International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

CYBER SECURITY DETECTION USING ANN

Harsha[1], A. Rishika[2] ,Dr.D.Shravani[3]

Department of Computer Science and Engineering, Stanley College of Engineering and Technology for Women, Telangana, India

Abstract

One of the major challenges in cyber security is the provision of an automated and effective cyber-threats detection technique. 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.

Keywords:

Machine Learning, Neural Networks, pre-processing, Supervised Learning Algorithm, Unsupervised Learning Algorithm, Support Vector Machine, K-Nearest Neighbor, Random Forest, Naive Bayes, Decision Tree, Convolution Neural
Networks

1. Introduction

1.1 About Project

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. 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.

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 learningdriven 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 learningbased 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.

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. 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, CICIDS2017, and Kyoto-Honeypot. 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. 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. 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.

1.2 Objectives of the Project

It is highly challenging to protect IT systems against threat and malicious behaviors in network. One of the major challenges in cyber security is the provision of an automated and effective cyber-threats detection technique. 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. Thus, we are introducing a system that will help us to identify the thread and improve analysis. 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 machinelearning 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 machinelearning methods

1.3 Scope of the Project

This method is developed in an efficient way to predict an organizations or assets from damage. It also helps us to detect any malicious activity that could cause negative impact on an asset or organization. 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. Thus, we are introducing a system that will help us to identify the thread and improve analysis

2. Literature Survey

2.1 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

2.2 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 intrusiondetection.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

3. Proposed Architecture

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 TFIDF mechanism, and event profiling stage are consecutively performed in the AISIEM 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.

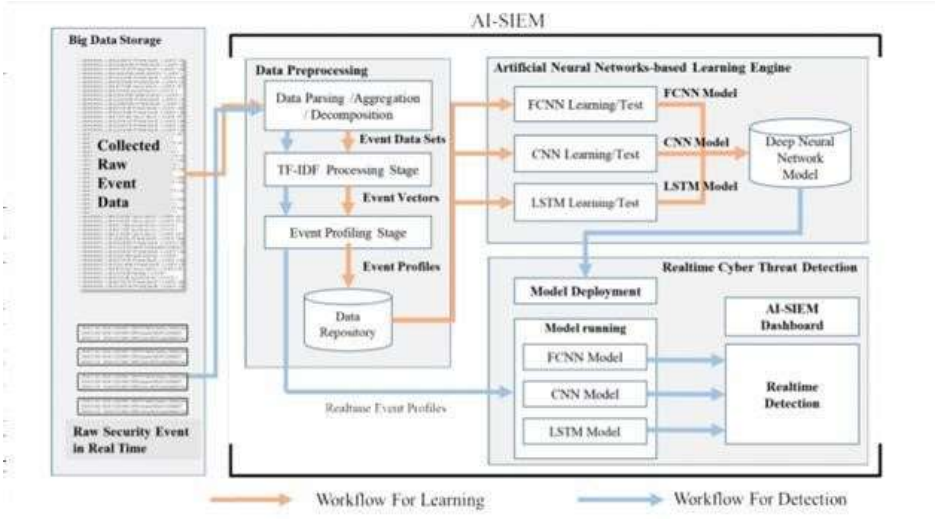
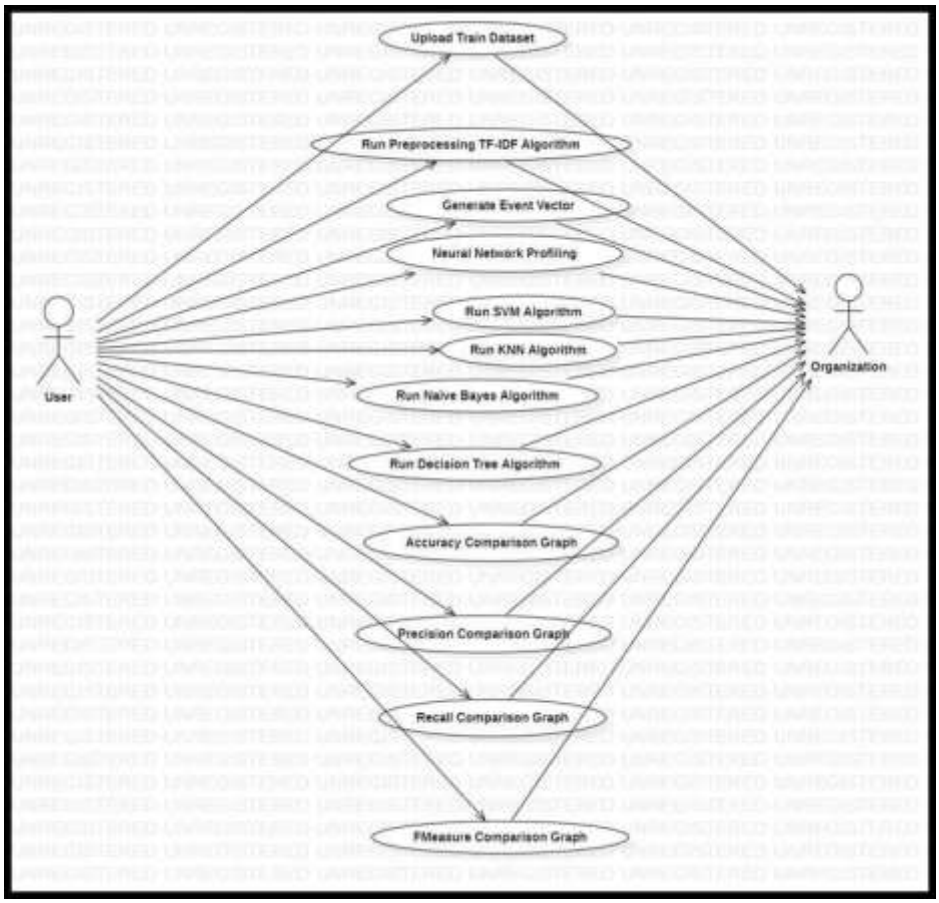


Fig 3.1: Proposed Architecture for Cyber Threat Detection

USE CASE DIAGRAM

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org



4. Implementation

4.1 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
- F-Measure Comparison Graph

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

4.2 MODULES DESCRIPTION

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 FMeasure.

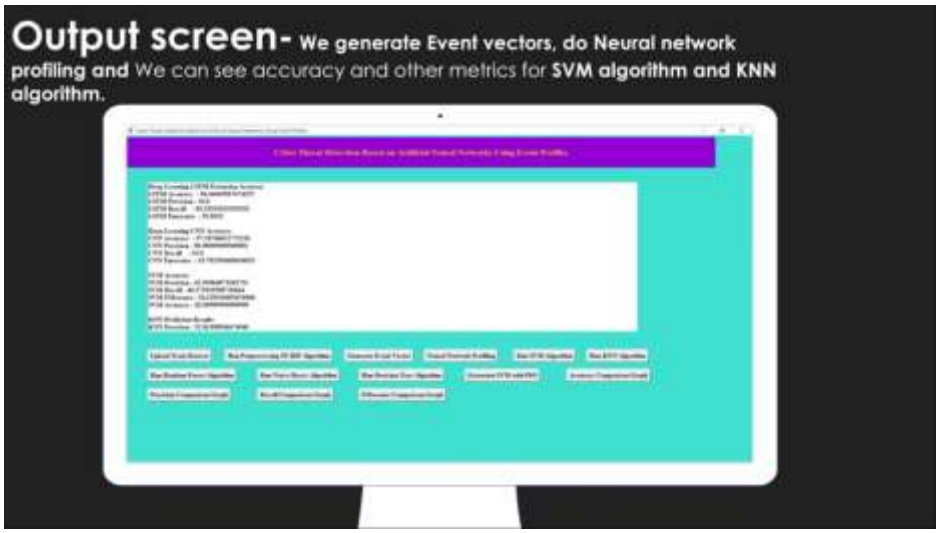
Algorithm will learn perfectly will yield better accuracy result and that model will be selected to deploy on real system for attack detection. Datasets which we are using for testing are of huge size and while building model it's going to out of memory error but kdd_train.csv dataset working perfectly but to run all algorithms it will take 5 to 10 minutes

5. Result

In above screen click on 'Upload Train Dataset' button and upload dataset, dataset contains 9999 records and now click on 'Run Preprocessing TF-IDF Algorithm' button to convert raw dataset into TF-IDF values.

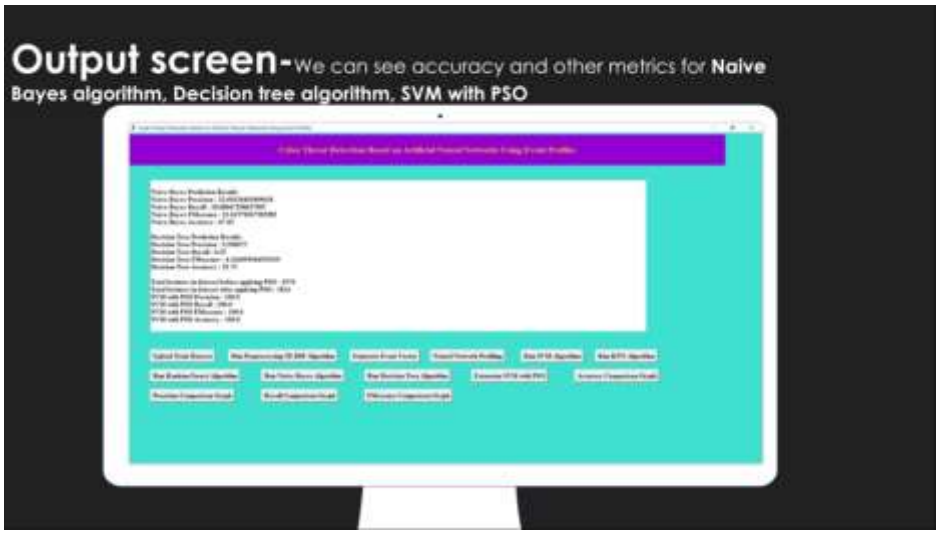
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

In above screen we can see total 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.



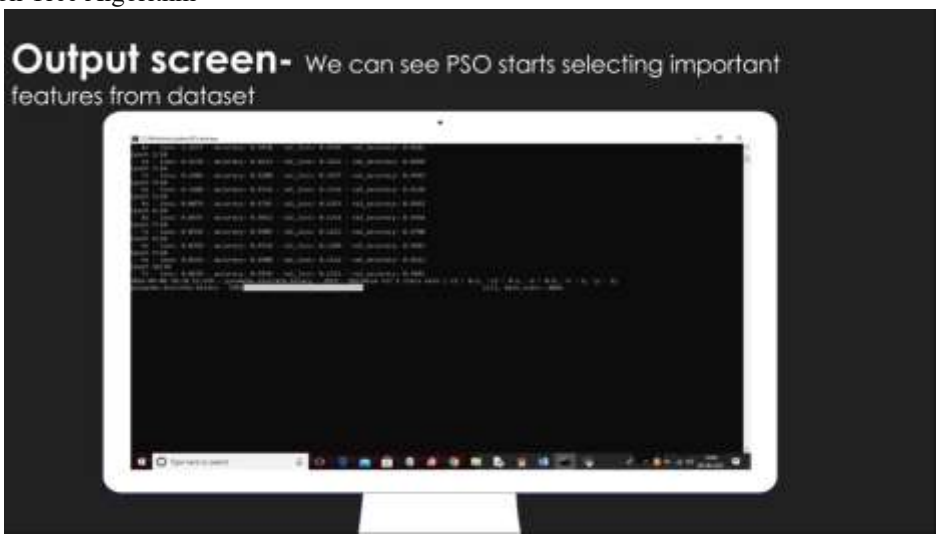
In above screen LSTM model is generated and its epoch running also started and its starting accuracy is 0.94. Running for entire dataset may take time so wait till LSTM and CNN training process completed. Here dataset contains 7999 records and LSTM will iterate all records to filter and build model. In above selected text we can see LSTM complete all iterations and in below lines we can see CNN model also starts execution. 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. In above screen we can see both algorithms accuracy, precision, recall and F-Measure values. Now click on Run SVM Algorithm button to run existing SVM algorithm. In above screen we can see SVM algorithm output values and now click on Run KNN-Algorithm to run KNN algorithm

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org



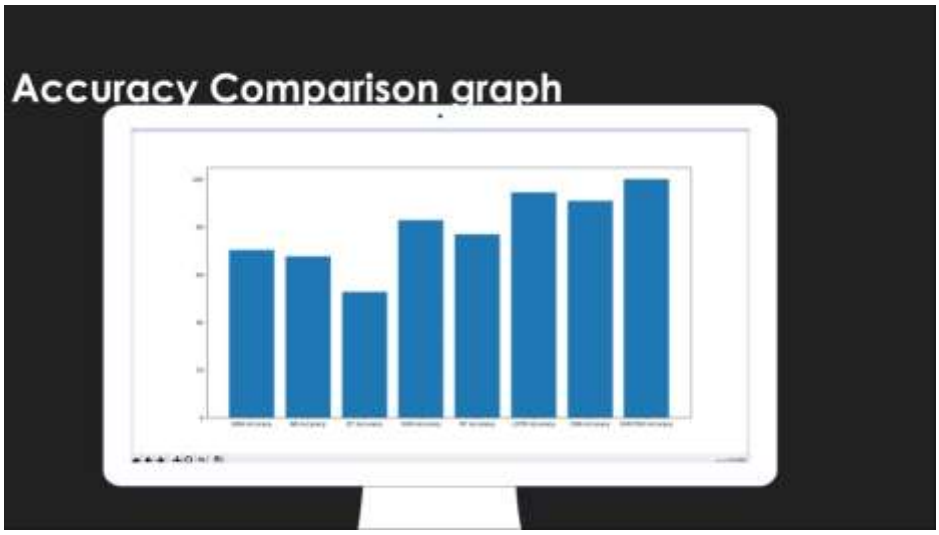
In above screen we can see Random Forest algorithm output values and now click on 'Run Naïve Bayes Algorithm' to run Naive Bayes algorithm. In above screen we can see Naïve Bayes algorithm output values and now click on 'Run Decision Tree

Algorithm' to run Decision Tree Algorithm



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

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

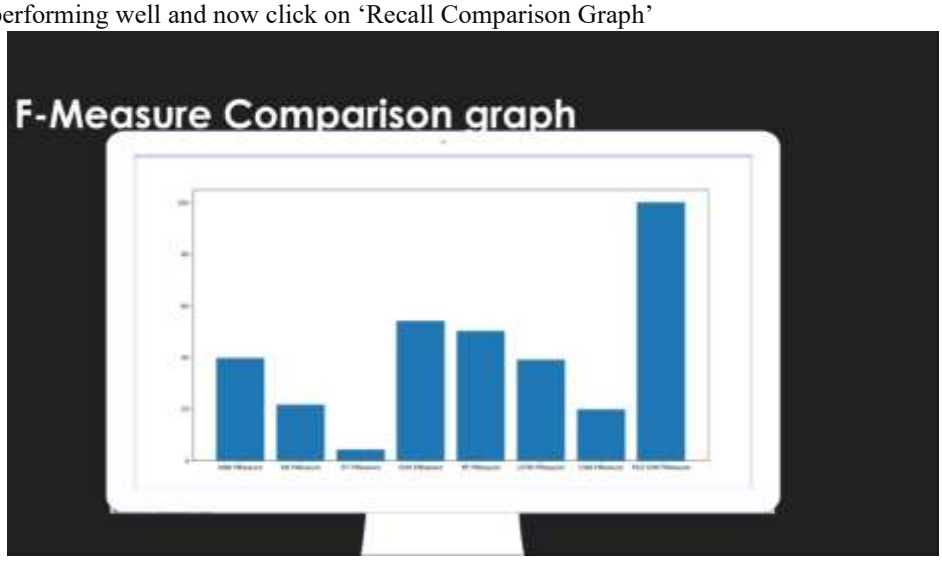


In above graph x-axis represents algorithm name and y-axis represents accuracy of those algorithms and from above graph we can conclude that LSTM and CNN perform well. Now click on Precision Comparison Graph' to get below graph

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

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

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



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

6. Conclusion

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.

7. Future Scope

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.

8. References

1. 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.
2. 39, no. 4, pp. 592-604, Aug. 2017
3. W. Wang, Y. Sheng and J. Wang, "HAST-IDS: Learning hierarchicalspatialtemporal 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. 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.