Learning Network Behavior Patterns for Predictive Intrusion Recognition

Mr.Jajjara Bhargav
Cyber Security
Chalapathi
bhargavchalapathi@gmail.com
Mungamuri Venkata Gopiraju
Cyber Security
Chalapathi
mungamurivenkatagopiraju@gmail.com

Cherukuri Joy Jonathan
Cyber Security
Chalapathi
joyjoel12345@gmail.com
Gudari gopiraju
Cyber Security
Chalapathi
gopirajugudari@gmail.com

Dandu Rohith Reddy
Cyber Security
Chalapathi
rohithreddydandu@gmail.com

***Abstract*— Current Network Intrusion Detection Systems (NIDSs) face sustainability and viability issues in modern networks, with concerns about reduced detection accuracy and increased human intervention. This project introduces a novel intrusion detection method based on deep learning. A Deep Neural Network (DNN) model is created and trained using the NSLKDD Dataset, focusing onessential features out of the available. Utilizing the NSL-KDD Dataset, the project emphasizes a streamlined approach by selecting key features. This approach aims to enhance the efficiency of intrusion detection. The project underscores the significant potential of deep learning in NIDS, showcasing the capability of the proposed method to address identified challenges more effectively than existing approaches. The efficiency of the proposed deep learning method is demonstrated through a comparative analysis with previous studies. Performance metrics such as precision, accuracy, recall, and F-measure values are used to validate and highlight the method's effectiveness. And the project for enhancing the intrusion detection's accuracy , CNN and hybrid method combining Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architectures are employed. This amalgamation aims to leverage the strengths of both CNN and LSTM, providing a more robust and accurate prediction model for network intrusion detection. The CNN achieved impressive 99% accuracy. Furthermore, we extended the usability of our system by developing a user-friendly front end using the Flask framework, ensuring seamless user testing, and incorporating secure user authentication for enhanced network security during intrusion detection.**

***Keywords*— NIDSs, Deep Learning, NSL-KDD.**

I. INTRODUCTION

Illegal and hostile users can always break, infiltrate, or penetrate a computer network. The network, which stores materials for authorized users to share, attracts illegitimate users who abuse it. Many protective policies are rare and difficult to implement. Security breaches or incursions are serious issues for any company. Therefore, preemptive steps are needed to defend the organization's interests from the different forms of attacks it faces[1].

Intruder detection systems identify malicious network activities. It should detect many external assaults and security breaches. The system should monitor unethical or abusive insiders. There are three categories of intruders: Unauthorized users use valid user accounts to masquerade. Insider misfeasors abuse their rights and access unauthorized resources. An insider or outsider attempting administrative access to the system is clandestine action [2].

Network intrusion detection systems find malicious packets. Real-time traffic monitoring detects network anomalies. Real-time jobs are needed due to big data, robust computing, and network expansion. Thus, unlike old approaches, NIDS monitoring should be cautious, correct, and precise. The rapid precision development of machine learning algorithms is impressive. Its birth was driven by the need for higher performance across diverse network types due to the increased scope and diversity of security threats [3,4,15].

Software defined network (SDN) implementation has expanded network-based intrusion detection system deployment. Large amounts of network data and connected devices pose security risks. The rise of AI, quantum computing, and IoT has increased the hazard level [4].

Network security (NIDS) challenges include creating a reliable and effective network intrusion detection system [2,3,4]. NIDS technology has advanced, although most systems use signature-based methods instead of anomaly detection. The system's behavior dynamics, high false mistake rate (and costs), difficulty gathering quality training data, and length of training data all contribute to its resistance to change. Such techniques will eventually lead to unreliable and inaccurate detection[4].

II. RELATED WORK

Network intrusion detection systems (NIDSs) play a crucial role in defending computer networks[3,4,15]. However, there are concerns regarding the feasibility and sustainability of current approaches when faced with the demands of modern networks. More specifically, these concerns relate to the increasing levels of required human interaction and the decreasing levels of detection accuracy. This paper presents a novel deep learning technique for intrusion detection, which addresses these concerns. We detail our proposed nonsymmetric deep autoencoder (NDAE) for unsupervised feature learning. Furthermore, we also propose our novel deep learning classification model constructed using stacked NDAEs. [1] Our proposed classifier has been implemented in graphics processing unit (GPU)-enabled TensorFlow and evaluated using the benchmark KDD Cup '99 and NSL-KDD datasets. Promising results have been obtained from our model thus far,

demonstrating improvements over existing approaches and the strong potential for use in modern NIDSs.

Software Defined Networking (SDN) is a networking model that allows for greater dynamic control of a networking environment [3,4,15]. With today's increasingly complex networking environment, SDN networks allow for a greater degree of control and flexibility of a network. This is accomplished through the separation of the control and data planes, as well as the implementation of a global programmable controller. A Network Intrusion Detection Systems (NIDS) can work very well with SDN networks as it can help monitor the overall security of a network by analyzing the network as a whole and making choices to defend the network based on data from the entire network. Using a Hidden Markov Model (HMM), a NIDS could monitor a network and learn from the evolving network activity of the present and react accordingly[2]. This machine-learning NIDS could improve the efficiency of security applications and increases the range of activities that they are able to accomplish. In this paper we plan to demonstrate the possibility of using Hidden Markov models to develop an adaptive NIDS for use in the new emerging technology of SDN.

Network intrusion detection systems (NIDSs) play a crucial role in defending computer networks. However, there are concerns regarding the feasibility and sustainability of current approaches when faced with the demands of modern networks. More specifically, these concerns relate to the increasing levels of required human interaction and the decreasing levels of detection accuracy. This paper presents a novel deep learning technique for intrusion detection, which addresses these concerns. We detail our proposed nonsymmetric deep autoencoder (NDAE) for unsupervised feature learning. Furthermore, we also propose our novel deep learning classification model constructed using stacked NDAEs. Our proposed classifier has been implemented in graphics processing unit (GPU)-enabled TensorFlow and evaluated using the benchmark KDD Cup '99 and NSL-KDD datasets [15]. Promising results have been obtained from our model thus far, demonstrating improvements over existing approaches and the strong potential for use in modern NIDSs.

This paper proposes a novel scheme that uses robust principal component classifier in intrusion detection problem where the training data may be unsupervised. Assuming that anomalies can be treated as outliers, an intrusion predictive model is constructed from the major and minor principal components of normal instances [5]. A measure of the difference of an anomaly from the normal instance is the distance in the principal component space. The distance based on the major components that account for 50% of the total variation and the minor components with eigenvalues less than 0.20 is shown to work well. [4,6,12] The experiments with KDD Cup 1999 data demonstrate that our proposed method achieves 98.94% in recall and 97.89% in precision with the false alarm rate 0.92% and outperforms the nearest neighbor method, density-based local outliers (LOF) approach, and the outlier detection algorithms based on Canberra metric.

During the last decade the analysis of intrusion detection has become very important, [7] the researcher focuses on various dataset to improve system accuracy and to reduce false positive rate based on DAPRA 98 and later the updated version as KDD cup 99 dataset which shows some statistical issues, it degrades the evaluation of anomaly detection that affects the performance of the security analysis which leads to the replacement of KDD dataset to NSL-KDD dataset [3,4,15]. This paper focus on detailed study on NSL- KDD dataset that contains only selected record. This selected dataset provide a good analysis on various machine learning techniques for intrusion detection.

III. Materials And Methods

The proposed system introduces a novel intrusion detection method using a Deep Neural Network (DNN), trained on the NSLKDD Dataset with essential features [11,14]. Leveraging deep learning, it aims to enhance NIDS efficiency by categorizing patterns in network activity, providing a data-driven approach for early detection and decision-making. The use of a neural network model facilitates handling numerous parameters for generalized results. And also in the project for enhancing the intrusion detection's accuracy , CNN and hybrid method combining Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architectures are employed . This amalgamation aims to leverage the strengths of both CNN and LSTM, providing a more robust and accurate prediction model for network intrusion detection. The CNN achieved impressive 99% accuracy. Furthermore, we extended the usability of our system by developing a user-friendly front end using the Flask framework, ensuring seamless user testing, and incorporating secure user authentication for enhanced network security during intrusion detection.

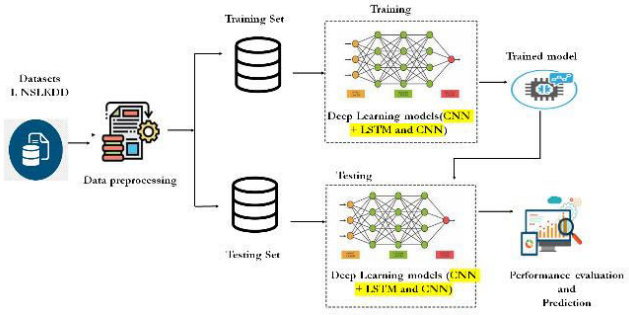

Fig. 1. System Architecture

This study aims to provide a deep-learning-based method for detecting network intrusions. The method uses a deep network to educate itself on anomaly patterns to discriminate between ordinary connections and intrusions. The method also seeks to reduce the number of false alarms to an absolute minimum. The method is adaptable enough to consider new incursion patterns and potential changes in the subject's behavior. The suggested system uses a deep network system trained using the NSL-KDD dataset (encoder with Label Encoder), as shown in Figure 1.

A) Dataset Collection

The public NSL-KDD dataset was created from the KDD cup99 dataset. A statistical investigation of the cup99 dataset revealed flaws that greatly affect intrusion detection accuracy and misjudge AIDS. The massive number of duplicate packets in KDD is the fundamental issue. This large number of duplicate instances in the

training set would bias machine-learning systems toward normal instances and prevent them from learning irregular examples, which are more harmful to the computer system. The NSL-KDD dataset in 2009 from the KDD Cup'99 dataset to eliminate duplicates. The NSL-KDD train dataset has 125,973 records and the test dataset 22,544. The NSL-KDD dataset is large enough to use without sampling randomly. Multiple studies have yielded similar and comparable results. The NSL_KDD dataset has 22 training intrusion attacks and 41 features. This dataset has 21 connection attributes and 19 host-specific attributes.

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_srv_count	dst_host_same_srv_rate	dst_host_d
0	0	tcp	http	SF	181	5450	0	0	0	0	...	9	1.0	
1	0	tcp	http	SF	239	486	0	0	0	0	...	19	1.0	
2	0	tcp	http	SF	235	1337	0	0	0	0	...	29	1.0	
3	0	tcp	http	SF	219	1337	0	0	0	0	...	39	1.0	
4	0	tcp	http	SF	217	2032	0	0	0	0	...	49	1.0	

5 rows × 42 columns

Fig 2 NSL KDD dataset

B) Data Processing:

Data processing involves transforming raw data into valuable information for businesses. Generally, data scientists process data, which includes collecting, organizing, cleaning, verifying, analyzing, and converting it into readable formats such as graphs or documents. Data processing can be done using three methods i.e., manual, mechanical, and electronic. The aim is to increase the value of information and facilitate decision-making. This enables businesses to improve their operations and make timely strategic decisions. Automated data processing solutions, such as computer software programming, play a significant role in this. It can help turn large amounts of data, including big data, into meaningful insights for quality management and decision-making.

C) Feature selection

Feature selection selects the most consistent, non-redundant, and relevant features for model construction. As databases grow in quantity and variety, methodically lowering their size is crucial. The basic purpose of feature selection is to increase predictive model performance and reduce computing cost.

One of the key parts of feature engineering is picking the most significant characteristics for machine learning algorithms. To decrease input variables, feature selection approaches eliminate redundant or unnecessary features and narrow the set to those most relevant to the machine learning model. Instead than allowing the machine learning model choose the most significant features, feature selection beforehand has several advantages.

D) Algorithms

A Deep Neural Network (DNN) is a type of artificial neural network with multiple layers between the input and output layers. It excels in learning intricate patterns and representations from data through the use of hidden layers. DNN is employed in the project for its capability to handle complex patterns and parameters efficiently. It is well-suited for intrusion detection tasks, providing a robust framework for learning and categorizing various network activity patterns [11,14].

$$y = f(Wn \cdot f(Wn-1 \cdot \ldots f(W1 \cdot x + b1) \ldots + bn-1) + bn) \quad (1)$$

Where:

- x – Input features (network traffic data).
- Wn,Wn−1,...,W1 – Weight matrices for each layer.
- bn,bn−1,...,b1– Bias terms for each layer.
- f – Activation function (e.g., ReLU, Sigmoid, Softmax).

A *Convolutional Neural Network (CNN)* is a deep neural network specialized in image processing and pattern recognition, using convolutional layers to learn spatial hierarchies of features. CNN is employed for its efficacy in extracting hierarchical features from network traffic data, particularly for intrusion detection, where its ability to recognize spatial patterns proves effective in identifying complex intrusion patterns.

$$y = f(\sum i = 1n(xi * ki) + b)(2)$$

Where:

- xi – Input feature map (network traffic representation).
- ki– Convolutional kernel (filter for extracting features).
- b – Bias term.
- f – Activation function (e.g., ReLU, Softmax).
- y – Output (classified network intrusion or normal activity).

CNN + LSTM refers to the fusion of a Convolutional Neural Network (CNN) with a Long Short-Term Memory (LSTM) network, combining spatial feature extraction with sequential learning for comprehensive data analysis. This hybrid approach is utilized to capture both spatial and temporal aspects of network data, enhancing the model's capability in recognizing complex patterns and relationships. The CNN + LSTM combination is particularly effective in intrusion detection tasks involving spatial and sequential dependencies in network activities.

IV. Experimental Results

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (4)$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (5)$$

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$F1\ Score = 2 * \frac{Recall\ \text{X}\ Precision}{Recall + Precision} * 100 \qquad (6)$$

Table.1 Performance Evaluation

Algorithms	Accuracy	F1 - score	Recall	Precision
DNN	0.976	0.668	0.686	0.665
Extension CNN	0.993	0.993	0.993	0.994
Extension CNN+LSTM	0.981	0.987	0.981	0.994

Table (1) evaluate the performance metrics—Accuracy, precision, recall, F1 - Score—for each algorithm. Across all metrics, the CNN consistently outperforms all other algorithms. The tables also offer a comparative analysis of the metrics for the other algorithms.

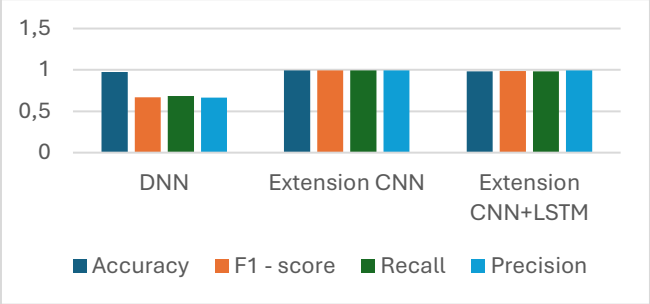

Fig. 3. Comparison Graph

Accuracy is represented in blue, precision in orange, recall in green and Recall in Sky blue Graph (1). In comparison to the other models, the CNN shows superior performance across all metrics, achieving the highest values. The graphs above visually illustrate these findings.

V. Conclusion

The project showcases the transformative impact of Deep Neural Networks (DNN) on intrusion detection, benefiting network administrators and fortifying computer networks against evolving cyber threats. By reducing reliance on manual intervention, the project enhances operational efficiency, saving time and resources for organizations and enabling security teams to focus strategically. The use of the NSLKDD Dataset refines intrusion detection training and contributes to broader cybersecurity knowledge, fostering a proactive stance against emerging threats. The CNN which is extension algorithm excelled in network intrusion detection, showcasing remarkable accuracy in discerning complex features. Its seamless integration into the user-friendly Flask front end, with real feature values for testing, highlights its practical effectiveness in both accuracy enhancement and user experience. The project's findings guide cybersecurity professionals in developing more effective Network Intrusion Detection Systems (NIDS). Future scope involves refining deep learning approaches to address novel cyber threats, ensuring continual improvement in network security measures.

The project can evolve to incorporate real-time threat response mechanisms, enabling automated actions based on detected intrusions, thereby enhancing the overall security posture. Future developments could involve the continual enrichment of the training dataset with evolving network threats, ensuring the intrusion detection system stays updated and effective against emerging attack vectors. Extending the project to integrate with cloud security frameworks can enhance scalability and adaptability, ensuring effective intrusion detection in dynamic and distributed cloud environments. Future enhancements may focus on improving the interpretability of the model's decisions, making it more transparent and understandable for security analysts to interpret and act upon the generated alerts. As quantum computing technologies advance, the project can explore ways to adapt intrusion detection methods to handle the unique challenges and threats posed by quantum computing, ensuring the system remains robust in the face of evolving technologies.

References

[1] Gurung, S & et al., "Deep Learning Approach on Network Intrusion Detection System using NSL-KDD Dataset." Int. Journal of Computer Network and Information Security, 11(3), 8–14,2019.

[2] D. Mukhopadhyay, Forouzan, B., "Cryptography and network security." V12: McGraw Hill Education, NY, USA,2015.

[3] Hurley, T. & et al., "HMM-Based Intrusion Detection System for Software Defined Networking. 15th IEEE Int. Conf. on Machine Learning and Applications (ICMLA), 2016.

[4] Shone, N. & et al., "A Deep Learning Approach to Network Intrusion Detection." IEEE Tran. on Emerging Topics in Computational Intelligence, 2(1), 41–50, 2018.

[5] Shyu, M. & et al., "A Novel Anomaly Detection Scheme Based on Principal Component Classifier." Int. Con. on Data Mining, 172–179, 2003.

[6] KDD Cup 1999 Data. (2018, December 11). Kaggle. https://www.kaggle.com/datasets/galaxyh/kdd-cup-1999-data: accessed on oct-2022.

[7] Revathi, S., & Malathi, A., "A Detailed Analysis of NSL-KDD Dataset for Various Machine Learning Techniques for Intrusion Detection." Int. Journal of Engineering Research and Technology, 2(12), 2013.

[8] Chen, Y. & et al., "Feature Selection and Intrusion Detection Using Hybrid Flexible Neural Tree." Adv. in Neural Networks – ISNN 2005, 439–444.

[9] Zachary T. & et al., "Learning to Detect: A Data-driven Approach for Network Intrusion Detection". IEEE Int. Performance, Computing, and Communications Conference (IPCCC), 2021.

[10] Alzahrani, A. O., & Alenazi, M. J. F., "Designing a Network Intrusion Detection System Based on ML for Software Defined Networks." Future Internet, 13(5), 111. 2021.

[11] Tang, C.& et al., "SAAE-DNN: Deep Learning Method on Intrusion Detection." Symmetry, 12(10), 1695, 2020.

[12] Tavallaee, M.,& et al., "A detailed analysis of the KDD CUP 99 data set", the IEEE Symposium on Comp. Intelligence for Security and Defense Applications, 2009.

[13] Wu, P. & et al., "A Transfer Learning Approach for Network Intrusion Detection." IEEE 4th Int. Conf. on Big Data Analytics (ICBDA), 2019.

[14] Tang, T. A. & et al., "DL approach for Network Intrusion Detection in Software Defined Networking," the Int. Conf. on Wireless Networks and Mobile Communications (WINCOM), 2016.

[15] Niyaz, Q. & et al.," A Deep Learning Based DDoS Detection System in Software-Defined Networking (SDN)." ICST Transactions on Security and Safety, 4(12), 153515, 2017.

[16] Kim, J. & et al., "Long Short-Term Memory Recurrent Neural Network Classifier for Intrusion Detection." Int. Conf. on Platform Technology and Service (PlatCon), 2016.

[17] Gao, N., "An Intrusion Detection Model Based on Deep Belief Networks," Second Int. Conf. on Advanced Cloud and Big Data, 2014.

[18] McHugh, J., "Testing Intrusion detection systems," ACM Transactions on Info. and System Security, 3(4), 262–294, 2000.

[19] Deng, L, "Deep Learning: Methods and Applications." Foundations and Trends® in Signal Processing, 7(3–4), 197–387, 2014.

[20] Dong, B., & Wang, X., "Comparison of deep learning method to traditional methods used for network intrusion detection." The 8th IEEE Int. Conf. on Communication Software and Networks (ICCSN), 2016.

[21] Zhao, R., et al., "Deep learning and its applications to machine health monitoring." Mechanical Systems and Signal Processing, 115, 213–237.

[22] Alrawashdeh, K., & Purdy, C., "Toward an Online Anomaly Intrusion Detection System Based on Deep Learning." 15th IEEE Int. Conf. on Machine Learning and Applications (ICMLA), 2016.