International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Detecting the Cyber Attacks for Distributed Systems Using Machine Learning Algorithms

Geetha.T[1], Biridepalli Mounisha[2], Jasthi Jyothi[3], Karumudi Meghana[3], Kavuri Swathi[3]

[1]Assistant professor, Department of Computer Science and Engineering, Dhanalakshmi Srinivasan Engineering College, Perambalur

[2]Student, Department of Computer Science and Engineering, Dhanalakshmi Srinivasan Engineering College, Perambalur

[3]Student, Department of Computer Science and Engineering, Dhanalakshmi Srinivasan Engineering College, Perambalur

ABSTRACT

Cyber-physical system security for electric distribution systems is critical. In direct switching attacks, often coordinated, attackers seek to toggle remote-controlled switches in the distribution network. Due to the typically radial operation, certain configurations may lead to outages and/or voltage violations. Existing optimization methods that model the interactions between the attacker and the power system operator (defender) assume knowledge of the attacker's parameters. This reduces their usability. Furthermore, the trend with coordinated cyber attack detection has been the use of centralized mechanisms, correlating data from dispersed security systems. This can be prone to single point failures. In this paper, novel mathematical models are presented for the attacker and the defender. The models do not assume any knowledge of the attacker's parameters by the defender. Instead, a machine learning (ML) technique implemented by a multi-agent system correlates detected attacks in a decentralized manner, predicting the targets of the attacker. Furthermore, agents learn optimal mitigation of the communication level through Q-learning. The learned attacker motive is also used by the defender to determine a new configuration of the distribution network. Simulations of the technique have been performed using the IEEE 123-Node Test Feeder. The simulation results validate the capability and performance of the algorithm."Detecting Cyber Attacks in Distributed Systems Using Machine Learning Algorithm" presents a solution to the problem of identifying cyber attacks in distributed systems. As distributed systems are becoming increasingly complex, traditional methods of detecting cyber attacks have become insufficient. In this paper, the authors propose a machine learning-based approach that can identify cyber attacks in

distributed systems. The approach involves collecting data from various sources, preprocessing the data, and using machine learning algorithms to classify the data as normal or an attack. The proposed approach was evaluated on a real-world datasets and showed promising results. The findings of this paper have implications for improving the security of distributed systems, which are critical for the functioning of many organizations.

Keywords: the IEEE 123-Node Test Feeder , Logistic Regression.

I.INTRODUCTION

With the integration of advanced communication technology, the power grid is increasingly remotely monitored and controlled. Nevertheless, the advancement has also made the smart grid more vulnerable to cyber attacks. In December 2015, six distribution utilities in Ukraine suffered cyber attacks. The ensuing outage affected about 225,000 customers [1]. Significant research has been conducted in the area of distribution system cyber security, and several techniques have been proposed for different applications.

Cyber attacks on distributed systems have become increasingly sophisticated and difficult to detect. Machine learning algorithms have emerged as a promising approach to improving the accuracy and speed of cyber attack detection in distributed systems. In this context, machine learning algorithms can be used to analyze large volumes of data generated by distributed systems, identify patterns and anomalies, and automatically detect cyber attacks.

The main challenge in using machine learning algorithms for cyber attack detection in distributed systems is the need to develop effective models that can accurately distinguish between normal and malicious network traffic. This requires the use of large and diverse datasets, as well as sophisticated algorithms that can adapt to changing attack patterns.

In recent years, there has been significant research in the area of machine learning-based cyber attack detection for distributed systems. Techniques such as deep learning, reinforcement learning, and ensemble learning have been applied to improve the accuracy and robustness of these models. Additionally, the use of feature selection and dimensional reduction techniques has been shown to improve the efficiency of these algorithms.

Overall, the use of machine learning algorithms for cyber attack detection in distributed systems holds great promise for improving the security of these systems. However, there are still challenges that need to be addressed, such as the need for high-quality datasets and the development of more robust and efficient algorithms.

II. RELATED WORK

In this section, we evaluated some of the research similar to our project that has been conducted by various authors and researchers utilizing Machine learning techniques to predict the cyberattacks on distributed system.

Kousik Barik,Sanjay Misra,Karabi Konar,Luis Fernandez-Sanz,Murat Koyuncu, they proposed the system Cyber attacks are increasing rapidly due to advanced digital technologies used by hackers. In addition, cyber criminals are conducting cyber attacks, making cyber security a rapidly growing field. Although machine learning techniques worked well in solving large-scale cyber security problems, an emerging concept of deep learning (DL) that caught on during this period caused information security specialists to improvise the result. The deep learning techniques analyzed in this study are convolution neural networks, recurrent neural networks, and deep neural networks in the context of cyber security.A framework is proposed, and a real-time laboratory setup is performed to capture network packets and examine this captured data using various DL techniques. A comparable interpretation is presented under the DL techniques with essential parameters, particularly accuracy, false alarm rate, precision, and detection rate. The DL techniques experimental output projects improvise the performance of various real-time cyber security applications on a real-time datasets. CNN model provides the highest accuracy of 90.64% with a precision of 90% with binary class. The RNN model offers the second-highest accuracy of 90.75%. CNN model provides the highest accuracy of 90.42 with multi class. The study shows that DL techniques can be effectively used in cyber security applications. Future research areas are being elaborated, including the potential research topics to improve several DL methodologies for cyber security applications.

Laraib Sana, Muhammad Mohsin Nazir, Muddesar Iqbal ,Lal Hussain ,Amjad Ali they proposed the system From past few years, the Internet of things (IOT) is an emerging and encouraging technology that has gained prominence in the industries. Due to its increasing usages, a huge amount of data are exchanged within IOT architecture using the internet, which is why privacy and cyber-security are major issues. The heterogeneous nature of various technologies that are combined using IOT makes it problematic to provide security using prescriptive networking. The future of secure IOT depends on privacy issues. The research intends to improve security mechanisms based on intrusion and anomaly detection for IOT using deep learning. In this context, a systematic literature review (SLR) is conducted to identify ‘How to perform data transformation analysis of IOT datasets to detect anomaly detection for cyber IOT attacks? The SLR result found 24 datasets used for IOT analysis, 35 performance metrics to evaluate IOT problems, 6–42 features identified for detection, 42 preprocessing techniques have been used for transforming data,and 26 different methods and models were used to process the given problem. The SLR highlights further

enhancement for the issue and identification of cyber-security in IOT. Anomaly detection can be done based on reinforcement deep learning after a thorough analysis of SLR.

Jun Zhang , Lei Pan ,Qing-Long Han ,Chao Chen ,Sheng Wen ,Yang Xiang they proposed the system With the booming of cyber attacks and cyber criminals against cyber-physical systems (CPSs), detecting these attacks remains challenging. It might be the worst of times, but it might be the best of times because of opportunities brought by machine learning (ML), in particular deep learning (DL). In general, DL delivers superior performance to ML because of its layered setting and its effective algorithm for extract useful information from training data. DL models are adopted quickly to cyber attacks against CPS systems. In this survey, a holistic view of recently proposed DL solutions is provided to cyber attack detection in the CPS context. A six-step DL driven methodology is provided to summarize and analyze the surveyed literature for applying DL methods to detect cyber attacks against CPS systems. The methodology includes CPS scenario analysis, cyber attack identification, ML problem formulation, DL model customization, data acquisition for training, and performance evaluation. The reviewed works indicate great potential to detect cyber attacks against CPS through DL modules. Moreover, excellent performance is achieved partly because of several high-quality datasets that are readily available for public use. Furthermore, challenges, opportunities, and research trends are pointed out for future research.

N.Vadivelan, K.Bhargavi, Sarangam kodati, M.Nalini they proposed the system Cyber security professionals pay greater regard to risk evaluation and propose techniques for mitigating. Throughout the area of cyber defense, designing successful strategies was a plan set. Machine learning also increasingly become an important concern in data protection although machine learning is successful in cyber defense. The rapid expansion in Cloud Computing, networking and evolutionary computation has been the result of unprecedented developments in computing, storage and computational technology. The planet is rapidly being digitized - there is a growing want of comprehensive and sophisticated information security and privacy issues And Strategies to fight security threats, which are becoming more complicated. Cyber terrorism is spreading worldwide using all kinds of computer weakness. Machine learning algorithms were used to address global computer security threats such as malware detection, ransom ware recognition, fraud detection and spoofing identification. It research analyzes how cyber training is used in defense as well as offence, providing details about cyber threats on machine learning techniques and The much more popular kinds of cyber security risks are evaluated using machine learning algorithms which describe how machine learning is used for computer defence such as the identification and avoidance of attacks, vulnerability scanning and recognition and public internet risk assessment.

Khalid Almullah he proposed the system Rapid shifting by government sectors and companies to provide their services and products over the internet,has immensely increased internet usage by individuals.Through extra-net to network services or corporate networks used for personal purposes. Computer hackers can lead to financial losses and manpower/time consumption.Therefore,it is vital to take all necessary measures to minimize losses by detecting attacks preemptively.Due to learning algorithms in cyberspace security challenges,deep learning-based cyber defense has lately become a hot topic.Penetration testing,malware categorization and identification ,spam filtering,and spoofing detection are just a few of the key concerns in cyber defense that were tackled using machine learning approaches.Result, effective adaptive approaches,such as machine learning approaches could result in increased response times,reduced probability of false alerts ,as well as cheaper computing and communication expenses.Our primary point is to demonstrate the problem of detecting malware is distinct from other technologies,making it far more difficult for the access control group to properly use machine learning.

III.DATASET

In this study, the model was trained using a hybrid dataset that contains the data src_bytes , dst_bytes , count , same_srv_rate , dst_ hst_srv_count , dst_host_same_srv_rate , dst_host_same_src_port_rate ,protocol_type , service , flags,etc . In this module the datasets will trained in the model by using the datasets which was directly loaded through the python code in backend as traindata.csv file. The total 10,88,304 data were loaded in the model as 25912 rows & 42 columns , By using the SVM model those data's will be trained and displays the normal and anomaly percentages of cyberattacks detected in the given datasets in the form of Pie-charts. The data which was trained in the model by using svm technique will be tested in this phase by using logistic regression for testing the data applying confusion matrix for detecting cyber attacks and classification report on detected anomaly and normal values for the accuracy report. Then the model will load totally 22544 rows & 41 columns of total 9,24,304 data's by eliminating the unwanted data from trained data's of train_data.csv by using the decision trees method and placing the new data in the test_data.csv. Then using Matploit-lib the detected cyber attacks normal and anomaly percentages will be displayed as a pie-charts.

IV. LOGISTIC ARCHITECTURE

Logistic regression is a supervised machine learning algorithm mainly used for classification tasks where the goal is to predict the probability that an instance of belonging to a given class or not. It is a kind of statistical algorithm, which analyse the relationship between a set of independent variables and the dependent binary variables. It is a powerful tool for decision-making. For example email spam or not.

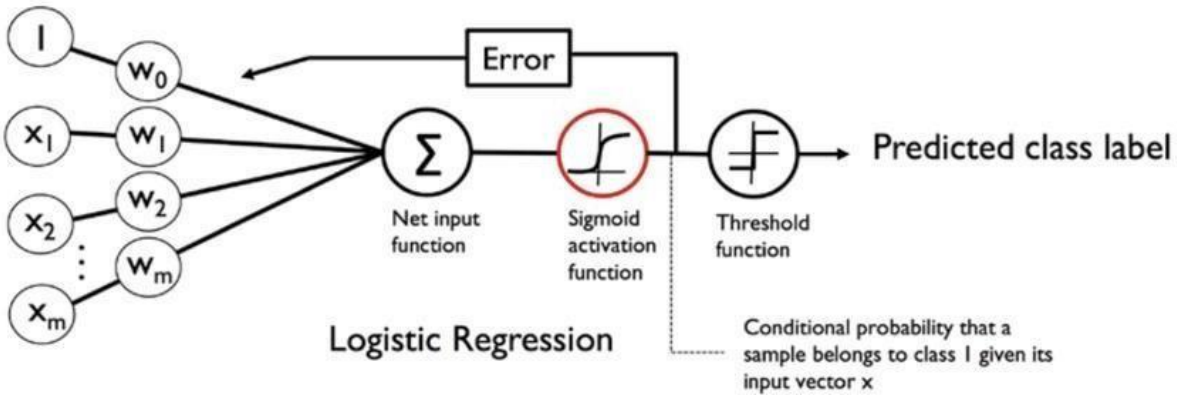

Figure 5: Architecture of LOGISTIC REGRESSION

V. METHODOLGY

In this section, we discussed the methods and modules of our proposed system logistic architecture, the project explain about the detection of cyber attacks for distributed systems using machine learning algorithms.Where This Module includes several steps.

- USER REGISTRATION
- USER LOGIN
- PROCESSING THE DATASETS
- DETECTING CYBER ATTACKS

USER REGISTRATION

In this module by using the web application , The user need to be register to web server by giving their details :

- Creating username
- Creating password
- Entering email
- Entering age
- Entering gender
- Entering phone number
- Entering address

By registering those user data the user will be registered to our server. By using the created username and password the user can login detect the cyber attacks in the next modules.

USER LOGIN

In this module the user will login to our server, By entering the registered username and password created in registration module . Then the home page of the user account will be opened. In this page by using the predict model option which was created in it can train all datasets in the model.

PROCESSING THE DATASETS

- In this module the datasets will trained in the model by using the datasets which was directly loaded through the python code in back-end as traindata.cs v file.
- It contains the data src_bytes , dst_bytes , count , same_srv_rate , dst_ hst_srv_count , dst_host_same_srv_rate , dst_host_same_src_port_rate ,protocol_type , service , flags,etc.
- The total 10,88,304 data were loaded in the model as 25912 rows & 42 columns.
- By using the SVM model those data will be trained and displays the normal and anomaly percentages ofCyberattacks detected in the given datasets in the form of pie-chart.

DETECTING CYBER ATTACKS

- In this module the data which was trained in the model by using SVM technique it will be tested in this phase by using logistic regression for testing the data applying confusion matrix for detecting Cyber attacks and classification report on detected anomaly and normal values for the accuracy report.
- It will load totally 22544 rows & 41 columns of total 9,24,304 data by eliminating the unwanted data from trained dataof train_data.cs v by using the decision trees method and placing the new data in the test_data.csv

- Then using mat-plot_lib the detected Cyber attacks normal and anomaly percentages will be displayed as a pie-charts.

VI. RESULT AND DISCUSSION

The pre-trained model that we employed in this study was imported from pre-proccesor . The lakhs of dataset were used to pre-train this model. The average accuracy using the modified proposed method is 96.08% using the logistic regression. The corresponding precision, recall, specificity and F1-score were 0.9620,0.9617,09921 and 0.9616 respectively.

VII. CONCLUSION AND FUTURE WORK

This paper presents a decentralized attack correlation technique and a hybrid mitigation. Compared to interdiction models in the literature, this work assumes no explicit knowledge of the attacker's parameters by the defenders, which in this case, are agents. The targets of an attack are predicted in a decentralized manner using a learning mechanism, and new NIDS thresholds optimally found from reinforcement learning are applied. When enough alerts are received, physical mitigation is triggered. The proposed technique is also superior as it is not prone to single point failures; should the central agent be compromised, communication level mitigation is still enforced by the dispersed agents. Currently, the NIDS implemented by the algorithm is anomaly-based and makes use of only communication level thresholds. It is therefore limited to only man-in-the-middle attacks.

References

[1] A. Gusrialdi and Z. Qu, ‘‘Smart grid security: Attacks and defenses,’’ in Smart Grid Control (Power Electronics and Power Systems), 1st ed. Cham, Switzerland: Springer, 2018

[2] C. Moya and J. Wang, ‘‘Developing correlation indices to identify coordinated cyber-attacks on power grids,’’ IET Cyber-Phys. Syst., Theory Appl., vol. 3, no. 4, pp. 178–186, Dec. 2018.

[3] Electricity Information Sharing and Analysis Center (E-ISAC). (Mar. 2016). Analysis of the Cyber Attack on the Ukrainian Power Grid, Electricity Information Sharing and Analysis Center (E-ISAC), [Online]. Available: https://nsarchive.gwu.edu/sites/default/files/ documents/3891751/SANS-and-Electricity-Information-Sharing-and.pdf

[4] H. Zhang, B. Liu, and H. Wu, ‘‘Smart grid cyber-physical attack and defense: A review,’’ IEEE Access, vol. 9, pp. 29641–29659, 2021.

[5] I.-S. Choi, J. Hong, and T.-W. Kim, ‘‘Multi-agent based cyber attack detection and mitigationfor distribution automation system,’’ IEEE Access, vol. 8, pp. 183495–183504, 2020.

[6] J. Appiah-Kubi and C.-C. Liu, ‘‘Decentralized intrusion prevention (DIP) against co-ordinated cyberattacks on distribution automation systems,’’ IEEE Open Access J. Power Energy, vol. 7, pp. 389–402, 2020

[7] K. Lai, M. Illindala, and K. Subramaniam, ‘‘A tri-level optimization model to mitigate coordinated attacks on electric power systems in a cyberphysical environment,’’ Appl. Energy, vol. 235, pp. 204–218, Feb. 2019.

[8] R. Deng, P. Zhuang, and H. Liang, ‘‘False data injection attacks against state estimation in power distribution systems,’’ IEEE Trans. Smart Grid, vol. 10, no. 3, pp. 2871–2881, May 2019.

[9] S. Lakshminarayana, J. Ospina, and C. Konstantinou, ‘‘Load-altering attacks against power grids under COVID-19 low-inertia conditions,’’ IEEE Open Access J. Power Energy, vol. 9, pp. 226–240, 2022.

[10] Y. Lin and Z. Bie, ‘‘Tri-level optimal hardening plan for a resilient distribution system considering reconfiguration and DG islanding,’’ Appl. Energy, vol. 210, pp. 1266–1279, Jan. 2018.