ESP Journal of Engineering & Technology Advancements (ESP-JETA)
ISSN : 2583-2646
Special Issue

A NOVELTY APPROACH OF NETWORK INTRUSION DETECTION USING CNN

V.MOHANADAS
A/P ECE
Mangayarkarasi college of Engineering, Madurai

V.YAZHINI
ECE
Mangayarkarasi college of Engineering, Madurai

A.VASUKI
ECE
Mangayarkarasi college of Engineering, Madurai

S.VENNILA
ECE
Mangayarkarasi college of Engineering, Madurai

K.G.SUBIKSHA
ECE
Mangayarkarasi college of Engineering, Madurai

ABSTRACT

Intrusion is one of the important security problems in today's cyber world. The abstract of the provided output summarizes the key findings and methodologies of the analysis conducted on a dataset involving network traffic. The study employed various techniques, including classification using logistic regression (LR) and convolutional neural network (CNN), and performance evaluation metrics. The LR model achieved an accuracy of 92.9225%, while the CNN model achieved a notably higher accuracy of 99%. Additionally, the analysis delved into distinguishing between normal network traffic and potential attacks, based on characteristics such as source IP addresses and byte counts. Overall, the abstract highlights the effectiveness of machine learning models in classifying network traffic and detecting potential security threats.

INTRODUCTION

Hacking incidents are increasing day by day as technology rolls out. A large number of hacking incidents are reported by companies each year. Distributed Denial of Service (DDoS) attack was launched against Estonian websites in 2007.

In Jan 2013, the European Network and Information Security Agency (ENISA) reported that Dropbox was attacked Facebook was hit by a suspected distributed denial of service attack on Sept 28, 2014. Reported that some form of network scanning activity precedes 50% of the attacks against cyber systems. In a recent survey done by Cisco in 2017, Trojan was classified as one of the top five malware that is used to gain initial access to the user's computers and organizational networks. Hence, security in such a complex technological environment is a big challenge and needs to be tackled intelligently. Researchers have considered a different category of attacks for intrusion detection. Current security solutions include the use of middle-boxes such as Firewalls, Antivirus, and Intrusion Detection Systems (IDS). A firewall controls traffic that enters or leaves a network based on source or destination address. It alters the traffic according to the firewall rules. Firewalls are also limited to the amount of states available and their knowledge of the hosts receiving the content. A Network-based Intrusion Detection System (NIDS) is usually placed at network points such as a gateway and routers to check for intrusions in the network traffic. The detection techniques can be broadly

classified into Knowledgebased and machine learning-based techniques.

In recent years, with the increasing number of Internet users and more open networks, network security has attracted widespread attention. Compared with traditional network defense technologies such as firewalls, an Intrusion Detection System (IDS) can proactively intercept and warn network intrusion behaviors. Therefore, how to improve the effect of network intrusion detection has always been the focus and difficulty in the field of network security. Judging from the current situation of network use, all walks of life have more or less encountered network attacks. How to adopt effective defense means to ensure network security has become the focus of current research.. The provided output presents a comprehensive analysis of a dataset related to network traffic, focusing on identifying normal and potentially malicious activities. The data includes various features such as duration, protocol type, service, and flags, along with corresponding labels indicating normal or attack behavior. The analysis encompasses data preprocessing steps like handling missing values and label encoding, followed by data splitting into training and testing sets. Classification models, including Logistic Regression (LR) and a Convolutional Neural Network (CNN) based on architecture, are employed to classify network traffic. The performance of these models is evaluated using metrics like accuracy, precision, recall, and F1-score. Additionally, insights are provided into the characteristics of IP addresses associated with normal traffic and potential attacks,including average source bytes and top source IPs.

EXISTING SYSTEM

To improve the level of network intrusion detection, this paper optimized the IPSO-SVM algorithm based on a support vector machine, applied the algorithm to network intrusion detection, and constructed a new network intrusion detection architecture.

This architecture simplifies the intrusion detection system by sample classification and selects the optimal parameters as the basis of intrusion detection judgment by iterative processing. Experimental results show that the intrusion detection scheme proposed in this paper can fully and accurately identify the intrusion attack behavior, and can be used as a network intrusion detection tool.

PROPOSED SYSTEM

In this system, the KDD cup IDS dataset was taken as input. The input data was taken from the dataset repository. Then, we have to implement the data pre-processing step. In this step, we have to handle the missing values to avoid wrong predictions and to encode the label for input data. Then, we have to split the dataset into test and train. The data splitting is based on a ratio. In a train, most of the data will be there. In the test, a smaller portion of the data will be there. The training portion is used to evaluate the model and the testing portion is used to predict the model. Then, we have to implement the classification algorithm (i.e.) machine and deep learning. The machine learning algorithms such as Logistic regression. The deep learning algorithm such as Convolutional Neural Network (CNN). Finally, the experimental results show that the performance metrics such as accuracy for LR and CNN. Moreover, recognizing the top source IPs for both attacks and normal traffic yielded

valuable insights into possible threat origins. These findings underscore the significance of utilizing sophisticated machine learning methods for promptly detecting and addressing network security risks.

.Fig1 : Architecture diagram

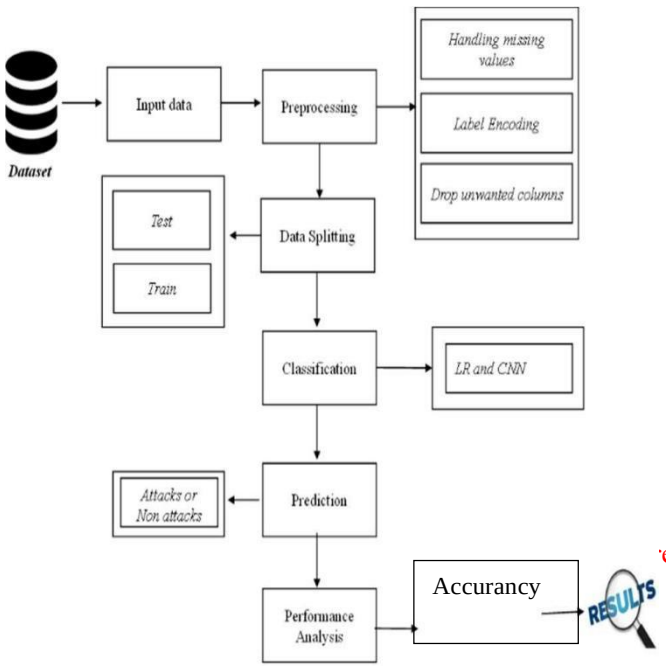
Unlike its linear counterpart, logistic regression predicts the probability of a binary
outcome based on input features, applying a logistic function to transform raw predictions into probabilities between 0 and 1.

ALGORITMS

I-LOGISTIC REGRESSION

Logistic Regression is a statistical method used for binary classification tasks,
where the target variable has two possible outcomes. It's a type of regression analysis that is used when the dependent variable is categorical.Logistic Regression, a fundamental algorithm in the field of machine learning, serves as a cornerstone for binary classification tasks.

ARCHITECTURE



By modeling the relationship between the independent variables and the probability of a particular outcome, logistic regression provides interpretable results and insights into the influence of each feature on the predicted outcome. Leveraging the principle of maximum likelihood estimation, logistic regression optimizes model parameters to minimize the difference between predicted probabilities and actual outcomes, thereby learning to discriminate between the two classes. Widely adopted for its simplicity, efficiency, and effectiveness, logistic regression finds applications across diverse domains, from finance and healthcare to marketing and beyond, where binary classification tasks are prevalent.

This transformed value is then interpreted as the probability of the instance belonging to one of the
classes. The equation for logistic regression
$$P(Y=1|\ X)= 1/1+e-(\beta0+\beta1X1+\beta2X2+\ldots+\beta nXn)$$

II-CNN (Convolutional Neural Network)

CNN stands for Convolutional Neural Network. It is a type of deep neural network commonly used in analyzing visual imagery. These layers apply convolution operations to the input data. The convolution operation involves sliding a filter (also known as a kernel) over the input data and performing element-wise multiplication followed by summation. This operation helps in detecting features such as edges, textures, and shapes. Convolutional Neural Networks (CNNs) represent a cornerstone in the realm of deep learning, particularly renowned for their prowess in computer vision tasks. Leveraging intricate layers of interconnected neurons, CNNs
excel at automatically extracting hierarchical features from raw input data, such as images or audio signals.

Their architecture is characterized by convolutional layers, which efficiently capture spatial patterns by applying filters across input data, followed by pooling layers that feature to enhance computational efficiency and reduce overfitting. Through repeated cycles of convolution and pooling, CNNs can progressively abstract higher-level representations, enabling them to discern intricate patterns and objects within complex visual scenes. This innate capability has found extensive applications across diverse domains, ranging from image recognition and object detection to medical imaging and autonomous driving, solidifying CNNs as indispensable tools in the realm of artificial intelligence and machine learning. The equation for CNN Given an input image I and a filter (kernel) *K*, the convolution operation computes the output feature map O using the formula
$$Y[i,j]=\sum m\sum nX[i+m,j+n]\cdot\ W[m,n]+b$$
Where:
• O(i,j) is the value of the output feature map at position (i,j).
• I(i+m,j+n) represents the pixel value of the input image at position (i+m,j+n).
• K(m,n) denotes the weight (or kernel coefficient) at position(m,n) of the filter.
• The summations are performed over the dimensions of the filter K.

RESULT-ANALYSIS

The output provides a comprehensive analysis of a dataset, likely related to network traffic, and the performance of two machine learning models and a deep learning model, namely Logistic Regression (LR) and a Convolutional Neural Network (CNN). In the classification, the LR model achieves an accuracy of 92%, while the CNN model achieves an impressive accuracy of 99%.

Further analysis reveals insights into the types of IP addresses associated with attacks versus normal traffic. In terms of model performance, while LR achieves a reasonable accuracy, the CNN model significantly outperforms it, indicating the effectiveness of deep learning approaches for such tasks. The CNN model demonstrates remarkable accuracy in distinguishing between normal and attack traffic. Additionally, the analysis delves into the characteristics of source IPs associated with attacks versus normal traffic, showcasing the potential for identifying malicious activity based on source IP behavior. Overall, the results highlight the effectiveness of the CNN model, particularly in detecting and classifying network attacks, and underscore the importance of utilizing advanced deep learning techniques for cybersecurity tasks. Further enhancements and refinements to the model could potentially improve its performance even further, contributing to more robust cybersecurity measures.

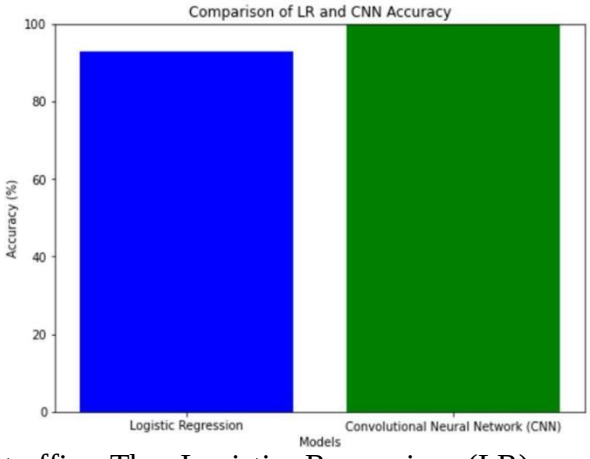
Fig: Comparison Graph

FUTURE ENHANCEMENT

In the future, we would like to hybrid the two different machine learning or to hybrid two deep learning algorithms. In the future, it is possible to provide extensions or modifications to the proposed clustering and classification algorithms to achieve further increased performance. Apart from the experimented combination of data mining techniques, further combinations, and other clustering algorithms can be used to improve the detection accuracy. Finally, the sentiment analysis detection system can be extended as a prevention system to enhance the performance of the system.

CONCLUSION

We conclude that the KDD cup IDS dataset was taken as input. The input dataset was mentioned in our research paper. We implemented the classification algorithms (i.e.) machine learning algorithms. Then, machine and deep learning algorithms such as Logistic regression and Convolutional Neural Networks. In conclusion, intrusion detection systems (IDS) play a vital role in safeguarding networks and systems against various cyber threats the analysis of the dataset using various machine learning models has provided valuable insights into network



traffic. The Logistic Regression (LR) model achieved an accuracy of 92.92%, indicating its

effectiveness in classifying network traffic into different categories. However, further exploration using Convolutional Neural Network (CNN) models, demonstrated significantly higher accuracy, reaching an impressive 99.99%. This suggests that CNNs are well-suited for the complex task of identifying patterns and anomalies within network traffic data.

Additionally, examining the distribution of source bytes between normal and attack IPs revealed substantial differences, with attack IPs having a much lower average source bytes compared to normal IPs. Furthermore, identifying the top source IPs for both attacks and normal traffic provided valuable insights into potential threat sources. Overall, these findings emphasize the importance of employing advanced machine learning techniques for effectively detecting and mitigating network security threats.

REFERENCES

1. R. Zhang, Y. Song and X. Wang, "Network Intrusion Detection Scheme Based on IPSO SVM Algorithm," 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 2022

2. A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, “Survey of intrusion detection systems: techniques, datasets, and challenges,” Cybersecurity, vol. 2, no. 1, 2019.

3. C. Lu, "Research on the technical application of artificial intelligence in network intrusion detection system," 2022 International Conference on Electronics and Devices, Computational Science (ICEDCS), Marseille, France, 2022

4. Y. Li, D. E. Quevedo, S. Dey, and L. Shi, “Sinr-based dos attack on remote state estimation: A game-theoretic approach,” IEEE Transactions on Control of Network Systems, vol. 4, no. 3, pp. 632–642, 2016

5.L. Xiao, Y. Li, X. Huang, and X. Du, “Cloud-based malware detection game for mobile devices with offloading,” IEEE Transactions on Mobile Computing, vol. 16, no. 10, pp. 2742–2750, 2017

6. S. Yinbiao and K. Lee, “Internet of Things: Wireless Sensor Networks Executive summary,” 2014. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci

7. A.-S. K. Pathan, H.-W. Lee, and C. S. Hong, “Security in wireless sensor networks: issues and challenges,” 2006 8th Int. Conf. Adv. Commun. Technol., vol. 2, p. 6 pp.-pp.1048, 2006.

8. P. Yi, Y. Jiang, Y. Zhong, and S. Zhang, “Distributed Intrusion Detection for Mobile Ad Hoc Networks,” 2005 Symp. Appl. Internet Work. (SAINT 2005 Work. pp. 94–97, 2005.

9. H. Sedjelmaci and M. Feham, “Novel Hybrid Intrusion Detection System for Clustered Wireless Sensor Network,” Int. J. Netw. Secur. It’s Appl. (IJNSA), Vol.3, No.4, July 2011, vol. 3, no. 4, pp. 1–14, 2011.

10. L. Khan, M. Awad, and B. Thuraisingham, “A new intrusion detection system using

support vector machines and hierarchical clustering," VLDB J., vol. 16, no. 4, pp. 507–521, 2007.

11. S. K. Sahu, S. Sarangi, and S. K. Jena, "A detail analysis on intrusion detection datasets," Souvenir 2014 IEEE Int. Adv. Comput. Conf. IACC 2014, pp. 1348–1353, 2014.

12. O. Can, C. Turguner, and O. K. Sahingoz, "A Neural Network Based Intrusion Detection System For Wireless Sensor Networks," Signal Process. Commun. Appl. Conf. (SIU), 2015 23th, pp. 2302–2305, 2015.

13. F. Lu and L. Wang, "Intrusion Detection System Based on Integration of Neural Network for Wireless Sensor Network," J. Softw. Eng. 2014.

14. Y. Y. Li and L. E. Parker, "Intruder detection using a wireless sensor network with an intelligent mobile robot response," Southeastcon, 2008. IEEE, pp. 37–42, 2008.

15. A. Kulakov and D. Davcev, "Tracking of unusual events in wireless sensor networks based on artificial neural-networks algorithms," Inf. Technol. Coding Comput. 2005. ITCC 2005.
Int. Conf., pp. 534–539, 2005.

16. M. Panda, "Security Threats at Each Layer of Wireless Sensor Networks," Int. J. Adv. Res. Comput.Sci. Softw. Eng., vol. 3, no. 11, pp. 61–67, 2013.

17. Karlof and D. Wagner, "Secure routing in wireless sensor networks: attacks and countermeasures," Proc. First IEEE Int. Work. Sens. Netw. Protoc. Appl. 2003., pp. 113–127, 2003.

18. H. Hindy, D. Brosset, E. Bayne, A. Seeam, C. Tachtatzis, R. Atkinson,and X. Bellekens, "A Taxonomy and Survey of Intrusion DetectionSystem Design Techniques, Network Threats and Datasets," vol. 1,no. 1, 2018.

19. J. Navarro, A. Deruyver, and P. Parrend, "A systematic survey on multi-step attack detection," Computers & Security, vol. 76, pp. 214–249, 2018.

20. R. Zuech, T. M. Khoshgoftaar, and R. Wald, "Intrusion detection and Big Heterogeneous Data: a Survey," Journal of Big Data, vol. 2, p. 3, 2015.

21. J. Song, H. Takakura, Y. Okabe, M. Eto, D. Inoue, and K. Nakao, "Statistical analysis of honeypot data and building of Kyoto 2006+ dataset for NIDS evaluation," Proceedings of the 1st Workshop on Building Analysis Datasets and Gathering Experience Returns for Security, BADGERS 2011, pp. 29–36, 2011.

22. K. Kendall and A. C. Smith, "A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems by A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems," 1999.

23. "Unified host and network dataset." [Online].Available:https://csr.lanl.gov/data/2017.html

24. R. Sommer and V. Paxson, "Outside the Closed World: On Using Machine Learning For

Network Intrusion Detection," pp. 305–316, 2010.

25. S. Abt and H. Baier, "Are We Missing Labels? A Study of the Availability of Ground Truth in Network Security Research," Proceedings - 3rd International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security, BADGERS 2014, pp. 40–55, 2016.

26. M. Baykara and R. Das¸, "A Survey on Potential Applications of Honeypot Technology in
Intrusion Detection Systems," International Journal of Computer Networks and Applications (IJCNA, vol. 2, no. 5.

27. I. Yahya, M. Al, P. Chauhan, S. Shukla, and M. B. Potdar, "Review on efficient log analysis to evaluate multiple honeypots using ELK," no. 6, pp. 492–504, 2016.

28. A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges," Cybersecurity, vol. 2, no. 1, 2019.

29. A. Sperotto, G. Schaffrath, R. Sadre, C. Morariu, A. Pras, and B. Stiller, "An overview of ip flow-based intrusion detection," IEEE Communications Surveys Tutorials, vol. 12, no. 3, pp. 343–356, 2010.

30. P. Sangkatsanee, N. Wattanapongsakorn, and C. Charnsripinyo, "Practical real-time intrusion detection using machine learning approaches," Computer Communications, vol. 34, no. 18, pp. 2227–2235, 2011.