Innovative Approaches to Detect and Attribute Cyber-Attacks in IoT-Enabled Cyber–Physical Infrastructures

Mr.SYED.AKHTAR BASHA[1], ANNAPAREDDY MADHU BABU[2]

[1] Associate Professor,Dept of MCA, Audisankara college of Engineering and Technology (AUTONOMOUS),Gudur (M), Tirupati (Dt),AP

[2]PG Scholar, Dept of MCA,Audisankara college of Engineering and Technology (AUTONOMOUS)Gudur (M), Tirupati (Dt),AP

ABSTRACT: Securing cyber-physical systems (CPS), particularly those enabled by the Internet of Things (IoT), presents unique challenges due to the limitations of traditional security measures designed for general information/operational technology (IT/OT) systems. To address this, this article introduces a novel two-level ensemble attack detection and attribution framework tailored specifically for CPS, with a focus on industrial control systems (ICS). At the first level of the framework, a decision tree combined with an innovative ensemble deep representation-learning model is developed to detect attacks in unbalanced ICS environments effectively. Then, at the second level, an ensemble deep neural network is employed to attribute detected attacks, providing valuable insights for response and mitigation strategies. To validate the effectiveness of the proposed model, real-world datasets from water treatment systems and gas pipelines are utilized for evaluation. Results demonstrate that the suggested framework outperforms rival strategies with comparable computational complexity, highlighting its efficacy in enhancing security for IoT-enabled CPS environments

1.INTRODUCTION

Internet of Things (IoT) devices are increasingly being incorporated into cyber-physical systems (CPS), including vital infrastructure components like dams and power plants. In these settings, IoT devices, which are also referred to as Industrial IoT (IIoT), are frequently a part of an industrial control system (ICS), which is in charge of ensuring that the infrastructure functions properly. Distributed control systems (DCS), supervisory control and data acquisition (SCADA) systems, and systems with programmable logic controllers (PLCs) and Modbus protocols are all examples of ICS. However, the association of ICS or IIoT-based frameworks with public organizations increases their attack surfaces and the likelihood of being

targeted by digital attackers. One prominent example is the Stuxnet campaign, which reportedly targeted Iranian centrifuges for nuclear enrichment in 2010 and severely damaged the equipment [1], [2]. Another example is the 2011 episode about a siphon that caused a water plant in Illinois to be disappointed [3]. In 2015, BlackEnergy3, a separate campaign, targeted Ukraine's power grids, resulting in approximately 230 000 people losing power [4]. In addition, successful cyberattacks on three gas pipeline companies in the United States were reported in April 2018 [1], making electronic customer communication systems unavailable for several days. Security solutions designed for operational technology (OT) and information technology (IT) systems may not be applicable to ICS, despite their relative maturity. Because of the close integration between the controlled physical environment and the cyber systems, for instance, this might be the case. System-level security measures are therefore required for physical behavior analysis and system availability maintenance [1]. ICS security objectives are prioritized in that order, in contrast to the majority of IT/OT systems, which typically place an emphasis on availability, integrity, and confidentiality [5]. (Effective) digital attacks on ICS have the potential to have severe and even lethal consequences for the general public and our current situation due to the close coupling that exists between aspects of the criticism control circle and actual cycles. This demonstrates how crucial it is to create extremely robust safety and security measures for ICS intrusion detection and prevention [1]. It is common practice to use signature- and anomaly-based techniques to identify and attribute attacks. To overcome the known drawbacks of signature-based and anomaly-based detection and attribution methods, hybrid-based approaches have been proposed [6]. Although crossover-based methods are effective at identifying unusual acts, they are unreliable due to ongoing organizational changes that result in a variety of interruption identification framework (IDS) typologies. Beyond this, the primary component of conventional attack detection and attribution strategies is network metadata analysis, which includes things like IP addresses, transmission ports, traffic duration, and packet intervals. Consequently, interest in solutions for attack detection and attribution based on machine learning (ML) or deep neural networks (DNN) has recently resurfaced. Furthermore, there are network-based and have-based approaches to assault identification. The methods of supervised clustering, single-class or multiclass support vector machine (SVM),

fuzzy logic, artificial neural network (ANN), and DNN are frequently utilized for the purpose of detecting attacks in network traffic. These techniques analyze real-time traffic data to promptly identify malicious attacks. However, attack detection that only considers the host and network data may miss sophisticated attacks as well as insider threats. Unsupervised models that incorporate process/physical data have the potential to enhance the monitoring of a system because they do not necessitate an in-depth understanding of the cyberthreats. A sophisticated attacker with sufficient knowledge and time, such as a nation-state advanced persistent threat actor, can generally circumvent robust security measures. The majority of currently available methods also disregard the imbalanced property of ICS data by modeling only the normal behavior of a system and reporting deviations from normal behavior as anomalies. This may be due to limited assault tests in actual situations and existing informational collections. The trained model will not be able to recognize the patterns in the attack samples, despite the fact that using majority class samples is a good way to avoid issues brought on by data sets that are not evenly distributed. At the end of the day, such a method has a high rate of false positives and fails to identify subtle assaults [7]. As a result, efforts to use DL draws have been made, for example, to use computerized include (portrayal) to distinguish complex ideas from simpler ones [8] without relying on human-made highlights [9]. Inspired by the aforementioned observations, we present our novel two-stage ensemble deep-learning-based attack detection and attribution framework for imbalanced ICS data sets in this article. In the first stage, attacks in an unbalanced environment are detected with the help of an ensemble representation learning model and a decision tree (DT). During the second stage, several one-versus-all classifiers will join to form a larger DNN to classify the attack attributes with a confidence interval. In addition, the proposed framework is capable of locating attack samples that were not previously observed. A synopsis of our approach to this study is provided below. 1) A novel two-phase ensemble ICS attack detection method that can differentiate between known and unknown attacks is developed by us. In addition, we will demonstrate that the proposed approach outperforms the competition in terms of accuracy and f-measure. The proposed deep representation learning makes this method tolerant of data with imbalances. 2) We propose a novel self-tuning two-phase attack attribution method that ensembles

several deep one-versus-all classifiers using a DNN architecture to lower false alarm rates. The proposed approach makes it possible to accurately attribute attacks with high similarity. At the time of this examination, this is the first ML-based assault attribution strategy. 3) To demonstrate that, despite its superior performance, the proposed attack detection and attack attribution framework is comparable to other DNN-based methods described in the literature, we examine the computational complexity of the framework. The remainder of this article will be arranged as follows::.

2.LITERATURE SURVEY

2.1 Girish L, Rao SKN (2020) "Quantifying sensitivity and performance degradation of virtual machines using machine learning.",Journal of Computational and Theoretical Nanoscience , Volume 17, Numbers 9-10, September/October 2020, pp.4055-4060(6) https://doi.org/10.1166/jctn.2020.901

Virtualized data centers bring lot of benefits with respect to the reducing the high usage of physical hardware. But nowadays, as the usage of cloud infrastructures are rapidly increasing in all the fields to provide proper services on demand. In cloud data center, achieving efficient resource sharing between virtual machine and physical machines are very important. To achieve efficient resource sharing performance degradation of virtual machine and quantifying the sensitivity of virtual machine must be modeled, predicted correctly. In this work we use machine learning techniques like decision tree, *K* nearest neighbor and logistic regression to calculate the sensitivity of virtual machine. The dataset used for the experiment was collected using collected from open stack cloud environment. We execute two scenarios in this experiment to evaluate performance of the three mentioned classifiers based on precision, recall, sensitivity and specificity. We achieved good results using decision tree classifier with precision 88.8%, recall 80% and accuracy of 97.30%.

2.2 Madala, S. R., & Rajavarman, V. N. (2018). Efficient Outline Computation for Multi View Data Visualization on Big Data. International Journal of Pure and Applied Mathematics, 119(7), 745-755

In Big data analysis, representation of data in different views with respect to visualization for handling large scale data. Continuous parallel co-ordinate framework is effective data visualization tool to analyze each attribute without any change or update in their values, without change in

continues information structures and present data in structural orientation based on attributes to handle high amount of data. To present data in multi attribute evaluation, traditionally use Similarity Measure Centered with Multi Viewpoint (SMCMV) approach and related clustering approaches to represent data based on multi view data visualization procedure with different attributes. For multi dimensional and large scale data have different types of attributes to process and evaluate data based on different values in high amount of data. For efficient data processing to evaluate each attribute in separate manner to represent data in different factor with respect to returning of interest points in large scale data. So that in this paper, we present and develop novel Hybrid machine learning with sorting algorithm to evaluate data based on different attributes with respect to interest points from high amount of data. Sorting algorithm consists two basic steps in evolution of data, first step evaluates sorted positional index, second step exploits sorted positional index and then evaluate computational with selective and sequential data into table formation. Our implemented approach performs on real world UCI repository mostly used data sets with sorting to exploit results comparison of existing algorithms with respect to time, memory and table index evaluation for sorted data.

2.3 Vivek, T. V. S., Rajavarman, V. N., & Madala, S. R. (2020). Advanced graphical-based security approach to handle hard AI problems based on visual security. International Journal of Intelligent Enterprise, 7(1-3), 250-266

Security is the main aspect to explore human data from different web oriented applications present in artificial intelligence (AI). It is very difficult to use different web applications without security to access data in various places. So that various types of security related approaches were introduced to use services in securely in outside environment, but they have some limitations to protect data from outside attackers (hackers). So that in this paper, we propose and introduce a novel and advanced security model to provide security from outside attackers in AI related web oriented applications. In this approach, we follow the basic features related to Captcha as a graphical password to enable security services in our proposed approach. Using Captcha graphical passwords in our approach, we describe pushing attacks, pass-on attacks and guessing attacks in web applications with random selection of Captcha passwords to

use web services. Our experimental results show efficient security relations when compare to existing security approaches in terms of Captcha generation, time and other parameters present in web security applications

3.PROPOSEDWORK

1) We develop a novel two-phase ensemble ICS attack detection method capable of detecting both previously seen and unseen attacks. We will also demonstrate that the proposed method outperforms other competing approaches in terms of accuracy and f-measure. The proposed deep representation learning results in this method being robust to imbalanced data.

2) We propose a novel self-tuning two-phase attack attribution method that ensembles several deep one-versusall classifiers using a DNN architecture for reducing false alarm rates. The proposed method can accurately attribute attacks with high similarity. This is the first ML-based attack attribution method in ICS/IIoT at the time of this research.

3) We analyze the computational complexity of the proposed attack detection and attack attribution framework, demonstrating that despite its superior performance, its computational complexity is similar to that of other DNN-based methods in the literature.

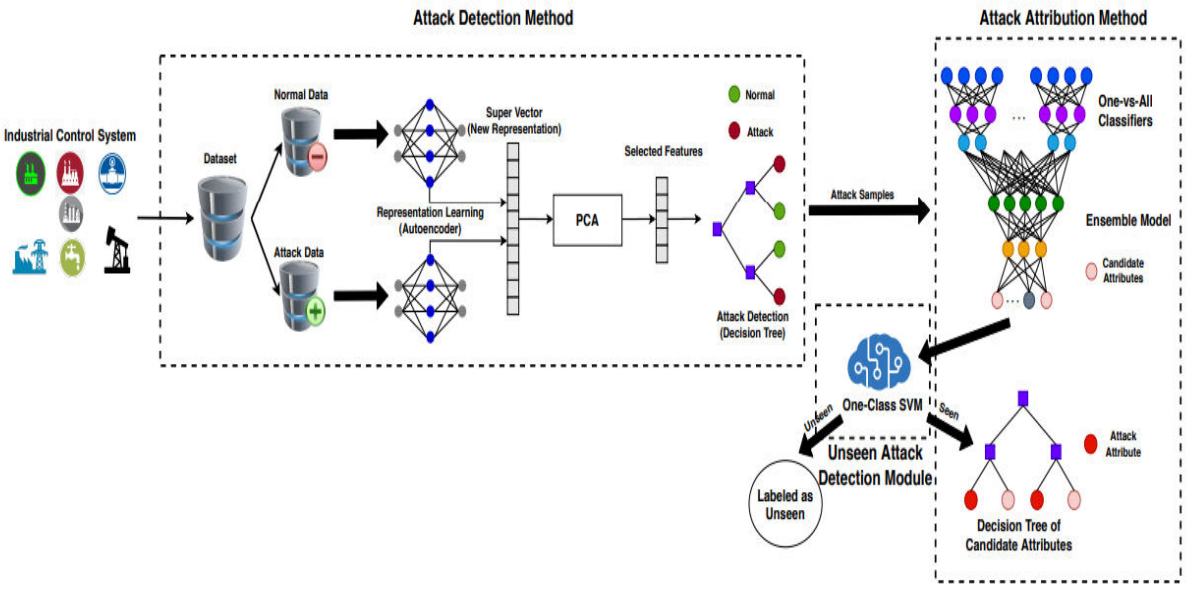


Fig 1:Working Architecture

3.1 IMPLEMENTATION

3.1.1 Upload SWAT Water Dataset: using this module we will upload dataset to application and then read dataset and then find different attacks found in dataset

3.1.2 Preprocess Dataset: using this module we will replace all missing values with 0 and then apply MIN-MAX scaling algorithm to normalized features values and then split dataset into train and test where application used 80% dataset for training and 20% for testing

3.1.3 Run AutoEncoder Algorithm: using this module we will trained AutoEncoder deep learning algorithm and then extract features from that model.

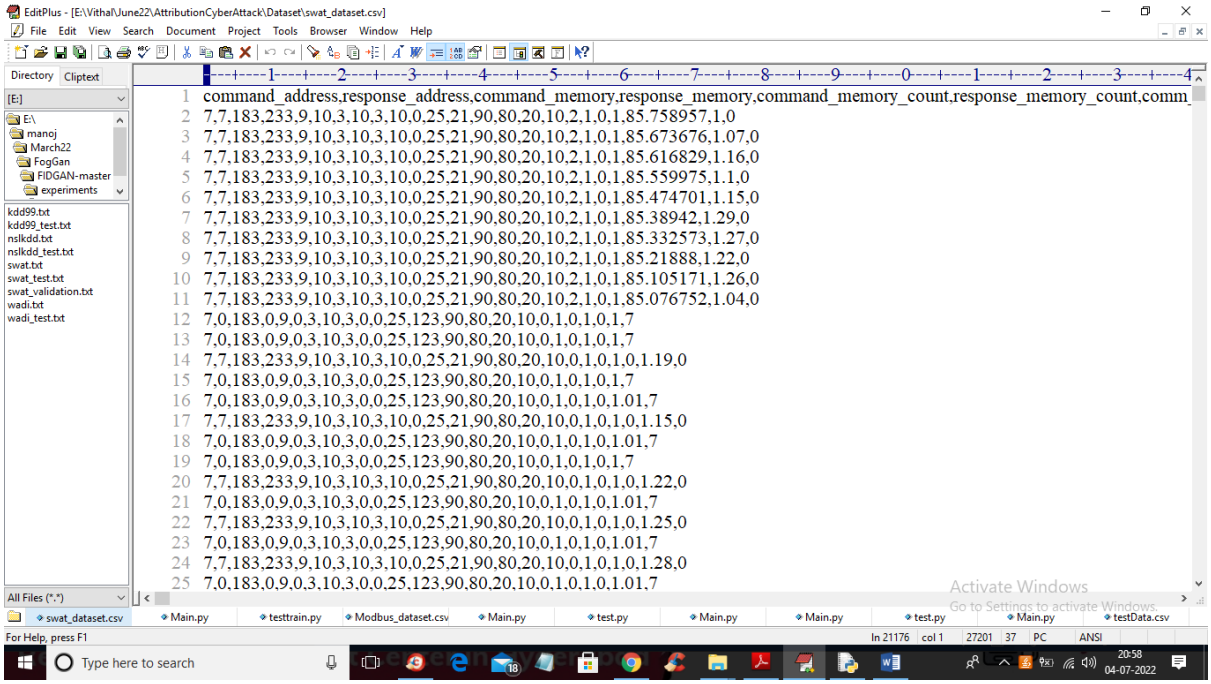
3.1.4 Run Decision Tree with PCA: extracted features from AutoEncoder will get transform using PCA to reduce features size and then retrain with Decision tree. Decision tree will predict label for each record based on dataset signatures

3.1.5 Run DNN Algorithm: predicted decision tree label will further train with DNN (deep neural network) algorithm to detect and attribute attacks

3.1.6 Detection & Attribute Attack Type: using this module we will upload unknown or un-label TEST DATA and then DNN will predict attack type

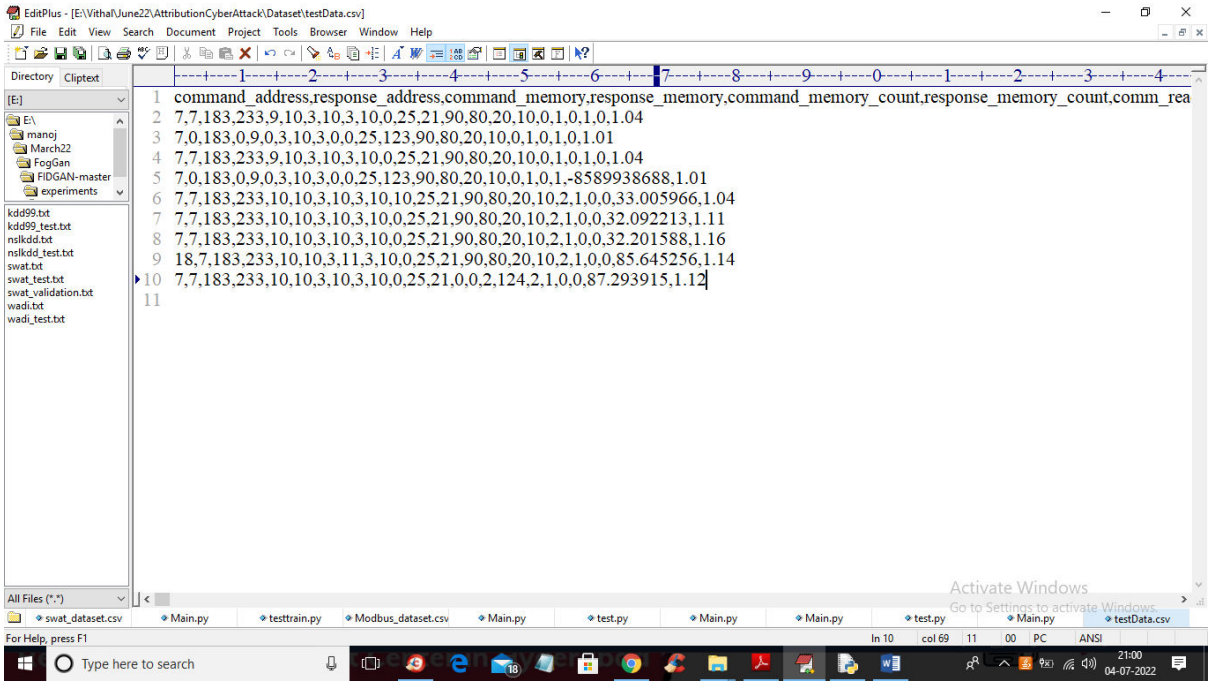
3.1.7 Comparison Graph: using this module we will plot comparison graph between all algorithms

3.2 ABOUT DATASET



In above dataset screen first row contains dataset column names and remaining rows contains dataset values and in last column we have attack type from label 0 to 7. We will used above dataset to train propose Auto Encoder, decision tree and DNN algorithms.

In below screen we are using NEW test data which contains only signature and there is no class label and propose algorithm will detect and attribute class labels.



In above test data we have IOT request signature without class labels.

4.RESULTS AND DISCUSSION

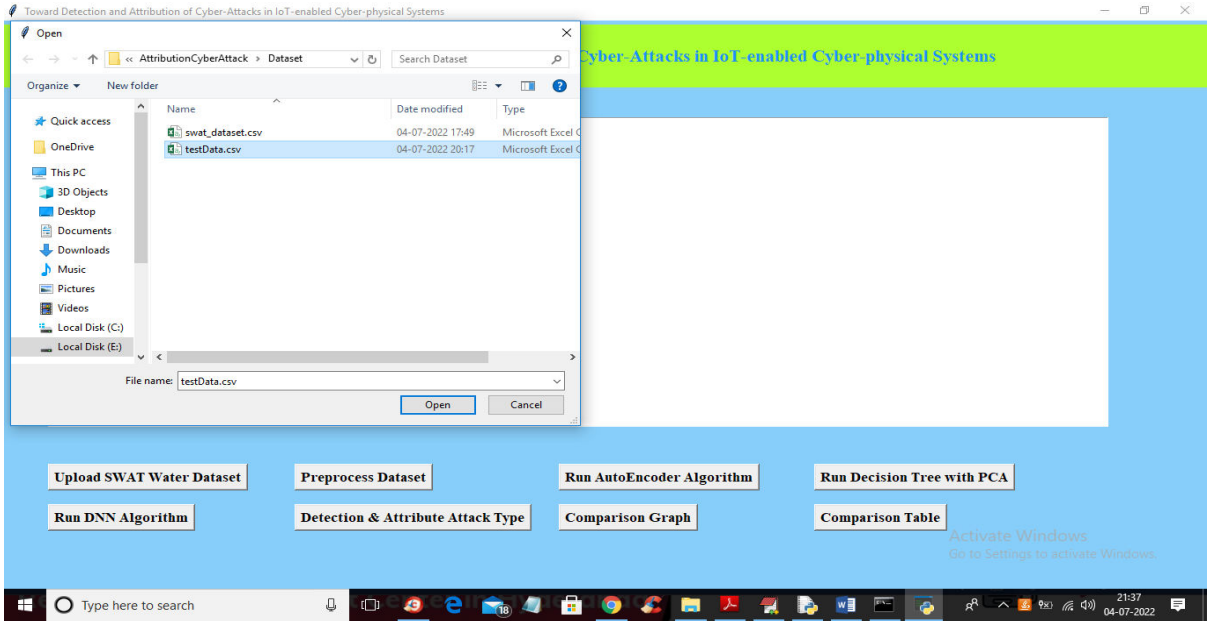


Fig 2:In above screen selecting and uploading 'TEST DATA' file and then click on 'Open' button to get below output

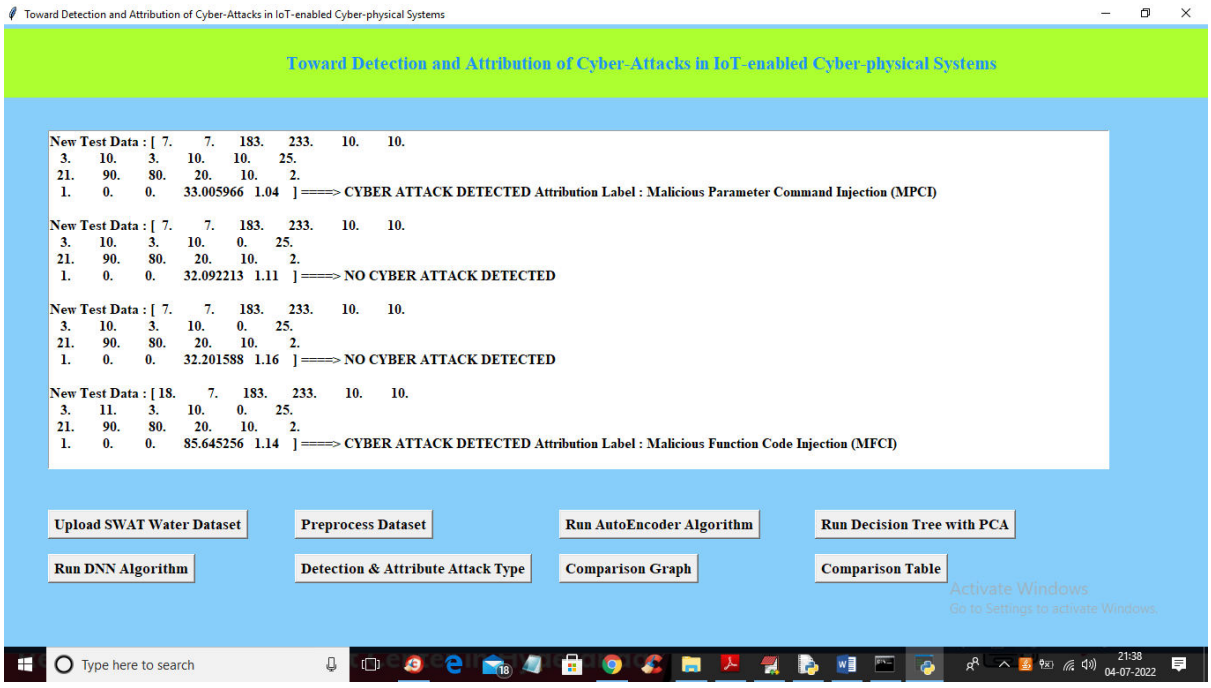


Fig 3:In above screen we can see detected various attacks and now click on 'Comparison Graph' button to get below graph

Fig 4:In above graph x-axis represents algorithms names and y-axis represents different metric values such as precision, recall, accuracy and FSCORE with different colour bars and in all algorithms DNN got high accuracy and now close above graph and then click on 'Comparison Table' to get below comparison table of all algorithms

5.CONCLUSION

Cyber-physical systems (CPS) with Internet of Things (IoT) capabilities might be challenging to secure since security techniques built for general information/operational technology (IT/OT) systems may not be as effective in a CPS context. This article presents a two-level ensemble attack detection and attribution framework for CPS, specifically an industrial control system (ICS). A decision tree is paired with a novel ensemble deep representation-learning model to detect attacks in unbalanced ICS environments at the first level. An ensemble deep neural network is built to aid in assault attribution at the second level. Real-world data from water treatment systems and gas pipelines are utilized to assess the suggested model. Results show that the suggested model performs better than other rival strategies with comparable computational complexity.

REFERENCES

[1] K. Graves, Ceh: Official certified ethical hacker review guide: Exam 312-50. John Wiley & Sons, 2007.

[2] R. Christopher, "Port scanning techniques and the defense against them," SANS Institute, 2001.

[3] M. Baykara, R. Das¸, and I. Karado ˘gan, “Bilgi g ¨uvenli ˘gi sistemlerinde kullanilan arac¸larin incelenmesi,” in 1st International Symposium on Digital Forensics and Security (ISDFS13), 2013, pp. 231–239.

[4] S. Staniford, J. A. Hoagland, and J. M. McAlerney, “Practical automated detection of stealthy portscans,” Journal of Computer Security, vol. 10, no. 1-2, pp. 105–136, 2002.

[5] S. Robertson, E. V. Siegel, M. Miller, and S. J. Stolfo, “Surveillance detection in high bandwidth environments,” in DARPA Information Survivability Conference and Exposition, 2003. Proceedings, vol. 1. IEEE, 2003, pp. 130–138.

[6] K. Ibrahimi and M. Ouaddane, “Management of intrusion detection systems based-kdd99: Analysis with lda and pca,” in Wireless Networks and Mobile Communications (WINCOM), 2017 International Conference on. IEEE, 2017, pp. 1–6.

[7] N. Moustafa and J. Slay, “The significant features of the unsw-nb15 and the kdd99 data sets for network intrusion detection systems,” in Building Analysis Datasets and Gathering
Experience Returns for Security (BADGERS), 2015 4th International Workshop on. IEEE, 2015, pp. 25–31.

[8] L. Sun, T. Anthony, H. Z. Xia, J. Chen, X. Huang, and Y. Zhang, “Detection and classification of malicious patterns in network traffic using benford’s law,” in Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017. IEEE, 2017, pp. 864–872.

[9] S. M. Almansob and S. S. Lomte, “Addressing challenges for intrusion detection system using naive bayes and pca algorithm,” in Convergence in Technology (I2CT), 2017 2nd International Conference for. IEEE, 2017, pp. 565–568.

[10] M. C. Raja and M. M. A. Rabbani, “Combined analysis of support vector machine and principle component analysis for ids,” in IEEE International Conference on Communication and Electronics Systems, 2016, pp. 1–5

Author’s Profiles

Mr.SYED.AKHTAR BASHA currently he is working Associate professor in Audisankara college of Engineering and

Technology Gudur (M),Tirupati (Dt) he is done MCA from SANA INSTITUTIONS NELLORE 2005,M.Tech from QUBA COLLEGE OF ENGINEERING AND TECHNOLOGY 2012

ANNAPAREDDY MADHU BABU is pursuing MCA from Audisankara college of Engineering and Technology Gudur, Affiliated to JNTUA in 2024,Andhrapradesh, India.