ISSN PRINT 2319 1775 Online 2320 7876

 Journal Volume 11 , Iss 11, Nov 2022

Deep Neural Networks Using Event Profiles Based on Cyber Threat Detection

Supriya Sawwashere[1], Kiran Bode[2], Sujata Helonde[3], Marshnil Patel[4]

[1,2,3]Assistant Professor, Department of CSE J D College of Engineering & Management, Nagpur
[4]Research Scholar Department of CSE J D College of Engineering & Management, Nagpur

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 cyberthreat 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.

1. INTRODUCTION

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

ISSN PRINT 2319 1775 Online 2320 7876

information security solutions generally fall into two categories: analyst-driven and machine learning-driven solutions. Analystdriven 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 cyberattacks 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 well known 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.

Outline of the Project:

With the emergence of artificial intelligence (AI) techniques, learning-based approaches for detecting cyberattacks, have become further improved, and they have achieved significant results in many studies. However, owing to constantly evolving cyberattacks, 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, 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 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. 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 cyberattacks in systems and 11 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

ISSN PRINT 2319 1775 Online 2320 7876

 Journal Volume 11 , Iss 11, Nov 2022

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 well known 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.

Domain Introduction

The Deep Neural Network (DNN) is a neural network with a certain level of complexity, a neural network with more than two layers. Deep neural networks use sophisticated mathematical modeling to process data in complex ways.

A neural network, in general, is a technology built to simulate the activity of the human brain – specifically, pattern recognition and the passage of input through various layers of simulated neural connections. Many experts define deep neural networks as networks that have an input layer, an output layer and at least one hidden layer in between. Each layer performs specific types of sorting and ordering in a process that some refer to as “feature hierarchy.” One of the keys uses of these sophisticated neural networks is dealing with un labeled or unstructured data. The phrase “deep learning” is also used to describe these deep neural networks, as deep learning represents a specific form of machine 12 learning where technologies using aspects of artificial intelligence seek to classify and order information in ways that go beyond simple input/output protocols.

Benefits of Deep Neural Network

Neural networks use randomness by design to ensure they effectively learn the function being approximated for the problem. Randomness is used because this class of machine learning algorithm performs better with it than without. The most common form of randomness used in neural networks is the random initialization of the network weights. Although randomness can be used in other areas, here is just a short list:

- Randomness in Initialization, such as weights.
- Randomness in Regularization, such as dropout.
- Randomness in Layers, such as word embedding.

2. 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.

ISSN PRINT 2319 1775 Online 2320 7876

 Journal Volume 11 , Iss 11, Nov 2022

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 NSLKDDTest21. All experiments in this paper are performed by authors on a GPUbased 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 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

ISSN PRINT 2319 1775 Online 2320 7876

 Journal Volume 11 , Iss 11, Nov 2022

published approaches in terms of accuracy, detection rate, and FAR, which successfully demonstrates its effectiveness in both feature learning and FAR reduction.

Aim

The aim of the project is Threat detection and response is the most important aspect of cyber security for IT organizations that depend on cloud infrastructure. Threat detection, therefore, describes the ability of IT organizations to quickly and accurately identify threats to the network or to applications or other assets within the network.

Scope of the Project

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.

Objectives

Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant Thus the objective of input design is to create an input layout that is easy to follow.

3. METHODOLOGY

Existing System

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 logs . 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.

- Disadvantages:

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

These learning-based approaches require to learn the attack model from historical threat data and use the trained models to detectintrusions for unknowncyber threat.

ISSN PRINT 2319 1775 Online 2320 7876

 Journal Volume 11 , Iss 11, Nov 2022

Proposed System

In order to solve the above problem, all Our proposed system aims at converting a large amount of security events to individual event profiles for processing very large scale data. We developed a generalizable security event analysis method by learning normal and threat patterns from a large amount of collected data, considering the frequency of their occurrence. In this study, we specially propose the method to characterize the data sets using the basepoints in data preprocessing step. This method can significantly reduce the dimensionality space, which is often the main challenge associated with traditional data mining techniques in log analysis.

• Our event profiling method for applying artificial intelligence techniques, unlike typical sequence-based pattern approaches, provides featured input data to employ various deep-learning techniques. Hence, because our technique is able to facilitate improved classification for true alerts when compared with conventional machine-learning methods, it can remarkably reduce the number of alerts practically provided to the analysts.

• For the applicability, we evaluate our system with real IPS security events from a real security operations center (SOC) and validate its effectiveness through performance metrics, such as the accuracy, true positive rate (TPR), false positive rate (FPR) and the F-measure. Moreover, 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). And we also perform an evaluation by applying our method to two benchmark datasets (i.e., NSLKDD, CICIDS2017), which are most commonly used in the field of network intrusion detection research.

- Advantages:

For cyber-threat detection, the SIEM analysts spend an immense amount of effort and time to differentiate between true security alerts and false security alerts in collected events.

System Architecture

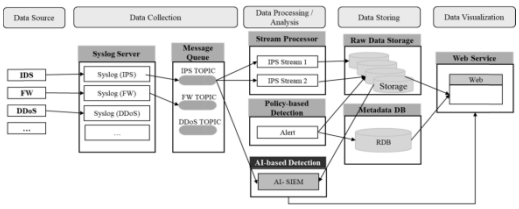


Fig 4.1 Architectural design

Architectural design

Modules

Propose algorithms consists of following module are.

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 spitted into train and test model based on profiling events.

ISSN PRINT 2319 1775 Online 2320 7876

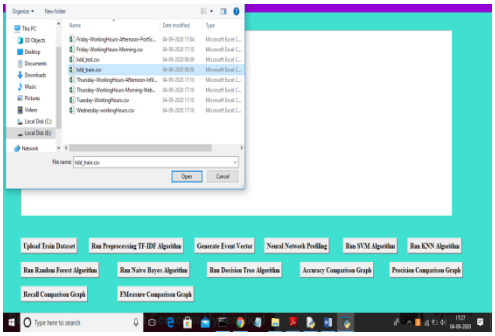
 Journal Volume 11 , Iss 11, Nov 2022

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 FMEA sure. Algorithm will learn perfectly will yield better accuracy result and that model will be selected to deploy on real system for attack detection.

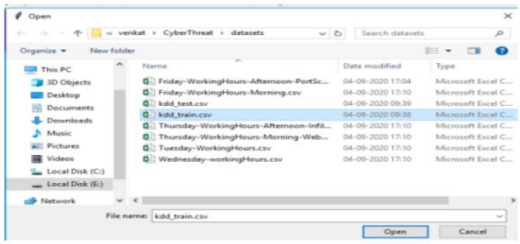
Module Description

To build the complete prediction model, it is essential to carry out step-by-step execution of all required modules. To run project double, click on 'run.bat' file to get below screen.

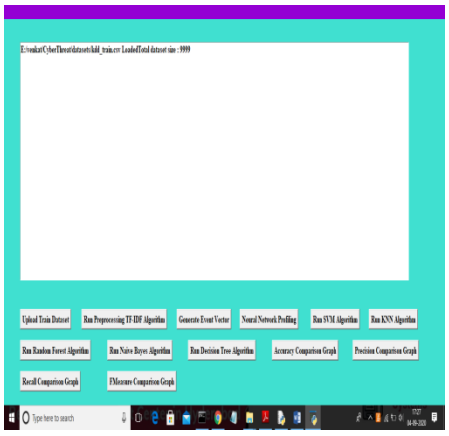
Upload Datasets



In above screen click on 'Upload Train Dataset' button and upload dataset. In above screen uploading 'kdd_train.csv' dataset and after upload will get below screen.



Tf-Idf Algorithm



ISSN PRINT 2319 1775 Online 2320 7876

In above screen we can see dataset contains 9999 records and now click on 'Run Preprocessing TF-IDF Algorithm' button to convert raw dataset into TF-IDF values.

In above screen TF-IDF processing completed and now click on 'Generate Event Vector' button to create vector from TF-IDF with different events

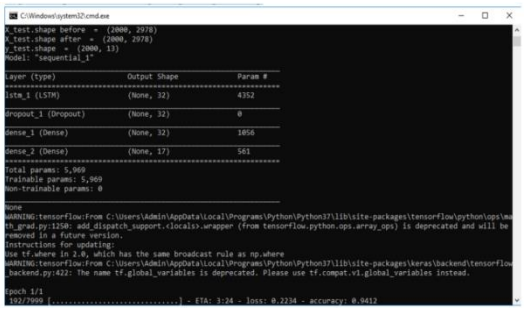
Genrate Event Vector:

In above screen we can see totally different unique events names and in below we can see dataset total size and application using 80% dataset (7999 records) for training and using 20% dataset (2000 records) for testing. Now dataset train and test events model ready and now click on 'Neural Network Profiling' button to create LSTM and CNN model.

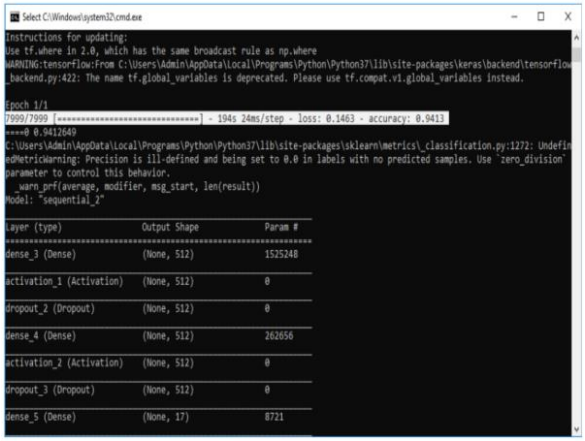
ISSN PRINT 2319 1775 Online 2320 7876

 Journal Volume 11 , Iss 11, Nov 2022

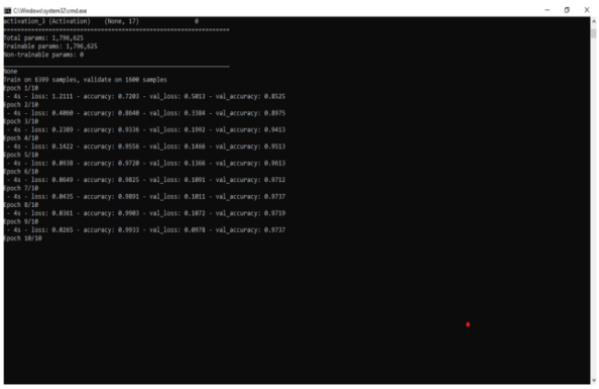
Neural Networks Profring:



In above screen LSTM model is generated and its epoch running also started and its starting accuracy is 0.94. Running for entire dataset may take time so wait till LSTM and CNN training process completed. Here dataset contains 7999 records and LSTM will iterate all records to filter and build model.



In above selected text we can see LSTM complete all iterations and in below lines we can see CNN model also starts execution.



IJFANS International Journal of Food And Nutritional Sciences
Official Publication of International Association of Food and Nutrition Scientists

ISSN PRINT 2319 1775 Online 2320 7876

Cnn Algorithm

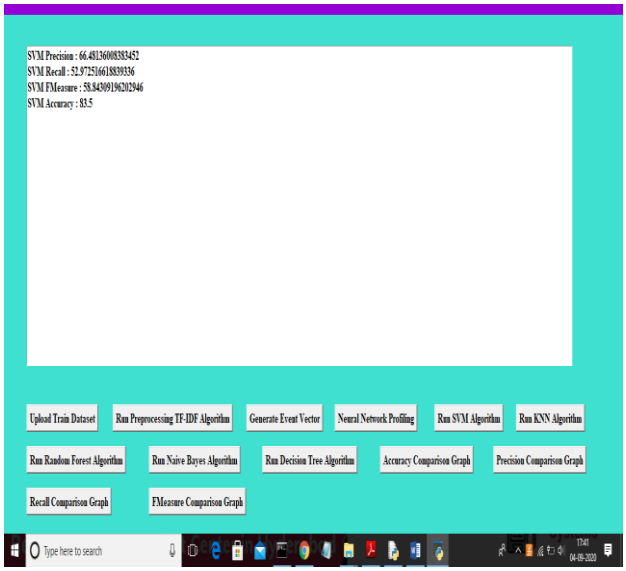
In above screen CNN also starts first iteration with accuracy as 0.72 and after completing all iterations 10 we got filtered improved accuracy as 0.99 and multiply by 100 will give us 99% accuracy. So, CNN is giving better accuracy compare to LSTM and now see below GUI screen with all details.

Knn Algorithm

In above screen we can see both algorithms accuracy, precision, recall and FMEA sure values. Now click on 'Run SVM Algorithm' button to run existing SVM algorithm.

ISSN PRINT 2319 1775 Online 2320 7876

 Journal Volume 11 , Iss 11, Nov 2022



In above screen we can see SVM algorithm output values and now click on 'Run Random Forest Algorithm' to run Random Forest algorithm.

Random Forest Algorithm

In above screen we can see Random Forest algorithm output values and now click on 'Run Naïve Bayes Algorithm' to run Naïve Bayes algorithm.

Naïve Bayes Algorithm

In above screen we can see Naïve Bayes algorithm output values and now click on 'Run Decision Tree Algorithm' to run Decision Tree Algorithm

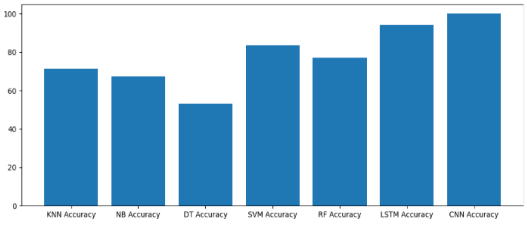
Decision Three Algorithm

IJFANS International Journal of Food And Nutritional Sciences
Official Publication of International Association of Food and Nutrition Scientists

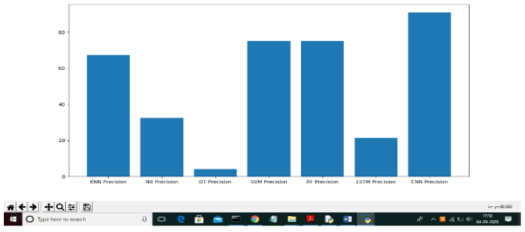
ISSN PRINT 2319 1775 Online 2320 7876

 Journal Volume 11 , Iss 11, Nov 2022

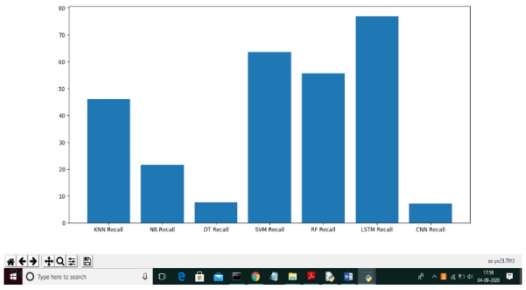
Now click on 'Accuracy Comparison Graph' button to get accuracy of all algorithm



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. Now click on Precision Comparison Graph' to get below graph



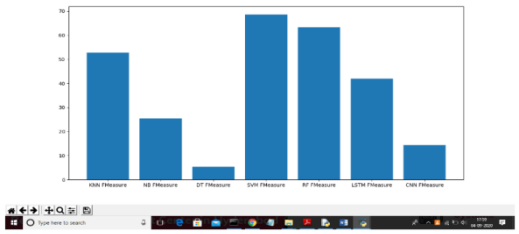
In above graph CNN is performing well and now click on 'Recall Comparison Graph'



IJFANS International Journal of Food And Nutritional Sciences
Official Publication of International Association of Food and Nutrition Scientists

ISSN PRINT 2319 1775 Online 2320 7876

In above graph LSTM is performing well and now click on FMeasure Comparison Graph button to get below graph



From all comparison graph we can see LSTM and CNN performing well with accuracy, recall and precision.

Objectives:

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.
2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

Input and Output Design: Input Design:

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy.

Output Design

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that

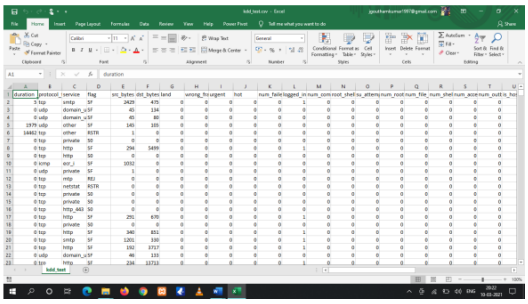
ISSN PRINT 2319 1775 Online 2320 7876

 Journal Volume 11 , Iss 11, Nov 2022

people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system.
The output form of an information system should accomplish one or more of the following objectives.
• Convey information about past activities, current status or projections of the Future.
• Signal important events, opportunities, problems, or warnings.
• Trigger an action.
• Confirm an action.

4. RESULTS

Datasets



Our dataset has been collected from two large enterprise sys-terms, namedESX1andESX-2. The security raw events were collected over 5 months for ESX-1, over 30 days for ESX-2, respectively, in which the detecting threat information was separately recorded by the SOC security analysts whenever network intrusion occurred. The list of threat detection information contains threat occurrence time, related attacks, category of attack, respond contents, attack IP address, and victim network information. In our datasets, we investigated 798 detecting cyber threats in ESX-1, which are dispersed across the entire collection period. Looking at the type of occurred attacks in recorded cyber threats, there are 240 scanning, 547 system hack-king, and 11 worm attacks. Similarly, in ESX-2 there are941 scanning, 3,077 system hacking, and 51 worm attacks. This categorizing of attack type was manually performed by SOC analysts. By category, the system hacking attack includes a cross site script, DDoS, brute force attack, and injection attack. A trojan and backdoor attack belongs scanning attack. Overall, the number of attacks were found 4,079 cyber-threats.

Data Visualization

The t-SNE is not only commonly utilized for vector data visualization but also considered as embedding tools to visualize high-dimensional data. The t-SNE is able to visual-zed high-dimensional data into two-dimensional maps by learning twodimensional embedding vectors that preserves neighbour structures among highdimensional data. The N data rows in dataset are randomly selected, which are visualized by performing analysis in t-SNE represent the maps that are visualized by t-SNE for CICIDS 2017 and ESX-2, respectively. The t-SNE plots in the figure show that the normal and attack data points located nearby in the same space, which makes it very hard to classify them into either normal or attack. Although the t-SNE plots of normal and attack data are clustered, it clearly finds out that those are not

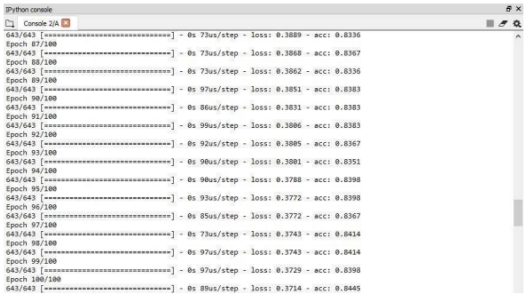
ISSN PRINT 2319 1775 Online 2320 7876

 Journal Volume 11 , Iss 11, Nov 2022

linearly separated. In general, it is known that deep learning is then effective at dealing with high-dimensional data with non-linearity [50], which is one of the reasons we employ deep learning approaches to detect cyber threats.

Experimental Results

Based on the results of this experiment, we are able to arrive at two meaningful conclusions. First, our mech-amiss are capable of being employed as learningbased models for network intrusion detection. When the performance evaluations were conducted using two well-known benchmark datasets such as NSLKDD and CICIDS2017, the result proved as capable as the conventional machine-learning models. This means that our proposed methods, employed in the AI-SIEM system, have applicability for learning-based network intrusion detection. Second, when the conventional learning-based methods, which accomplish good result by bench mark dataset, are employed in the real world, the performance of overall accuracy is not as reliable as those of benchmark datasets. Never the less, the accuracy performance of our three EP-ANN models were not significantly degraded, despite the large amount of data and a lack of benchmark dataset features, such as seen in the result for ESX-2. By contrast, the accuracy of conventional methods had degraded from approximately 0.90 to 0.85.



Epoch

In neural networks generally, an epoch is a single pass through a full dataset. It is the iterations constituting one forward pass and one backward pass. A confusion matrix is a technique for summarizing the performance of a classification algorithm. The number of correct and incorrect predictions are summarized with count values and broken down by each class. This is the key to the confusion matrix. Classification accuracy is the ratio of correct cyber threats. Computers do not generally store arbitrarily large numbers. Instead, each number stored by a computer is allotted a fixed amount of space. Therefore, when the number of time units that have elapsed since a system's epoch exceeds the largest number that can fit in the space allotted to the time representation, the time representation overflows, and problems can occur. While a system's behavior after overflow occurs is not necessarily predictable, in most systems the number representing the time will reset to zero, and the computer system will think that the current time is the epoch time again.

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

ISSN PRINT 2319 1775 Online 2320 7876

 Journal Volume 11 , Iss 11, Nov 2022

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.

Future Work

In the future, to address the evolving problem of cyber-attacks, we will focus on enhancing earlier threat predictions through the multiple deep learning approach to discovering the long-term patterns in history data. In addition, to improve the precision of labeled dataset for supervised-learning and construct good learning datasets, many SOC analysts will make efforts directly to record labels of raw security events one by one over several months. For testing, we constructed the purpose-built test bed where for conducting performance evaluations. This test bed consists of the big data platform and the AI-SIEM system. Moreover, in the SOC, we also had collected real-world IPS data over several months.

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 Deep Neural Networks," IEEE Access, vol. 6, pp. 48231-48246, 2018.
2. B. Zhang, G. Hu, Z. Zhou, Y. Zhang, P. Qian, 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 spatialtemporal 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 Sekaran, K. Kandasamy, "Profiling SIEM tools and correlation engines for security analytics," In Proc. Int. Conf. Wireless Com., Signal Prove. and Net. (Wisp NET), 2017, pp. 717-721.
6. Hubbell and V.Surya narayana False alarm minimization techniques in signaturebased intrusion detection systems: A survey,'' Compute. Common., vol. 49, pp. 1- 17, Aug. 2014.
7. A. Naser, M. A. Majid, M. F. Zolile 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. Marconi, P. Verviers, and Gianluca Stringham, "Tiresias: Predicting Security Events Through Deep Learning," In Proc. ACM CCS 18, Toronto, Canada, 2018, pp. 592-605.

ISSN PRINT 2319 1775 Online 2320 7876

9. Kyle Soaks 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. Veerama channid, I. Arnaldo, V. Koraput, C. Basis, K. Li, "AI2: training a big data machine to defend," In Proc. IEEE Bigdata Security HPSC IDS, New York, NY, USA, 2016, pp. 49-54