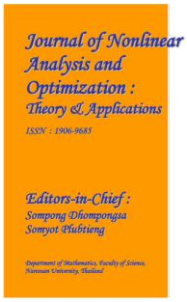
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.15 : 2024
ISSN : **1906-9685**



Paper ID: ICRTEM24_159 **ICRTEM-2024 Conference Paper**

CYBER THREAT DETECTION USING MACHINE LEARNING ALGORITHMS

[#1]SK. YAKOOB, ***Associate Professor & HOD,***
Department of CSE,
[#2]V. NARESH, Assistant Professor, CSE (AI&ML),
SAI SPURTHI INSTITUTE OF TECHNOLOGY, SATHUPALLI, KHAMMAM

ABSTRACT: One of the major challenges in cybersecurity is the provision of an automated and effective cyber-threats detection technique. In this paper, we present an AI technique for cyber-threats detection, based on artificial neural networks. The proposed technique converts multitude of collected security events to individual event profiles and use a deep learning-based detection method for enhanced cyber-threat detection. For this work, we developed an AI-SIEM system based on a combination of event profiling for data preprocessing and different artificial neural network methods, including FCNN, CNN, and LSTM. The system focuses on discriminating between true positive and false positive alerts, thus helping security analysts to rapidly respond to cyber threats. All experiments in this study are performed by authors using two benchmark datasets (NSLKDD and CICIDS2017) and two datasets collected in the real world. To evaluate the performance comparison with existing methods, we conducted experiments using the five conventional machine-learning methods (SVM, k-NN, RF, NB, and DT). Consequently, the experimental results of this study ensure that our proposed methods are capable of being employed as learning-based models for network intrusion-detection, and show that although it is employed in the real world, the performance outperforms the conventional machine-learning methods.

Keywords- Machine Learning, Artificial Neural Networks, cyber threats.

I. INTRODUCTION

With the emergence of artificial intelligence (AI) techniques, learning-based approaches for detecting cyber attacks have become further improved, and they have achieved significant results in many studies. However, owing to constantly evolving cyber attacks, it is still

highly challenging to protect IT systems against threats and malicious behaviors in networks. Because of various network intrusions and malicious activities, effective defenses and security considerations were given high priority for finding reliable solutions [1], [2], [3], [4].

Traditionally, there are two primary systems for detecting cyber-threats and network intrusions. An intrusion prevention system (IPS) is installed in the enterprise network, and can examine the network protocols and flows with signature-based methods primarily. It generates appropriate intrusion alerts, called the security events, andreports the generating alerts to another system, such as SIEM. The security information and event management (SIEM) has been focusing on collecting and managing the alerts of IPSs. The SIEM is the most common and dependable solution among various security operations solutions to analyze the collected security events and logs [5]. Moreover, security analysts make an effort to investigate suspicious alerts by policies and threshold, and to discover malicious behavior by analyzing correlations among events, using knowledge related to attacks.

Nevertheless, it is still difficult to recognize and detect intrusions against intelligent network attacks owing to their high false alerts and the huge amount of security data [6], [7]. Hence, the most recent studies in the field of intrusion detection have given increased focus to machine learning and artificial intelligence techniques for detecting attacks. Advancement in AI fields can facilitate the investigation ofnetwork intrusions by security analysts in a timely and automated manner. These learning-based approaches require to learn the attack model from historical threat data and use the trained models to detect intrusions for unknown cyber threats [8], [9].

A learning-based method geared toward determining whether an attack occurred in a large amount of data can be useful to analysts who need to instantly analyze numerous events. According to [10], information security solutions generally fall into two categories: analyst-driven and machine learning-driven solutions. Analyst-driven solutions rely on rules determined by security experts called analysts. Meanwhile, machine learning-driven solutions used to detect rare or anomalous patterns can improve detection of new cyber threats [10]. Nevertheless, while learning-based approaches are useful in detecting cyber attacks in systems and networks, we observed that existing learning-based approaches have four main limitations.

First, learning-based detection methods require labeled data, which enable the training of the model and evaluation of generated learning models. Furthermore, it is not straightforward to obtain such labeled data at a scale that allow accurate training of a model. Despite the need for labeled data, many commercial SIEM solutions do not maintain labeled data that can be applied to supervised learning models [10].

Second, most of the learning features that are theoretically used in each study are not generalized features in the real world, because they are not contained in common network security systems [3]. Hence, it makes difficult to utilize to practical cases. Recent efforts on intrusion detection research have considered an automation approach with deep learning technologies, and performance has been evaluated using wellknown datasets like NSLKDD [11], CICIDS2017 [12], and Kyoto-Honeypot [13]. However, many previous studies

used benchmark dataset, which, though accurate, are not generalizable to the real world because of the insufficient features. To overcome these limitations, an employed learning model requires to evaluate with datasets that are collected in the real world.

Third, using an anomaly-based method to detect network intrusion can help detect unknown cyber threats; whereas it can also cause a high false alert rate [6]. Triggering many false positive alerts is extremely costly and requires a substantially large amount of effort from personnel to investigate them. Fourth, some hackers can deliberately cover their malicious activities by slowly changing their behavior patterns [10], [14]. Even when appropriate learning-based models are possible, attackers constantly change their behaviors, making the detection models unsuitable. Moreover, almost all security systems have been focused on analyzing short-term network security events. To defend consistently evolving attacks, we assume that over long-term periods, analyzing the security event history associated with the generation of events can be one way of detecting the malicious behavior of cyber attacks.

These challenges form the primary motivation for this work. To address these challenges, we present an AI-SIEM system which is able to discriminate between true alerts and false alerts based on deep learning techniques.

Our proposed system can help security analysts rapidly to respond cyber threats, dispersed across a large amount of security events. For this, the proposed the AI-SIEM system particularly includes an event pattern extraction method by aggregating together events with a concurrency feature and correlating between event sets in collected data. Our event profiles have the potential to provide concise input data for various deep neural networks. Moreover, it enables the analyst to handle all the data promptly and efficiently by comparison with long term history data.

LITERATURE SURVEY

2.1 Enhanced Network Anomaly Detection Based on Deep Neural Networks

Due to the monumental growth of Internet applications in the last decade, the need for security of information network has increased manifolds. As a primary defense of network infrastructure, an intrusion detection system is expected to adapt to dynamically changing threat landscape. Many supervised and unsupervised techniques have been devised by researchers from the discipline of machine learning and data mining to achieve reliable detection of anomalies. Deep learning is an area of machine learning which applies neuron-like structure for learning tasks. Deep learning has profoundly changed the way we approach learning tasks by delivering monumental progress in different disciplines like speech processing, computer vision, and natural language processing to name a few. It is only relevant that this new technology must be investigated for information security applications. The aim of this paper is to investigate the suitability of deep learning approaches for anomaly-based intrusion detection system. For this research, we developed anomaly detection models based on different deep neural network structures, including convolutional neural networks, autoencoders, and recurrent neural networks. These deep models were trained on NSLKDD training data set and evaluated on both test data sets provided by NSLKDD, namely NSLKDD Test+ and

NSLKDD Test21. All experiments in this paper are performed by authors on a GPU-based test bed. Conventional machine learning-based intrusion detection models were implemented using well-known classification techniques, including extreme learning machine, nearest neighbor, decision-tree, random-forest, support vector machine, naive-bays, and quadratic discriminant analysis. Both deep and conventional machine learning models were evaluated using well-known classification metrics, including receiver operating characteristics, area under curve, precision-recall curve, mean average precision and accuracy of classification. Experimental results of deep IDS models showed promising results for real-world application in anomaly detection systems.

2.2 Network Intrusion Detection Based on Directed Acyclic Graph and Belief Rule Base

Intrusion detection is very important for network situation awareness. While a few methods have been proposed to detect network intrusion, they cannot directly and effectively utilize semi-quantitative information consisting of expert knowledge and quantitative data. Hence, this paper proposes a new detection model based on a directed acyclic graph (DAG) and a belief rule base (BRB). In the proposed model, called DAG-BRB, the DAG is employed to construct a multi-layered BRB model that can avoid explosion of combinations of rule number because of a large number of types of intrusion. To obtain the optimal parameters of the DAG-BRB model, an improved constraint covariance matrix adaption evolution strategy (CMA-ES) is developed that can effectively solve the constraint problem in the BRB. A case study was used to test the efficiency of the proposed DAG-BRB. The results showed that compared with other detection models, the DAG-BRB model has a higher detection rate and can be used in real networks.

2.3 HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection

The development of an anomaly-based intrusion detection system (IDS) is a primary research direction in the field of intrusion detection. An IDS learns normal and anomalous behavior by analyzing network traffic and can detect unknown and new attacks. However, the performance of an IDS is highly dependent on feature design, and designing a feature set that can accurately characterize network traffic is still an ongoing research issue. Anomaly-based IDSs also have the problem of a high false alarm rate (FAR), which seriously restricts their practical applications. In this paper, we propose a novel IDS called the hierarchical spatial-temporal features-based intrusion detection system (HAST-IDS), which first learns the low-level spatial features of network traffic using deep convolutional neural networks (CNNs) and then learns high-level temporal features using long short-term memory networks. The entire process of feature learning is completed by the deep neural networks automatically; no feature engineering techniques are required. The automatically learned traffic features effectively reduce the FAR. The standard DARPA1998 and ISCX2012 data sets are used to evaluate the performance of the proposed system. The experimental results show that the HAST-IDS outperforms other published approaches in terms of accuracy, detection rate, and FAR, which successfully demonstrates its effectiveness in both feature learning and FAR reduction.

2.4 Data security analysis for DDoS defense of cloud based networks

Distributed computing has become an effective approach to enhance capabilities of an institution or organization and minimize requirements for additional resource. In this regard, the distributed computing helps in broadening institutes IT capabilities. One needs to note that distributed computing is now integral part of most expanding IT business sector. It is considered novel and efficient means for expanding business. As more organizations and individuals start to use the cloud to store their data and applications, significant concerns have developed to protect sensitive data from external and internal attacks over internet. Due to security concern many clients hesitate in relocating their sensitive data on the clouds, despite significant interest in cloud-based computing. Security is a significant issue, since data much of an organizations data provides a tempting target for hackers and those concerns will continue to diminish the development of distributed computing if not addressed. Therefore, this study presents a new test and insight into a honeypot. It is a device that can be classified into two types: handling and research honeypots. Handling honeypots are used to mitigate real life dangers. A research honeypot is utilized as an exploration instrument to study and distinguish the dangers on the internet. Therefore, the primary aim of this research project is to do an intensive network security analysis through a virtualized honeypot for cloud servers to tempt an attacker and provide a new means of monitoring their behavior

III.EXISTING SYSTEM

Cyber security has recently received enormous attention in today's security concerns, due to the popularity of the Internet-of-Things (IoT), the tremendous growth of computer networks, and the huge number of relevant applications. Thus, detecting various cyber-attacks or anomalies in a network and building an effective intrusion detection system that performs an essential role in today's security is becoming more important.However, many previous studies used benchmark dataset, which, though accurate, are not generalizable to the real world because of the insufficient features. To overcome these limitations, an employed learning model requires to evaluate with datasets that are collected in the real world. Third, using an anomaly-based method to detect network intrusion can help detect unknown cyber threats; whereas it can also cause a high false alert rate.

IV PROPOSED SYSTEM:

we present an AI technique for cyber-threats detection, based on artificial neural networks. The proposed technique converts multitude of collected security events to individual event profiles and use a deep learning-based detection method for enhanced cyber-threat detection. For this work, we developed an AI-SIEM system based on a combination of event profiling for data preprocessing and different artificial neural network methods, including FCNN, CNN, and LSTM. The system focuses on discriminating between true positive and false positive alerts, thus helping security analysts to rapidly respond to cyber threat. we conducted experiments using the five conventional machine-learning methods (SVM, k-NN, RF, NB, and DT). Consequently, the experimental results of this study ensure that our proposed methods are capable of being employed as learning-based models for network intrusion-detection

IV.SYSTEM ARCHITECTURE

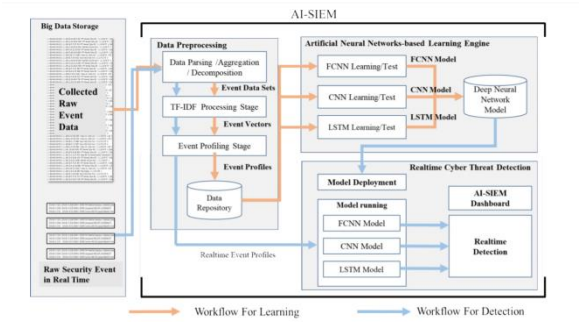


Fig1: Architecture of cyber threat detection using machine learning algorithms.

MODULES:

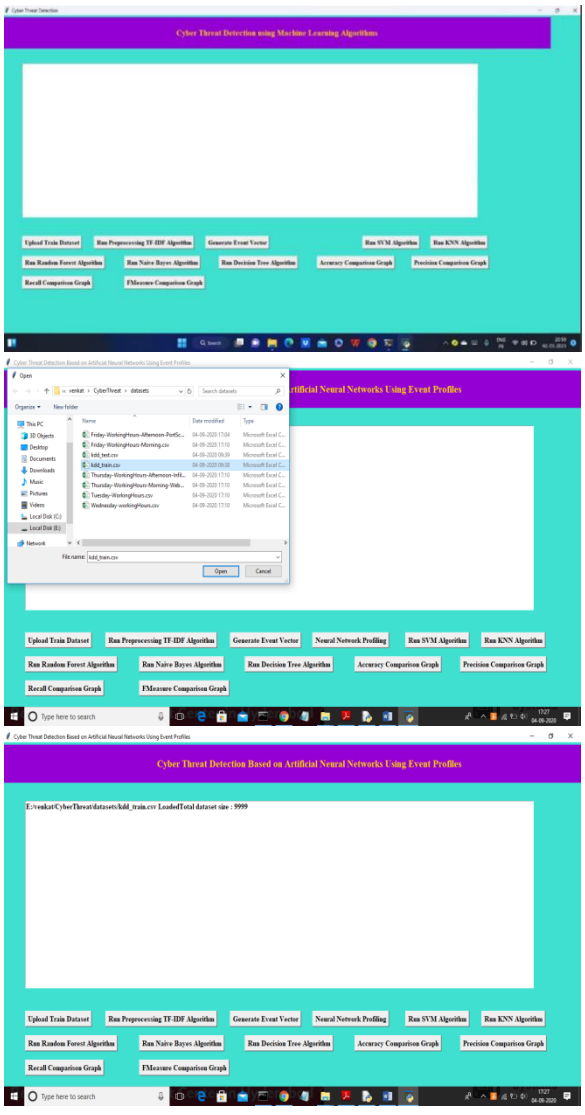
- ❖ upload Train Dataset
- ❖ Run Preprocessing TF-IDF Algorithm
- ❖ Generate Event Vector
- ❖ Neural Network Profiling
- ❖ Run SVM Algorithm
- ❖ Run KNN Algorithm
- ❖ Run Naive Bayes Algorithm
- ❖ Run Decision Tree Algorithm
- ❖ Accuracy Comparison Graph
- ❖ Precision Comparison Graph
- ❖ Recall Comparison Graph
- ❖ FMeasure Comparison Graph

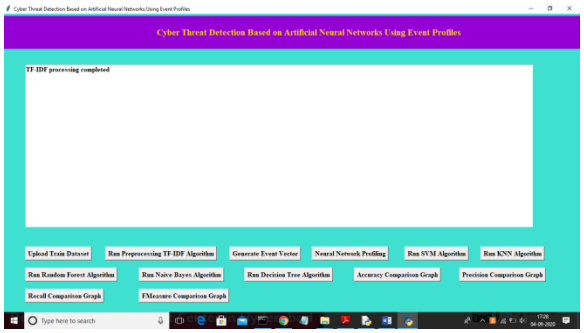
Propose algorithms consists of following module

1) Data Parsing: This module take input dataset and parse that dataset to create a raw data event model
2) TF-IDF: using this module we will convert raw data into event vector which will contains normal and attack signatures
3) Event Profiling Stage: Processed data will be splitted into train and test model based on profiling events.

4) Deep Learning Neural Network Model: This module runs CNN and LSTM algorithms on train and test data and then generate a training model. Generated trained model will be applied on test data to calculate prediction score, Recall, Precision and F-Measure. Algorithm will learn perfectly will yield better accuracy result and that model will be selected to deploy on real system for attack detection. Datasets which we are using for testing are of huge size and while building model it's going to out of memory error but kdd_train.csv dataset working perfectly but to run all algorithms it will take 5 to 10 minutes. You can test remaining datasets also by reducing its size or running it on high configuration system.

RESULT:





VII. CONCLUSION

In this paper, we have proposed the AI-SIEM system using event profiles and artificial neural networks. The novelty of our work lies in condensing very large-scale data into event profiles and using the deep learning-based detection methods for enhanced cyber-threat detection ability. The AI-SIEM system enables the security analysts to deal with significant security alerts promptly and efficiently by comparing long term security data. By reducing false positive alerts, it can also help the security analysts to rapidly respond to cyber threats dispersed across a large number of security events.

For the evaluation of performance, we performed a performance comparison using two benchmark datasets (NSLKDD, CICIDS2017) and two datasets collected in the real world. First, based on the comparison experiment with other methods, using widely known benchmark datasets, we showed that our mechanisms can be applied as one of the learning-based models for network intrusion detection. Second, through the evaluation using two real datasets, we presented promising results that our technology also outperformed conventional machine learning methods in terms of accurate classifications.

VIII. REFERENCES

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

[2] B. Zhang, G. Hu, Z. Zhou, Y. Zhang, P. Qiao, L. Chang, "NetworkIntrusion Detection Based on Directed Acyclic Graph and Belief RuleBase", *ETRI Journal*, vol. 39, no. 4, pp. 592-604, Aug. 2017

[3] W. Wang, Y. Sheng and J. Wang, "HAST-IDS: Learning hierarchicalspatial-temporal features using deep neural networks to improveintrusion detection," *IEEE Access*, vol. 6, no. 99, pp. 1792-1806,2018.

[4] M. K. Hussein, N. Bin Zainal and A. N. Jaber, "Data security analysisfor DDoS defense of cloud based networks," *2015 IEEE StudentConference on Research and Development (SCOReD)*, KualaLumpur, 2015, pp. 305-310.

[5] S. Sandeep Sekharan, K. Kandasamy, "Profiling SIEM tools andcorrelation 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 minimizationtechniques in signature-based intrusion detection systems: Asurvey,’’ *Comput. Commun.*, vol. 49, pp. 1-17, Aug. 2014.

[7] A. Naser, M. A. Majid, M. F. Zolkipli and S. Anwar, "Trusting cloudcomputing for personal files," *2014 International Conference onInformation 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. ACMCCS 18, Toronto, Canada, 2018, pp. 592-605.

[9] Kyle Soska and Nicolas Christin, "Automatically detectingvulnerable websites before they turn malicious,", *In Proc. USENIXSecurity 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. IEEEBigDataSecurity HPSC IDS*, New York, NY, USA, 2016, pp. 49-54

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