INTRUSION DETECTION IN A NETWORK USING MACHINE LEARNING AND NEURAL NETWORK APPROACHES

Shivam Dangi
Vellore Institute of Technology
Shivam.dangi2021@vitstudent.ac.in

Akshay Reddy
Vellore Institute of Technology
Akshay.reddy2021@vitstudent.ac.in

Omprakash P
Vellore Institute of Technology
omprakash.pugazhendhi@gmail.com

Sai Pranav Reddy
Vellore Institute of Technology
Pranav.reddy@vitstudent.ac.in

ABSTRACT - With advancing in tech-stacks, cyber threats are also evolving at exponential rate. In a recent statistics review over cyber-attacks, it has been observed that there are 97 cyber-attack victims are there every hour. To encounter this threat, we require a robust and sophisticated intrusion detection system that can detect the malicious activity in a real-time network traffic. This kind of proficient system can be designed with the help of machine learning and neural network-based technologies. These methods can help to detect the possibility of an unauthorised intervention in the network and report. This project aims to compare multiple approaches in order to identify the anomalies.

KEYWORDS

Intrusion, Machine Learning, Cyber Security, CNN

INTRODUCTION

The tech-enabled era described by the network of digital interconnectivity; network security has become the paramount under-consideration. Intrusion Detection Systems provides sophistication defence mechanism against huge-array of cyber threat activities, which includes but not limited to, unauthorised access attempts, malware intervening, and many more. As these attacking mechanisms are evolving with revolutionary landscape of data sharing, traditional rule-based IDS are facing challenges to maintain effective perseverance against these attack vectors and methodologies.

The constraints of the conventional IDS techniques have spurred a parallel shift towards more intelligent, scalable and adaptability-oriented algorithms. This brought the inclusion of Machine Learning and Neural Network based technology aspects into Intrusion detection frameworks. Anomaly Detection by identifying patterns from usual network activities, and potentially uncovering variations; enhancing scalability as and how the network size increases while maintaining efficiency; adaptability with different kind of data flow in networks with constant updates; and performing contextual analysis for more nuanced and accuracy-reliable threat assessment.

However, the challenges like collection of huge, diverse and real-time collection of network infrastructure dataset; balancing model complexity with handling computation cost with real-time operations;

and ensuring decisions with respect to security critical networking environment related networks are some major hinderances.

Proposed methodology aims to bring contribution toward ongoing developments in domain of Machine Learning and Neural Network applications in domain of cyber-security by bringing up a comprehensive comparative analysis of multiple possible approaches. The study encorporate a range of tecnhlogies like conventional ML algorithm like Support Vector Machine, Random Forests and more deep learning architectures like Long Short Term Memory networks and Convolution Neural Networks.

With evaluation of these different approaches with various metrics like accuracy of detection, decrease false positive rate, enhanced computational efficiency and adaptability towards new pattern threat, we seek to provide insights of these approaches in real-time networking environment.

LITERATURE SURVEY

Abdulatif and Hussain [1], in their research uses the Kitsune dataset to enhance machine-learning techniques for network intrusion detection. It assesses how well different tree algorithm variations detect various network assaults and contrasts several optimizers, including Bayesian, Grid, and Random search. The study concludes that the Grid search optimizer is the best choice for hyperparameter tuning and that the tree method works well for detecting network attacks.

Khalid et. al. [2] in their study focuses on deep learning techniques for Internet of Things (IoT) intrusion-detection systems (IDSs). It reviews deep learning versus conventional methods for anomaly-based intrusion detection systems. To categorize assaults in Internet of Things environments, the study assesses how well machine learning methods such as Logistic Regression, Support Vector Machine, Decision Trees, and Artificial Neural Networks perform. The study also addresses the difficulties associated with IoT security and the significance of anomaly-based intrusion detection systems (IDS) in identifying malevolent intrusions.

Vanin et. al. [3] demonstrated that Data transfer has increased dramatically as a result of the Internet's rapid expansion, making these data targets for hackers who devise new methods of stealing or corrupting them. The rise in attacks is posing significant challenges to intrusion detection systems (IDS), which monitor network traffic for intrusions. Research on intrusion detection systems (IDS) is underway, but more needs to be done to improve detection precision and reduce false alarms, particularly in the case of zero-day attacks. Recently, machine learning techniques have become more and more popular as practical tools for identifying network intrusions. This paper analyzes recent machine learning-based IDS research, reviews key metrics for evaluating IDS performance, looks at the concept of IDS, and categorizes machine learning.

Gyamfi and Jurcut [4] has depicted the swift proliferation of Internet of Things (IoT) applications has resulted in amplified network data and computational demands on linked devices. Since these devices frequently possess restricted resources, they are vulnerable to cyber-attacks. Multi-access edge computing (MEC) provides an answer by relocating labour-intensive tasks from Internet of Things devices. With an emphasis on MEC and machine learning

techniques, this paper examines current network intrusion detection systems (NIDS) and IoT security practices. Additionally, it contrasts deployment methods, assessment metrics, and datasets that are freely accessible in NIDS design. Lastly, the paper suggests an MEC-based NIDS architecture designed especially for IoT networks.

Micha and Pawlicki [5] in their article subject the new concept for hyperparameter optimization used in artificial neural networks. Other (ANN) Intrusion detection methods are part of network security applications. In this study, the plan is presented in detail and the results are shown on two criteria, its effectiveness is demonstrated by showing comparative results. Various ANN topologies, different hidden layers are trained and tested to enhance effectiveness.

Omer et. al. [6] presents a new framework for analysing effects and monitoring lethal behaviour: Firefly Optimization and Probabilistic Neural Networks (FFO-PNN). It uses various techniques, including feature removal, prioritization, and authentication. During prioritization, the min-max normalization method is used to improve the detection of attacks. In the feature extraction step, the firefly optimization method is used to select the best features from the dataset. Desired features can be improved by creating more accurate measurements. The research used optimization techniques that reduce the training time while increasing the scalability of neural networks.

Ahsan et. al. [7] reports the results of a comparative study on different data using 10 machine learning algorithms along with our iterative process to control network interference. Data deconvolution can have a significant impact on performance when used in conjunction with classical machine learning algorithms. Deep learning models will suffer when under sampling techniques are applied to raw data. We believe that replacing the structure of FCN and autoencoders with FCN can lead to better results. According to this review, the concept of IDS and different classification systems are explained in detail. Then, the methodology of each paper is discussed and the strengths and weaknesses of each paper in terms of access to discovery potential and model difficulties are discussed.

Ahmad et. al. [8] in their study, reports the recent changes propose the use of deep learning to improve the efficiency and effectiveness of NIDS based on accurate identification and reduction of low FAR. Approximately 80% of the solutions are based on DL methods; AE and DNN are the most commonly used algorithms. Although deep learning methods have better performance than machine learning based methods in terms of their own learning characteristics and strong model fitting.

PROPOSED ARCHITECTURE

The designed architecture for our intrusion detection system integrates the machine learning and deep learning architectures to detect possible network anomalies by analysing potential threats in the network traffic-flow, and respond to unusual behaviours. Figure 1 describes a high-level view of how the process looks like:

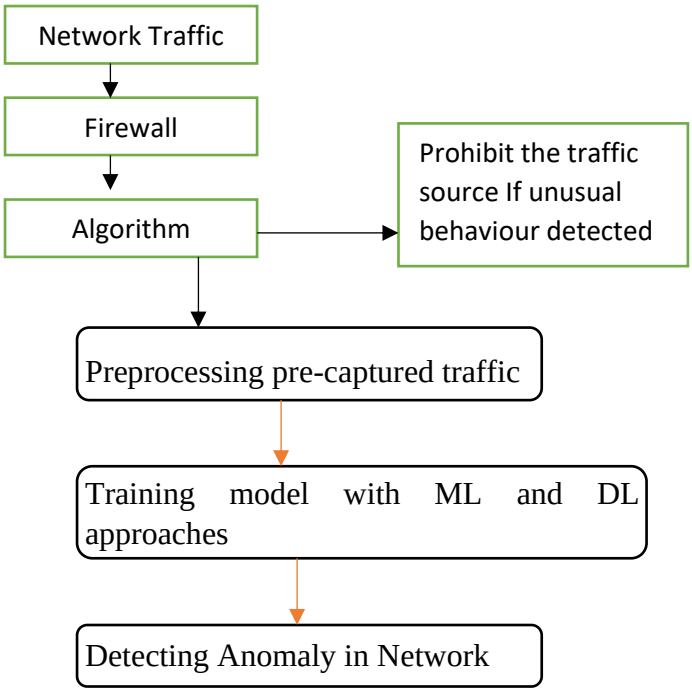

Image 1: Functioning of Model

Components of the architecture includes:

1. Network Traffic: The entry point of the detection is the network-traffic. It includes inspection of all the data-packets passing through between various nodes in the network, in a duplex manner. This includes, but not limited to, the HTTP requests of the participating nodes of networks, file-transfer and sharing requests and more.
2. Firewall: It is the most conventional and robust security mechanisms, a firewall serving as one of the first steps towards defence mechanism of a networking system by filtering network traffic. It inspects the traffic like data packets passing through it based on rules and parameters configured by the network administrator.
3. Algorithm: The core component block of the architecture is algorithms. This is the ventral-intelligent part of the system, where pre-trained models are used to implement. The key responsibilities of the algorithm block includes pre-processing the captured traffic data for doing feature extraction and encoding, now these pre-processed traffic passes through the training models and then results in favour or against of possible anomalies. This component can be include with firewall block in real-time configuration, and can handle the response mechanism as similar to firewall, and incorporate continuous learning mechanism using feedback loop to learn and mitigate the new pattern threats.

This proposed mechanism can operate simultaneously with the existing measures like firewalls to provide additional layer of security. By leveraging the advancements of machine learning and neural networks, this adaptive and intelligent approach can potentially recognise subtle patterns and anomalies against evolving sophisticated cyber-threats.

COMPARATIVE STUDY

In this section, a comparative study of the three approaches of ML and NN, which includes Support Vector Machine, Random Forest and CNN models which are implemented, is done. Each of the method is compared and evaluates on the basis of the performance on training and test datasets, which focus on parameters or metrics like accuracy, f1-score, training accuracy and test accuracy. The following table 1 gives the metrics score of each architecture used in this study.

Architecture		
Support Vector Machine	Training Accuracy	97.10%
	Test Accuracy	97.34%
Random Forest	Training Accuracy	99.38%
	Test Accuracy	99%
CNN	Accuracy	99%
	F1-Score	98.20%

Table 1: Results

Analysing the performance metrics, following are certain observations regarding the inclusion of some of the above architectures in the algorithm block:

1. Support Vector Machine shows solidify performance with good generalisation when train and test scores are compared. This kind of generalisation behaviour indicates that the model does not undergoes neither overfitting nor underfitting to a significant point.
2. Random Forest model derives exceptional integral performance, pertaining maximum accuracy score when compared to others. With this highly accurate predictive capacity, the drop in test accuracy indicates the generalization of the model with respect to unseen test data.
3. CNN model also perform well with high F1-score which signifies the incorporation of the architecture for maintaining a better scale of balance between precision and recall, which indirectly are crucial parameters to settle about intrusion detection systems.

The parameters used to compare the three architectures includes the accuracy, generalisation, model complexity and interpretability, and computational requirement.

1. Accuracy: Random forest model has achieved the highest accuracy amongst the three with both the training and test dataset. Followed by CNN which has nearly similar on par accuracy with Random Forest, with SVM have slightly lower compared to two others.
2. Generalisation: All the three models show better generalisation having there accuracy-test scores close to each other and slightly lower than training scores. Random forest however, has the smallest gap between the training and test score indicating good generalisation rate.
3. Model Complexity and Interpretability: SVM and Random Forest are found to be more interpretable than CNN. Random Forest helps to provides better feature extraction and importance, which values as the key indicator toward the intrusion. CNN however automatically learns new complex pattern in the traffic data.
4. Computational Requirement: Random Forest and SVM, in general, has required less computational powers as compared to CNN. CNN is resource intensive but has better potential stake towards understanding and capturing complex and different attack vectors and patterns in the network traffic.

The following bar-graph representation indicates the comparison of the different methodologies:

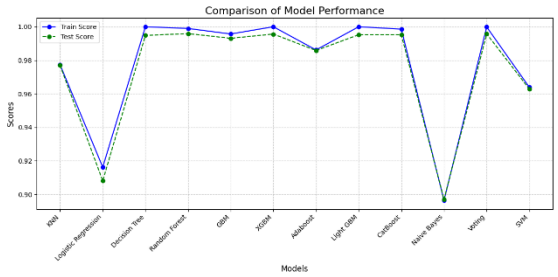

Image 2: Comparison of multiple approaches

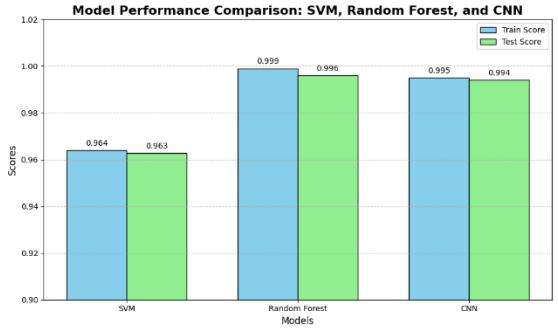

Image 3

Based on this understanding of the study done, the three methods used, depicts and demonstrates their effectivity towards detecting intrusion in networking environment. This brings us an understanding of the fact, that the use-case requirement of the network based upon its architecture, the choice can be made as follows:

- For maximum accuracy and interpretability, the best option to choose be Random Forest
- With computational limited, less sensitive and less threat-prone networking environment, SVM is better choice.
- For situations where scalability and adaptability are key concern and understanding of complex traffic is crucial, CNN based architecture is useful.

The high-performance rate amongst all the multiple architectures indicates how beneficial and effective are the application of the modernised technologies like ML and Neural Networks possess. These methods can indeed provide enhanced security mechanism along with conventional methodologies existing.

CONCLUSION

The proposed research paper has explored the significant application and advantages of machine learning and neural network methodologies in intrusion detection in digital networking. With the help of this comprehensive and detailed comparative study of implementation of Support Vector Machines, Random Forest and Convolution Neural Networks, we have demonstrated the efficiency of these advance and modern computational architectures in detecting, recognising and mitigating possible potential security threats.

This study brings up understanding of how the trade-off should be done for selecting the model as per specific use-case scenario, need of automatic learning of the models, undiscovered potential of deep learning and neuron based highly adaptable model designs, and high-performance capabilities.

With integration of these advance algorithms, a robust framework can be designed that will serve as a next-gen intrusion detection system by leveraging the capabilities of ML and Neural Networks which will help in detecting subtle anomalies and potential threats and attack vectors.

FUTURE WORK

While this research has generated promising insights, there are multiple other avenues that can advance this domain of ML/NN based intrusion detection. This majorly includes ensemble modelling, adversarial modelling, using more enriched and developed GenAI methods, privacy

concern studies, and long-term efficacy assessments.

The advancements and developments are rapidly increasing up the pace of the networking, and so is the urge of designing and developing more dedicated and sophisticated IDS algorithms to handle and manage the cyber-threats.

RELATED WORKS

1. Network intrusion detection system using an optimized machine learning algorithm Abdulatif Alabdulatif , Syed Sajjad Hussain Rizvi. 2023
2. IoT Intrusion Detection Taxonomy, Reference Architecture,and Analyses Khalid Albulayhi , Abdallah A. Smadi, Frederick T. Sheldon and Robert K. Abercrombie 2021
3. A Study of Network Intrusion Detection Systems Using Artificial Intelligence/Machine Learning Patrick Vanin, Thomas Newe, Lubna Luxmi Dhirani, Eoin O'Connell, Donna O'Shea, Brian Lee, and Muzaffar Rao 2022
4. Intrusion Detection in Internet of Things Systems: A Review on Design Approaches Leveraging Multi-Access Edge Computing, Machine Learning, and Datasets Eric Gyamfi and Anca Jurcut 2022
5. Intrusion detection method based on optimized artificial neural network: Micha Chora, Marek Pawlicki, 2021
6. A novel optimized probabilistic neural network approach for intrusion detection and categorization: Nadir Omer, Ahmed H. Samak, Ahmed I. Taloba, Rasha M. Abd El-Aziz, 2023
7. Network intrusion detection using machine learning approaches: Addressing data imbalance: Rahbar Ahsan, Wei Shi, Jean-Pierre Corriveau ,2022
8. Network intrusion detection system: A systematic study of machine learning and deep learning approaches: Zeeshan Ahmad, Adnan Shahid Khan, Cheah Wai Shiang, Johari Abdullah, Farhan Ahmad , 2022.
9. Talukder, M. A., Islam, M. M., Uddin, M. A., Hasan, K. F., Sharmin, S., Alyami, S. A., & Moni, M. A. (2024). Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction. Journal of big data, 11(1), 33.
10. Paya, A., Arroni, S., García-Díaz, V., & Gómez, A. (2024). Apollon: a robust defense system against adversarial machine learning attacks in intrusion detection systems. Computers & Security, 136, 103546.
11. Turukmane, A. V., & Devendiran, R. (2024). M-MultiSVM: An efficient feature selection assisted network intrusion detection system using machine learning. Computers & Security, 137, 103587.
12. Nabi, F., & Zhou, X. (2024). Enhancing intrusion detection systems through dimensionality reduction: A comparative study of machine learning techniques for cyber security. Cyber Security and Applications, 100033.
13. Brinkley, Y., Thompson, D., & Simmons, N. (2024). Machine learning-based intrusion detection for zero-day ransomware in unseen data.

14. Nabi, F., & Zhou, X. (2024). Enhancing intrusion detection systems through dimensionality reduction: A comparative study of machine learning techniques for cyber security. *Cyber Security and Applications*, 100033.
15. Ioannou, I., Nagaradjane, P., Angin, P., Balasubramanian, P., Kavitha, K. J., Murugan, P., & Vassiliou, V. (2024). GEMLIDS-MIOT: A Green Effective Machine Learning Intrusion Detection System based on Federated Learning for Medical IoT network security hardening. *Computer Communications*, *218*, 209-239.