IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Detailed Analysis Of Intrusion Detection With Machine Learning

Dhanesh Prasad Saket[1], Prof. Chetan Gupta[2]

M. Tech. Scholar, Dept. of CSE, SIRTS, Bhopal, India[1], Assistant Professor, Dept. of CSE, SIRT, Bhopal, India[2],

Abstract- The significance of intrusion detection systems (IDS) in network security and the need for accurate and effective detection methods are emphasised by research on the subject. Research emphasises the use of statistical methods in host-based systems, web-based data mining systems, and modern intrusion detection systems such as firewalls. It offers a thorough synopsis and evaluation of the body of knowledge about this topic. Using web-based data mining and machine learning algorithms—such as Rough Set Theory and Support Vector Machine—as well as an optimised framework and a two-layer mechanism, the article investigates intrusion detection systems. Effective feature selection and representation are the foundation of machine learning-based intrusion detection systems (IDS). For feature extraction, anomaly detection, abuse detection, and hybrid models, new approaches are required. It's also necessary to have domain adaptability, interpretable models, visualisation, and strong defensive mechanisms. Compared to conventional techniques, the suggested learning-based intrusion detection system may identify network traffic patterns more precisely, minimising false positives and logging network activity. Although it can be flexible and scalable to many kinds of attacks and network circumstances, it could have issues with interpretability, resource needs, and possible hostile manipulation vulnerabilities.

Key Terms—IDS, Network Security, Support Vector Machine, Rough Set Theory.

I. INTRODUCTION

An intrusion detection system (IDS) monitors network traffic, keeps an eye out for odd behaviour, and notifies users when it's detected. While anomaly detection and reporting are an intrusion detection system's primary responsibility, certain intrusion detection systems have the ability to respond to the discovery of hostile activity or anomalous traffic.

An intrusion detection system (IDS) monitors network traffic in order to spot potentially harmful transactions and promptly notifies users when one is detected. It is software that scans a system or network for nefarious activity or infractions of policies. IDS keeps an eye out for harmful behaviour on a network or system and guards against users, including potential insiders, gaining unauthorised access to a computer network. The goal of the intrusion detector learning job is to create a prediction model, or classifier, that can discriminate between "good (normal) connections" and "bad connections," or intrusions or assaults.

By allowing it to learn from past data and adjust to new, undiscovered dangers, machine learning may improve intrusion detection systems. It does this by looking for trends and abnormalities that might point to malevolent activity. There are two types of intrusion detection systems: network-based IDS (NIDS) and host-based IDS (HIDS). Network administrators use host-based intrusion detection systems to track and examine activity on a specific computer. AI and ML are examples of new technologies that are developing and producing excellent outcomes.

These technologies include a variety of algorithms and methods that can operate on big data sets and provide the best outcomes. When ML algorithms are used in conjunction with intrusion detection systems, massive networks of computers may be simply and effectively monitored and analysed to identify intrusion assaults.

II. BACKGROUND

The literature review of IDS using machine learning techniques mentioned below:

(Almasoudy et al., 2020) [1], The number of moving packets and network stress have grown recently due to the growing need of using the Internet for all applications and domains. As a result, even with several network security mechanisms in place, including the firewall system, which is a powerful preventative and protective mechanism, the most sensitive data is still vulnerable. The firewall systems stop unauthorised users from accessing the systems, but they are unable to monitor the monitoring after the data has been sent through. It won't be able to recognise any assault that gets past it. Therefore, the intrusion detection system (IDS) has to be encircling the network in order to keep it under observation. An intrusion is characterised as a danger to the availability, confidentiality, and integrity of data that arises from either authorised users abusing their rights or unapproved individuals exploiting security holes in the system to gain unauthorised access.

(Nalluri et al., 2005) [2], A Web-based data mining system for analyzing intrusions, implemented using freeware available in the public domain, capable of differentiating intrusions from normal activity and generating rules for behaviour capture, suitable for educational and training purposes. (Siahaan et al., 2017) [3], It provides a brief summary of the purpose and function of an Intrusion Detection System (IDS) in protecting the network from threats and attacks, emphasizing the need for network administrators to stay updated with new types of attacks and highlighting the benefits of using IDS for network security.

(Vokorokos et al., 2010) [4], It presents an intrusion detection system that informs the system administrator about potential intrusion incidents and employs statistical methods of data evaluation for detection based on user activity deviation from learned profiles representing standard user behavior. (Liu et al., 2014) [5], It provides an overview of the importance of network security, the concept of intrusion, the construction of intrusion detection systems, and details about existing intrusion detection techniques. (Prabhu et al., 2014) [7], Attacks on computer and data networks are increasing in frequency and sophistication, leading to a shift in the focus of intrusion detection towards networks, with the aim of identifying misuse and unauthorized use through statistical anomaly and rule-based misuse models.

(Junedul Haque et al., 2012) [8], discusses intrusion detection based on data mining and proposes a framework to detect intrusions and demonstrate the improvement of the k-means clustering algorithm. (Das et al., 2010) [9], discusses the importance of network and system security, the use of Intrusion Detection Systems (IDS), and the application of Rough Set Theory (RST) and Support Vector Machine (SVM) for detecting network intrusions, with a focus on reducing the number of features and improving accuracy.

(Sulaiman et al., 2021) [10], provides a brief overview of ongoing research and techniques for detecting attacks in Intrusion Detection Systems.

(Ariafar & Kiani, 2017) [11], It paper presents an optimized framework for network attack detection using data mining techniques, achieving a 99.1% detection rate and a 1.8% false alarm rate on the NSL-KDD 2009 dataset. (Dhage et al., 2011) [12], an architecture for detecting intrusions in a distributed cloud computing environment and safeguarding it from security breaches by deploying separate instances of IDS for each user and using a single controller to manage the instances, with the IDS capable of using both signature-based and learning-based methods. (Cisar & Maravic Cisar, 2008) [13], provides a general overview of intrusion detection systems and emphasizes the need for these systems to detect a wide range of malicious activities.

(Shu Wenhui & Tan, 2001) [14], It proposes a two-layer mechanism for detecting intrusions against a web-based database service, aiming to reduce error rates and improve the accuracy of intrusion detection and incident handling. (Shailaja Jadhav et al., 2022) [15], the use of machine learning algorithms, such as Support Vector Machine (SVM) and Naïve Bayes, for evaluating the accuracy and misclassification rate of an intrusion detection system. (Zala et al., 2020) [16], The paper discusses the use of machine learning techniques for intrusion detection, including preprocessing techniques, model comparisons, and evaluation techniques. (Almutairi et al., 2022) [17], discusses the use of machine learning techniques, such as Support Vector Machine, J48, Random Forest, and Naïve Bytes, for network intrusion detection. It evaluates the performance of these techniques using the NSL-KDD benchmark dataset.

III. COMPARATIVE STUDY

Intrusion detection systems (IDS) based on machine learning algorithms have been developed to monitor network activity and detect malicious behavior. These systems analyze large amounts of network traffic data to identify potential intrusions and improve detection accuracy. Various machine learning techniques such as Support Vector Machine (SVM), Naïve Bayes, J48, and Random Forest have been applied to evaluate IDS performance. The use of machine learning in IDS has shown promising results in quickly detecting intrusions and distinguishing between normal and anomalous patterns. These approaches have the potential to enhance network security and protect against evolving threats.

Table 1: Comparative Study of Various IDS Methods

Reference	Method	Conclusion	Limitation
Faezah et al., 2020	Differential Evolution with Extreme Learning	accuracy of 80.15 % and 87.53 for five and binary classification respectively with a reduction in training and testing time	Not perform in real time IDS
Shailaja Jadhav et al., 2022	SVM and Naïve Bayes	NSL-KDD dataset is used to calculate accuracy and misclassification rate with accuracy 88.31%	Accuracy Limited
Zala et al., 2020	Preprocessing techniques	IDS using Machine Learning can detect intruding attacks.	Not perform with imbalanced data
Almutairi et al., 2022	SVM, Random Forest	Accuracy 88.72%	Binay classification not perform

IV. RESEARCH GAP AND FUTURE IMPLICATIONS

Adversarial attacks pose a major challenge to the security and reliability of ML-based intrusion detection systems (Siahaan et al., 2017) [3]. Research is needed to develop effective defensive mechanisms against adversarial attacks, such as adversarial training, feature obfuscation, and model diversification, in order to increase the resilience of ML-based IDS in adversarial scenarios.

Machine learning-based intrusion detection systems taught in one network environment may not perform well in another due to domain shift or distributional changes (Vokorokos et al., 2010) [4]. To improve IDS performance in target domains where labelled data is scarce, further research is needed to look at transfer learning and domain adaptation techniques.

ML-based intrusion detection systems often face problems with efficiency and scalability when deployed in large-scale or high-speed networks (Das et al., 2010) [9]. Research is needed to develop efficient and scalable machine learning algorithms and infrastructures that can control the volume and velocity of network traffic data while maintaining real-time detection performance.

ML-based intrusion detection systems may be classified into two categories: abnormalities and abuses (Shailaja Jadhav et al., 2022) [15]. It will need research to evaluate the benefits and drawbacks of each tactic as well as hybrid models that mix the two in an effort to increase detection precision and lower false alarm rates.

Efficient feature selection and representation are essential for ML-based intrusion detection systems to function well (Zala et al., 2020) [16]. Investigation is needed into new feature extraction and feature selection procedures that might capture the unique characteristics of network data while reducing dimensionality and increasing computer performance.

Due to their "black-box" nature, interpretability and explainability problems may prohibit many of the ML models employed in IDS from being extensively deployed in practice (Almutairi et al., 2022) [17]. To develop interpretable machine learning models and visualisation techniques that can shed light on the decision-making process of IDS models and foster more user trust and transparency, research is necessary.

V. PROPOSED WORK

The proposed methodology works as follows

A. Import dataset: The dataset that is accessible on either Kaggle library or https://www.unb.ca/cic/datasets/nsl.html used by us.

B. Data Pre-processing: For pre-processing, we can use PCA (Principal Component Analysis). It speaks about removing stop words, tokenizing, and stemming data to prepare it for analysis. Enhancing the data's quality and adapting it to the particular data categorization job is the aim of data preparation.

C. Feature Selection/Extraction: The method of autonomously selecting pertinent features for a learning model using a collection of fictitious profiles is known as feature selection. Following feature selection, the dataset was split into training and testing labels.

D. Modeling with Train Data: The act of providing data to a proposed learning algorithm in order to assist it find and learn optimal values for all related attributes is known as model training.

E. Optimize Modeling: Finding the optimal outcome under the conditions is the act of optimisation. A prediction model's performance may be increased by fine-tuning a number of factors.

F. Evaluate Prediction: Use relevant assessment measures, such as accuracy, precision, recall, F1 score, etc., to assess the predictions based on the testing data.

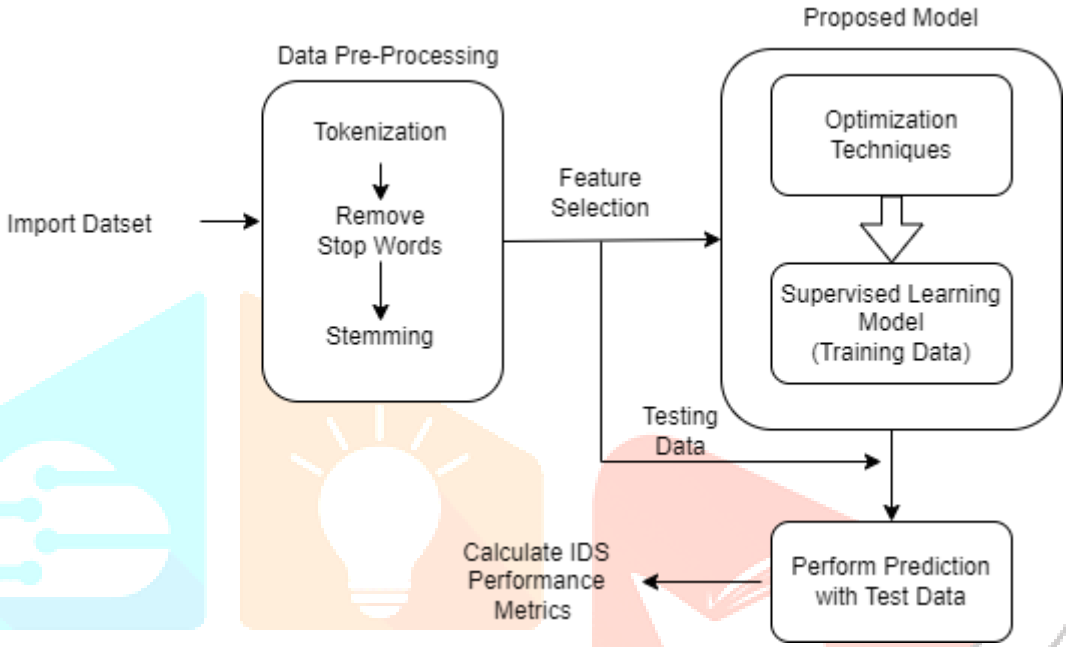

Figure 1: Proposed Learning Model

VI CONCLUSION

When compared to conventional rule-based or statistical approaches, the suggested method may be able to discover patterns and characteristics more intricately from network traffic or system logs, which might result in a greater detection accuracy. Compared to conventional IDSs, which often generate alarms based on preset rules that may not properly catch regular changes in network traffic, suggested learning-based IDSs can decrease false positives and better capture the usual behaviour of a network or system. Because the proposed model can be trained on big datasets and adjusted as new data becomes available, it can be scalable and adaptive to various network conditions and attack types. Proposed learning-based intrusion detection systems may potentially encounter difficulties with result interpretability, resource requirements for deployment and training, and possible hostile manipulation vulnerabilities.

REFERENCES

[1] Almasoudy, F. H., Al-Yaseen, W. L., & Idrees, A. K. (2020). Differential Evolution Wrapper Feature Selection for Intrusion Detection System. Procedia Computer Science, 167, 1230–1239. https://doi.org/10.1016/j.procs.2020.03.438

[2] Nalluri, A., & Kar, D.C. (2005). A web-based system for intrusion detection. Journal of Computing Sciences in Colleges, 20, 274-281.

[3] Siahaan, A.P. (2017). Intrusion Detection System in Network Forensic Analysis and Investigation.

[4] Vokorokos, L., & Balaz, A. (2010, May). Host-based intrusion detection system. 2010 IEEE 14th International Conference on Intelligent Engineering Systems. https://doi.org/10.1109/ines.2010.5483815

[5] Liu, G. G. (2014, July). Intrusion Detection Systems. Applied Mechanics and Materials, 596, 852–855. https://doi.org/10.4028/www.scientific.net/amm.596.852

[6] Almasoudy, F. H., Al-Yaseen, W. L., & Idrees, A. K. (2020). Differential Evolution Wrapper Feature Selection for Intrusion Detection System. Procedia Computer Science, 167, 1230–1239. https://doi.org/10.1016/j.procs.2020.03.438

[7] Prabhu, G., Jain, K., Lawande, N., Zutshi, Y., Singh, R., & Chinchole, J. (2014). Network Intrusion Detection System.

[8] Junedul Haque, M., Magld, K., & Hundewale, N. (2012, May). An intelligent approach for Intrusion Detection based on data mining techniques. 2012 International Conference on Multimedia Computing and Systems. https://doi.org/10.1109/icmcs.2012.6320182

[9] Das, V., Pathak, V., Sharma, S., Sreevathsan, Srikanth, M., & Gireesh Kumar, T. (2010, December 23). Network Intrusion Detection System Based On Machine Learning Algorithms. International Journal of Computer Science and Information Technology, 2(6), 138–151. https://doi.org/10.5121/ijcsit.2010.2613

[10] Sulaiman, N. S., Nasir, A., Othman, W. R. W., Wahab, S. F. A., Aziz, N. S., Yacob, A., & Samsudin, N. (2021, May 1). Intrusion Detection System Techniques : A Review. Journal of Physics: Conference Series, 1874(1), 012042. https://doi.org/10.1088/1742-6596/1874/1/012042

[11] Ariafar, E., & Kiani, R. (2017, December). Intrusion detection system using an optimized framework based on datamining techniques. 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI). https://doi.org/10.1109/kbei.2017.8324903

[12] Dhage, S. N., Meshram, B. B., Rawat, R., Padawe, S., Paingaokar, M., & Misra, A. (2011). Intrusion detection system in cloud computing environment. Proceedings of the International Conference & Workshop on Emerging Trends in Technology - ICWET '11. https://doi.org/10.1145/1980022.1980076

[13] Cisar, P., & Maravic Cisar, S. (2008, September). Intrusion detection-one of the security methods. 2008 6th International Symposium on Intelligent Systems and Informatics. https://doi.org/10.1109/sisy.2008.4664926

[14] Shu Wenhui, & Tan, T. (n.d.). A novel intrusion detection system model for securing web-based database systems. 25th Annual International Computer Software and Applications Conference. COMPSAC 2001. https://doi.org/10.1109/cmpsac.2001.960624

[15] Shailaja Jadhav, Vinaya Bhalerao, Varsha Yadav, Snehal Kamble, & Bhavana Shinde. (2022, January 10). Network Intrusion Detection System Using Machine Learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 74–81. https://doi.org/10.32628/cseit22819

[16] Zala, J., Panchal, A., Thakkar, A., Prajapati, B., & Puvar, P. (2020, May 1). Intrusion Detection System using Machine Learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 61–71. https://doi.org/10.32628/cseit2062166

[17] Almutairi, Y., Alhazmi, B., & Munshi, A. (2022, July 1). Network Intrusion Detection Using Machine Learning Techniques. Advances in Science and Technology Research Journal, 16(3), 193–206. https://doi.org/10.12913/22998624/149934