CYBER THREAT DETECTION IN INTERNET OF THINGS USING DEEP LEARNING APPROACH

Pokala Mounika
Department of Computer Science And System Engineering, Andhra University College of Engineering (A), Visakhapatnam, Andhra Pradesh

Abstract:- Developing an automated method for detecting cyberthreats is one of the biggest difficulties in cybersecurity. In this study, we describe an artificial intelligence (AI) method based on artificial neural networks for detecting cyberthreats. The suggested technique uses a deep learning-based detection method to improve cyber-threat detection by breaking down a large volume of collected security events into individual event profiles. For this project, we created an AI-SIEM system that combines event profiling for data pretreatment with various artificial neural network techniques, such as FCNN, CNN, and LSTM. The approach focuses on separating true positive warnings from false positive alerts to assist security analysts in quickly responding to cyber attacks. The authors run every experiment in this paper on two benchmark datasets (NSLKDD and CICIDS2017) as well as two real-world datasets. We carried out experiments utilizing the five traditional machine-learning techniques to assess the performance comparison with existing approaches (SVM, k-NN, RF, NB, and DT). The experimental findings of this study confirm that our suggested methods may be used as learning-based models for network intrusion detection and demonstrate that, even when applied in the real world, they beat traditional machine-learning techniques.

Index Terms— SVM, k-NN, RF, NB, DT, AI-SIEM, FCNN, CNN, LSTM

I Introduction

Learning-based systems for identifying cyber assaults have developed further with the development of artificial intelligence (AI) capabilities, and they have shown considerable outcomes in numerous studies. However, protecting IT systems from threats and criminal network behaviour is still very difficult because cyberattacks are always changing. Effective defences and security concerns were given top importance for finding dependable solutions because of various network invasions and criminal activities. For identifying network breaches and cyber threats, there are typically two main systems. The company network has an intrusion prevention system (IPS) installed, which uses signature-based techniques to primarily inspect network protocols and flows. It produces the necessary intrusion alerts, also known as security events, and reports the alert generation to another system, like SIEM. The gathering and administration of IPS alerts has been the primary focus of security information and event management (SIEM). Among the different security operations solutions, the SIEM is the most popular and reliable option for analysing the gathered security data. Additionally, security analysts work to evaluate suspicious alerts based on policies and thresholds and to find malicious behaviour by looking at correlations between events and applying attack-related knowledge. Due to their high false alarm rates and the vast volume of security data, intrusions against intelligent network attacks are still challenging to distinguish and detect. Therefore, machine learning and artificial intelligence algorithms for identifying attacks have received more attention in the most recent studies in the field of intrusion detection. The development of AI-SIEM related domains can help security experts investigate network attacks quickly and automatically. These approaches that rely on learning necessitate learning the attack model.

The trained models are used to detect incursions for unidentified cyber threats and historical threat data. For analysts who need to assess a huge number of

events, a learning-based approach targeted toward evaluating whether an attack occurred in a big amount of data can be helpful. Information security solutions, in general, can be divided into two categories: those driven by analysts and those driven by machine learning. Analyst-driven solutions rely on guidelines created by analysts, or security specialists. While this is going on, the detection of emerging cyber threats can be enhanced by machine learning-driven solutions used to spot unusual or anomalous patterns. We found that existing learning-based techniques have four fundamental drawbacks, despite the fact that they are useful for identifying cyber assaults in systems and networks.

2 Literature survey

S. Naseer etc al. [1] The necessity for information network security has multiplied due to the enormous expansion of Internet applications during the past ten years. An intrusion detection system is supposed to respond to the dynamically changing threat landscape as the main line of defence for network infrastructure. Researchers in the fields of machine learning and data mining have developed a variety of supervised and unsupervised algorithms to reliably detect anomalies. Deep learning is a kind of machine learning that uses structures resembling neurons to perform learning tasks. The way we approach learning tasks has been dramatically changed by deep learning. bringing about enormous advancement in a number of fields, including speech processing, computer vision, and natural language processing, to mention a few. It only matters that this new technology be looked into for information security uses. The purpose of this research is to examine whether deep learning techniques are appropriate for an intrusion detection system that uses anomalies. Convolutional neural networks, autoencoders, and recurrent neural networks are just a few of the deep neural network topologies that we used to create anomaly detection models for this study. These deep models were tested using the provided test data sets after being evaluated on the training data set.

The authors do each experiment in this study on a GPU-based test system. Extensive learning machine, closest neighbour, decision tree, random forest, support vector machine, naive bays, and quadratic discriminant analysis were used to create conventional machine learning-based intrusion detection models. Utilizing well-known classification criteria such receiver operating parameters, area under the curve, precision-recall curve, mean average precision, and accuracy of classification, both deep and standard machine learning models were assessed. Deep IDS models' experimental findings offered encouraging signs for their practical use in anomaly detection systems.

Zhang et al [2] For network situational awareness, intrusion detection is crucial. Although a few approaches for detecting network intrusion have been presented, they cannot directly and effectively use semi-quantitative information made up of quantitative data and expert knowledge. The proposed model, known as the, is used to build a multi-layered model that may avoid explosion of combinations of rule number due to a high number of forms of incursion. It is built on a directed acyclic graph and a belief rule base. An enhanced constraint covariance matrix adaptation evolution approach is created in order to efficiently solve the constraint problem in the model and acquire the ideal parameters. A case study was employed to evaluate the effectiveness of the suggested

W. Wang et al [3] One of the main research directions in the field of intrusion detection is the creation of an anomaly-based intrusion detection system (IDS). An IDS can identify novel and undiscovered attacks by studying network data to determine what is normal and abnormal behaviour. However, the design of an IDS's features has a significant impact on how well it performs, and developing a feature set that can precisely describe network traffic is currently an open research question. Moreover, anomaly-based IDSs have its practical uses are severely constrained by the issue of a high false alarm rate (FAR). Using deep convolutional neural networks (CNNs) to first learn low-level spatial features of network traffic and long short-term memory networks to learn high-level

temporal features, we propose a novel IDS in this paper called the hierarchical spatial-temporal features-based intrusion detection system (HAST-IDS). The entire process of feature learning is accomplished by the deep neural networks automatically; no feature engineering approaches are required. The FAR is significantly decreased by the automatically learned traffic features. The proposed system's performance is assessed using the standard and data sets. The experimental results indicate the HAST-performance IDS's in both feature learning and FAR reduction by showing that it beats other published techniques in terms of accuracy, detection rate, and FAR.

3 Implementation Study

For identifying network breaches and cyber threats, there are typically two main systems. The company network has an intrusion prevention system (IPS) installed, which uses signature-based techniques to primarily inspect network protocols and flows. It produces the necessary intrusion alarms, also known as security events, and reports the alerts to another system, like SIEM. The gathering and administration of IPS alerts has been the primary focus of security information and event management (SIEM). Security analysts work to investigate suspicious alerts by policies and threshold, and to uncover malicious behaviour by analysing correlations among events, using knowledge related to attacks. The SIEM is the most popular and reliable solution among various security operations solutions to analyse the collected security events..

3.1 Proposed Methodology

By grouping events with a concurrency feature and comparing event sets in the gathered data, the suggested AI-SIEM system comprises an event pattern extraction approach in particular. Our event profiles may offer condensed input data for a variety of deep neural networks. Additionally, it allows the analyst to quickly and effectively handle all the data by comparison with long-term history data..

The developed artificial intelligence (AI)-based SIEM system's workflow and architecture. The data preparation, artificial neural network-based learning engine, and real-time threat detection phases make up the three core components of the AI-SIEM system. The system's first preprocessing stage, known as event profiling, tries to turn raw data into condensed inputs for various deep neural networks. Data aggregation includes parsing, data preparation, and data normalization stage using In the AI-SIEM system, the TF-IDF mechanism and the event profiling stage are executed one after the other. As indicated in Figure, each step produces event data sets, event vectors, and event profiles, and uses the result in the stage after it. When the system operates on detecting network intrusions in real time, this phase not only comes before the data learning stage but also comes before the conversion of raw security events to the deep-learning engine's input data. Three artificial neural networks are used in the second AI-based learning engine's modelling process. The three artificial neural networks (ANNs) receive the preprocessed data for the data learning step, and each ANN performs learning to identify the most precise model. Last but not least, each ANN model automatically categorizes each security input in real-time threat detection., each ANN model mechanically classifies each security raw event using the trained model, and the dashboard shows the only recognized true alerts to security analysts for reducing false ones.

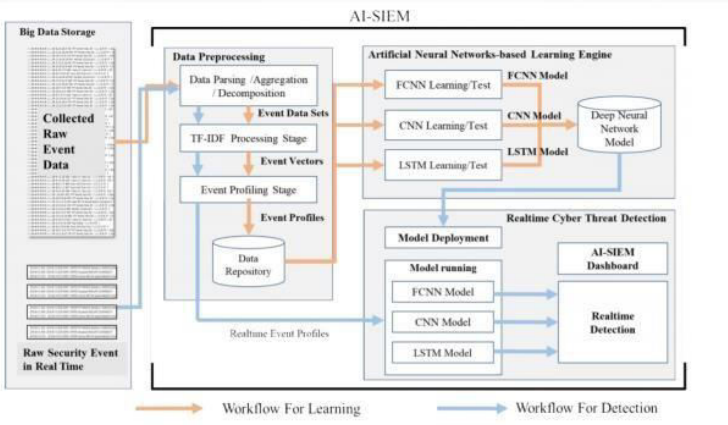

Fig 1: - Flow of Proposed System

3.2 Methodology

- 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
- Extension SVM with PSO
- Accuracy Comparison Graph
- Precision Comparison Graph
- Recall Comparison Graph
- FMeasure Comparison Graph

1) Data Parsing: This module takes 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 FMeasure. 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.

4. Results and Evolution Metrics

Fig 2:_ Main screen

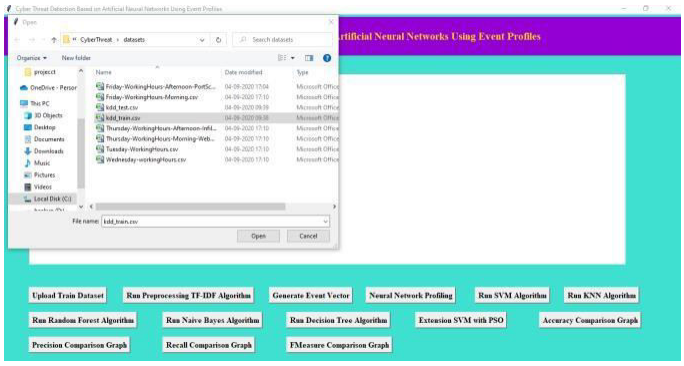

Fig 3: In above screen uploading 'kdd_train.csv' dataset and after upload will get below screen

Fig 4: In above screen we can see dataset contains 9999 records and now click on 'Run Preprocessing

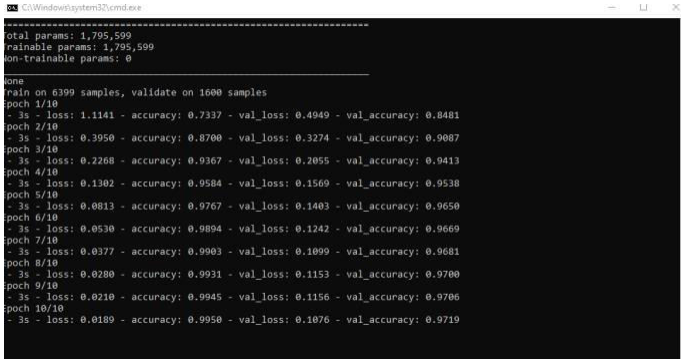

Fig 5: In above screen CNN also starts first iteration

with accuracy as 0.72 and after completing all iterations 10 we got filtered improved accuracy as 0.99 and multiply by 100 will give us 99% accuracy. So, CNN is giving better accuracy compare to LSTM and now see below GUI screen with all details

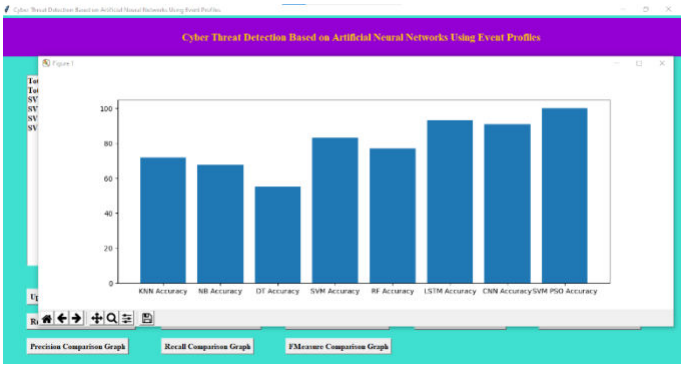

Fig 6:- In above graph x-axis represents algorithm name and y-axis represents accuracy of those algorithms and from above graph we can conclude that LSTM and CNN perform well.

5. 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.

6. 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