IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Machine Learning Approach For Intelligent Intrusion Detection System.

Karishma Chavan, Dhanshree Gandas, Trimbkeshwar Jagtap, Mayur Sutar, Sujeet More (Assistance professor)

(1234 UG Students, Guide Name, Department of Information Technology, Trinity College of Engineering & Research Pune, India – 411048)

ABSTRACT

Because of its effects on the various communication and security domains, intrusion detection in computer networks is crucial. It can be difficult to find network intrusions. Furthermore, network intrusion detection is still a difficult task because it takes a tremendous amount of data to train modern machine learning models to recognize network intrusion threats. Recently, a number of methods have been put out for network intrusion detection.

Keywords:

NSL-KDD, machine learning, feature scaling, network intrusion detection system (NIDS), software defined network (SDN), AdaBoost.

1. INTRODUCTION

Intrusion detection in computer networks is essential due to the way it impacts numerous communication and security sectors. Finding network intrusions can be challenging. Furthermore, given the enormous volume of data produced, network intrusion detection is still a challenging process needed to train cutting-edge machine learning models to recognise threats to network infiltration. Several techniques for detecting network intrusions have recently been published. But because there are so many new threats popping up all the time that our current systems can't handle, they struggle mightily.

In this study, multiple approaches to developing a network intrusion detection system are evaluated. The best characteristics in the dataset are selected based on the correlation between the features. We also provide an AdaBoost-based technique for identifying network breaches.

Machine learning methods based on network monitoring have been applied in many different sectors. The analysis and detection of road accidents is proposed using a social media network
monitoring system that uses bi-directional long-short-term memory neural networks. The suggested solution uses query-based crawling to gather sentences about any traffic-related occurrences, such as traffic jams, road closures, etc., from social media (Facebook and Twitter). Following that, a number of pre-processing techniques, including steaming, tokenization,

POS segmentation and tagging are used to organise the obtained data into an organised format. Using a latent Dirichlet allocation (LDA) technique, the data are then automatically classified as "traffic" or "nontraffic." Three categories of traffic-labeled data are analysed: positive, negative, and neutral. The result of this stage is a sentence that is labelled with the polarity of the traffic sentence (positive, negative, or neutral), as well as whether it is a traffic or non-traffic sentence. Each sentence is then converted into a one-hot encoding representation using the bag-of-words (Bow) method before being fed to the Bi-directional LSTM neural network (Bi-LSTM).

Following the learning process, the sentences are classified according to location, traffic event, and polarity types using multi-class classification using the SoftMax layer by the neural networks. The suggested method evaluates various traditional machine learning and cutting-edge deep learning approaches based on parameters such as accuracy, F-score, and others.

The second defence line of a system includes intrusion detection systems (IDS). To better protect the systems from cyberattacks, IDSs can be installed in conjunction with other security measures including access control, authentication procedures, and encryption methods. IDSs can identify between lawful and unlawful behaviour using patterns of good traffic, typical behaviour, or particular criteria that describe an assault.

Dewa and Mulgaras assert that knowledge discovery, also known as data mining, can be utilised to build and deploy IDSs with more accurate and resilient behaviour than conventional IDSs, which might not be as successful against contemporary sophisticated cyber-attacks.

2 LITERATURE REVIEW

2.1 A wireless sensor network routing technique with intrusion detection and prevention

Wireless sensor networks (WSNs) are vulnerable to a variety of security attacks due to characteristics including scarce resources, wireless connectivity, and harsh conditions. Therefore, intrusion detection and prevention techniques are required in WSNs. When the two types of systems are used, there is a significant increase in communication overhead and node energy consumption as a result.Machine Learning Based.

2.2 A Tool Base2d Verification for a Novel Intrusion Detection and Prevension System.

In recent years, the military has used wireless sensor networks in a variety of other industries, including healthcare, manufacturing, and many more. As is generally known, WSNs have some unique features, such as a constrained power source, a constrained bandwidth, and a constrained amount of energy. As a result, there are many different security measures that may be taken to protect traditional networks, but we cannot use the same measures to protect WSNs. So, new methods and ideas were required to enhance WSN security generally. Because intrusion detection is already at saturation point in WSNs, intrusion prevention is typically the main issue

2.3 A Localised Appoach for intrusion in heterogenous wireless sensors that uses little energy.

Wireless sensor networks are widely 668tilized in a variety of settings to carry out various monitoring duties, including target tracking, search and rescue operations, disaster relief, and many more functions in

smart environments. Many of these tasks use node Localization is a system parameter by definition. Node localization is necessary to report the cause of occurrences, help group sensor querying, route traffic, and respond to inquiries about network coverage.

2.4 A Service-Oriented Vehicular Networks Intrusion Detection Mechanism that is Effective and Lightweight

Vehicular ad hoc networks (VANETs) are wireless networks that enable high-speed data transmission between vehicles and between vehicles and roadside infrastructure. It is believed that VANETs (ITS) will serve as the main wireless communication platforms for intelligent transportation systems. Service-oriented vehicular networks are defined as VANETs that offer a range of infrastructure-based commercial infotainment services, including as Internet access, real-time traffic monitoring and centralised defence against threats, such as a firewall, intrusion detection system, or proxy. Ad hoc networks are particularly practical because of their self-organizing, self-maintaining, andwireless communication capabilities.

3 METHODOLOGY

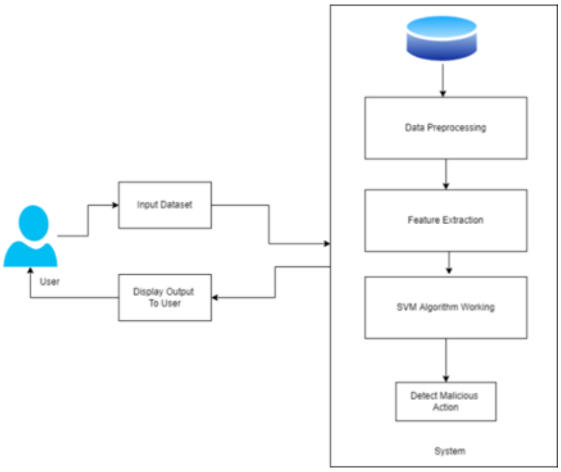


Figure 1. Block Diagram

Data collection, preprocessing, feature extraction, model selection, and evaluation are typical steps in an ML-based IDS technique.

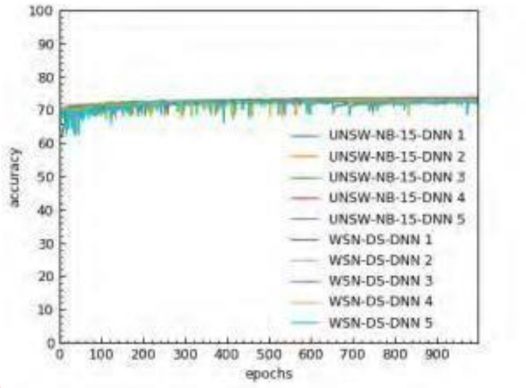
Data Collection: The first step in developing an ML-based IDS is to collect a dataset that contains both normal and malicious traffic. The dataset can be collected from various sources, such as public datasets or from an organization's own network traffic. The dataset should be representative of the network traffic that the IDS is intended to monitor, and should cover a wide range of normal and malicious behaviors.

Data preprocessing: Following the collection of the dataset, preprocessing the data is the next stage. In order to do this, any noise or irrelevant data from the dataset must be removed, including any duplicate records or erroneous data points. To make sure that the data is appropriate for the ML techniques employed by the IDS, it can also be necessary to normalise or alter it.

Feature Extraction: The following phase involves extracting pertinent features from the preprocessed data. In feature extraction, the most informative and discriminative features are chosen so that the ML algorithms may use them to discriminate between legitimate and malicious communication. Different methods, including statistical analysis, signal processing, or machine learning, can be used to extract the features.

Model Selection: After the characteristics have been retrieved, the IDS will be built using a suitable ML technique that has been chosen. Different ML algorithms, including decision trees, Naive Bayes, SVMs, KNNs, and neural networks, can be used. The properties of the dataset and the IDS's performance requirements determine which algorithm is used.

Model Training: The classifier must be trained using the training set after the ML algorithm has been chosen. The classifier's performance is enhanced by using the training set to refine its parameters. Several metrics, including accuracy, precision, recall, and F1 score, can be used to assess the classifier's performance.



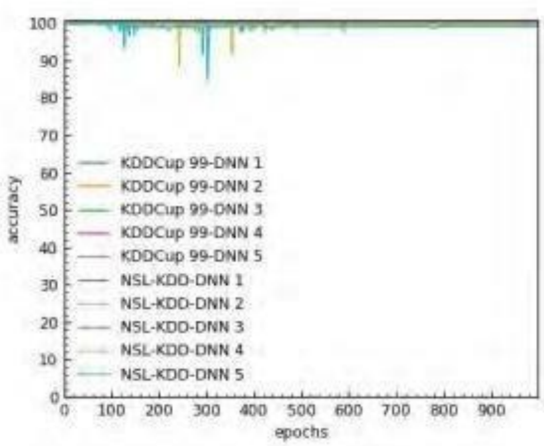
(b)

Model Evaluation: After the classifier has been trained, the testing set is used to measure its effectiveness. The testing set is used to assess the classifier's robustness and accuracy when applied to fresh, untested data. Several metrics, including detection rate, false alarm rate, and accuracy, can be used to assess the classifier's performance.

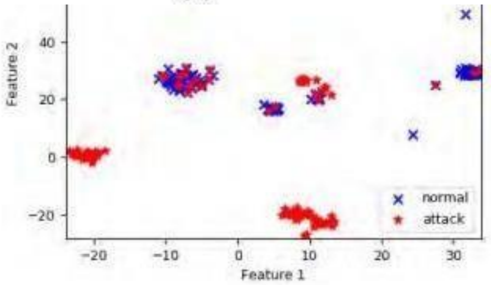
Model tuning: If the classifier's performance is unsatisfactory, the model may need to be adjusted or improved. Model tuning is the process of changing the classifier's parameters to enhance its performance on the testing set. Techniques like cross-validation, ensemble learning, or feature selection might be used in this. When the model has been optimised, it is ready for deployment in a real-world setting.

4 RESULTS

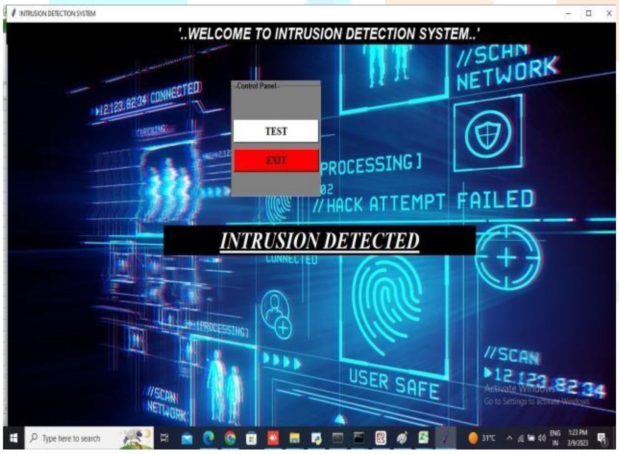
The effectiveness of traditional machine learning and DNNs was assessed using publicly accessible NIDS and HIDS datasets in order to establish a baseline technique. The L2 normalisation method was used to divide these datasets into train and test datasets. Machine learning models were taught using train datasets, and the learned models were evaluated using test datasets. Train multi-class accuracy using DNN for UNSW NB-15, WSN-DS, and KDDCup 99, NSL-KDD.



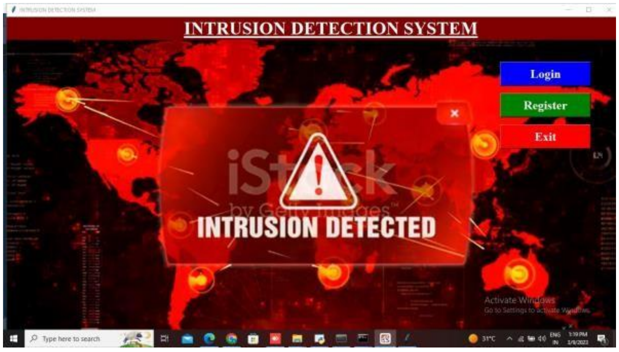
(a)



(c)



(d)



(e)

5 CONCLUSION

The correlation matrix between all of the characteristics in the UNSW-NB 15 dataset was used in this work to suggest a method for choosing features from it. Additionally, based on the selected features and utilising them, we suggest a technique for identifying network intrusions.

the decision tree-based classifier built on AdaBoost. We also discussed the challenges that the current NIDS (network intrusion detection systems) approaches face. The suggested method starts by choosing features using a correlation matrix. The traits that scored highly were connected to other input criteria and had little effect on the outcome label. AdaBoost and the decision tree classifier form the foundation of the strategy.

6 REFERENCES

[1] Maglaras LA, Kim K-H, Janicke H, Ferrag MA, Rallis S, Fragkou P, et al. Cyber security of critical infrastructures.ICT Express 2018;4(1):42–
[2] Ahmim A, Derdour M, Ferrag MA. An intrusion detection system based on combining probability predictions of a tree of classifiers. Int. J. Commun. Syst. 2018;31(9):e3547.
[3] Ahmim A, Maglaras L, Ferrag MA, Derdour M, Janicke H. A novel hierarchical intrusion detection system based on decision tree and rules-based models. In: 15th International Conference on Distributed Computing in Sensor Systems (DCOSS). IEEE; 2019. p. 228–33. https://ieeexplore.ieee.org/abstract/ document/8804816/.
[4] Dewa Z, Maglaras LA. Data mining and intrusion detection systems. Int. J. Adv.

Comput. Sci. Appl. 2016;7(1):62–71.
[5] Stewart B, Rosa L, Maglaras LA, Cruz TJ, Ferrag MA, Simões P, et al. A novel intrusion detection mechanism for scada systems which automatically adapts to network topology changes.. EAI Endorsed Trans. Ind. Netw. Intell. Syst. 2017;4(10):e4

[6] Chavez, A., Lai, C., Jacobs, N., Hossain-McKenzie, S., Jones, C.B., Johnson, J. and Summers, A., 2019, April. Hybrid intrusion detection system design for distributed energy resource systems. In 2019 IEEE CyberPELS (CyberPELS) (pp. 1-6). IEEE.

[7] Chiba, Z., Abghour, N., Moussaid, K. and Rida, M., 2019. Intelligent approach to build a Deep Neural Network based IDS for cloud environment using combination of machine learning algorithms. computers & security, 86, pp.291- 317.

[8] Darwish, A., 2018. Bio-inspired computing: Algorithms review, deep analysis, and the scope of applications. Future Computing and Informatics Journal, 3(2), pp.231- 246.