ISSN: 0974-5823 Vol. 7 No. 6 June, 2022

International Journal of Mechanical Engineering

ANN based cyber threat detection by utilising event profiles

[1]Dr. Abdul Rasool MD, [2]Lameah Noorin, [3]Md Arham Saleem, [4]Mohammed Shakeel, [5]Abubakr Syed Masood

[1]Associate Professor, Department of Computer Science and Engineering, Lords Institute of Engineering and Technology, Hyderabad.

[2,3,4,5]Research Scholar, Department of Computer Science and Engineering, Lords Institute of Engineering and Technology, Hyderabad.

ABSTRACT—the requirement of an automatically generated as well as efficient method of detecting information security is indeed a massive challenge for the researchers of cyber security. Therefore in document, we introduce an artificial neural network-based AI algorithm to identify cyber attacks. The proposed approach uses a deep learning-based approach to detect to better identify cyber-threats by converting a large variety of security occurrences into individual event profiles.We built an AI-SIEM system that relies on event profiling [1] and various artificial neural network methodologies, such as FCNN, CNN, and LSTM. The system tends to focus on distinguishing among false positive and true constructive notices, thus offering to help intelligence experts quickly react to cyberattacks. All experimental studies in this research are managed to perform by researchers utilising two databases (NSLKDD and CICIDS2017) as well as 2 different sets of data gathered in the real world. Five traditional machine-learning techniques were used in experimental tests to compare the effectiveness of the algorithm to that of other previous techniques (SVM, k-NN, RF, NB, and DT). Consequentially, the exploratory outcomes of this study focus on ensuring that our proposed approaches are responsible of someone being utilised as attempting to learn models for network intrusion-detection, as well as demonstrate that because it was utilised in the actual life, the effectiveness outperforms the conventional machine-learning techniques.

Index Terms—artificial intelligence, deep neural networks and intrusion detection are just a few of the technologies being used in cyber security.

I.INTRODUCTION

Continuing to learn methodologies for identifying cyberthreats had also improved significantly since the occurrence of AI techniques, thanks in large part to the positive outcomes of recent research. It is, moreover, even now extremely impossible to prevent Information systems processes from risks as well as fraudulent activity in connections because of the constantly evolving of cyber attacks. Since different types of network incursions and suspicious attacks have taken place, the importance of proper defences and safety measures has been given high priority in the search for a long-term solution. Information security and security breaches have historically been detected using 2 main structures. The corporate network is protected by an intrusion detection and prevention (IPS), which primarily uses cryptography techniques to analyze internet protocol and flows. Intrusion alerts, known as threat intelligence, are generated and sent to a further framework, such as SIEM, for further investigation and reporting of the generated alerts. The security event mangers (SIEM) [5] have really been concentrating on collecting and examining the alerts from the IPS. In terms of security operations solutions, the SIEM has been the most prevalent as well as reputable tool for analysing gathered events and records. Aside from investigating suspicious alerts in accordance with policy and threshold, cybersecurity experts also use attack-related knowledge to analyse correlation coefficients by many events and uncover intrusion attempts. The high number of false alarms and the enormous amount of secret information [6], [7] make it extremely difficult to recognise and respond to security incidents against advanced network attacks. Computer science as well as artificial intelligence has therefore been given increased importance in the modern studies in the area of vulnerability scanning. Security analysts can more quickly and efficiently

investigate network intrusions with the help of AI advancements. It is necessary to use historical threat data to train the attack model and then use that model to detect intrusions from unknown cyber threats in these learning-based approaches[8],[9]. If a large amount of data needs to be instantly examined, analysts may benefit from a learning-based approach for defining whether an attack occurred. According to [10], there are two types of information security solutions: those driven by analysts and those driven by machine learning. Security experts, referred to as analysts, determine the rules that govern the implementation of analyst-driven solutions. The detection of new cyber threats can be aided by machine learning-driven solutions that look for unusual or anomalous patterns. Our research observed that the current learning-based methodologies have four main limitations, which limit their ability to detect attacks on systems and networks. To begin with, learning-based detection methods necessitate labelled data for training and evaluation of learning models. Obtaining such labelled data on a large enough scale to allow for accurate model training is also not straightforward. Numerous advertisement SIEM alternatives do not maintain dataset that could be implemented to controlled learning models, despite the fact that labelled data is required. Second, because they're not included in most standard network security systems, the training characteristics which are mathematically utilised in every research aren't applicable in the real world [3]. Consequently, it is difficult to implement in real-world situations. The performance of recent intrusion detection research has been evaluated using well-known datasets such as NSLKDD [11], CICIDS2017 [12], and Kyoto-Honeypot [13]. Recent efforts have recognised an automated testing strategy with deep learning technologies. However, previous studies have relied on benchmark datasets, which, while accurate, lack sufficient features to be able to be generalized to the real world. In order to overcome these drawbacks, a learning model in use must be evaluated using real-world datasets. An additional benefit of using an anomaly-based approach to network intrusion detection is its ability to identify previously unknown cyber threats while also increasing the likelihood of false alarms [6]. False positive alerts are enormously expensive and consume a lot of effort from staff to investigate. It is possible for some hackers to cover up their malicious activities with a gradual shift in how they behave [10, 14]. Hackers continuously change their behaviour, making detection models unsuitable even when learning-based models are available. Most security systems have also been designed to focus on detecting and analysing incidents involving network security in the near term. Over time, we believe that analysing the vulnerability management background generated in the process of incidents could be used to identify malicious behaviour in cyberattacks. This research is motivated by these issues. We've developed an AI-SIEM system that uses deep learning techniques to tell the difference between legitimate and erroneous alerts. Cyber threats that are spread across a large number of security events can be quickly responded to by our proposed system. A suggested AI-SIEM system provides a method for extracting patterns in acquired data by aggregating and correlating events with such a concurrent feature. It's possible that our event profiles could be used as input data for a variety of deep neural networks. When compared to historical data, it allows the analyst to quickly and efficiently process all of the data. What we've accomplished can be summarised as follows: In order to process massive amounts of data, we suggest a method that can break down enormous numbers of security events into smaller, more manageable pieces. Analysis of security events can be generalised by learning patterns from huge amounts of collected data, taking into account the frequency with which they occur. Data pre-processing base points are used in this work to characterise the data sets in particular. In log analysis, typical data mining approaches often struggle to minimise the dimensionality of the data, which is always the main problem with this technique. We have developed an event profiling method for using artificial intelligence algorithms that is distinct from traditional sequence-based pattern approaches. In other words, we may significantly reduce the amount of alerts that analysts receive by using our method's ability to assist better classification for true warnings when compared to current machine-learning methods. Real IPS security events from a SOC are used to test our system's applicability by evaluating its precision, prediction performance (TPR), false positive rate (FPR), and the F-measure, among other performance metrics. Furthermore, we tested our new method against the five most commonly used machine-learning algorithms to see how well it performed (SVM, KNN, RF, NB and DT). To test our method, we applied it to two commonly used datasets in the field of network - based intrusion detection research (NSLKDD and CICIDS2017). The TF-IDF mechanism is used in this research to break down a large number of gathering events in and out of autonomous agent event that happens profiles. For each TF-IDF event set, we compute the similarity value between that set and the designated base points. There are three layers of models in AI-SIEM: FCNN, CNN, and LSTM. The produced event profiles feed into these layers. As a result, we plan to demonstrate the effectiveness of our system by comparing it to 2 different well-known standard datasets and 2 different data sources gathered from trying to operate an IPS. However, NSLKDD and CICIDS2017 data

sources are widely used as performance metrics for comparing machine learning methodologies, despite their inherent limitations. The genuine records and two new standard performance datasets are used to conduct a comparison of the performance with traditional technologies. Machine learning approaches that perform well on benchmark datasets must also perform well on real-world datasets in order to be useful.

II.RELATED WORK

Cloud-based network DDoS defence using data security analysis

The use of distributed computing has shown to be a successful strategy for increasing an institution's capabilities while reducing the need for extra resources. In this aspect, distributed computing aids in enhancing the IT capabilities of educational institutions. Distributed computing is currently an essential aspect of the IT industries most rapidly increasing sectors. It's a novel and efficient way to expand your business, according to some. Concerns about the security of sensitive data have grown as more organisations and individuals begin to manage their accounts and applications in the cloud [4]. Many clients are reluctant to move their sensitive data to the cloud because of security concerns, despite the widespread adoption of cloud computing. Distributed computing advancement will be hampered if security concerns aren't addressed, as data from most organisations is a tantalising vulnerable to hackers. New tests and insights about honeypots are provided by this study. Two types of honey pots exist, one for handling and one for conducting research. Real-world threats can be mitigated by handling honeypots in a safe manner. Researchers use a research honeypot as an exploratory tool to examine and differentiate between the threats that lurk on the internet. That's why we want to conduct a thorough network security investigation by setting up a virtualization phishing attack on cloud servers that will lure an attacker and allow us to see their activity in an entirely different light.

Analyzing security data using SIEM technologies and correlation engines

The amount of data generated by today's IT organisations is enormous. In the world of IT, simply being able to deal with such large amounts of data is essential. So by centralising the log management program, a business can better protect its data. High-profile profiling tools are needed in order to enhance the standard of security in such organisations. Organizational security can be better understood through the use of managed security services (SIEM) [5]. Logs are centralised in SIEM tools by analysing and normalising all files and data coming from various devices. With an emphasis on the most famous SIEM techniques as well as open platform rule-based correlation engines, this paper presents an abstract of SIEM tools and event causal connection engines. It also contains a review of their technical comparison.

III.METHODOLOGY

There is an AI-SIEM (Artificial Intelligence-Security Data and Systems Involving) method described in this paper that utilises a combination of deep learning techniques like FCNN, CNN (Convolution Neural Networks), and LSTM (long short term memory) to identify attacks. SVM, Decision Tree, Random Forest, KNN, and Naive Bayes are used to evaluate the proposed work's effectiveness. Here, I'm working to implement CNN and LSTM methodologies.

The following modules make up a proposed algorithm.

Data Parsing:This module creates an original data ability to establish by parsing an input dataset.

TF-IDF: In order to create an event vector that contains both normal and attack signatures, we will use this module.

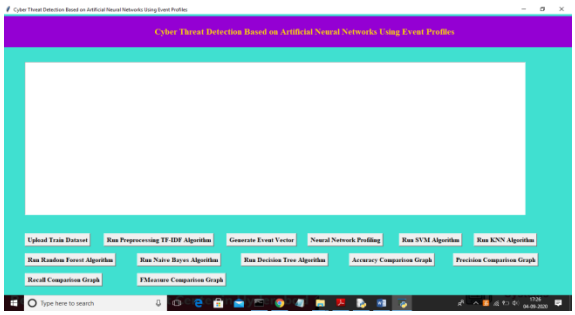
Event Profiling Stage: Based on profile events, the information generated would be split into training and testing model.

Deep Learning Neural Network Model: On the basis of information from the training and test datasets, this module generates a framework for use in further analysis [14]. The resulted training set would be used to measure the important outcome, recall, precision, and FMeasure on test data. A methodology that learns properly would then produce a better accuracy result, and that concept would be chosen to be deployed on an actual system for attack detection.

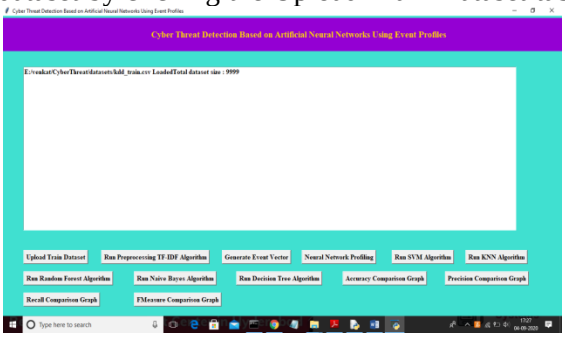
In ability to execute all techniques, this will start taking 5 to 10 minutes to run all the data sources, but the KDD train analysis of the data runs perfectly. Other datasets could be tested by reducing their size or running them on a system with a higher configuration specification.

IV.RESULT AND DISCUSSION

Run project to get below screen

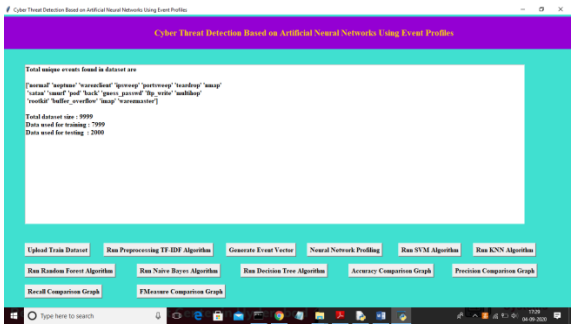


It's easy to upload your train dataset by clicking the Upload Train Dataset tab on the upper right of the result, then selecting KDD train.csv.

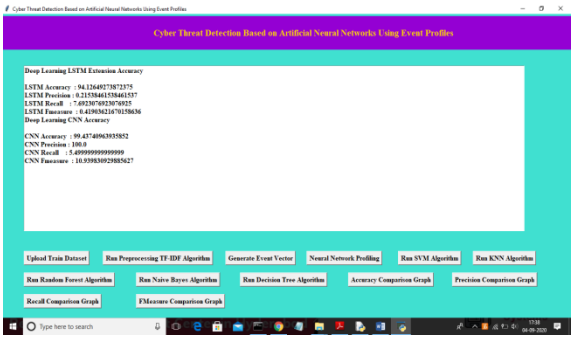


Now that we've seen that the dataset comprises 9999 entries, we can click on the 'Run Pre-processing TF-IDF Algorithm' tab to transform the raw dataset into TF-IDF values.

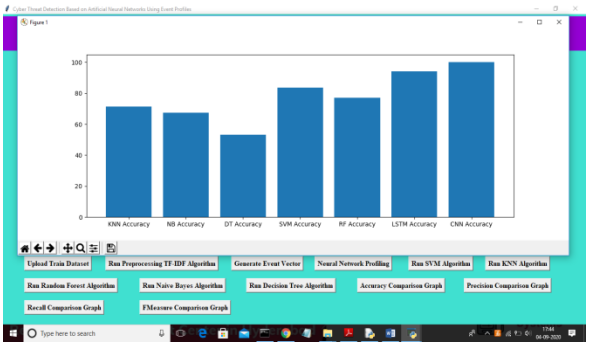
If you want to produce an event vector from TF-IDF, you can click on the "Generate Event Vector" tab.



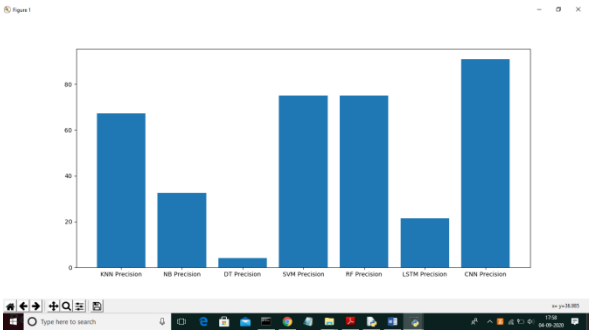
In the above result, we can see the total number of unique event names, as well as the overall size of the dataset and the percentage of the dataset used for training and testing. To generate an LSTM and CNN model, click the "Neural Network Profiling" button after the data has been prepared and the training and testing events models have been completed.



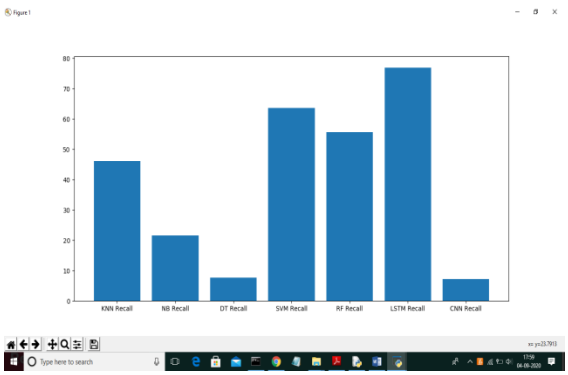
We can see the accuracy, precision, recall, and FMeasure values in the results shown above. Afterwards, Run all Algorithms and to see the accuracy of all algorithms, click the 'Accuracy Comparison Graph' button.



We can see from the graph above that LSTM and CNN perform well on the y-axis, which depicts algorithm name and the x-axis, which depicts accuracy. To get the graph below, click on 'Precision Comparison Graph'.



The graph above shows how well CNN performs; now select 'Recall Comparison Graph' to see how well it compares to the competition.



The LSTM is showing good results in the graph above; to see the comparison graph shown below, select the FMeasure Comparison Graph button.

We can observe from the comparison graphs that LSTM and CNN perform well in terms of accuracy, recall, and precision.

V.CONCLUSION

The AI-SIEM system, based on event profiles and artificial neural networks, is presented in this study. Our study is unique in that it uses deep learning-based detection approaches to improve computer hackers identification by condense very large amounts of information in to the event profiles. By comparing long-term security data, the AI-SIEM [5] system enables security analysts to respond quickly and efficiently to critical security alarms. It could also assist cybersecurity experts respond more quickly to cybercrime that are scattered throughout a large wide range of security events by lowering false positive alerts.Two benchmark datasets (NSLKDD and CICIDS2017) and two datasets from the actual world were used for the performance appraisal process. First, we demonstrated that our techniques may be used as being one of the continuing to learn approaches for intrusion detection [14] systems by comparing them to other approaches utilising benchmark datasets. Secondly, we demonstrated that our technique beat standard machine learning techniques in terms of correctly classified instances by evaluating two real datasets.

In the future, we'll use numerous deep learning approaches to uncover long-term connections in historical information to increase previous attack forecasts in order to handle the growing challenge of cyber threats. In addition, many SOC analysts may attempt to record the labels of raw threat intelligence individually across several months in order to improve the precision of the particular dataset for carefully monitored and generate good learning datasets.

VI.REFERENCES

[1] S. Naseer, Y. Saleem, S. Khalid, M. K. Bashir, J. Han, M. M. Iqbal, K. Han, "Enhanced Network Anomaly Detection Based on Deep Neural Networks," IEEE Access, vol. 6, pp. 48231-48246, 2018.

[2] B. Zhang, G. Hu, Z. Zhou, Y. Zhang, P. Qiao, L. Chang, "Network Intrusion Detection Based on Directed Acyclic Graph and Belief Rule Base", ETRI Journal, 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, no. 99, 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," 2015 IEEE Student Conference on Research and Development (SCOReD), Kuala Lumpur, 2015, pp. 305-310.

[5] S. Sandeep Sekharan, K. Kandasamy, "Profiling SIEM tools and correlation engines for security analytics," In Proc. Int. Conf. Wireless Com., Signal Proce. and Net.(WiSPNET), 2017, pp. 717- 721.

[6]N.Hubballiand V.Suryanarayanan,‘‘False alarm minimization techniques in signature-based intrusion detection systems: A survey,’’ Comput. Commun., vol. 49, pp. 1-17, Aug. 2014.

[7] A. Naser, M. A. Majid, M. F. Zolkipli and S. Anwar, "Trusting cloud computing for personal files," 2014 International Conference on Information and Communication Technology Convergence (ICTC), Busan, 2014, pp. 488-489.

[8] Y. Shen, E. Mariconti, P. Vervier, and Gianluca Stringhini, "Tiresias: Predicting Security Events Through Deep Learning," In Proc. ACM CCS 18, Toronto, Canada, 2018, pp. 592-605.

[9] Kyle Soska and Nicolas Christin, "Automatically detecting vulnerable websites before they turn malicious,", In Proc. USENIX Security Symposium., San Diego, CA, USA, 2014, pp.625-640. [10] K. Veeramachaneni, I. Arnaldo, V. Korrapati, C. Bassias, K. Li, "AI2: training a big data machine to defend," In Proc. IEEE BigDataSecurity HPSC IDS, New York, NY, USA, 2016, pp. 49-54

[11] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu and Ali A. Ghorbani, "A detailed analysis of the kdd cup 99 data set," In Proc. of the Second IEEE Int. Conf. Comp. Int. for Sec. and Def. App., pp. 53-58, 2009.

[12] I. Sharafaldin, A. H. Lashkari, A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization", Proc. Int. Conf. Inf. Syst. Secur. Privacy, pp. 108- 116, 2018.

[13] [online] Available: http://www.takakura.com/Kyoto_data/

[14] 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, pp. 41-50, Feb. 2018

[15] R. Vinayakumar, Mamoun Alazab, K. P. Soman, P. Poornachandran, Ameer Al-Nemrat and Sitalakshmi Venkatraman, "Deep Learning Approach for Intelligent Intrusion Detection System," IEEE Access, vol. 7, pp. 41525-41550, Apr. 2019.