs' internal communications are safe against passive assaults.

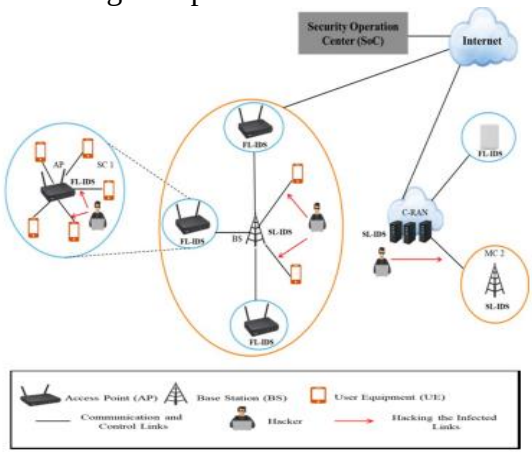

Figure 2. Overview of 5G wireless network security and IDS installations

II LITERATURE SURVEY

The paper's findings are summarised as follows. The section under "Related Works" compares and contrasts previous research on the topic of DDOS assaults. Following a discussion of the suggested approaches for using a numerical model and an AI model to recognize DDOS assaults, we give the actual models. The aftereffects of the exploratory models are momentarily summed up in the Part named "End and Degree for Additional Improvement." The article concludes with a comprehensive summary of the research presented above in the "Declarations" section, along with suggestions for further research and useful resources.

A Review of the Literature DDoS (Distributed Denial of Service) attacks are very common and may severely impair a server's performance, among other bad outcomes. By taking use of either known or unknown security flaws, in a distributed denial of service (DDoS) attack, the attacker exploits compromised computers to send a flood of requests to their target. Transmitting data or processing the attack would take a lot of time and energy in the virtualized environment. This research mitigated DDoS attacks by developing a detection algorithm based on the C.4.5 approach.

When combined with trademark recognition methods, this method yields a classification tree that accurately and quickly detects telltale signs of forgery during Dos and ddos attacks [1]. In [2], we see the first appearance of a deep neural network–based DDoS warning system (Deep Defence). Recent breakthroughs in technology have made it possible to model and infer complex systems by identifying and isolating key high-level qualities from more mundane ones. To better identify patterns in network activity cycles and to track down the sources of network attacks, a continuous recurrent neural network has been set up. Results show progress towards more conventional AI models. When contrasted with an ordinary AI procedure, the pace of

blunder diminishes from 7.517 percent to 2.03 percent with a bigger informational index. Fragrance is a self-sufficient DDoS protection system that makes use of the more centralised management options afforded by the Software Defined Networking (SDN) architecture, as detailed in [3]. ArOMA, in particular, has the potential to eliminate the need for non-trivial human input and the separation of various security processes like monitoring systems, anomaly detection, and mitigation measures. By logically decentralising important security operations throughout an ISP, it enables service providers to deal with DDoS traffic in a way that best meets the demands of their customers, thereby encouraging customers and ISPs to collaborate on DDoS mitigation. Our research shows that even in the face of frequent distributed denial of service (DDOS) assaults, ArOMA is able to keep streaming server performance to an acceptable level. In a lab experiment [4], AI methodologies were tried for distinguishing swarm DDoS assaults. The UNBS-NB 15 and Privation openness datasets, which are both notable for distinguishing malware DDoS assaults, were utilized in this examination.

Machine learning methods like Support Vector Machine (SVM), Artificial Neural Network (ANN), Nave Bayes (NB), Decision Tree (DT), and Reinforcement Classification (USML) are used to initially investigate the accuracy, number of false positives (FAR), sensitivity, specificity, false positive rate (FPR), area under the curve (AUC), and Matthews correlation coefficient (MCC) of datasets. It has been exhibited that the UNBS-NB 15 dataset performs more terrible than the KDD99 dataset. In the field of PC security and a number of related fields, this kind of evaluation is crucial. [5] recommends employing the DBod binary mixture-based botnet detection strategy in light of an assessment of the questioning example of organization action. The proposed strategy exploits the way that hosts contaminated with similar halfway satisfaction of the prerequisites ransomware inquiry similar arrangements of spaces all through the area list, with by far most of these requests falling flat on the grounds that main a little piece of these areas are really connected with a functioning C&C.

Using DNS data gathered over the course of 26 months from a school network, the viability of the suggested technique is assessed. The results demonstrate DBod's efficacy and precision in continually recognising classic and future-looking decision-outcome botnets in real-world connections. In an experimental study, [6] looked at classification methods for detecting Cyberattacks on networks. The UNBS-NB 15 and KDD99 publicity datasets were used in the investigation because of their reputation for detecting malware DDoS attacks. Machine learning methods like Convolutional Neural Network (SVM), Multilayer Perceptron (ANN), Naive Bayes (NB), Decision Tree (DT), and Reinforcement Classification (USML) are used to investigate the accuracy, sensitivity, specificity, false positive rate (FPR), area under the curve (AUC), and Matthews correlation coefficient (MCC) of datasets. Compared to the UNBS-NB 15 dataset, the KDD99 dataset has been shown to be more effective. Its verification is essential in cryptography and other related fields. By combining common machine learning techniques with Domain Name System query data, the authors of [7] present a botnet detection model and assess its efficacy. The irregular woodland calculation has amazing by and large discovery aftereffects of more than 90%, demonstrating that AI calculations might be helpful in botnet identification. Fow table overloading DDoS assaults may cripple SDN-based clouds, therefore [8] suggests a new method of sharing fow tables to forestall this. By utilizing the inactive fow-tables of other OpenFlow switches in the organization, this technique keeps the change from becoming overloaded. With our support, the cloud infrastructure is less vulnerable to data breaches, and the SDN controller has a smaller role to play. So, the burden of communication is reduced.

Many studies in the field of 3D computing corroborate our findings, making our predictions even more solid. In the last three months of 2017, malware assaults on credentials increased dramatically, as seen by an analysis of 7.3 trillion bot queries, which suggested that 40% of password resets were fraudulent. Recently, attackers' preferred method of enslaving computers for botnets has been exploiting vulnerabilities that allow remote code execution, especially in corporate applications. Both the Access The Cloud Server and the Go Ahead embedded HTTP server, which together represent a potential attack surface of 700,000, are among the most frequently targeted targets in the world of cybercrime [9]. To be more specific, [10] is built on the research line "Why? What? How?" We'll start by discussing the significance of aggressive interaction. After that, we talk about the ideas, varieties, and perils of enemy warfare. We conclude by discussing the most common attack techniques and their foundations across all of our different use cases. This paper centers around the spaces of adulterated, photographs, and texts, as well as the crown arraignment classification and techniques of these different data sources, to help specialists in rapidly choosing the sort of examination that best meets their prerequisites even with the rising intricacy of the brain network model. The review concludes with a discussion of its significance and a reference to other studies that have addressed comparable topics. [11] They suggested using Dempster's tandem rule to look at contradictory evidence of multiple cloud strikes in the IT industry of Pakistan. [12, 13] present a primary translation learning structure that utilizes AI procedures to characterize botnets and harmless projects as per botnet-explicit examples of status to completely and characteristics. The picked designs have been completely examined, and the findings show

that, in comparison to other patterns, they have a high detection accuracy and a very low rate of false positives. The support vector machine classifier has shown superior to other methods of classification in both theoretical and empirical studies. [13, 14] and [15] use the Extra-Trees method, network entropy gauge, organizer, data gain proportion, and an internet based continuous semi truck ML answer for DDoS discovery. The unsupervised nature of the method reduces the amount of baseline traffic data utilised in DDoS detection, hence enhancing accuracy and decreasing false positive rates.

In order to estimate how many hackers will attempt to breach a network, the suggested framework combines game theory with hidden markov. However, because to a dearth of training data, the markovprocess-generated training model was unable to identify sophisticated attackers that constantly switch up their methods of assault. In [16], the authors developed a deep learning-based system for detecting anomalies. The purpose of this study is to assess how many botnet assaults can be spotted using anomaly detection systems. In order to identify both common and rare harmful activities, deep learning methods use neural networks and LSTMs. The authors looked at how changing the number of hidden layers in neural networks affected how well graphics processing units (GPUs) and central processing units (CPUs) handled computation. It's especially true when the number of hidden layers is considerable, since this causes an increase in the computational cost. The authors of this research do not elaborate on the deployment of anomaly detection systems inside the network or the targets that these systems keep an eye on. For a software-defined 5G network architecture, see [17] [18] for a proposal of a machine learning detection system. The detection system has three primary layers: a forwarding layer, a control layer, and an intelligence layer. Together, these three work to gather, monitor, and identify dangerous data. The authors compute the most important security metrics, such as detection and prediction rates, false positive and negative rates, and provide an analysis of the system's performance based on these numbers. The essential blemish of the work is that the creators neglect to determine which fundamental parts of the 5G design are being followed by the identification framework. To safeguard cell phones and the 5G center organization from sticking assaults, a drone-assisted 5G communication network is suggested [19]. A reinforcement learning algorithm is used in the framework to make the 5G system resilient against attackers who break connections between mobile and stationary devices. In order to learn and classify drones, the machine learning algorithm relies heavily on the intensity of their signals. The scientists used mathematical and simulation research to demonstrate that the signal-to-interference ratio of drone-assisted 5G networks is improved and the error rate is reduced, even in the face of jamming assaults. Jamming attacks may be avoided, but the authors have not yet created a detection mechanism to prevent them [20].

III MATERIALS AND METHODS

A type of cyberattack known as a denial-of-service attack (DDoS) involves a large number of compromised computers working together to prevent users of a targeted resource from accessing it. This paper specifically proposes a model for detecting distributed denial of service (DDoS) attacks that includes the following components: To use an ML model. An additional throughput study was performed to detect DDoS assaults. Statistical techniques like Logistic Regression, Support Vector Machine, KNN, and Naive Bayes are utilized in the development of machine learning models for the purpose of detecting DDoS attacks.

A.SUPPORT VECTOR MACHINE:

A regulated AI approach known as a SVM can be applied to characterization and relapse issues. Notwithstanding, arrangement issues are its essential application. In this methodology, a solitary direction is utilized to address every data of interest in a n-layered space, where n is the all out number of elements. The data are then categorized by locating the hyperplane that effectively divides the two groups (see the snapshot below).

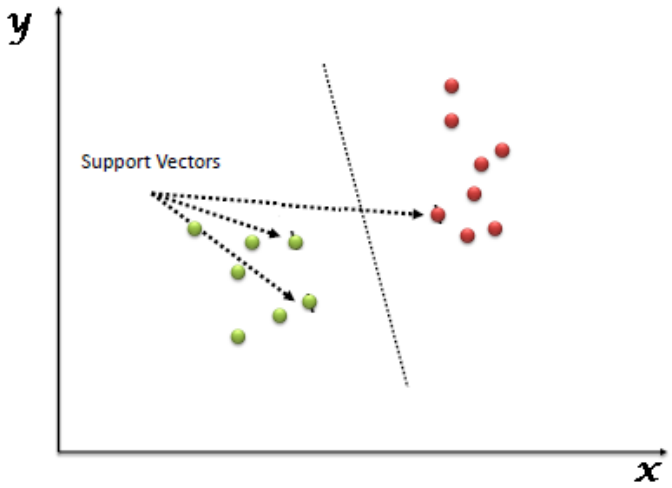

Figure3. Support Vector

The coordinates of each each observation constitute what we call a "support vector." A frontier that excels in distinguishing between these two groups (hyper-plane and line) is the Support Vector Machine.

B.K-NEAREST NEIGHBOR (KNN)

In several fields, including machine learning and pattern recognition, KNN is the most popular algorithm utilised. For difficulties of categorization, KNN is a useful tool. You may know this method by its other name, "instance based" (lazy learning) algorithm. There is a delay between collecting training data and creating a model or classifier to use on future observations. Lazy learning algorithms are superior than eager learning methods because they do not need to build a classifier in advance of classifying a new observation. The importance of this technique increases when frequent modifications and updates to dynamic data are necessary. KNN was used with a range of distance measures. The KNN method uses the following phases, which are based on the formula for the distance between two points in Euclidean space.

Class memberships are determined using k-closest training samples in this non-parametric classification approach. The k-nearest neighbour approach is the most basic and a kind of lazy learning. All calculation is postponed until classification is complete, and classification itself continues to be approximated locally. As a result, in this research, cross-validation is used to give a label to EEG data based on the label most often used by its k closest neighbours.

- Step I: In order to train the system, the feature space must be sent to KNN.
- Step II: The Euclidean distance is the standard measure of long-distance travel.
- Step III: Euclidean distance values may be sorted as follows: d i ≤ d i + 1, where i = 1, 2, 3, …, k
- Step IV: Adjust voting procedures and methods to suit the kind of information being voted on.
- Step V: The size and kind of data input into KNN determines the optimal value of K (the number of closest Neighbors). When working with big datasets, k is maintained at a high value, whereas when working with small datasets, k is likewise kept at a low value.

IV MODULE DESCRIPTION

1. Data loading

The act of moving data from one location to another is referred to as data loading, often a database. Data is often copied from one location and pasted into another, or loaded into a data storage or processing tool, to apply this method. Database-based extraction and loading methods rely on data loading. It is common practise to convert such information into a format compatible with the receiving programme before loading it.

When moving information from a word processor to a database programme, for instance, the original.doc or.txt file is converted to a more suitable.CSV or DAT format. This is often done during or as the final step of the Extract, Transform, and Load procedure. In order to load the data into the target application, it must first be converted from its original format.

2. Data Preprocessing

In order to ensure that all algorithms could deal with missing data, they were imputed. However, unlike imputation, certain algorithms (like XGBoost) can handle missing data on their own. In order to simplify the comparison, the missing values were imputed based on the data type. When dealing with numeric data, the

median value of the remaining items is used to fill in the blanks. When dealing with categorical information, the missing values were substituted with the median of the remaining data points.

3. Data cleaning

Information is polished off in this subsystem. After the data has been cleaned, it is sorted into groups based on what is needed. Data clustering refers to this kind of organisation of information. Then, you may verify whether or not the data set has any blanks. When a value is absent, it might be replaced with a given default. At that point, any necessary re-formatting of the data is completed. Data pre-processing refers to all the steps taken before a prediction may be made. The information is then put to use in the forecasting and predicting process.

4. Data splitting

For every preliminary, we separated the whole dataset into a preparation set of 70% and a test set of 30%. The preparation set was utilized to prepare the model, and the test set was utilized to test the model's adequacy. Resampling and hyper parameter adjustment were also performed on the training set. We assured that the data would be divided in the same way each time the programme ran by specifying a random seed (any random integer).

5. DATA TRAINING

All algorithms are data learners. From the information they've been taught to analyse, they draw conclusions, make choices, and gauge their level of confidence. The precision of the model is straightforwardly corresponded with the nature of the preparation information.

Having sufficient and high-quality training data is just as important as using the right methods for your data project.

Even if you have a large quantity of clean data, the labels may not be ideal for training your model. In addition to raw language, chatbots, for instance, require entity extraction and careful syntactic analysis; Labels are needed for projects that use sentiment analysis to help an algorithm figure out when someone is using slang or sarcasm; autonomous vehicles also require annotated images of every vehicle, pedestrian, street sign, and more.

To rephrase, it is common practise to enrich or label training data before using it. Data's possible you merely need more of it to run your algorithms, and that means you may start collecting it now. However, it's likely that the data you've collected isn't yet suitable for training your classifiers.

For the simple reason that good training data is essential for building a good model. And we have some experience with it as well. After all, we have classified more than 5 billion rows of data for some of the most cutting-edge companies in the world. Your data can be in any format, including images, text, audio, or any combination thereof; We can assist you in creating the training set that will guarantee the success of your models.

6. Prediction and forecasting based on collected data:

It is at this stage that the preprocessed data is used to make a forecast. Any of the aforementioned procedures is appropriate for making this prediction. However, Linear Regression outperforms the competition in terms of prediction accuracy. Therefore, the linear regression technique is employed for forecasting purposes in this project. To do this, the prepared dataset is partitioned into train and test sets. Then, using the learned value as a basis for prediction, a predictive object is constructed. After then, the item is put to work predicting information for the next years.

7. KEY PERFORMANCE INDICATORS

Three quarters of the data were utilized for testing, while seventy percent were utilized for training. Enthought Canopy was used to apply all six algorithms on the same dataset.

$$\text{Accuracy} = (TP+TN) / (P + N)$$

The major metric we utilised to assess the success of our effort was the quality of its predictions. Equation may be used to circumvent accuracy. An algorithm's accuracy may be measured by how often it produces desired results.

8. CONFUSION MATRIX:

For the simple reason that it can be used to calculate other crucial metrics like accuracy, recall, precision, etc., it has become the de facto standard for evaluating predictive analysis. An N-by-N matrix is used to describe a model's performance in a classification problem, where N is the number of class labels.

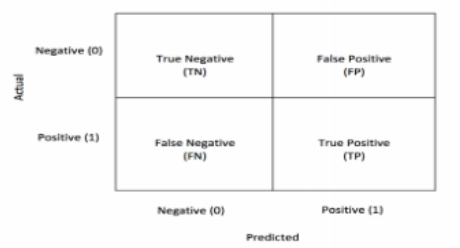

Figure4. Confusion Matrix

Divided by the total number of both positive and negative predictions. A table lists the expected True Positive (TP), True Negative (TN), False Negative (FN), and False Positive (FP) outcomes for each algorithm.
There are a plethora of metrics that may be used to evaluate a system's efficiency. To evaluate an intrusion detection system, repeated tests are performed.

- True positive (TP): Total number of occasions when an intrusion was found and subsequently fixed.
- True Negative (TN): how many times an incident was wrongly labelled as an invasion.
- False-positive (FP): How many intrusions were mistakenly labelled as benign.
- False-negative (FN): How often a typical occurrence was mistakenly labelled as a security breach.

V.SYSTEM DESIGN

A system is designed when its user interface, components, and data are defined in order to meet a set of predetermined requirements. One way in which system theory is put to use is in the process of designing systems. The primary objective of system design is to produce the system architecture by providing the knowledge and data required for system implementation.

1.ARCHITECTURE DIAGRAM:

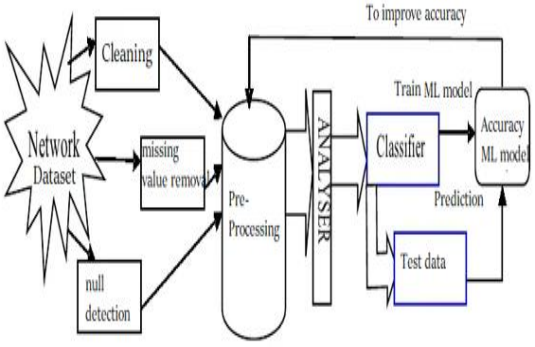

Figure5. Architectural Diagram for Machine Learning Model

2.DATA FLOW DIAGRAM:

The data flow in an enterprise IT system may be represented visually with the help of data flow diagrams. Data Flow Diagram (DFD) is a method for explaining how a system gathers information from its input and stores and uses that information to produce outputs like files and reports. Data flow diagrams are divided into two categories: physical and logical. A visual representation of the information flows required to complete a given task is the logical data flow diagram of a business. For further information on how the logical data flow is realised, have a look at the corresponding physical data flow diagram.

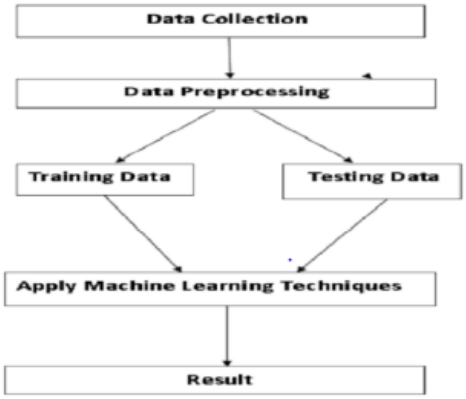

Figure6. DataFlow Diagram

3. USECASE DIAGRAM:

The use cases, the actors who engage with the system, and the connections between the two are all shown in a use case diagram.

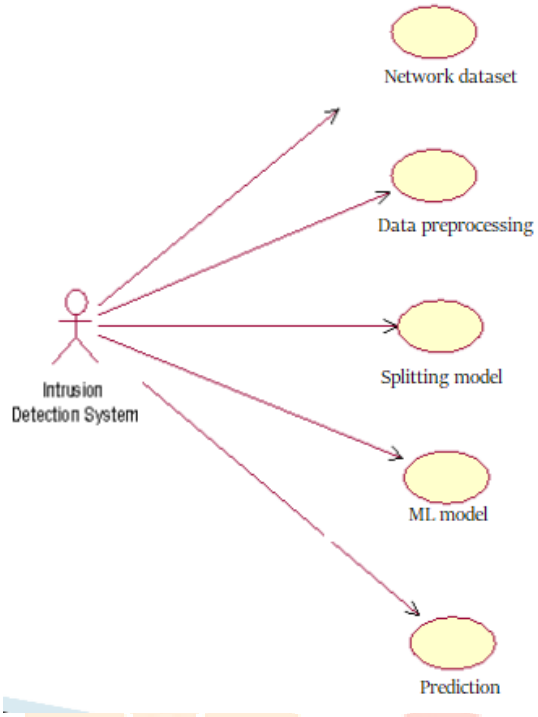

Figure7. Usecase Diagram

4. CLASS DIAGRAM:

There is no dynamic information conveyed by a class diagram. It stands in for the unchanging perspective of a programme. You may use a class diagram to build the code that runs an application as well as to visualise, describe, and document its components.

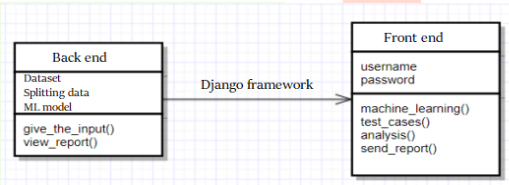

Figure8.Class Diagram

VI. CONCLUSION AND FUTURE ENHANCEMENT

The current system employs a variety of counterattack strategies, such as CAPTCHA puzzles, which provide a straightforward method for preventing attacks, but have been proved to be ineffectual in recent research. While the model's enhanced accuracy in intrusion detection was promising, it was unable to identify more common forms of denial-of-service (DoS) attack, thus a digital signature for network flow analysis utilising meta-heuristic approaches was developed to analyse the suspicious activity.

DRAWBACKS:

- Low precision
- Data from networks is gathered, but only in small quantities.
- To make matters worse, we're still relying on antiquated data sets that can't be utilised to foresee the kinds of assaults that
- Model training is a time-intensive process.

Methodology Suggested:

IDS (anomaly base or misuse base) (anomaly base or misuse base). Attacks are detected using a system that looks for deviations from the usual in order to prevent further damage. Due to its effectiveness in detecting novel forms of infiltration, this sort of intrusion detection system is increasingly in demand. In the suggested setup, the existence of an attack is detected using a support vector machine (SVM) and a KNN classifier.

ADVANTAGES:

- The time needed to train the model is minimal.
- Very precise
- When training a model, a lot of network data is often used.

- It is possible to foresee the emergence of a new kind of assault.

Here, we provide a novel dataset that includes contemporary forms of attack that have not been included in previous studies. There are 27 outliers and 5 categories in this dataset. The data collected has been catalogued for use in future attacks against the System and Application layers. The collected dataset was put through three different machine learning algorithms (support vector machine, random forest, and k-nearest neighbour) in order to better define the four distinct types of DDoS attacks (Smurf, UDP-Flood, HTTP-Flood, and SIDDOS).

As this paper's study demonstrates,

In making the detection of DDoS assaults easier and more effective. The usefulness of machine learning techniques is demonstrated through a comparison of logistic regression and naive bayes. In order to do effective analysis, real-time data sets are utilised. Statistical methods like logistic regression and naive bayes were favoured above other options because of the positive outcomes they produced. Similarly, mathematical models need to be put through their paces by simulating real-world assaults with machine learning. The methods described in this study may also be used to defend against assaults on firewalls and memory management. As such, the proposed model is limited by the fact that it was trained on a single dataset. Therefore, it is possible to analyse a decentralised dataset to determine where improvements should be made

REFERENCES

[1] C. Yao, X. Wang, Z. Zheng, G. Sun, L. Song, “Edge Flow: Open-source multi-layer data flow processing in edge computing for 5G and beyond”, IEEE Network Magazine, 2018, pp.1-8.

[2] F.Z. Yousaf, M. Bredel, S. Schaller, F. Schneider, “NFV and SDN key technology enablers for 5G networks”, IEEE Journal on Selected Areas in Communications, Vol 35, Issue 11, 2017, pp. 2468-2478.

[3] A. Gupta, R. Kumar Jha, P. Gandotra, S. Jain, “Bandwidth spoofing and intrusion detection system for multi stage 5G wireless communication network”, IEEE Transactions on Vehicular Technology, Vol 67, Issue 1, 2018, pp.618-632.

[4] N. Adem, B. Hamdaoui, A. Yavuz, “Mitigating jamming attacks in mobile cognitive networks through time hopping”, Wireless Communications and Mobile Computing, Vol 16, Issue 17, 2016, pp. 3004- 3014.

[5]L.F.Maimó, A.L.P. Góme, F.J.G. Clemente, M.G. Pérez, G.M. PÉREZ, “A self-adaptive deep learning-based system for anomaly detection in 5G networks”, IEEE Access, Vol 6, 2018, pp. 7700-7712.

[6] J. Li, Z. Zhao, R. Li, “Machine learning-based IDS for software-defined 5G network”, IET Networks, Vol 7, Issue 2, 2018, pp. 53-60.

[7] X. Lu, L. Xiao, C. Dai, “Uav-aided 5G communications with deep reinforcement learning against jamming”, arXiv preprint arXiv:1805.06628, 2018.

[8]H.Sedjelmaci, F. Guenab, A. Boudguiga, “Cooperative security framework for CBTC network”, IEEE ICC, Kansas City, MO, USA, 2018.

[9] A. Servin, D. Kudenko, “Multi-Agent Reinforcement Learning for Intrusion Detection”, European Symposium on Adaptive and Learning Agents and Multi-Agent Systems, Springer, 2008, pp. 211-223.

[10]B.Tomasik, “Ethical issues in artificial reinforcement learning”. https://reducing-suffering.org/ethical-issues-artificial-reinforcementlearning/, 2017.

[11] T. Alpcan and S. Buchegger, “Security games for vehicular networks", IEEE Trans. Mobile Comput., Vol. 10, No. 2, 2011 pp. 280–290.

[12] H. Sedjelmaci, SM. Senouci, N. Ansari, "Intrusion detection and ejection framework against lethal attacks in UAV-aided networks: a Bayesian game-theoretic methodology," IEEE Transactions on Intelligent Transportation Systems, Vol 18, Issue 5, 2017, pp. 1143 – 1153.

[13] 3GPP TS 09.60, “General Packet Radio Service (GPRS); GPRS Tunnelling Protocol GPT) across the Gn and Gp Interface”, 2015

[14] 3GPP TS 29.281, “General Packet Radio System (GPRS) Tunnelling Protocol User Plane (GTPv1-U)”, 2015.

[15] ETF RFC 7540: "Hypertext Transfer Protocol Version 2 (HTTP/2)".

[16]. Zekri M, Kafhali S, Aboutabit N, Saadi Y, "DDoS attack detection using machine learning techniques in cloud computing environments", 3rd international conference of cloud computing technologies and applications (CloudTech), pp 1–7,2017. hthttps://doi.org/10.1109/cloudtech.2017.8284731.
[17]Xiaoyong Yuan, Chuanhuang Li, Xiaolin Li, "DeepDefense: Identifying DDoS Attack via Deep Learning", IEEE International Conference on Smart Computing (SMARTCOMP), 2017.
[18] Sahay R, Blanc G, Zhang Z, Debar H. Aroma: an SDN based autonomic DDoS mitigation framework. Computer Security. 2017;70:1–18. https://doi.org/10.1016/j.cose.2017.07.008.
[19]Antonakakis M, April T, Bailey M, Bernhard M, Bursztein E, Cochran J, Kumar D, .Understanding the miraiBotnet. USENIX security symposium, 2017.
[20] Wang TS, Lin HT, Cheng WT, Chen CY. "DBod: Clustering and detecting DGA-based botnets using DNS trafc analysis. Computer Security. 2017;64:1–15.