Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1, January 2024
Paper: 68

Time Efficient and Improved Probalistc Mechanism for Detecting Intrusion Using a Set of Ml Learning Algorithm

Dr. D. Sasireka[1], Ishaan Sharma[2], Janhavi Singh[3], Souvic Nanda[4]
SRM Institute of Science and Technology. Ramapuram Campus, Chennai, India

***Abstract*— As the Network Industry continues to grow, various technologies such as Cloud Computing and Internet of Things have emerged, leading to the sharing of large amounts of data. However, with the increase in connected devices, the risk of attacks and intrusions has also risen. The paper aims to classify network traffic as normal or abnormal and predict any potential abnormal activity using machine and deep learning algorithms. Intrusion Detection Systems (IDS) play a crucial role in network security, as they help detect and classify any abnormal actions. Therefore, it is important for IDS to stay up-to-date with the latest hacker attack signatures to keep services safe and available. The proposed novel distributional features, combined with techniques that enable modeling complex input feature spaces, result in highly accurate predictions by our trained models, which are validated by matching predictions to existing denylists of published malicious IP addresses and deep packet inspection.**

Keywords– Intrusion Detection System, Ensemble Learning, Machine Learning, Network Traffic, Malicious IP Address, Abnormal Traffic Classification, Hacker Attacks, Network Security

I. INTRODUCTION

The security of computer networks is becoming increasingly important as the number of cyber-attacks and security breaches grows. Network Intrusion Detection (NID) is an essential mechanism that helps detect unauthorized access to computer networks by analyzing network traffic for signs of malicious activity. Intrusion Detection Systems (IDS) are a critical element in network defense, detecting successful and unsuccessful intrusion attempts and reporting all irregular device activity. To construct an appropriate machine learning model for IDS, feature extraction and feature selection are two common approaches for dimensionality reduction.

As the volume of internet traffic increases and a broader range of devices connect to the internet, there is a growing need to monitor internet traffic for malicious activity. This is particularly crucial because internet security breaches can cause lasting harm to both individual devices and carrier networks. IP traffic security monitoring is an essential issue that is growing in importance. There are many challenging statistical issues in network security, such as identifying malicious events like scanning, Distributed Denial of Service (DDoS) attacks, malware injection, and spam attacks

The paper stresses that developing dependable and adaptable security strategies for computer networks is critical due to the constant emergence of new types of threats. It discusses the need for a more dynamic dataset to improve the ability of an IDS to detect intrusions. Using deep learning techniques such as Generative Adversarial Networks (GANs), additional data can be fabricated using existing datasets to increase the classification accuracy of an IDS, especially for rare attack categories.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1, January 2024
Paper: 68

In conclusion, the paper emphasizes the need for IDS to secure network systems and protect important data from attackers. The different types of IDS and the importance of developing adaptable security strategies are also discussed

II. LITERATURE SURVEY

(Sultan Zavrak, 2020) This study called "Anomaly-Based Intrusion Detection From Network Flow Features Using Variational Autoencoder" was proposed in which the researchers evaluated the effectiveness of deep learning methods such as AE and VAE, in combination with OCSVM, for detecting anomalies using a semi-supervised learning approach. During model creation, only normal flow-based data was used. Furthermore, both normal and anomaly data were used for testing the models. To put it differently, the study aimed to assess the anomaly detection capabilities of VAE and AE deep learning techniques, along with OCSVM, using a semi-supervised learning method based on network flow features. Only normal flow-based data was used in building the models, while both normal and anomalous data were employed for testing purposes.

(Yong Zhang, 2019) published the article PCCN: Parallel Cross Convolutional Neural Network for Abnormal Network Traffic Flows Detection in Multi-Class Imbalanced Network Traffic Flows, which aims to propose a new intrusion detection network based on deep learning, named parallel cross convolutional neural network (PCCN), to improve the detection performance of imbalanced abnormal flows. By fusing the flow features learned from the two branch convolutional neural networks (CNN), PCCN can better learn the flow features with fewer samples, to improve the detection results of the imbalanced abnormal flows. We proposed an improved feature extraction method of the original flow to extract multi-class flow features at the same time.

(Xiaojuan Wang, 2019) This work titled "Network Intrusion Detection: Based on Deep Hierarchical Network and Original Flow Data." The study aimed to extract raw data information directly from network flow for analysis, eliminating the need for manual feature design. The authors introduced a novel network intrusion detection model called the deep hierarchical network, which combined the improved LeNet-5 and LSTM neural network structures to learn spatial and temporal features of the flow. A reasonable network cascading method was designed to enable simultaneous training of the hierarchical network, instead of training two separate networks. The performance of the proposed hierarchical network model was evaluated using the CICIDS2017 and CTU datasets, which contain numerous and diverse types of flows with relatively new attack types. The experimental results revealed that the proposed model outperformed other network intrusion detection models and achieved the highest detection accuracy. Additionally, the authors presented a traffic feature analysis method that contributed significantly to abnormal traffic detection by providing insights into the actual meanings of important features.

(Ulya Sabeel, 2021) Building an Intrusion Detection System to Detect Atypical Cyberattack Flows is the title for paper . The suggested system for binary attack flow identification was discussed in this research, and the CICIDS2017 dataset was used to train and validate the AI models. The system is then tested using synthetic attack flows that are designed to resemble real-world circumstances. Several Deep Learning and

Machine Learning models, including DNN, Linear-SVC, and Stacked Decision Tree Classifier (S-DTC), are used to show the usefulness of the suggested unusual attack flow detection approach. According to simulation data, the suggested defensive AI engine greatly raises the True Positive Rate (TPR) of AI models against numerous unusual attacks.

(Satish Kumar, 2021) Research Trends in Network-Based Intrusion Detection Systems was the title of a study that was published. The analysis discussed in this paper is based on the total number of intrusion detection-related articles published each year, the number of citations each article received after it was published, and the most frequently cited research papers about intrusion detection systems in journals and conferences, separately. This paper also covers the state-of-the-arts of NIDS, commonly used NIDS, citation-based analysis of benchmark datasets, and NIDS approaches utilised for intrusion detection based on the published articles in the intrusion detection field for the last 15 years. This report also includes a comparative analysis based on publications and citations to measure the popularity of various approaches. The research presented in this article might be useful for beginners and researchers who want to assess current NIDS research trends and their applications.

(Giuseppina Andresini, 2020) Based on Multi-Channel Deep Feature Learning for Intrusion Detection was proposed, An intrusion detection mechanism for networks called MINDFUL was discussed in this paper. A convolution neural network that has been trained on a multi-channel representation of network ows is used to learn an intrusion detection model. In both the normal and attack ows, two autoencoders are trained. With the feature vectors created by these autoencoders, they are employed to provide the original feature vector representation of the network ows. The fundamental contention is that patterns, derived from the original features and their autoencoder-based equivalents, may be found throughout the channels..

(Aimin Yang, 2019) This paper provided a concept for Design of Intrusion Detection System for Internet of Things Based on Improved BP Neural Network. The LM-BP neural network model is proposed in this article. An intrusion detection system is supplied after the LM-BP neural network model has been applied to it, together with the LM-BP algorithm's flow for intrusion detection. The weight threshold of a conventional BP neural network is optimized using the LM method, which possesses the properties of quick optimization speed and good robustness. The KDD CUP 99 intrusion detection data set is imported into an LM-BP neural network classifier, and the best results are attained by continual training. Finally, the results of the experimental simulation demonstrate that this model has a greater detection rate and a lower false alarm rate compared to the conventional BP neural network model and the PSO-BP neural network model for DOS, R2L, U2L, and probing. As a consequence, this modified model has some promotion value.

(Tongtong Su, 2020) BAT: Deep Learning Methods on Network Intrusion Detection Using NSL-KDD Dataset was a model that was put up. A traffic anomaly detection model called BAT is suggested in this model to address the issues of low accuracy and feature engineering in intrusion detection. BLSTM (Bidirectional Long Short-Term Memory) and attention mechanisms are combined in the BAT model. The network flow vector, which is made up of packet vectors produced by the BLSTM model and may extract the essential properties for classifying network traffic, is screened using an attention mechanism.

We also use many convolutional layers to capture the regional characteristics of the traffic data. We refer to the BAT model as the BAT-MC since several convolutional layers are used to process data samples. Network traffic classification is done using the softmax classifier. The suggested end-to-end model may automatically pick up the most important features of the hierarchy without the need for feature engineering expertise. It can efficiently explain the behaviour of network traffic and enhance the ability to detect anomalies. We evaluate our model using a publicly available benchmark dataset, and the experimental findings show that it performs better than existing comparison techniques.

(Ao Liu, 2019) the model was proposed A system for detecting intrusions based on the quantitative model of port interactions. The model provides a quantitative description of Port Interaction Mode in Data Link Layer (PIMDL), with an emphasis on enhancing the precision and effectiveness of intrusion detection by considering the traffic's arrival time distribution. With the help of the phase space reconstruction and visualisation technique, the proposed model's viability is demonstrated. A neural network built on CNN and LSTM is intended to mine the differences between normal and pathological models based on the traits of long and short sessions. In order to categorise sessions in model space, a better intrusion detection system built on a multimodel scoring mechanism is developed. Session classification in model space is based on a multimodel scoring system. And the results demonstrate that the suggested enhanced algorithm and quantitative model can not only successfully prevent camouflaging identification information but also boost computing efficiency and improve small sample anomaly detection accuracy.

(Manuel Lopez-Martin, 2021) Network Intrusion Detection Based on Extended RBF Neural Network with Offline Reinforcement Learning was published. By incorporating the traditional radial basis function (RBF) neural network as a policy network in an offline reinforcement learning method, this work expands its capabilities. With this method, all radial basis function parameters (together with the weights of the network) are learned end-to-end by gradient descent without the use of external optimisation.

(Yousef Abuadlla, 2019) Flow-Based Anomaly Intrusion Detection System Using Two Neural Network Stages is the title of an article by. In this study, two levels of neural networks are used to develop flow-based anomaly IDS. The implemented system in numerous earlier studies [18], [8], [19] is a neural network based on DARPA [20] or KDD [21] dataset with the ability to recognise normal or aberrant traffic. Using data that was retrieved from the labelled DARPA dataset, our study classifies the types of assaults that have occurred as well as their existence. The limitation on the type of data that may be extracted from the router NetFlow data was placed on the extracted data. This labelled data will be referred to as labelled NetFlow dataset or simply NetFlow dataset in the material that follows, and the training procedure for neural network is based on it.

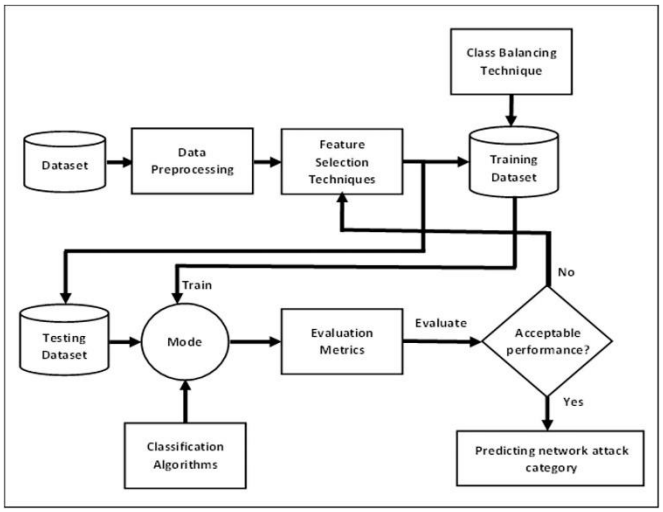
(Khac-Hoai Nam Bui, 2020)In order to enhance autonomous learning and cut down on time-consuming operations, we concentrate on hyperparameter tuning for training large traffic datasets in this paper. As a case study for the experiment, we use information from the Vehicle Detection System (VDS) to gauge the effectiveness of our strategy. Data has been gathered specifically from 21 sensors that are situated in an urban area. Our suggested method for training various traffic datasets has shown promising results in trials.

(Soosan Naderi Mighan, 2018) has published a paper named "Deep learning based latent feature

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1, January 2024
Paper: 68

extraction for intrusion detection." In this study, the network-based IDS system that employs anomaly detection techniques is the main topic. Stack Auto-encoder Network (SAE), followed by an SVM classifier, is the method we suggest for intrusion detection. A feature reduction technique called SAE is applied. On the UNB ISCX 2012 IDS dataset, the effectiveness of combining SVM classifier with deep learning techniques is assessed. On a classification task, our suggested method's precision exceeds SVM.

III. METHODS



Data Pre-Processing:

To prepare data for intrusion detection using machine learning algorithms, it's crucial to apply appropriate preprocessing steps. Here are the key points to keep in mind:

- Categorical data must be transformed into numerical data before applying scaling techniques. However, using techniques such as Min-Max scaling may not be appropriate due to the unknown correlation between the original and the transformed data. This can result in misleading or inaccurate results.

- To safely convert categorical data into numerical data, the one-hot encoding technique is recommended. This technique involves creating a binary column for each category, and assigning a value of 1 if the category is present, and 0 if it's not. This ensures that each category is treated independently, without introducing any artificial correlation between them.

- Proper data preprocessing is critical for the success of intrusion detection using machine learning algorithms. By applying appropriate techniques such as one-hot encoding, we can ensure that our models are trained on accurate and meaningful data, and can provide reliable results in identifying potential intrusions.

Feature Extraction:

- There are three basic classes of feature selection methods i.e. wrapper, filter, and embedded methods. Wrapper methods iteratively evaluate selected feature subsets using a learning algorithm to determine their accuracy, while filter methods use dataset characteristics to measure the relevance between a feature and the target label using measures such as distance and consistency. Embedded methods involve embedding the learning algorithm with no iterative evaluations for the classification accuracy of the feature subset.

- Wrapper methods are known to be more accurate in feature selection, but are computationally more expensive, while filter methods are easily scalable to high dimensions and can be performed in one iteration.

- During the training phase, feature coefficients are set by minimizing fitting miscalculations, resulting in selected features based on their coefficients. This approach is suitable for high-dimensional feature selection domains.

- The ability of the selected feature space to provide pertinent information is critical to the accuracy of the subsequent machine learning step. By selecting relevant features and removing irrelevant ones, we can ensure that our predictive model is trained on the most accurate and meaningful data, leading to more reliable and effective detection of botnets.

IDS Prediction Evaluation Metrics:

ML Model Creation:

The SVM classifier is based on statistical learning theory and is used to isolate a class of positive instances from a class of negative instances by creating a hyperplane using structural risk minimization rules. It determines the class to which each data point belongs by maximizing the margin between support vectors.

- SVM is a popular learning technique due to its high classification accuracy and ability to solve regression and classification tasks. It was initially designed for binary classification but was later extended to multi-class scenarios.

- SVM uses various basic functions, including linear, polynomial, sigmoid, and RBF kernels. The RBF kernel, also known as the Gaussian kernel, is used in our research.

- The hyperplane created by SVM can effectively separate classes of data points, and by maximizing the margin between support vectors, it can improve the accuracy of classification. This makes SVM a powerful tool in detecting and isolating positive instances from negative instances, allowing for effective identification of potential threats.

Evaluation Metrics:

To see how well our method is working, we use different measurements called performance metrics. We calculate these measurements using the values that we get from training and testing the NSL-KDD dataset.
Here are the important terms and what they mean:

- True Positive (TP): When we correctly identify an anomaly.
- False Positive (FP): When we wrongly identify a normal instance as an anomaly.
- True Negative(TN): When we correctly identify a normal instance.
- False Negative (FN): When we wrongly identify an anomaly instance as a normal instance.

We use these terms to calculate different performance metrics. Important metric used in this research are accuracy, precision, recall and F-measure.

Accuracy:

Accuracy is a measure that tells us how often our model or classifier is correct in predicting the correct class or label of a given record in the testing set. It represents the ratio of the number of correct predictions to the total number of predictions made.

(TP + TN) / (TP + FP + TN + FN)

Precision:

Precision measures the accuracy of a model in predicting true positive instances. It calculates the ratio of correctly predicted intrusions to the total number of predicted intrusions. In simple terms, precision evaluates the correctness of the model in identifying actual attacks among all the instances classified as attacks.

TP / (TP + FP)

Recall:

Recall is the measure of the proportion of correctly predicted intrusion instances out of the total actual intrusion instances present in the testing dataset.

TP / (TP + FN)

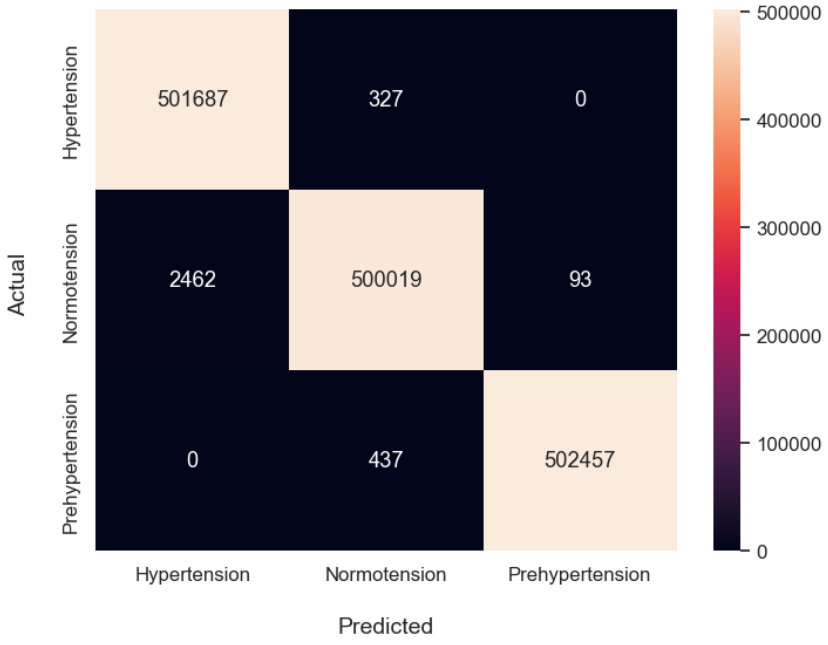
F-Score:

The F-measure is an important metric used in network intrusion detection, as it takes into account both precision and recall. It is calculated as the harmonic mean of precision and recall, representing how well the model distinguishes between different classes. A decrease in precision or recall will result in a decrease in the F1-score. For multiclass classification, both unweighted and weighted F1-scores are used, with the weighted F1-score considering label imbalance in the calculation.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1, January 2024
Paper: 68

2 * ((Precision * Recall) / (Precision + Recall))

RESULTS AND CONCLUSION



The aim of this study was to develop a reliable and effective network intrusion detection system using a hybrid machine learning approach that included data balancing and feature extraction. ML algorithms were used to evaluate the proposed method and performance metrics were used to determine the effectiveness of the binary and multiclass attack algorithms. The proposed approach performed well and can be implemented in real-time.

V. REFRENCES

[1] S. Caton and R. Landman, ‘‘Internet safety, online radicalization and young people with learning disabilities,’’ Brit. J. Learn. Disabilities, vol. 50, no. 1, pp. 88–97, Mar. 2022.
[2] J. Lewis, ‘‘Economic impact of cybercrime, no slowing down,’’ Center Strategic Int. Stud., McAfee, San Jose, CA, USA, 2018, vol. 13, p. 2019.
[3] P. Devan and N. Khare, ‘‘An efficient XGBoost– DNN-based classification model for network intrusion detection system,’’ Neural Compute. Appl., vol. 32, no. 16, pp. 12499–12514, 2020.

[4] H. Wang, Z. Cao, and B. Hong, ‘‘A network intrusion detection system based on convolutional neural network,’’ J. Intel. Fuzzy Syst., vol. 38, no. 6, pp. 7623–7637, Jun. 2020.
[5] S. Al and M. Dener, ‘‘STL-HDL: A new hybrid network intrusion detection system for imbalanced dataset on big data environment,’’ Compute. Secur., vol. 110, Nov. 2021, Art. no. 102435.
[6] Z. Wu, J. Wang, L. Hu, Z. Zhang, and H. Wu, ‘‘A network intrusion detection method based on semantic re-encoding and deep learning,’’ J. Newt. Compute. Appl., vol. 164, Aug. 2020, Art. no. 102688.
[7] Y. Li, J. Xia, S. Zhang, J. Yan, X. Ai, and K. Dai, ‘‘An efficient intrusion detection system based on support vector machines and gradually feature removal method,’’ Expert Syst. Appl., vol. 39, no. 1, pp. 424– 430, Jan. 2012.
[8] S. Tahseen and C. A. Kumar, ‘‘An analysis of supervised tree-based classifiers for intrusion detection system,’’ in Proc. Int. Conf. Pattern Recognit., Informat. Mobile Eng., Feb. 2013, pp. 294–299.
[9] M. Ge, N. F. Syed, X. Fu, Z. Baig, and A. RoblesKelly, ‘‘Towards a deep learning-driven intrusion detection approach for Internet of Things,’’ Comput. Netw., vol. 186, Feb. 2021, Art. no. 107784.
[10] M. C. Belavagi and B. Muniyal, ‘‘Performance evaluation of supervised machine learning algorithms for intrusion detection,’’ Proc. Comput. Sci., vol. 89, pp. 117–123, Jan. 2016.
[11] G. Loukas, T. Vuong, R. Heartfield, G. Sakellari, Y. Yoon, and D. Gan, ‘‘Cloud-based cyber-physical intrusion detection for vehicles using deep learning,’’ IEEE Access, vol. 6, pp. 3491–3508, 2018.
[12] S. Otoum, B. Kantarci, and H. T. Mouftah, ‘‘On the feasibility of deep learning in sensor network intrusion detection,’’ IEEE Netw. Lett., vol. 1, no. 2, pp. 68–71, Jun. 2019.
[13] R. Vinayakumar, M. Alazab, K. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, ‘‘Deep learning approach for intelligent intrusion detection system,’’ IEEE Access, vol. 7, pp. 41525–41550, 2019.
[14] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, ‘‘A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction,’’ J. Appl. Sci. Technol. Trends, vol. 1, no. 2, pp. 56–70, May 2020.
[15] M. Ring, S. Wunderlich, D. Scheuring, D. Landes, and A. Hotho, ‘‘A survey of network-based intrusion detection data sets,’’ Comput. Secur., vol. 86, pp. 147– 167, Oct. 2019.
[16] L. Ruiz, F. Gama, and A. Ribeiro, ‘‘Gated graph recurrent neural networks,’’ IEEE Trans. Signal Process., vol. 68, pp. 6303–6318, 2020.
[17] M. Xia, W. Liu, K. Wang, W. Song, C. Chen, and Y. Li, ‘‘Non-intrusive load disaggregation based on composite deep long short-term memory network,’’ Expert Syst. Appl., vol. 160, Dec. 2020, Art. no. 113669.
[18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, ‘‘BERT: Pre-training of deep bidirectional transformers for language understanding,’’ 2018, arXiv:1810.04805.
[19] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, ‘‘An image is worth 16×16 words: Trans formers for image recognition at scale,’’ 2020, arXiv:2010.11929.
[20] S. Huang, Y. Liu, C. Fung, R. He, Y. Zhao, H. Yang, and Z. Luan, ‘‘Hit Anomaly: Hierarchical transformers for anomaly detection in system log,’’ IEEE Trans. Newt. Service Manage., vol. 17, no. 4, pp. 2064–2076, Dec. 2020.