Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Intrusion Detection System Using Pca

Junaid Ahmed Khan, [2] Syed Essa Shah [3] Syed Adnan Aftab,

[4] Mr. Syed Omer Farooq

BE Student, Dept. of Computer Science Engineering, ISL Engineering College

BE Student, Dept. of Computer Science Engineering, ISL Engineering College

BE Student, Dept. of Computer Science Engineering, ISL Engineering College

Assistant Professor, Dept. of Computer Science Engineering, ISL Engineering College

Article Info
Page Number: 1484-1490
Publication Issue:
Vol. 72 No. 1 (2023)

Article History
Article Received: 15 October 2022
Revised: 24 November 2022
Accepted: 18 December 2022

ABSTRACT
Software-Defined Networking is an innovative architecture approach in the networking field. This technology allows networks to be centrally and intelligently managed by unified applications such as traffic classification and security management. Traditional networks' static nature has a minimal capacity to meet organizations' business requirements. Software-Defined Networks (SDNs) are emerging architectures that address a range of networking challenges with new solutions. Nevertheless, these centralized and programmable techniques face various challenges and issues that require contemporary security solutions such as Intrusion Detection Systems. Recently, the majority of this type of security solution has been developed using Machine Learning techniques. Deep Learning algorithms have recently been used to provide more accuracy and efficiency. This paper presents a new detection approach based on Convolutional Neural Network (CNN). The experiments proved that the proposed model could be successfully implemented in a Software-Defined Network controller to detect various attacks with 100% accuracy, and achieved a low degradation rate of 2.3% throughput and 1.8% latency when executed in a large-scale network.

1. INTRODUCTION

With the progressive development and improvement of Internet technology, the Internet is providing various convenient services for people. However, we are also facing various security threats. Network viruses, eavesdropping, and malicious attacks are on the rise, causing network security to become the focus of attention of society and government departments. Fortunately, these problems can be well solved via intrusion detection. Intrusion detection plays an important part in ensuring network information security.

However, with the explosive growth of Internet business, traffic types in the network are increasing day by day, and network behavior characteristics are becoming increasingly complex, which brings great challenges to intrusion detection [1], [2]. How to identify various malicious network traffics, especially unexpected malicious network traffics, is a key problem that cannot be avoided. In fact, network traffic can be divided into two categories (normal traffics and malicious traffics). Furthermore, network traffic can also be divided into five categories:

Normal, DoS (Denial of attacks), R2L (Root to Local attacks), U2R (User to Root attack), and Probe (Probing attacks). Hence, intrusion detection can be considered as a classification problem. By improving the performance of classifiers in effectively identifying malicious traffics, intrusion detection accuracy can be largely improved. Machine learning methods [3][8] have been widely used in intrusion detection to identify malicious traffic. However, these methods belong to shallow learning and often emphasize feature engineering and selection.

They have difficulty in feature selection and cannot effectively solve the massive intrusion data classification problem, which leads to low recognition accuracy and a high false alarm rate. In recent years, intrusion detection methods based on deep learning have been proposed successively. In [9], the authors propose a malware traffic classification method based on convolution.

1.1 Scope of the Project

The scope of using machine learning techniques for detecting cyber-attacks in a network is quite extensive. Machine learning algorithms can be applied to various stages of the network security process, from preprocessing and feature selection to the actual detection and response. Here are some key aspects to consider within this scope Data Collection and Preprocessing: Gathering network traffic data, system logs, and other relevant data sources.

2. LITERATURE REVIEW

In the research of network intrusion detection based on machine learning, scholars mainly distinguish normal network traffic from abnormal network traffic by dimensionality reduction, clustering, and classification, to realize the identity fiction of malicious attacks [10], [11].

Pervez proposed a new method for feature selection and classification merging of the multi-class NSL-KDD Cup99 dataset using Support Vector Machine (SVM) and discussed the classification accuracy of classifiers under different dimension features [12]. Shiraz studied some new technologies to improve CANN intrusion detection methods' classification performance and evaluated their performance on the NSL-KDD Cup99 dataset [13]. He used the K Farthest Neighbor (KFN) and the K Nearest Neighbor (KNN) to classify the data and used the Second Nearest Neighbor (SNN) of the data when the nearest and farthest neighbors have the same class label. The result shows the CANN detection rate and reduces the failure the alert rate is improved or provides the same performance. Bhattacharya proposed a machine learning model based on hybrid Principal Component Analysis (PCA)-Firefly [14].

The dataset used was the open dataset collected from Kaggle. Firstly, the model performs one key coding for transforming the IDS dataset, then uses the hybrid PCA-Firefly algorithm to reduce the dimension, and the XGBoost algorithm classifies the reduced dataset. In recent years, with the powerful ability of automatic feature extraction, deep learning has made remarkable achievements in the fields of Computer Vision (CV), Autonomous driving(AD), and Natural Language Processing(NLP). Many scholars apply deep learning to intrusion detection for traffic classification, which has become a hot spot of current research. The method of deep learning is to mine the potential characteristics of high-dimensional data through a training model and transform network traffic anomaly detection into a classification problem

[15]. Through a large number of sample data training, adaptive learning between normal network traffic and abnormal network traffic effectively enhances real-time intrusion processing. Torres et al. [16] first converted network traffic characteristics into a series of characters and then used Recurrent Neural Network (RNN) to learn their temporal characteristics, which were further used to detect malicious network traffic. Wang et al. [17] proposed a malicious software traffic classification algorithm based on Convolutional Neural Network(CNN). By mapping the traffic characteristics to pixels, the network traffic image is generated, and the image is used as the input of the CNN to realize traffic

3. PROPOSED SYSTEM

The proposed system aims to provide a We use the classic NSL-KDD and the up-to-date benchmark datasets and conduct detailed analysis and data cleaning. (2) This work proposes a machine learning algorithm, reducing the majority samples and augmenting the minority samples in the difficult set, tackling the class imbalance problem in intrusion detection so that the classifier learns the differences better in training. (3) The classification model uses Random Forest (RF), Support Vector Machine (SVM), XGBoost, NLP with other methods, we divide the experiment into 30 methods.

We propose an end-to-end deep learning model with ml models that is composed of logistic regression and attention mechanism. CNN can well solve the problem of Software Defined Networks and provide a new research method for Early Warning Proactive System.

We compare the performance of ML Modes with traditional deep learning methods, the model can extract information from each packet. By making full use of the structure information of network traffic, the logistic regression model can capture features more comprehensively. 4) We evaluate our proposed network with a real NSL-KDD dataset. The experimental results show that the performance of the algorithm is better than the traditional methods

4. IMPLEMENTATION

To conduct studies and analyses of an operational and technological nature, and To promote the exchange and development of methods and tools for operational analysis as applied to defense problems. The logical design of a system pertains to an abstract representation of the data flows, inputs and outputs of the system. This is often conducted via modeling, using an over-abstract (and sometimes graphical) model of the actual system. In the context of systems design are included.

The logical design includes ER Diagrams i.e. Entity Relationship Diagrams. The physical design relates to the actual input and output processes of the system. This is laid down in terms of how data is input into a system, how it is verified/authenticated, how it is processed, and how it is displayed as output. In Physical design, the following requirements about the system are decided.

Detecting cyber-attacks in a network using machine learning techniques involves training a model to recognize patterns and anomalies in the nearest traffic data.

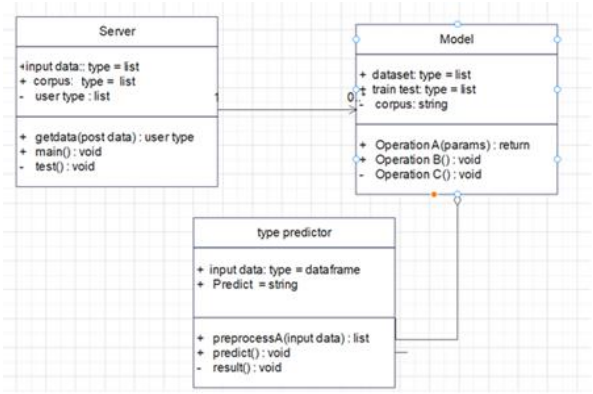

Fig 1: Class Diagram

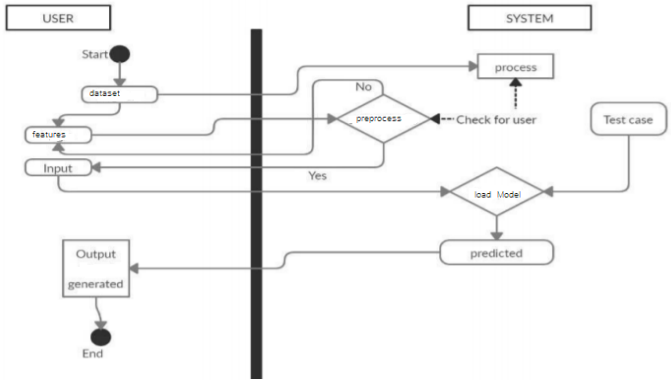

Fig 2: Activity Diagram

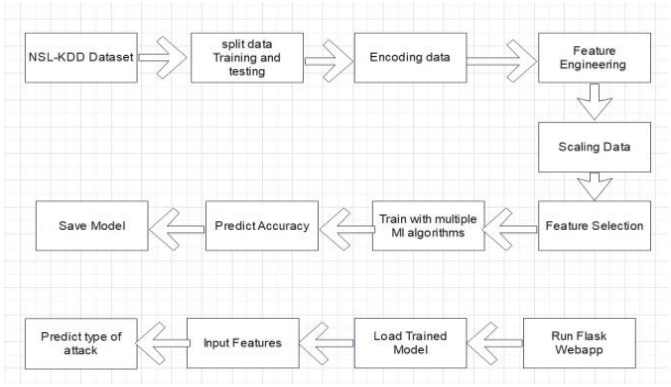

Fig 3: System Architecture

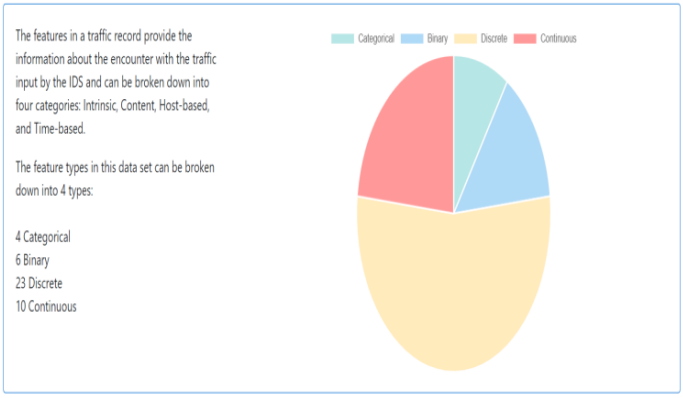

Fig 4: Pie Chart

Fig 5: Analysis

5. Conclusion and Future Enhancement

As network intrusion continues to evolve, the pressure on network intrusion detection is also increasing. In particular, the problems caused by imbalanced network traffic make it difficult for intrusion detection systems to predict the distribution of malicious attacks, making cyberspace security face a considerable threat. This paper proposed a novel Difficult Set Sampling Technique, which enables the classification model to strengthen imbalanced network data learning.

A targeted increase in the number of minority samples that need to be learned can reduce the imbalance of network traffic and strengthen the minority's learning under challenging samples to improve classification accuracy. We used six classical classification methods in machine learning and deep learning and combined them with other sampling techniques. Experiments show that our method can accurately determine the samples that need to be expanded in the imbalanced network traffic and improve the attack recognition more effectively.

In the experiment, we found that deep learning performance is better than machine learning after sampling the imbalanced training set samples through the MLP algorithm. Although the neural networks strengthen data expression, the current public datasets have already extracted the data features in advance, which is more limited for deep learning to learn the preprocessed features and cannot take advantage of its automatic feature extraction.

Therefore, in the next step, we plan to directly use the deep learning model for feature extraction and model training on the original network traffic data, performance the advantages of deep learning in feature extraction, reduce the impact of imbalanced data and achieve more

REFERENCES

1. D. E. Denning, ‘‘An intrusion-detection model,’’ IEEE Trans. Softw. Eng., vol. SE-13, no. 2, pp. 222–232, Feb. 1987.
2. Anupong, W., Yi-Chia, L., Jagdish, M., Kumar, R., Selvam, P. D., Saravanakumar, R., & Dhabliya, D. (2022). Hybrid distributed energy sources providing climate security to the agriculture environment and enhancing the yield. Sustainable Energy Technologies and Assessments, 52 doi:10.1016/j.seta.2022.102142
3. Aoudni, Y., Donald, C., Farouk, A., Sahay, K. B., Babu, D. V., Tripathi, V., & Dhabliya, D. (2022). Cloud security based attack detection using transductive learning integrated with hidden markov model. Pattern Recognition Letters, 157, 16-26. doi:10.1016/j.patrec.2022.02.012
4. Dhabliya, D. (2021). Delay-tolerant sensor network (DTN) implementation in cloud computing. Paper presented at the Journal of Physics: Conference Series, , 1979(1) doi:10.1088/1742-6596/1979/1/012031 Retrieved from www.scopus.com
5. Dhabliya, D. (2019). Security analysis of password schemes using virtual environment. International Journal of Advanced Science and Technology, 28(20), 1334-1339. Retrieved from www.scopus.com
6. N. B. Amor, S. Benferhat, and Z. Elouedi, ‘‘Naive Bayes vs decision trees in intrusion detection systems,’’ in Proc. ACM Symp. Appl. Comput. (SAC), 2004, pp. 420–424.
7. M. Panda and M. R. Patra, ‘‘Network intrusion detection using Naive Bayes,’’ Int. J. Comput. Sci. Netw. Secur., vol. 7, no. 12, pp. 258–263, 2007.
8. M. A. M. Hasan, M. Nasser, B. Pal, and S. Ahmad, ‘‘Support vector machine and random forest modeling for intrusion detection system (IDS),’’ J. Intell. Learn. Syst. Appl., vol. 6, no. 1, pp. 45–52, 2014.
9. N. Japkowicz, ‘‘The class imbalance problem: Significance and strategies,’’ in Proc. Int. Conf. Artif. Intell., vol. 56, 2000, pp. 111–117.
10. Y. LeCun, Y. Bengio, and G. Hinton, ‘‘Deep learning,’’ Nature, vol. 521, no. 7553, pp. 436–444, 2015.
11. Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, ‘‘Deep learning for visual understanding: review,’ Neurocomputing, vol. 187, pp. 27–48, Apr. 2016.
12. T. Young, D. Hazarika, S. Poria, and E. Cambria, ‘‘Recent trends in deep learning based natural language processing [review article],’’ IEEE Comput. Intell. Mag., vol. 13, no. 3, pp. 55–75, Aug. 2018.
13. N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, ‘‘A deep learning approach to network intrusion detection,’’ IEEE Trans. Emerg. Topics Comput. Intell., vol. 2, no. 1, pp. 41–50, Feb. 2018.
14. D. A. Cieslak, N. V. Chawla, and A. Striegel, ‘‘Combating imbalance in network intrusion datasets,’’ in Proc. IEEE Int. Conf. Granular Comput., May 2006, pp. 732–73.
15. Syed Shehriyar Ali, Mohammed Sarfaraz Shaikh, Syed Safi Udin, Dr. Mohammed Abdul

Bari, “Saas Product Comparison and Reviews Using Nlp”, Journal of Engineering Science (JES), ISSN NO:0377-9254, Vol 13, Issue 05, MAY/2022

16. Hafsa Fatima, Shayesta Nazneen, Maryam Banu, Dr. Mohammed Abdul Bar,” Tensorflow-Based Automatic Personality Recognition Used in Asynchronous Video Interviews”, Journal of Engineering Science (JES), ISSN NO:0377-9254, Vol 13, Issue 05, MAY/2022
17. Mohammed Shoeb, Mohammed Akram Ali, Mohammed Shadeel, Dr. Mohammed Abdul Bari, “Self-Driving Car: Using Opencv2 and Machine Learning”, The International journal of analytical and experimental modal analysis (IJAEMA), ISSN NO: 0886-9367, Volume XIV, Issue V, May/2022
18. Mr. Pathan Ahmed Khan, Dr. M.A Bari,: Impact Of Emergence With Robotics At Educational Institution And Emerging Challenges”, International Journal of Multidisciplinary Engineering in Current Research(IJMEC), ISSN: 2456-4265, Volume 6, Issue 12, December 2021,Page 43-46
19. Mohammed Abdul Bari, Shahanawaj Ahamad, Mohammed Rahmat Ali,” Smartphone Security and Protection Practices”, International Journal of Engineering and Applied Computer Science (IJEACS) ; ISBN: 9798799755577 Volume: 03, Issue: 01, December 2021
20. Shahanawaj Ahamad, Mohammed Abdul Bari, Big Data Processing Model for Smart City Design: A Systematic Review “, VOL 2021: ISSUE 08 IS SN : 0011-9342 ;Design Engineering (Toronto) Elsevier SCI Oct
21. Mohammed Abdul Bari, Shahanawaj Ahamad, Mohammed Rahmat Ali,” Smartphone Security and Protection Practices”, International Journal of Engineering and Applied Computer Science (IJEACS) ; ISBN: 9798799755577 Volume: 03, Issue: 01, December 2021 (International Journal,U K)
22. Dr. M.A.Bari, ”EffectiveIDS To Mitigate The Packet Dropping Nodes From Manet “, JACE, Vol -6,Issue -6,June 2019
23. M.A.Bari & Shahanawaj Ahamad,” Process of Reverse Engineering of Enterprise InformationSystem Architecture” in International Journal of Computer Science Issues (IJCSI), Vol 8, Issue 5, ISSN: 1694-0814, pp:359-365,Mahebourg ,Republic of Mauritius , September 2011
24. M.A.Bari & Shahanawaj Ahamad, “Code Cloning: The Analysis, Detection and Removal”, in International Journal of Computer Applications(IJCA),ISSN:0975-887, ISBN:978-93-80749-18-3,Vol:20,No:7,pp:34-38, NewYork,U.S.A.,April 2011