# A Novel Intrusion Detection Model for Detecting Known and Novel Cyber-attacks Using Convolutional Neural Network

POOJA K S
4th Sem, MCA, AMCEC
pooja2018ks@gmail.com

Sravanthi kala
prof, MCA, AMCEC
Sravanthi.kalal@gmail.com

**Abstract**

*Today's cyber world faces a serious security problem with the use of new breach control techniques. a wide variety of tactics Learning approaches for developing a new intrusion prevention system have been developed. Therefore, we developed machine learning techniques to prevent infiltration. Developing a New Investigation Methodology. By using the approach, we may detect infiltration and identify the attacker's specifics. IDS are divided into two categories: host-specific and network-based. A host-based intrusion detection system (HIDS) monitors certain hosts or networks and alerts the user if suspicious activity is found. such as altering or deleting an internal document, making a sudden flurry of system calls, or making unforeseen changes to the settings. Network equipment like firewalls and bridges frequently have a network-based security mechanism (NIDS) linked to it to detect and prevent network threats. The KDD cup IDS dataset from the current research was obtained from an information source. The pre-processing techniques must then be used. The next step is to use a variety of deep learning and machine learning approaches, such as logistic regression (LR) and traditional neural networks (CNN). The query data indicates the accuracy of the aforementioned algorithms. After that, FLASK may be used to deploy the project as an internet application.*

**KEYWORDS:** *Cyber security, HIDS NIDS Machine Learning, GCR-MN, CNN.*

## I.INTRODUCTION

The frequency of hacking incidents increases as technology progresses. Each year, businesses report a sizable number of hacking incidents. A service interruption distributed attack targeting Estonian organizations was attempted in 2007. On June 17, 2008, Aws started getting a lot of verified requests from clients at one of its locations. Servers all across the world started to slow down noticeably as the quantity of as requests started to rise quickly. Facebook was purportedly the target of a distributed denial of service attack on September 28, 2014, according to information provided by ENISA, the European Institute for Network and Information Security, in January 2013. Cyberattacks start with a warning in 50% of cases. A firewall controls network traffic based on where the point of origin or reception is located. It adjusts collection in line with the barrier's rules. Other restrictions include the total quantity of data that routers may access and their understanding of the hosts that are receiving the content. Before alerting a network or system management, an intrusion detection system (IDS) monitors network traffic and evaluates the operating system for anomalous activity. Connected devices, such as firewalls and Connectors, are frequently configured as NIDSs to detect unauthorized network activity. At the highest level, these IDS's employ three different categories of detection methods: surveillance comes in many forms, including surveillance of violence, deviance, and hybridization. IDS maintains a set of guidelines for its abuse detection method. Our project's main objectives are to: • Identify or anticipate assaults effectively; and • Implement multiple categorization techniques for increased effectiveness. to improve the overall performance of neural networks used for categorization. to activate the online application.

## II. Literature Survey:

Due to the addition of adjustable components, Software Defined Networking Technology (SDN) has the possibility to effectively identify and track security issues. Machine learning (ML) techniques are currently being used in SDN-based Network Intrusion Detection Systems (NIDS) to safeguard servers and satisfy privacy concerns. Under the context of SDN, deep learning technology (DL) is emerging as a stream of sophisticated machine learning approaches. In this investigation, we looked at the majority of current techniques to computational learning (ML) use Ids to generate NIDS. many specific, we examined deep learning methods in order to determine the justification for developing SDN-based NIDS. Meanwhile, the main disadvantages of many feature learning algorithms are their complexity and high implementation costs. [2] a detailed examination of assaults with several steps Since the beginning of the Internet, online attacks have been a risk to people and organizations. Their complexity has increased in tandem with computer networks. Attackers now have to go through several invasive processes in order to accomplish their final purpose. The assortment of the terms "multi-step attack," "multi-stage attack," and "attack scenario" refer to different processes. Finding intrusions can be difficult because of the multi-step structure of the assault, which makes it necessary to correlate several activities in order to comprehend the attack strategy and assess the danger. Beginning in the first decade of the 2000s, the security research community has worked hard to create ways for detecting that sort of threat. threat and forecasting future actions. The goal of this survey is to collect all articles that propose multi-step assault detection systems I concentrate at methods that examine the attack and the connections among Its stages rather not purely looking for symptoms. We do thorough bibliographic research to find relevant information. The result of our effort is a corpus of 181 papers that includes 119 approaches, which we classify and characterize. After reading the articles, they may draw some judgments on the state of the multi-step identification of assaults research. The advantage of this method is that it monitors dangerous network events as they occur and employs IDS features to identify them. It does this by searching for matches based on an IP address or channel. The attacker can carry out a multi-step operation in any order they want because it is not required. The full cast of alternate action scenes might be rather complicated. [3] Using grouping to find abnormalities in real time Real-time network data has increased significantly recently as a result of the expanding use of linked Internet of Everything sensors. Network assaults cannot be prevented, hence real-time information on networks anomaly detection has become crucial. Critical comparison analysis is carried out using harmonic aggregation, exclusion trees, k-means, hierarchical density-based spatial clustering of applications with noise (HDBSCAN), and agglomerative clustering. The assessment results showed the suggested framework's usefulness with a much better accuracy rate of 96.51% when compared to other algorithms. In terms of memory utilization and execution time, the proposed structure performs better than the alternatives. The predicted technique also enables analysts to [4] Automated Methods for Connect Awareness Evaluate intrusion detection Anomaly detection is essential for identifying and thwarting security breaches because unusual network activity may indicate a possible network intrusion. Signature-based monitors are used in many of the early investigations in this field and in commercial Intrusion Detection Systems (IDS). Since the historical profile must be updated when fresh attack signatures become known, signature-based approaches are inefficient for real-time net anomaly diagnosis. Anomaly detection has recently seen a rise in use of data mining approaches. In deep learning, we use entropy computing to apply and evaluate seven different methods. Kyoto 2006+ data set research. Our findings suggest that on this data set, such specialized machine learning algorithms outperform. [5] Investigating Shodan, for instance, from the viewpoint of Industrial Control Systems (ICS), a vital piece of key infrastructure that is becoming more open to cyberattacks. The threat increased with Shodan's search feature's launch. A popular tool in the toolboxes of attackers and penetration testers, Shodan is an exploration tool that can locate and index industrial control equipment that is linked to the Internet. In this study, we use honeypot technology to conduct comprehensive research. A lookup engine is a Shodan. Six dispersed honeypot systems are first set up, and three months' worth of traffic data is gathered. We build hierarchies.

**SYSTEM ARCHITECTURE**

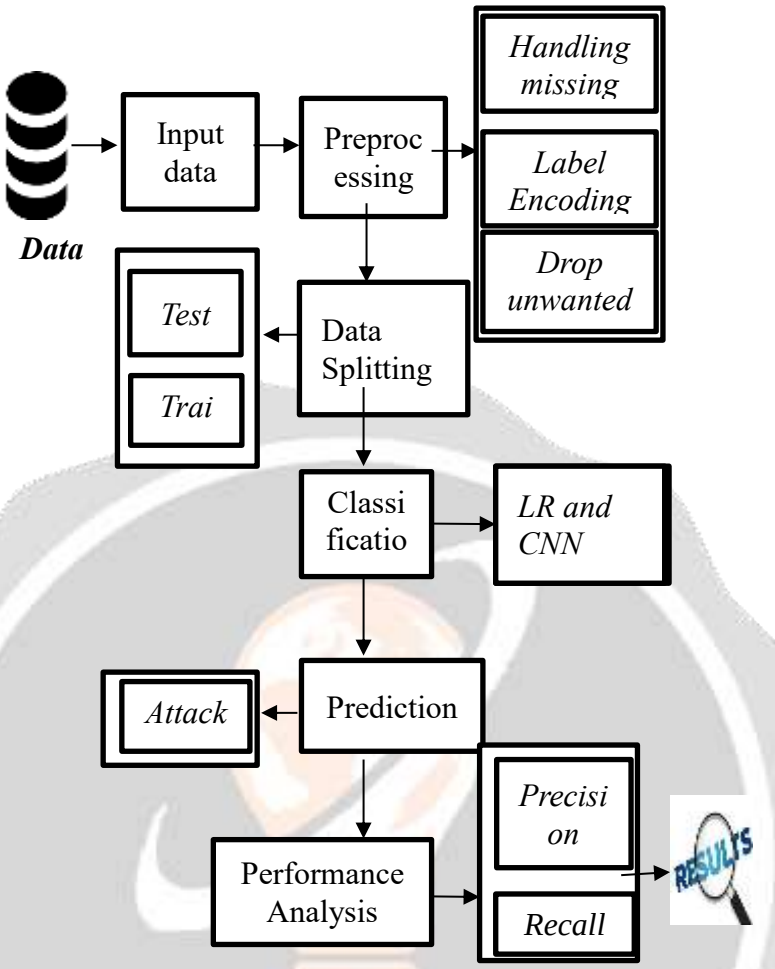


**Fig.1. System Architecture**

## III. Existing System:

The implementation of Network Intrusion Detection Systems (NIDS) is our primary area of interest in the present, hence this study evaluates the tools in this field. Along with the settings, free and open-source network sniffer software is currently used for NIDS development. In terms of design, sensing approach, testing techniques, handled dangers, and algorithm implementations, it evaluates, contrasts, and finally examines state-of-the-art NIDS principles in an Internet of Things environment. In addition to discussing prospective possibilities, this review also compares and contrasts traditional and machine learning (ML) NIDS methods. We want to focus on IoT NIDS implemented through neural networks in the current study since learning algorithms have a very high success rate in terms of security and privacy. Unlike previous top polls that concentrate on conventional methods, this research offers a thorough analysis of NIDSs using multiple learning algorithms for the Internet of Things (IoT). We believe that the investigation will benefit academic and industrial research in three different ways: first, it will help them identify IoT risks and issues; second, they can use their own NIDS; and third, it will help them propose creative, clever solutions to IoT limitations. Security experts will benefit from the poll's ability to distinguish between traditional and IoT NIDS. They believe the inquiry Researchers from academia and industry will benefit from it in the following ways: first, it will help them identify IoT risks and issues; second, it will help them put their own NIDS into practice; and third, it will help them propose creative, clever solutions to IoT constraints. The survey will help security professionals distinguish between IoT NIDS and traditional NIDS.

DISADVANTAGES: When compared to the suggested algorithm, the results are poor; For massive amounts of data, it is ineffective;

## IV. Proposed System:

The individual in question used the IDS's dataset as KDD feed. The dataset source was used to get the data for the input. Next, we must complete the step of data pre-processing. We must handle the missing numbers and encrypt the label for the given data in order to prevent erroneous predicting at this level. The dataset then has to be split into two sections: Check your preparation. Information is segregated based on a magnitude relationship. Most of the information will be available in-train. In the test, just a portion of the data will be available. A tool for evaluation is provided by the teaching portion. Using the model and testing components, the model is anticipated. The installation of a computerized, deep learning-based classification technique comes next. Logistic regression is one of the machine learning techniques. Deep learning methods like Convolution Neural Network (CNN) are used. Lastly, accuracy.

ADVANTAGES: It handles a variety of information sets well, and the experiment's results outperform the present approach. It is quick to complete.

## V. IMPLEMENTATION

**MODULES:**

- Data selection
- Preprocessing
- Data splitting
- Classification
- Result generation

**[1] Data Selection**

A list of numbers served as the source of the provided data. As part of our strategy, we use the KDD Cup IDS dataset. Data selection is the first step in the detection of attacks. An example of where the input data was obtained was from the UCI storage facility. The collection of values includes details on the protocol, duration, host error rate, label, and other variables.

**[2] Pre-processing**

Pre-processing data is a technique for deleting unnecessary data from a dataset. Using data conversion techniques for preliminary processing, the collection of observations is changed into a structure that is suitable for machine learning. This process also involves cleaning the collection to increase its efficiency by getting rid of any unnecessary or broken items that might jeopardize the correctness of the dataset. Remove any omitted information

**[3] Data splitting**

For teaching to take place, the machine learning process has to have information on its side. Results from tests are also needed to assess how well an algorithm performs and determine how effective it is in addition to what is needed for training. Our strategy involved treating 30% of the input dataset as test data and 70% as training data. The technique of separating a set of readily available data into two halves, often for cross-validator reasons, is known as data division. The purpose of one set of data is to create a prediction model, whereas the purpose of the other is to assess the effectiveness of the machine. While researching data mining approaches, it is crucial to divide the data into training and testing sets.

**[4] Classification**

A number of machines learning techniques, including LR and CNN, must be included along our procedure. Confronting classification issues can be aided by artificial intelligence. It uses the logistics regression methodology as a method of analysis for predictions that are based on the idea of probability. According to the principle of linear regression, the expenditure variable can only have a value between 0 and 1D. CNN can tell from accelerometer data if a person is standing, walking, or leaping. This understanding has two facets. The first dimension is the number of time steps, and the second is the magnitude of the velocity on three axes. The movement of the kernel is shown in the charts below based on accelerometer data.

**[5] Result generation**

The last method for determining the overall outcome will be to apply the overall classification and projection. This recommended strategy's effectiveness is assessed using metrics like, Precision The classifier's capabilities are described by its accuracy. How effectively a certain predictor can forecast the value of a predicted feature for fresh data is described by the classifier's accuracy. It predicts the class name correctly. Precision a: AC = (TP+TN)/(TP+TN+FP+FN) Clarity is calculated by dividing the total of real positives and false positives by the number of true positives.

## V1. CONCLUSION

Our conclusion is that the source was the Strategy for the Knowledge discovery and extraction (Cup dataset). The information for inclusion was put out during the present piece. We put machine learning techniques for identification to use. Deep learning and machine learning methods such as logistic regression and convolutional neural networks are then used. Finally, the findings show that accuracy is known using the aforementioned approach and calculating performance metrics including accurateness, precision, recall, and f1 ranking. We'd want to quickly combine two different machine learning or deep learning algorithms. The suggested grouping and classification methods could be enhanced or changed in the future to get even better outcomes. Further combinations and methods of clustering may be employed to improve identification precision, in addition to a variety of tried-and-true data mining techniques. Finally, by including a preventative mechanism in the sentiment analysis detecting system, the equipment' efficiency may be improved.

## REFERENCES

[1] Anderson, James P., Computer Security Threat Monitoring and Surveillance, James P. Anderson Co., Washing, PA, 1980.

[2] V. V. R. P. V. Jyotsna, V. V. Rama Prasad, and K. Munivara Prasad, "A review of anomaly-based intrusion detection systems," International Journal of Computer Applications, vol. 28, no. 7, pp. 26-35, 2011.

[3] Ciaburro, G., and B. Venkateswaran, "Neural networks with R: smart models using CNN, RNN, deep learning, and artificial intelligence principles." Packt Publishing, Birmingham, UK, 2017.

[4] M. Roopak, G. Yun Tian, and J. Chambers, "Deep learning models for cyber security in IoT networks," in Proceedings of the IEEE 9th Annual Computing and Communication Workshop and Conference, pp. 0452-0457, Las Vegas, NV, USA, 2019.

[5] R. Vinayakumar, K. P. Soman, and Prabaharan Poorna Chandran are the authors of this paper. "Applying convolutional neural networks for network intrusion detection," Proceedings of the International Conference on Advances in Computing, Communications, and Informatics, pp. 1222-1228, 2017.

[6] Ahmim, Ahmed, et al. "A novel hierarchical intrusion detection system based on decision tree and rules-based models," In Proceedings of IEEE 15th International Conference on Distributed Computing in Sensor Systems, pp. 228-233, 2019.