CYBER THREAT DETECTION BASED ON ARTIFICIAL NEURALNETWORKS USING EVENT PROFILES

[1]M.Sravanthi, [2]Gundeti Suchithra, [3]Pavuloori Vennela

[1]Assistant Professor, Department of Information Technology, Bhoj Reddy Engineering College for Women, Hyderabad, India

[2,3] Student, Department of Information Technology, Bhoj Reddy Engineering College for Women, Hyderabad, India

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 usingtwo 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

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

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 behaviorsin networks. Because of various network intrusions and malicious activities, effective defenses and security considerations were given high priority for finding reliable solutions .

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 securityinformation 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. 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 . 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 of network 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.

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 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 . Nevertheless, while learning-based approaches are useful in detecting cyber attacks in systems and networks, we observed that existing learning-based approaches have fourmain limitations.

2 Literature survey

Enhanced Network Anomaly Detection Basedon Deep Neural Networks

Abstract: 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 chaged the way approach learning task by delivering monumental progress in different disciplines like speech processing computer vision, and natural languageprocessing 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 neuralnetworks. These deep models were trained on training data set and evaluated on both test data sets provided by.

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.

Network Intrusion Detection Based on Directed Acyclic Graph and Belief Rule Base Abstract: Intrusion detection is very important for network situation awareness. While a few methods have been proposed to detect network intrusion, theycannot directly and effectively utilize semi- quantitative information consisting of multi- layered 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 model, an improved constraint covariance matrix adaption evolution strategy is developed that can effectively solve the constraint problem in the A case study was used to test the efficiency of the proposed the results showed that compared with other detection models, the model has a higher detection rate and can be used in real networks.

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

Abstract: 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 accuratelycharacterize 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 and 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.

3 Implementation Study

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, and reports 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 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.

Proposed Methodology

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.

The workflow and architecture for the developed artificial intelligent (AI)-based SIEM system. The AI-SIEM system comprises three main phases: The data preprocessing, artificial neural networks-based learning engine, and real-time threat detection phase. The first preprocessing phase in the system, termed event profiling, aims at providing concise inputs for various deep neural networks by transforming raw data. In the data preprocessing phase, data aggregation with parsing, data normalization stage using TF-IDF mechanism, and event profiling stage are consecutively performed in the AI-SIEM system.Each stage generates event data sets, event vectors, and event profiles, respectively, and the output is utilized in next each stage, as shown in Figure. This phase not only precedes the data learning stage but also precedes the conversion of raw security events to the deep-learning engine's input data when the system operates on detecting network intrusions in real time. The second AI-based learning engine employs three artificial neural networks for modeling. For the data learning stage, the preprocessed data are fed into the three artificial neural networks, and each ANN performs learning to find the most accurate model. Finally, 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.

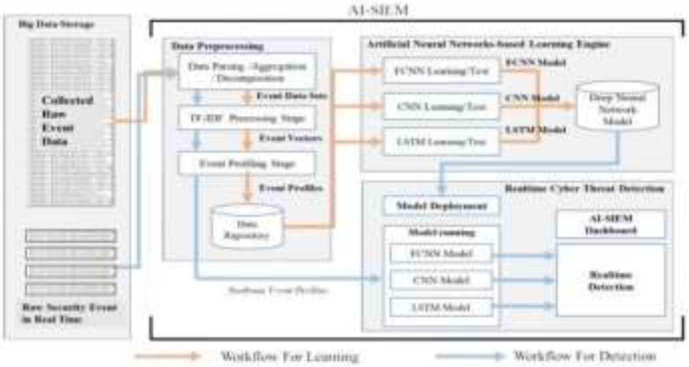

Fig 1: - flow of proposed system

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
- 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 containsnormal and attack signatures
3) Event Profiling Stage: Processed data will besplitted 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 byreducing its size or running it on high configuration system.

4 Results and Evolution Metrics

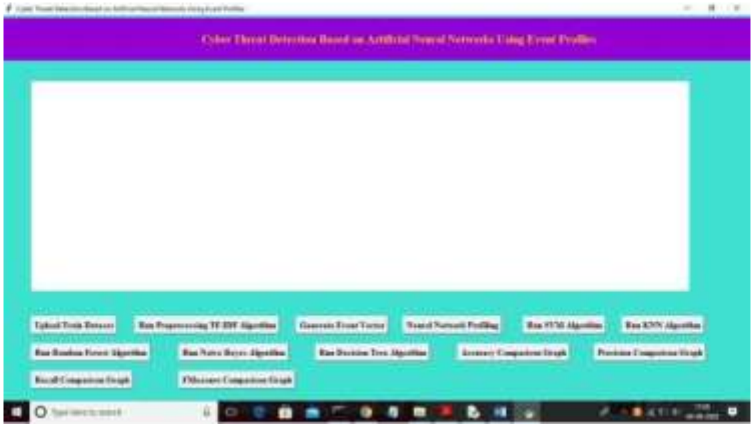
Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles
Upload Train Dataset
Run Preprocessing TF-IDF Algorithm
Generate Event Vector
Neural Network Profiling
Run SVM Algorithm
Run KNN Algorithm
Run Random Forest Algorithm
Run Naive Bayes Algorithm
Run Decision Tree Algorithm
Accuracy Comparison Graph
Precision Comparison Graph
Recall Comparison Graph
FMeasure Comparison Graph

Fig 2:_ Main screen

Fig 3: In above screen uploading 'kdd_train.csv'

dataset and after upload will get below screen

Fig 4: _ a In above screen we can see dataset contains 9999 records and now click on 'RunPreprocessing

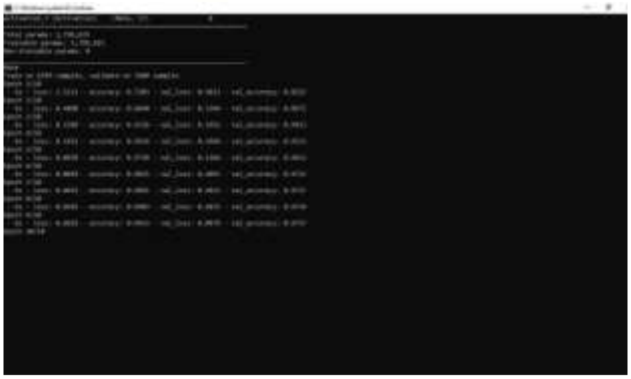

Fig 5: _ In above screen CNN also starts first iteration with accuracy as 0.72 and after completing all iterations 10 we got filtered improved accuracy as 0.99 and multiply by 100 will give us 99% accuracy. So, CNN is giving better accuracy compare to LSTM and now see below GUI screen with all details

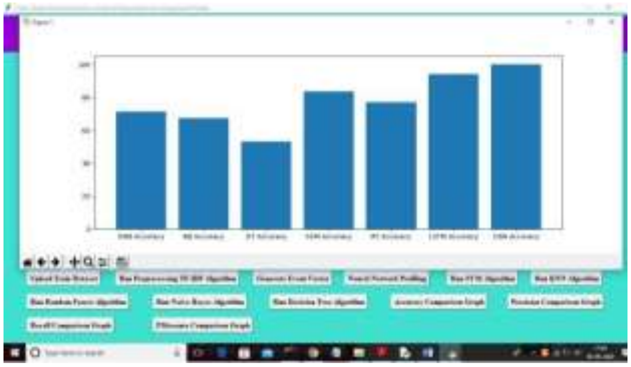

Fig 6:- In above graph x-axis represents algorithm name and y-axis represents accuracy of thosealgorithms and from above graph we can conclude that LSTM and CNN perform well. Now click on Precision Comparison Graph' to get below graph

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.

5 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", *ETRIJournal*, 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