CYBER THREAT DETECTION BASED ON ARTIFICIAL NEURAL NETWORKS USING EVENT PROFILES

A. Uday kiran, Assistant Professor, Department of CSE, CMR Technical Campus , Medchal, Hyderabad, Telangana, India, udaykiran.cse@cmrtc.ac.in

B. Sai Kumar, Department of CSE, CMR Technical Campus , Medchal, Hyderabad, Telangana, India, 197r1a05c9@cmrtc.ac.in

G. Srikanth, Department of CSE, CMR Technical Campus, Medchal, Hyderabad, Telangana, India, 197r1a05e0@cmrtc.ac.in

J. Sainath Reddy, Department of CSE, CMR Technical Campus ,Medchal, Hyderabad, Telangana, India, 197r1a05e2@cmrtc.ac.in

ABSTRACT: The formation of a brilliant and productive cyber threat location framework is one of the most troublesome aspects of network security.A artificial neural network-based artificial intelligence (AI) methodology for distinguishing cyberthreats is depicted in this review. The proposed innovation utilizes a deep learning-based recognition instrument to change countless recorded security occasions into individual occasion profiles for improved digital danger recognizable proof. We combined event profiling for information preprocessing with artificial neurall network techniques like FCNN, CNN, and LSTM to create an AI- SIEM framework for this endeavor.Security examiners can answer digital dangers all the more rapidly in light of the fact that the framework focuses on recognizing veritable positive signs and bogus positive signs. Using two real world datasets and two benchmark datasets (NSLKDD and CICIDS2017), the makers did all of the examinations in this work. SVM, k-NN, RF, NB, and DT, notwithstanding the Xgboost and Adaboosttechniques, were the five standard machine learning (ML) calculations that we assessed in contrast with the presentation of current procedures. Subsequently, the essential disclosures of this investigation exhibit that the proposed approaches may be used as learning-based models for recognizing network aggravations and defeat standard ML methods eventually.

Keywords *– Artificial intelligence, deep neural networks, network security, cyber security, and detection of intrusions.*

1. INTRODUCTION

As an outcome of the accessibility of artificial intelligence (AI) gadgets, learning-based structures for perceiving advanced dangers have improved and given extensive outcomes in a scope of tests. Be that as it may, it's still difficult to shield IT frameworks from dangers and criminal organization movement on the grounds that cyberattacks change continually. In order to find dependable solutions, effective defenses and security concerns were prioritized in response to a variety of network invasions and harmful actions. In the past, two fundamental systems have been utilized to identify network breaches and cyberthreats. In order to investigate network protocols and flows, an intrusion prevention system (IPS) may primarily employ signature-based methods in the company network. It sends relevant intrusion warnings, or security events, to another system, like SIEM, and reports them. The collection and management of IPS alarms has been the primary focus of

SIEM (security information and event management). For assessing gathered logs and security occasions, the SIEM is the most famous and reliable of the different security movement arrangements.Moreover, security investigators endeavor to examine dubious cautions in view of guidelines and limits and to recognize noxious way of behaving by breaking down associations among occasions and utilizing assault ability. However, it is still challenging to distinguish breaches from intelligent network attacks and to identify them due to the large volume of security data and the large number of false warnings. Consequently, approaches based on machine learning and artificial intelligence for identifying assaults have received more attention in current intrusion detection research. Security analysts may be able to speed up and automate their investigation of network attacks thanks to advancements in AI fields. In order to separate cyber disruptions from hidden dangers, these gaining-based processes need gaining the attack model using earlier danger information. Analysts who should survey numerous events simultaneously may profit from a learning-based method for recognizing whether an attack occurred in an immense amount of information. There are two sorts of information security game plans, according to: courses of action driven by examiners and by ML.Solutions that are analyst-driven are based on guidelines developed by analysts, who are security professionals. In the meantime, new cyber risks may be discovered with the assistance of machine learning-driven methods for detecting odd or aberrant patterns. However, despite the fact that learning-based methods are efficient at identifying cyberattacks in systems and networks, we discovered four major drawbacks.

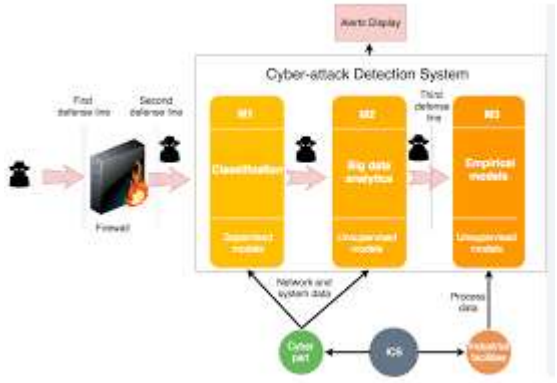

Fig.1: Example figure

It is anticipated that named data will initiate learning-based identification procedures for the creation of learning models and their evaluation. Notwithstanding, getting such stamped data at a scale that licenses exact model training is testing.Despite the fact that administered gaining models can profit from named information, numerous business SIEM frameworks don't save it. Second, most of the learning attributes used hypothetically in each study are absent in average organization security frameworks, making them unimportant to this present reality. It is along these lines challenging to apply to real circumstances. A notable dataset, like NSLKDD, CICIDS2017, and Kyoto-Honeypot, has been utilized to assess the presentation of a computerized interruption location strategy that utilizes deeplearning innovation. However, benchmark datasets used in a number of previous studies are inaccurate and cannot be applied to the real world due to insufficient characteristics. An adopted learning model must be evaluated with information obtained from the real world in order to circumvent these limitations. To wrap things up, utilizing an oddity based strategy to find network interruptions could help find new cyber threats, however it could likewise cause a ton of deceptions. A great deal of bogus positive admonitions cost huge load of cash and require a ton of work with respect to representatives to research. Fourth, a few programmers purposefully

disguise their criminal operations by continuously changing their way of behaving. Accordingly, recognition procedures are delivered inadequate in any event, while satisfactory learning-based models are accessible.Furthermore, the purpose of almost all security systems is to analyze short-term network security events. We think that one way to identify harmful cyber attack activity over long periods of time is to evaluate the security event history associated with event production in order to fight against attacks that change constantly. This attempt is fundamentally spurred by these issues. We offer a AI-SIEM framework that utilizes profound figuring out how to recognize authentic and bogus alerts to resolve these issues. Security experts might have the option to answer all the more rapidly to digital dangers spread across countless security occasions with the help of our proposed technique. The proposed AI-SIEM framework incorporates a technique for removing occasion designs by relating across occasion sets in assembled information and conglomerating occasions with a simultaneousness trademark. For various deep neural networks, our event profiles may provide straightforward input data. In addition, it enables the analyst to efficiently and quickly manage all of the data by comparing long-term historical data.

2. LITERATURE REVIEW

Enhanced Network Anomaly Detection Based on Deep Neural Networks:

The requirement for data network security has expanded at an outstanding rate throughout the course of recent years because of the dramatic development of Web applications. As a critical protect for the framework of an organization, an interruption identification framework is expected to answer a danger climate that is continually moving. For exact oddity recognition, specialists in ML and data mining have fostered various managed and solo techniques. A subfield of ML known as profound learning utilizes structures looking like neurons for the purpose of learning. Deep learning has fundamentally altered how we approach learning difficulties by making significant advancements in sound handling, PC vision, and regular language handling, to name a few. Examining this fresh out of the plastic new innovation for applications in data security is just essential. The reason for this exploration is to decide if inconsistency based interruption discovery frameworks can be executed utilizing deep learning calculations. Irregularity recognition models in view of autoencoders, convolutional neural networks, and recurrent neural networks were produced with the end goal of this examination. The NSLKDD preparing and test informational collections, NSLKDDTest+ and NSLKDDTest21, were utilized to prepare and assess these profound models. The developers completed all of the trials in this work using a GPU-based test seat. Customary ML-based interference acknowledgment models were assembled utilizing wide learning machines, decision trees, random forests, support vector machines, naive bayes, and quadratic discriminant evaluation. Deep and standard ML models were both tried utilizing huge gathering estimates, for example, advantageous working limits, area under curve, precision recall curve, mean normal precision, and arrangement accuracy. The exploratory discoveries of the deep IDS model uncovered promising outcomes for use in obvious idiosyncrasy area systems.

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

In order to keep a network's situational awareness up to date, intrusion detection is essential. Although a few methods for detecting network intrusion have been presented, they cannot use expert knowledge and

quantitative data in a direct and effective manner. Consequently, this work offers a novel belief rule foundation (BRB) and DAG-based detection model. The DAG is utilized to develop a diverse BRB model in the proposed engineering, which is called DAGBRB. This model might forestall rule number blast since there are various sorts of attack. To get the ideal boundaries for the DAGBRB model, a superior requirement covariance matrix adaption evolution method (CMAES) that is able to do successfully tending to the limitation issue in the BRB is proposed. The suggested DAGBRB's efficiency was had a go at using a logical examination. The outcomes showed that the DAGBRB model can be utilized in genuine organizations and has a higher location rate than different models.

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

The plan of a trademark based intrusion detection system (IDS) is a notable area of examination in the field of interference area. By concentrating on network traffic, an IDS finds out about typical and strange way of behaving and may find novel dangers. Regardless, the improvement of a list of capabilities that can successfully group network traffic stays an examination point on the grounds that the viability of an interruption discovery framework is vigorously subject to include plan. The high false alarm rate (FAR) of peculiarity based IDSs seriously limits their application. This study proposes an original intrusion detection system (IDS) called the hierarchical spatial-temporal features-based intrusion detection system (HAST-IDS). Long term memory networks are utilized to learn unquestionable level short lived properties, and profound convolutional neural networkss (CNNs) are utilized to learn low-level spatial elements of business traffic. The entire course of learning features is done normally by deep neural networks There is no prerequisite for incorporate planning philosophies. As an outcome of the improved traffic qualities, the FAR has been really decreased. The standard enlightening files DARPA1998 and ISCX2012 are utilized to survey the proposed structure's show. The principal discoveries propose that the HAST-IDS beats recently scattered procedures as far as precision, revelation rate, and FAR, showing its utility for feature learning and FAR decrease.

Data security analysis for DDoS defense of cloud based networks:

It has been exhibited that conveyed figuring is an incredible technique for expanding an association's abilities without requiring extra assets. In this sense, appropriated registering assists a foundation with further developing its IT capacities. It's vital to take note of that circulated registering is currently a fundamental piece of the IT business area, which is developing rapidly. It is viewed as a novel and productive technique for growing business. Concerns about the safety of sensitive data from both internal and external cyber threats have grown as more businesses and individuals begin to use the cloud to host their apps and data. Regardless of far reaching interest in distributed computing, security concerns keep numerous clients from relocating significant information there. Security is a major issue since programmers love to get at a ton of an association's information. In the event that these issues aren't fixed, the development of conveyed registering will keep on being hampered. Thus, this review gives a spic and span test and knowledge into a honeypot. It very well may be separated into two classifications: overseeing and investigating honeypots Certifiable dangers can be diminished by taking care of honeypots. An examination device called an exploration honeypot is utilized to dissect and track down internet based dangers. Subsequently, the essential goal of this

exploration project is to direct a thorough examination concerning the association's security by baiting a foe and giving a refined technique to observing their way of behaving.

3.METHODOLOGY

Because of the Internet of Things (IoT), the fast development of PC organizations, and the tremendous number of vital applications, cyber security has as of late gotten a ton of consideration in contemporary security challenges. Consequently, it is becoming increasingly important to develop a successful interruption detection framework that is essential to current security and identifies various hacks or anomalies within an organization. Be that as it may, benchmark datasets utilized in various past examinations are off base and can't be applied to this present reality because of lacking qualities. An embraced gaining model should be assessed with data got from this present reality to bypass these limits. To summarise, although using an inconsistency-based technique to identify network disruptions may aid in the detection of new digital threats, it may also result in a huge number of false alarms.

Disadvantages:

- Cyberattacks are the cause of data leaks.
- It is less protected from cyberattacks.

For cyber-threat detection, we describe an artificial neural network-based AI strategy. A depp learning-based area framework is utilized in the proposed development to change countless recorded security occasions into particular occasion profiles, coming about in improved digital danger ID evidence. To make a computer based intelligence SIEM system for this venture, we consolidated occasion profiling for data preprocessing with artificial neural network strategies like FCNN, CNN, and LSTM. Since the strategy centers around recognizing genuine and misleading positive signs, security specialists can answer computerized dangers all the more rapidly. Our examinations used SVM, k-NN, RF, NB, DT, Xgboost, and Adaboost, the five most normal ML calculations. Thus, the exploratory discoveries of this study show that the methodologies we propose can be utilized with the learning-based models for network interference location we give.

Advantages:

- When used in the real world, it performs better than traditional machine-learning techniques.
- The data in this project is protected from an attacker.

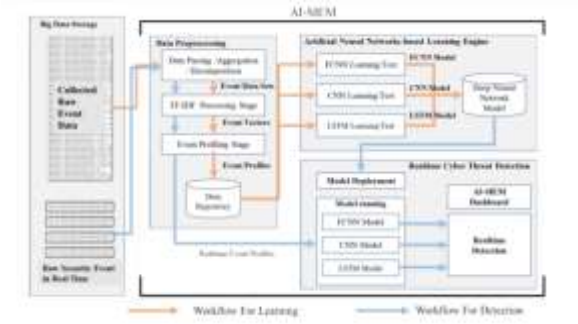

Fig.2: System architecture

The work process and engineering of the AI based SIEM framework that was created. There are three phases to the artificial intelligence SIEM framework: information readiness, danger distinguishing proof continuously, and a learning motor in light of artificial neural networks (ANN) Occasion profiling, the framework's underlying preprocessing step, attempts to change over crude information into brief contributions for different profound brain organizations. All through the information planning stage, the artificial intelligence SIEM framework performs occasion profiling, information standardization with the TF-IDF strategy, and information

conglomeration with parsing consecutively. As displayed in Figure, occasion informational collections, occasion vectors, and occasion profiles are made at each step and utilized in ensuing stages. This stage goes before the information learning stage as well as the exchange of crude security occasions to the info information of the deep learning motor when the framework identifies network breaks continuously. Three artificial neural networks are utilized in the displaying of the subsequent simulated intelligence based learning motor. For the information learning step, the preprocessed information are taken care of into the three counterfeit brain organizations, and each ANN figures out how to see as the most reliable model. To wrap things up, continuously danger recognition, each ANN model purposes the prepared model to arrange every security crude occasion consequently. The dashboard shows security analysts only genuine warnings, removing false ones.

MODULES:

The modules that make up propose algorithms are listed below.

1) Parsing Data: In order to produce a raw data event model, this module parses an input dataset.
2) TF-IDF: This module will be used to transform raw data into an event vector with normal and attack signatures.
3) Stage of Event Profiling: The processed data will be divided into train and test models based on profiling events.
4) The Model of a Deep Learning Neural Network: Using train and test data, this module constructs a training model with the help of CNN and LSTM algorithms. The trained model will be used to produce the prediction score, recall, precision, and FMeasure for the test data. When an algorithm learns correctly, it produces results with greater accuracy, and that model is chosen to be used for attack detection on a real-world system.

Due to the size of the datasets we are evaluating, we will encounter an out-of-memory issue while developing the model. Despite this, the kdd train.csv dataset functions properly; however, executing each method will take between 5 and 10 minutes. Reduce the size of the remaining datasets or run them on a machine with a high configuration to test them.

4. IMPLEMENTATION

Naive Bayes Classifier:

The classification strategy known as Naive Bayes is based on the idea that each characteristic stands alone and is unrelated to any other. It states that the status of other features in a class is unaffected by the state of one feature. It is regarded as a robust classification method because it is based on conditional probability. It works well with data with missing values and imbalances. The Bayes Theorem is used by the ML classifier Naive Bayes. The Bayes theorem can be utilized to compute the back likelihood P(C|X) for P(C), P(X), and P(X|C).

P(C|X) = (P(X|C) P(C))/P (X), where P(C|X) is the objective class' back probability.

The probability of the indicator class is shown by P(X|C).

The probability that class C is right is addressed by P(C).

The marker's prior probability is specified by P(X).

Decision Tree Classifier:

A regulated ML approach to order issues is Decision Tree. Using a decision rule derived from previous data,

this study aims to use Decision Tree to evaluate the objective class.Using internodes and nodes, it predicts and categorizes. Root hubs sort cases in view of their properties. The leaf node represents categorization, whereas the root node may have two or more branches. At each level, the Choice Tree chooses a hub in view of the greatest data gain among all characteristics. The Decision Tree method's efficacy was assessed.

Support Vector Machine (SVM):

In characterization, the managed ML model SVM is usually utilized. In a two-class planning test, the reason for a support vector machine is to find the ideal hyperplane with the greatest edge of parcel between two classes. For better speculation, the hyperplane shouldn't be closer to the other class's interesting data. Choose a hyperplane that is far away from each class's relevant data. The help vectors are the areas that are closest to the edge of the classifier. The examination's correctness is evaluated using the WEKA connection point. By increasing the distance between the two choices, the SVM selects the best division hyperplane. There will be an increase in the numerical distance between the hyperplane linked to wT x + b = 1 and the hyperplane comparable to wT x + b = - 1. The length of this distance is 2 w. This suggests that instead of solving for maximum 2 w, we should solve for min w | 2. All x(i) should be properly classified by the SVM, implying that yi (wT xi + b) is greater than 1, i, N. The SVM algorithm's ability to predict diabetes was tested.

K nearest neighbor algorithm:

In ML, K-Nearest Neighbors is a clear however pivotal grouping technique. Design acknowledgment, information mining, and interruption location all utilize this regulated learning calculation.

Since it is non-parametric, it can be used in a lot of real-world situations because it doesn't make any assumptions about how the data will be distributed (unlike other algorithms like GMM, which assume that the given data will be spread out Gaussian).

Some previous data, also known as training data, divides locations into groups according to a characteristic.

Xgboost:

Popular and effective is the open-source version of the gradient boosted trees method known as XGBoost (eXtreme Gradient Boosting). Gradient supporting is a technique for regulated discovering that attempts to anticipate an objective variable by joining gauges from an assortment of easier and more fragile models accurately.

Adaboost:

Each machine learning calculation might profit from the use of AdaBoost. It functions admirably with slow students. For a characterization task, these are models that arrive at exactness somewhat above irregular possibility. Random trees with one level are the most appropriate and thus most frequently utilized calculation with AdaBoost.

5. EXPERIMENTAL RESULTS

Fig.4: Home screen

Fig.5: Upload train dataset

Fig.6: Preprocessing TF-IDF algorithm

Fig.7: SVM algorithm

Fig.8: Random forest algorithm

Fig.9: Run Decision tree algorithm

Fig.10: Accuracy comparison graph

6. CONCLUSION

The AI-SIEM framework, which utilizes artificial neural networks and occasion profiles, is the subject of this review. The incorporation of significant learning-based area computations into event profiles and the utilization of extensive computerized risk acknowledgment capabilities are the novel aspects of our investigation. The artificial intelligence SIEM framework empowers security investigators to answer rapidly and successfully with basic security alerts by looking at long haul security information. By lessening misleading positive admonitions, it might likewise assist security experts with answering rapidly to digital dangers spread across countless security occasions. We utilized two benchmark datasets and two genuine world datasets (NSLKDD, CICIDS2017) to evaluate execution. Second, by standing out our frameworks from different methodologies and utilizing openly available benchmark datasets, we showed

the way that they can be utilized as one of the learning-based models for distinguishing network interferences. Second, our framework beat existing ML techniques regarding exact requesting, as shown by our investigation of two genuine world datasets.

7. FUTURE WORK

We will focus on further developing early danger estimates by utilizing a multi-deep learning way to deal with recognizing long haul patterns in verifiable information to resolve the developing issue of cyberattacks later on. Moreover, to work on the exactness of named datasets for coordinated learning and create reasonable learning datasets, a gathering of SOC inspectors might record individual characteristics of unrefined security events over various months.

8. ACKNOWLEDGMENTS

We thank CMR Technical Campus for supporting this paper titled " Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles", which provided good facilities and support to accomplish our work. Sincerely thank our Chairman, Director, Deans, Head Of the Department, Department Of Computer Science and Engineering, Guide and Teaching and Non- Teaching faculty members for giving valuable suggestions and guidance in every aspect of our work

REFERENCES

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

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

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

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

[5] S. Sandeep Sekharan, K. Kandasamy, "Profiling SIEM tools andcorrelation engines for security analytics," *In Proc. Int. Conf.Wireless Com., Signal Proce. and Net.(WiSPNET)*, 2017, pp. 717-721.

[6] N.Hubballiand V.Suryanarayanan,''False alarm minimizationtechniques in signature-based intrusion detection systems: Asurvey,'' *Comput. Commun.*, vol. 49, pp. 1-17, Aug. 2014.

[7] A. Naser, M. A. Majid, M. F. Zolkipli and S. Anwar, "Trusting cloudcomputing for personal files," *2014 International Conference onInformation and Communication Technology Convergence (ICTC)*,Busan, 2014, pp. 488-489.

[8] Y. Shen, E. Mariconti, P. Vervier, and Gianluca Stringhini, "Tiresias:Predicting Security Events Through Deep Learning," *In Proc. ACMCCS 18*, Toronto, Canada, 2018, pp. 592-605.

[9] Kyle Soska and Nicolas Christin, "Automatically detectingvulnerable websites before they turn malicious,", *In Proc. USENIXSecurity Symposium.*, San Diego, CA, USA, 2014, pp.625-640.

[10] K. Veeramachaneni, I. Arnaldo, V. Korrapati, C. Bassias, K. Li,"AI2: training a big data machine to defend," *In Proc. IEEEBigDataSecurity HPSC IDS*, New York, NY, USA, 2016, pp. 49-54

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

[12] I. Sharafaldin, A. H. Lashkari, A. A. Ghorbani, "Toward generatinga new intrusion detection dataset and intrusion trafficcharacterization", *Proc. Int. Conf. Inf. Syst. Secur. Privacy*, pp. 108-116, 2018.

[13] [online] Available: http://www.takakura.com/Kyoto_data/

[14] N. Shone, T. N. Ngoc, V. D. Phai and Q. Shi, "A deep learningapproach to network intrusion detection," *IEEE Trans. Emerg. TopicsComput. Intell.*, vol. 2, pp. 41-50, Feb. 2018

[15] R. Vinayakumar, MamounAlazab, K. P. Soman, P. Poornachandran,Ameer Al-Nemrat and Sitalakshmi Venkatraman, "Deep LearningApproach for Intelligent Intrusion Detection System," *IEEE Access*,vol. 7, pp. 41525-41550, Apr. 2019.