Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 2, August 2023
Paper : 31

INTRUSION DETECTION ON SMART GRID USING ENSEMBLE LEARNING

C. Kudiyarasudevi
Department of Computer Science Engineering
SRM Institute of Science and Technology
Ramapuram Chennai-600089,India
kudiyarc@srmist.edu.in

Merish Rohith B
Department of Computer Science Engineering
SRM Institute of Science and Technology
Ramapuram Chennai-600089,India
mb3663@srmist.edu.in

Dhanusri R
Department of Computer Science Engineering
SRM Institute of Science and Technology
Ramapuram Chennai-600089,India
dr3677@srmist.edu.in

Nithyan N
Department of Computer Science Engineering
SRM Institute of Science and Technology
Ramapuram Chennai-600089,India
nn9708@srmist.edu.in

Abstract - **One of the most unpredictable systems currently available to consumers throughout the world in the field of cyber-physical systems is smart grid. The main objective of smart grids is to use distribution and transmission frameworks to guarantee a supply of power from generators to end customers. While still susceptible to various forms of cyber attacks, smart grid technology has increased the capacity of conventional electricity grids. By exploiting these flaws, attackers can gain visibility into the smart grid network network, compromising integrity and security. In a smart grid context, a (IDS) thus becomes a crucial tool for offering secure and trustworthy services. The previously proposed system had little difficulties in deciding where the detector should be placed in our real-world environment..**

Index Terms— **Smart Grid, Cyber-physical systems, Ensemble Learning, Network traffic.**

I.INTRODUCTION

Transformers, circuit breakers, and switches are just a few of the parts that make up the power distribution system, which is in charge of providing electricity to consumers. The distribution system requires enhanced protection coordination between protective devices and automatic fault recovery through the use of an operational data acquisition system in order to maintain a stable and dependable power source (DAS). The development of recovery strategies and intrusion detection systems using fuzzy theory and decision theory has been studied in order to increase distribution reliability and defend against cyberattacks (IDS).

Anomaly-based and signature-based IDS are the two main types used. While signature-based IDS uses predefined signatures to find malware, anomaly-based IDS detects attacks by comparing current traffic to previously recorded normal traffic. Both IDS systems, however, have drawbacks, including the likelihood of new attacks going undetected for signature-based IDS and a high number of false positive warnings for anomaly-based IDS. Therefore, it is necessary to create a new system that can boost new attack detection rates and lower false alarm rates for predefined signatures.

Smart grids, which use information and communication technology to provide power on demand to end users and distribute it to centres, are being developed as a solution to these problems. Smart grid security, however, is a crucial issue because cyber-physical systems are susceptible to attacks from viruses, Trojans, and spam. Intrusion detection systems (IDS) are used to find, examine, and stop security breaches in the network in order to defend against these attacks.

IDS can be divided into two categories: network-based IDS and storage IDS. Network-based IDS examines and analyses network traffic in real-time, whereas storage IDS relies on previously discovered malware that has been stored in the database to identify attacks. IDS can be divided into anomaly-based and signature-based systems, respectively.

While signature-based IDS uses predefined signatures to find malware, anomaly-based IDS detects attacks by comparing current traffic to previously recorded normal traffic. The high rate of false positive alerts for anomaly-based IDS and the potential for new attacks to go undetected for signature-based IDS, respectively, are both shortcomings of these systems.

To overcome these limitations, a new system must be developed that can improve the detection rate of new attacks (zero-day malware) and reduce false alarm rates in predefined signatures. This requires the integration of advanced machine learning algorithms and data analytics techniques to enhance the performance of IDS. In addition, continuous research is needed to identify and address new security threats and vulnerabilities in smart grid systems, and to develop new strategies to protect against them.

Overall, the power distribution system and smart grid technology are critical components of the modern energy infrastructure. To ensure their reliable and secure operation, it is essential to continuously improve and update the protection and security measures, and to develop new technologies that can effectively detect and prevent cyber-attacks.

II. RELATED WORKS

The current method focuses mainly on transferring machine learning models between two specific situations, and determining whether differences in data require the use of a transfer learning method. To enhance distribution reliability and protect against cyber-attacks, a research study has been conducted on the use of fuzzy theory and decision theory to develop recovery strategies and intrusion detection systems (IDS).This paper suggests a novel method for analysing transfer learning portability in cyber-physical systems (CPS) by comparing the data distribution patterns of the source and target domains in order to overcome this limitation. The suggested method employs three distinct metrics to gauge these variations and trains two regression models to approximatively determine the relationship between the divergence and the model's accuracy. The target domains that call for transfer learning are then identified, and the models forecast the accuracy loss in unlabeled target regions. To maintain robust detection accuracy, the proposed approach uses Domain Adversarial Neural Networks as transfer learning models. The study uses false data injection (FDI) attacks to generate attack data with temporal, spatial, and spatiotemporal variations. Transfer learning is used as part of this process to evaluate the need for CPS monitoring security and intrusion detection in the smart grid. The proposed model evaluates accuracy drop and divergence in dynamic CPS operations using multiple metrics. The results show that transfer learning can be used to improve accuracy in situations where there is divergence due to attacks or changes in the CPS environment.

This research addresses the fundamental question of when transfer learning should be applied in machine learning-based intrusion detection in CPS. The suggested method offers insights into the relationship between divergence and accuracy and offers a systematic and thorough study of detector accuracy drop and divergence in various temporal, spatial, and spatiotemporal experiments. Using domain adversarial neural networks as transfer learning models, the proposed method maintains detection accuracy while predicting accuracy loss due to divergence with high precision. Supply chain security is becoming an increasingly important issue for cyber-physical power systems, but the NIST framework does not provide specific guidance on how to address this. The framework only recommends assessing supply chain risk and implementing appropriate controls. Additionally, although the framework recommends organizations have an incident response plan in place, it does not provide specific guidance on how to develop and implement such a plan for cyber-physical power systems. Incident response plans should take into consideration the unique characteristics of cyber-physical systems, such as the potential for physical damage or injury. The NIST framework is not fully integrated with other cyber security frameworks and standards, leading to inconsistencies in cyber security. It is essential to address these issues and improve the cyber-physical power system to ensure a safe and secure environment for all users.

III. PROPOSED METHODOLOGY

The system's approach to intrusion detection is based on the fact that many intrusion events have unique characteristics that can be identified through the analysis of network traffic data. By leveraging machine learning techniques, the system can automatically learn to identify these characteristics and detect malicious events in real-time.

Ensemble learning algorithms have several advantages, it combines the strengths of multiple learning models, making it more accurate and robust than any individual model. This can be especially useful in a dynamic network environment where threats and attacks can vary in their complexity and frequency.

In addition, the system's approach to data preprocessing is critical for the accuracy and efficiency of the overall system. Machine learning algorithms can quickly analyze raw data after preprocessing it. To get the data ready for analysis, the system employs a number of preprocessing techniques, including feature selection, normalisation, and scaling.

Overall, the proposed system provides an effective solution for network intrusion detection via leveraging machine learning techniques and an optimized algorithm for linear compression. It can help network administrators and security experts to detect and respond to potential threats more quickly and efficiently, reducing the risk of data breaches and other security incidents.

IV. LITERATURE REVIEW

[1] Incorporating machine learning into gradient boosting features: This paper, The difficulties of protecting SCADA systems in smart grids from cyber attacks are discussed in " Incorporating machine learning into gradient boosting features," along with using machine learning classifiers together with feature engineering-based preprocessing, this framework combines intrusion detection with intrusion detection. Before using decision-tree based machine learning algorithms for classification, the proposed approach uses gradient-boosting in order to choose the most promising features from the power grid dataset, feature selection (GBFS) was used. The advantages of the suggested approach are highlighted in the article, including enhanced detection rate, execution speed, and optimised false positive rate (FPR) when compared to cutting-edge methods.

[2] Systems that learn rules and detect intrusions in smart grids: This paper, we discuss classification of intrusion detection systems (IDS) into misuse-based, anomaly-based, and specification-based IDS in " Systems that learn rules and detect intrusions in smart grids." The authors contend that specification-based IDS, particularly when using rule learning techniques, may prove to be the most promising detection engine for cyber-physical systems for critical infrastructures like smart grids that may be subject to sophisticated unknown attacks in the future.

[3] An algorithm combining recursive feature elimination and majority vote ensembles for detecting intrusions in SCADA-based grids: This paper, according to the study "An algorithm combining recursive feature elimination and majority vote ensembles for detecting intrusions in SCADA-based grids," Power grids are being transformed into smart grids (SG) by technological advancements. It offers a number of advantages, including improved service quality, increased reliability, and effective use of existing infrastructure and renewable energy sources. Our approach combines the RFE-XGBoost feature selection method with the majority vote ensemble method (Recursive Feature Elimination-Extreme Gradient Boosting). Nine heterogeneous classifiers are used in the majority vote ensemble method to forecast the output label.

[4] Unified Deep Learning Approach for Smart Grid Anomaly Detection and Classification: This paper, an intrusion detection system (IDS) called MENSA is presented in "Unified Deep Learning Approach for Smart Grid Anomaly Detection and Classification" and is meant for smart grid (SG) settings that employ the Modbus/Transmission Control Protocol (TCP) and Distributed Network Protocol 3 (DNP3) protocols. Because of the SG ecosystem's extensive connectivity and heterogeneity, there are serious privacy and cyber security risks that could affect other crucial infrastructures.

[5] Defending distribution automation systems against coordinated cyber attacks with a decentralized intrusion prevention system (DIP): This paper, A multi-agent system approach for intrusion prevention at the distribution level is presented in " Defending distribution automation systems against coordinated cyberattacks with a decentralized intrusion prevention system (DIP)," which also discusses the need for intrusion detection systems for smart grids. Simulations on the IEEE 13-Node Test Feeder verify the viability of the suggested approach to thwart cyber intrusions.

[6] An in-depth analysis of wireless intrusions: Several factors are seriously threatening wireless local area network security in this paper "An in-depth analysis of wireless intrusions: An intelligent mechanism based on deep learning" with the advancement of wireless network technologies (WLAN) and the rise of cyber attacks (Cybercrime). Traditional intrusion detection technology has been a hot topic of research for many years, but it may not be very effective at real-time detection. Therefore, it is crucial to create a detection system that can quickly identify attacks. In this study, we use a real-time wireless network intrusion detection system that uses a CDBN (Conditional Deep Belief Network) to identify attack features.

[7] Detecting intrusions with cloud-based intrusion detection and blockchain applications: Approaches, challenges, and solutions: This paper, “Detecting intrusions with cloud-based intrusion detection and blockchain applications: Approaches, challenges, and solutions” discusses relying on intrusion detection and cloud systems using blockchain, utilising technologies like virtualization and containerization, collaborative anomaly detection systems for identifying insider and external attacks from cloud centres, and more. The paper also highlights the need for further research in this area, as cloud systems pose new security challenges, including the potential threat of live migration processes.

[8] Cyber-attacks and power system contingencies can be accurately classified using recurrent neural networks: In this paper, the author suggests using deep learning techniques to categories contingencies and cyberattacks and discusses the importance of accurate event and intrusion detection in power systems. in "Cyber-attacks and power system contingencies can be accurately classified using recurrent neural networks." Results from neural networks based on recurrences to classify binary and multiclass events using datasets from a power system testbed are presented in the paper. The sound operation of the power systems depends on prompt, accurate responses to abnormal conditions.

[9] Insuring the security of smart grids: Systems for Detecting and Preventing Intrusions: This paper, The smart grid has many advantages, but because it combines various technologies, it also presents security and privacy challenges, according to " Insuring the security of smart grids: Systems for Detecting and Preventing Intrusions." In spite of the fact that encryption and authorization mechanisms are useful, they might not be able to prevent every attack. The traditional electrical grid has undergone technological development with the smart grid (SG) paradigm. In addition to improving service quality and reliability as well as making efficient use of existing infrastructure and renewable energy sources, it also helps to protect the environment.

V. MODULE

Intrusion detection on smart grids is a critical issue as the power grid infrastructure is becoming more complex and vulnerable to cyber-attacks. The proposed module uses ensemble learning network intrusion detection for smart grids.

A. *Exploratory Data Analysis*

Exploratory data analysis (EDA) is a method that utilizes visual representation techniques to detect data patterns and to perform comparative analysis. EDA is widely preferred for feature engineering and selection processes in data science projects. There are various commonly used EDA techniques such as univariate analysis, bivariate analysis, multivariate analysis, bar charts, box plots, pie charts, line graphs, frequency tables, histograms, and scatter plots. EDA is also used to quickly identify any mistakes in the data. Univariate Analysis involves analyzing a single variable while Multivariate Analysis deals with comparative analysis between multiple variables. In machine learning and deep learning projects, identifying data correlations using visual representations is crucial to gain insights into the dataset. Therefore, exploring these insights through EDA helps in achieving the desired model prediction goals.

B. *Feature Selection*

In machine learning, selecting the most relevant and informative features from a dataset is crucial for building accurate models. An algorithm that selects features based on their importance and discards irrelevant or redundant ones is known as a feature selection algorithm. One commonly used metric for feature selection is Information Gain Ratio, which measures how much a given feature may reduce the entropy of a dataset. However, computing the information gain ratio values for all the features in a large dataset can be computationally expensive, and may not be practical in some cases. Our work proposes Optimal Feature Selection as a new algorithm for selecting features. The algorithm aims to reduce the time taken for computation while still identifying the most informative features. Based on the Information Gain Ratio calculated for each attribute in the dataset, the Optimal Feature Selection algorithm performs column reduction based on this value. This means that features with low Information Gain Ratio values will be removed, while those with high values will be retained.

C. *Training & Testing*

This module discusses a method frequently applied by Convolutional Neural Networks (CNNs) to perform image classification. From the input image, the convolution operation is used to extract features like edges. To reduce the number of features and improve the correlation between nearby pixels, pooling is typically used after convolution.

Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)
V

Volume 6- Issue 2, August 2023
Paper : 27

In the suggested approach, sample images are downsized and key features are extracted using max-pooling. By choosing the maximum value from each region of the feature map, max-pooling aids in preserving the most crucial features while eliminating the unnecessary ones. Rectified Linear Unit (ReLU) activation is used for each convolutional layer. ReLU is a non-linear activation function that increases the network's non-linearity and aids in avoiding the vanishing gradient problem during the training phase. The maximum dimension of a convolutional kernel is equal to the length of a row or column in an incidence matrix or an adjacency matrix. The idea behind this is that a specific row or column reflects the weights of the connections between the node represented by the row or column index and every other node in the network. Therefore, a convolutional step that includes all of the connected neighbours would ensure that crucial connections are not lost. The resulting matrices are flattened into a single column with all the pixel values after all the convolutions and max-pooling. The following artificial neural network uses this flattening layer as input to the hidden layer. A neural network's hidden layers use the flattened data to perform additional processing.

VI. RESULT

As a result, we examined various machine learning models over our dataset, and decision tree- based classifiers outperformed others based on testing accuracy. The majority of our dataset's attack and non-attack records could be distinguished using ensemble learning algorithm. This graph compares the accuracy and loss during training and validation.

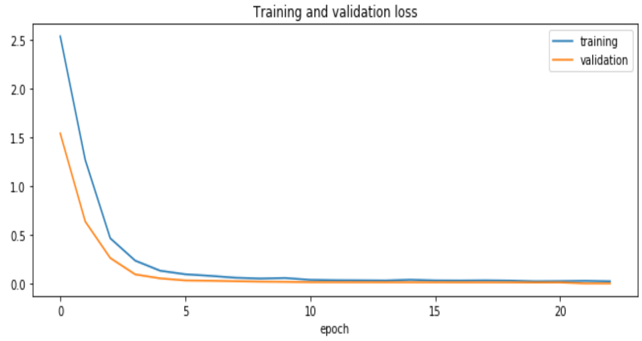

Fig. 1. Training and Validation Loss

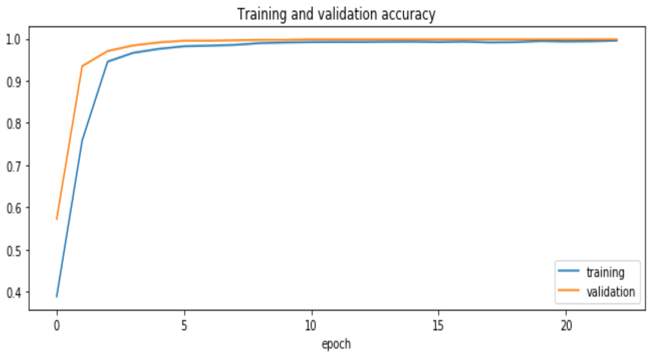

Fig. 2. Training and Validation Accuracy

VII. ARCHITECTURE DIAGRAM

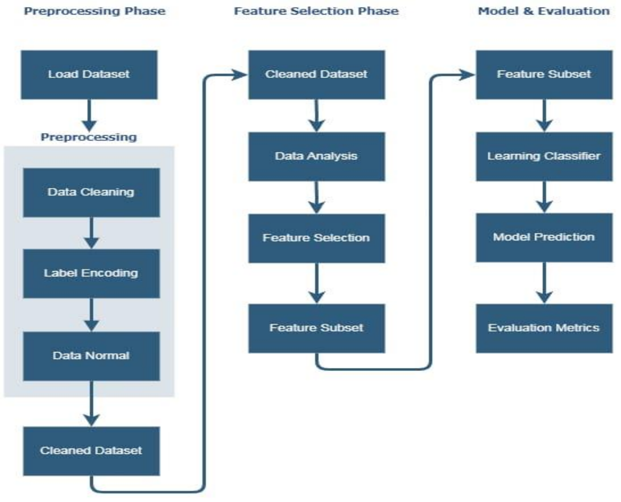

Fig. 3. Architecture Diagram

VIII. CONCLUSION

In order to prevent intrusions and respond to them before the system network is threatened, proactive security technology known as intrusion detection offers real-time protection against insider threats and external abuse. Intrusion detection from a multi-level defence perspective should grab everyone's attention from a three-dimensional deep network security perspective. Due to the shortcomings of current intrusion detection systems, this article introduces the concept of using ensemble learning methods in intrusion detection systems and establishes an intrusion detection system model. In order to realise a learning-based intrusion detection system, give a general overview of the system's process, including its key implementation steps.

IX. FUTURE ENHANCEMENTS

In the future, we will run more tests on our system and monitor any changes in accuracy. In order to increase the precision of IDS, we also hope to combine the RST method with the genetic algorithm. The current system only shows log data; it makes no attempt to analyse or otherwise gain knowledge from the data contained in the log records. Data mining techniques can be added to the system to analyse the data in the log records and support efficient decision-making. The system currently detects only known attacks. Self-knowledge gained from analysing increasing traffic and discovering fresh intrusion patterns.

REFERENCES

[1] C. Guo, Y. Ping, N. Liu and S. S. Luo, "A two-level hybrid approach for intrusion detection", Neurocomputing, vol. 214, pp. 391-400, 2016.

[2] A. A. Aburomman and M. Bin Ibne Reaz, "A novel SVM-kNN-PSO ensemble method for intrusion detection system", Applied Soft Computing Journal, vol. 38, pp. 360-372, 2016.

[3] S. O. Al-Mamory and F. S. Jassim, "On the Designing of Two Grains Levels Network Intrusion Detection System", Karbala International Journal of Modern Science Elsevier, vol. 1, pp. 15-25, 2015.

[4] W. Bul'ajoul, A. James and M. Pannu, "Improving network intrusion detection system performance through quality-of-service configuration and parallel technology", Journal of Computer and System Sciences, vol. 81, pp. 981-999, 2015.

[5] K. Zheng, Z. Cai, X. Zhang, Z. Wang and B. Yang, "Algorithms to speedup pattern matching for network intrusion detection systems", Computer Communications, vol. 62, pp. 47-58, 2015.

[6] S. Rastegari, P. Hingston and C. P. Lam, "Evolving statistical rulesets for network intrusion detection", Applied Soft Computing Journal, vol. 33, pp. 348-359, 2015.

[7] A. Khraisat, I. Gondal, P. Vamplew and J. Kamruzzaman, "Survey of intrusion detection systems: techniques datasets and challenges", Cybersecurity, vol. 2, no. 1, pp. 20, 2019.

[8] Z. El Mrabet, H. El Ghazi and N. Kaabouch, "A performance comparison of data mining algorithms-based intrusion detection system for smart grid", International Conference on Electro Information Technology (EIT), pp. 298-303, 2019.

[9] E. Anthi, L. Williams, M. Słowińska, G. Theodorakopoulos and P. Burnap, "A supervised intrusion detection system for smart home IoT devices", Internet of Things Journal, vol. 6, no. 5, pp. 9042-9053, 2019.

[10] S. Chesney, K. Roy and S. Khorsandroo, "Machine Learning Algorithms for Preventing IoT Cybersecurity Attacks", Intelligent Systems Conference, pp. 679-686, 2020.

[11] S. Ahmed, Y. Lee, S. H. Hyun and I. Koo, "Unsupervised Machine Learning-Based Detection of Covert Data Integrity Assault in Smart Grid Networks Utilizing Isolation Forest", Transactions on Information Forensics and Security, vol. 14, no. 10, pp. 2765-2777, 2019.

[12] M. Attia, S.M. Senouci, H. Sedjelmaci, E.H. Aglzim and D. Chrenko, "An efficient Intrusion Detection System against cyber-physical attacks in the smart grid", Computers and Electrical Engineering, vol. 68, pp. 499-512, 2018.