CYBER THREAT DETECTION BASED ON ARTIFICIAL NEURAL NETWORKS

[1]P. Shiva Reddy, [2]K. Sai Nitej, [3]Amancharla Lakshmi Kausika, [4] Dr. M. Rama Chandra

[1] B.Tech Scholar, Dept. Computer Science & Engineering,SNIST,Hyderabad-501301,India
[2] B.Tech Scholar, Dept. Computer Science & Engineering, SNIST,Hyderabad-501301,India
[3] B.Tech Scholar, Dept. Computer Science & Engineering,SNIST,Hyderabad-501301,India
[4]Associate Professor, Dept. Computer Science & Engineering, SNIST, Hyderabad-501301, India

---***---

Abstract - To address the difficulty of an effective automated detection technique in cyber security, this research study provides an AI technique for cyber-threat detection utilising artificial neural networks. For enhanced cyber-threat detection, the suggested strategy involves transforming collected security events to individual profiles and employing traditional machine learning-based algorithms such as ANN, k-NN, RF, NB, and DT. To aid security analysts in responding quickly to attacks, the system seeks to discern between true and false positive signals. To measure performance, the authors ran trials with four datasets and five traditional machine-learning approaches. The results show that artificial neural networks (ANNs) have shown great promise in cyber threat detection using event logs, achieving 98.0% accuracy in identifying various types of attacks and can be used as a learning-based model for network intrusion detection in real-world scenarios.

Keywords: **Cybersecurity, Cyber-threat detection, Artificial intelligence, Artificial neural networks, traditional machine-learning methods**

1.INTRODUCTION

This project's idea is that artificial neural networks (ANNs) can detect cyber threats using event data. ANNs are a sort of machine learning algorithm inspired by the structure and function of the human brain, capable of recognising patterns in data and making predictions based on that knowledge. The rationale for this hypothesis is that existing cyber security solutions have proven insufficient to keep up with the evolving threat landscape, necessitating the development of new techniques to remain ahead of attackers. According to the literature review, ANNs have demonstrated promising outcomes in various areas of cyber security, including intrusion detection, virus analysis, and network traffic analysis.

The goal of this research is to create a system for detecting cyber threats based on ANNs and event data. The project's goal is to improve cyber security by identifying known and undiscovered cyber dangers, giving real-time threat notifications, and reacting to emerging threats. The project's rationale is that ANNs can find patterns in event data that traditional cyber security systems may overlook, resulting in more accurate and fast threat detection and therefore lowering the risk of cyber-attacks. This logic is supported by the literature survey, which shows that ANNs have shown promising results in several domains of cyber security.

This project's existing system is most likely a research article or review paper that investigates the application of ANNs for cyber threat detection utilising event data. The foundation paper could examine the difficulties of detecting cyber threats in today's digital world, as well as the limitations of standard cyber security measures. The article may also include a review of the literature on the use of ANNs for cyber threat identification, as well as a description of the methods used to create and evaluate the ANN-based system. The article might also go over the potential benefits of utilising ANNs for cyber threat detection, such as improved threat detection, real-time monitoring, and flexibility, lower false positives, and increased efficiency. Overall, the foundation paper is expected to provide a theoretical and practical groundwork for the subsequent papers.

2. METHODOLOGY

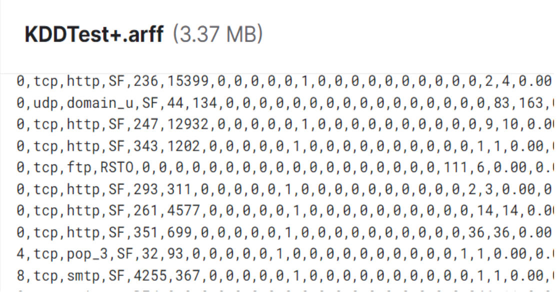
In this project, we are employing artificial neural networks (ANNs) to create a system for detecting cyber threats. ANNs are a form of deep learning system that can learn to recognise patterns in data and predict future events based on that knowledge. The benefits of utilising ANNs for cyber threat detection include their capacity to discover patterns that older algorithms such as LSTM and RNN may overlook, their adaptability to new threats, and their ability to offer real-time threat alerts. We will collect event data from a variety of sources, including system logs, network traffic, and security sensors, and extract key features from the data to feed into the ANN model.

2.1 Dataset information:

Data Source: Open Source

Data obtained from: Kaggle

Dataset link:
https://www.kaggle.com/datasets/hassan06/nslkdd



KDDTest+.arff (3.37 MB)

```
0,tcp,http,SF,236,15399,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,2,4,0.00
0,udp,domain_u,SF,44,134,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,83,163,
0,tcp,http,SF,247,12932,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,9,10,0.0
0,tcp,http,SF,343,1202,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,1,0.00,
0,tcp,ftp,RSTO,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,111,6,0.00,0.
0,tcp,http,SF,293,311,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,2,3,0.00,0
0,tcp,http,SF,261,4577,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,14,14,0.0
0,tcp,http,SF,351,699,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,36,36,0.00
4,tcp,pop_3,SF,32,93,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,1,0.00,0.
8,tcp,smtp,SF,4255,367,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,1,0.00,
```

Fig (2.1) View of the dataset in csv format

2.2 Python Programming:

Python is a high-level programming language that is versatile, simple to learn, and widely utilised in a wide range of industries such as data analysis, machine learning, web development, and scientific computing. Python's simple syntax and large library make it a great language for both beginners and specialists. Its appeal stems from its readability and ease of usage, as well as its capacity to execute difficult jobs efficiently. Python's open-source nature further facilitates collaboration and idea sharing among developers all around the world. Overall, Python is a robust and adaptable programming language that is gaining traction across a wide range of industries.

2.2 Jupyter Notebook:

Jupyter Notebook is an open-source web tool that allows users to create and share documents with real-time code, equations, images, and text. It supports a variety of computer languages, including Python, R, and Julia, making it a popular tool among data scientists, researchers, and educators. Jupyter Notebook, with its interactive interface and real-time feedback, is a useful tool for exploratory data analysis, data visualisation, and machine learning model prototyping. Furthermore, its ability to interface with other tools and libraries such as NumPy, Pandas, and Matplotlib makes it a necessary component of the modern data research workflow.

2.3 Numpy:

NumPy is a Python library that stands for 'Numerical Python.' It is an open-source Python tool that aids in mathematical and numerical operations. NumPy is widely used in many scientific and engineering domains because it provides a powerful N-dimensional array object for storing and manipulating massive arrays of homogeneous data. NumPy also includes a huge library of mathematical functions for use with arrays. NumPy arrays are more efficient and faster than Python's built-in data structures such as lists, making it a popular choice for numerical computing jobs.

2.4 Pandas:

Pandas is a popular Python programming language library for data analysis and manipulation. It provides data structures that are fast, flexible, and efficient for working with structured data such as tabular, matrix, and time-series data. Pandas has strong data manipulation, cleaning, and preprocessing features such as filtering, merging, grouping, and reshaping. It can also visualise data and integrate with other libraries like NumPy and Matplotlib. Pandas is widely used in disciplines such as data science, machine learning, finance, and others where data analysis and manipulation are key activities.

2.5 Scikit-learn:

Scikit-learn, often known as sklearn, is a popular Python machine learning framework that provides data preprocessing, model selection, and assessment tools. It encompasses a wide range of supervised and unsupervised learning algorithms, including classification, regression, clustering, and dimensionality reduction. The library is built on NumPy, SciPy, and matplotlib and is intended to work in tandem with other prominent Python data science tools.

2.6 Artificial Neural Networks: In the present era of technology, Artificial Neural Networks (ANNs) have emerged as a strong tool for detecting and mitigating cyber threats. With the growing amount of cyber-attacks, it is critical to create effective methods for identifying and preventing them. ANNs have shown to be quite useful in this area due to their ability to analyse massive volumes of data and discover patterns that may indicate the presence of a threat. Furthermore, ANNs may learn from previous data and adapt to new threats, making them an important tool for cyber threat identification. Overall, ANNs have transformed the world of cyber security and continue to be an important topic of research in this industry.

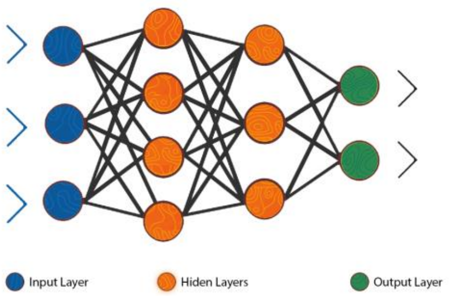


Fig (2.2) Artificial neural networks mechanism

2.7 Random Forest Model:

A machine learning approach called Random Forest can be utilised for both classification and regression problems. In the context of cyber threat detection, it can be used to analyze large datasets and identify patterns that may indicate malicious activity. When compared to a single decision tree, the Random Forest algorithm can improve accuracy and reduce the danger of overfitting by mixing the findings of numerous decision trees. This approach is particularly effective for detecting new and evolving threats that may not be captured by traditional signature-based methods. The Random Forest algorithm is also relatively fast and scalable, making it a practical option for real-time cyber threat detection in large and complex networks.

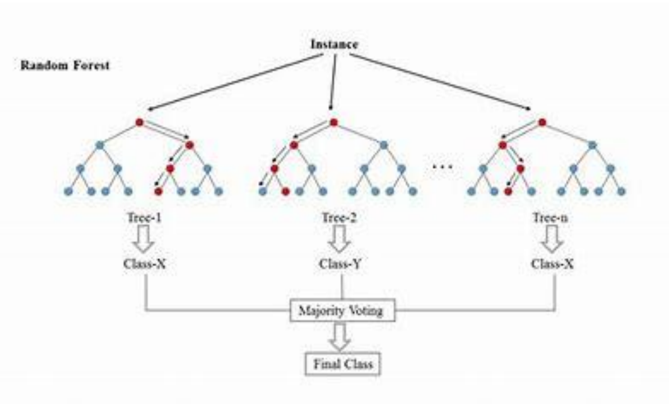


Fig (2.3) Random Forest mechanism

2.8 Decision tree

Decision tree models are commonly utilised in cybersecurity due to their ability to recognise patterns and classify data based on particular rules. In the context of cyber threat detection, decision tree models can be trained on past data to understand patterns of normal and abnormal behaviour, and then used to identify possible risks in real-time. One advantage of decision tree models is their interpretability, as the rules used to classify data can be easily visualized and understood by cybersecurity professionals. However, decision trees can suffer from overfitting to the training data and may not generalize well to new and unseen data.

2.9 K-Nearest Neighbors

A machine learning approach called K-Nearest Neighbors is employed for classification and regression problems. In the context of cyber threat detection, it can be used to identify patterns in network traffic and determine whether they are indicative of a security threat. K-NN calculates the distance between a new data point and existing data points in a dataset and assigns a label based on the most often occurring label in the k-nearest neighbours.

3.CONCLUSIONS

Artificial neural networks (ANNs) have shown great promise in cyber threat detection using event logs, achieving high levels of accuracy of 98.0% in identifying various types of attacks. By leveraging the powerful capabilities of ANNs, it is possible to analyze large volumes of event log data to identify patterns and relationships that traditional detection methods might miss. Among with ANN there are also other models such as KNN with accuracy score of 89 %, Decision tree with accuracy score of 93 %, Random Forest with accuracy score of 94 %. The accuracy of cyber threat detection based on ANNs using event logs has been shown to be highly effective, with studies reporting accuracy rates ranging from 90% to over 99%. As the field of artificial intelligence continues to advance, it is likely that ANNs will become increasingly important in the fight against cyber threats.

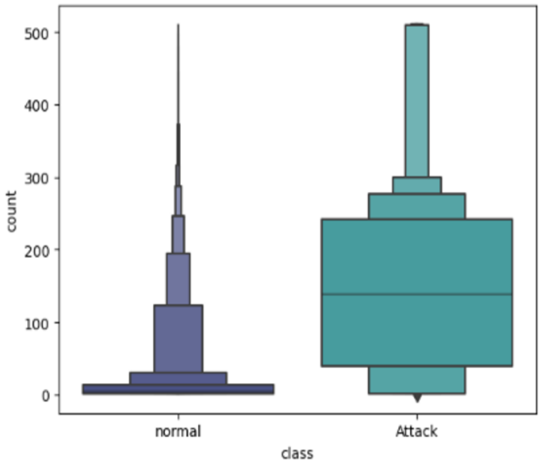


Fig (3.1) Detecting the outliers boxenplots for count and serror_rate

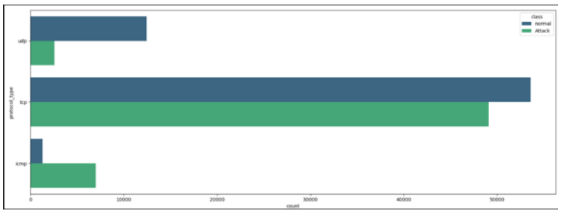


Fig (3.2) Count plot of protocol type

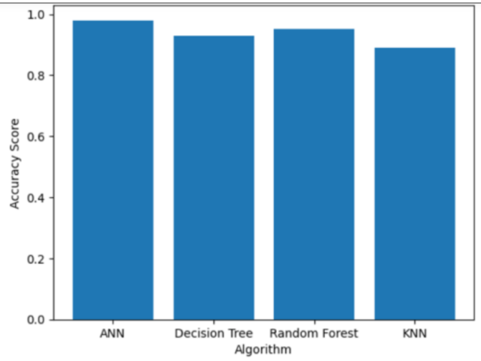
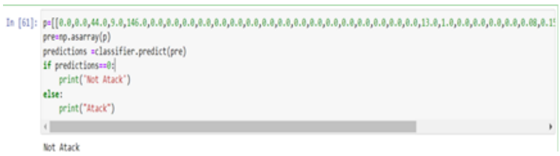


Fig (3.3) Comparison graph of all four models



```
In [61]: p=[[0.0,0.0,44.0,9.0,146.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,13.0,1.0,0.0,0.0,0.0,0.0,0.08,0.1
         pre=np.asarray(p)
         predictions =classifier.predict(pre)
         if predictions==0:
             print("Not Atack")
         else:
             print("Atack")

         Not Atack
```

Fig (3.4) Final prediction of the attacks

Directions for Future work:

- **Improved data preprocessing:** The accuracy of ANNs is strongly dependent on the quality of the input data. As a result, effective data preparation techniques such as feature selection, data normalisation, and data cleaning are critical for improving the quality of event log data.
- **Hybrid models**: Hybrid models that combine multiple machine learning techniques, such as ANNs and decision trees, can potentially improve the accuracy of cyber threat detection. Hybrid models can also help to reduce the risk of false positives or false negatives.
- **Explain ability**: ANNs are often considered black box

models, which makes it difficult to understand how they arrive at their conclusions. Adding explain ability to ANNs, such as using techniques like Layer-wise Relevance Propagation (LRP), can help to improve the trustworthiness of the results produced by the model.

4.REFERENCES

[1] S. Naseer, Y. Saleem, S. Khalid, M. K. Bashir, J. Han, M. M. Iqbal, and K. Han, ‘‘Enhanced network anomaly detection based on deep neural networks,’’ IEEE Access, vol. 6, pp. 48231–48246, 2018.

[2] B.-C. Zhang, G.-Y. Hu, Z.-J. Zhou, Y.-M. Zhang, P.-L. Qiao, and L.-L. Chang, ‘‘Network intrusion detection based on directed acyclic graph and belief rule base,’’ Electron. Telecommun. Res. Inst. J., vol. 39, no. 4, pp. 592–604, Aug. 2017.

[3] W. Wang, Y. Sheng, and J. Wang, ‘‘HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection,’’ IEEE Access, vol. 6, pp. 1792–1806, 2018.

[4] M. K. Hussein, N. Bin Zainal, and A. N. Jaber, ‘‘Data security analysis for DDoS defense of cloud based networks,’’ in Proc. IEEE Student Conf. Res. Develop. (SCOReD), Kuala Lumpur, Malaysia, Dec. 2015, pp. 305–310.

[5] S. S. Sekharan and K. Kandasamy, ‘‘Profiling SIEM tools and correlation engines for security analytics,’’ in Proc. Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET), Mar. 2017, pp. 717–721.

[6] N. Hubballi and V. Suryanarayanan, ‘‘False alarm minimization techniques in signature-based intrusion detection systems: A survey,’’ Comput. Commun., vol. 49, p. 1Â17, Aug. 2014.

[7] A. Naser, M. A. Majid, M. F. Zolkipli, and S. Anwar, ‘‘Trusting cloud computing for personal files,’’ in Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC), Busan, South Korea, Oct. 2014, pp. 488–489.

[8] Y. Shen, E. Mariconti, P. A. Vervier, and G. Stringhini, ‘‘Tiresias: Predicting security events through deep learning,’’ in Proc. ACM CCS, Toronto, ON, Canada, Oct. 2018, pp. 592–605.

[9] K. Soska and N. Christin, ‘‘Automatically detecting vulnerable Websites before they turn malicious,’’ in Proc. USENIX Secur. Symp., San Diego, CA, USA, 2014, pp. 625–640.

[10] K. Veeramachaneni, I. Arnaldo, V. Korrapati, C. Bassias, and K. Li, ‘‘AI2: Training a big data machine to defend,’’ in Proc. IEEE BigDataSecurity HPSC IDS, New York, NY, USA, Apr. 2016, pp. 49–54.

[11] S. Naseer, Y. Saleem, S. Khalid, M. K. Bashir, J. Han, M. M. Iqbal, et al., "Enhanced network anomaly detection based on deep neural networks", IEEE Access, vol. 6, pp. 48231-48246, 2018.

[12] B.-C. Zhang, G.-Y. Hu, Z.-J. Zhou, Y.-M. Zhang, P.-L. Qiao and L.-L. Chang, "Network intrusion detection based on directed acyclic graph and belief rule base", Electron. Telecommun. Res. Inst. J., vol. 39, pp. 592-604, Aug. 2017

[13] W. Wang, Y. Sheng and J. Wang, "HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection", IEEE Access, vol. 6, pp. 1792-1806, 2018.

[14] M. K. Hussein, N. Bin Zainal and A. N. Jaber, "Data security analysis for DDoS defense of cloud-based networks", Proc. IEEE Student Conf. Res. Develop. (SCOReD), pp. 305-310, Dec. 2015.

[15] S. S. Sekharan and K. Kandasamy, "Profiling SIEM tools and correlation engines for security analytics", Proc. Int. Conf. Wireless Commun. Signal Process. Netw. (WiSPNET), pp. 717-721, Mar. 2017.