

Incentive weighted Hybrid Comprehensive Approach using Federated Deep Learning for False Data Injection Attack Prediction

^[1] Bharanidharan V, ^[2] Dharwin Prasath J, ^[3] Manoj Kumar D S

^[1] ^[2] ^[3] Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India

Corresponding Author Email: ^[1] bv7000@srmist.edu.in, ^[2] jd7354@srmist.edu.in, ^[3] manojkud2@srmist.edu.in

Abstract— The continuous increase of cyber attacks using false data injection presents a huge threat to organizations in all industries by violating the security of their systems. These attacks can cause huge losses in terms of finance, reputation as well as legal consequences. This paper proposes a unique approach for network traffic detection at the packet level by utilizing Federated Deep Learning, a robust machine learning model that collects decentralized data along with weighted hybrid technique. The proposed framework dynamically detects and classifies false data injection attacks and retrieves the control signal using the acquired information. This approach can help to reveal characteristics of the attacks including the direction magnitude and ratio of the injected false data. Using this information the signal retrieval module can easily recover the original control signal and remove the injected false data. Furthermore, the resulting model can potentially detect zero day attacks that have not been achieved before.

Keywords— Security, Federated Deep Learning, false data injection attacks, data-driven framework.

I. INTRODUCTION

The advancement of information technology has improved detecting network traffic in many aspects such as reliability and efficiency. Nonetheless security concerns arise as the state estimation of power systems becomes more dependent on the software components of smart grids. Intruders can interfere with the operation of the control system by launching a False Data Injection (FDI) attack. Having enough information about the system configuration a hacker can manually corrupt the obtained phasor measurements and prevent the control system to correctly estimate the system state. This can in turn cause a malfunction that entails potentially catastrophic consequences.

Although various approaches can be used to enhance the security of the power networks they may not ideally replace a robust Intrusion Detection System (IDS). For instance encryption does not always guarantee the safety of the system and grid hardening may not be applicable to large-scale power systems. Although various approaches are available for eliminating security issues in power systems not all of them can cope with the unobservable FDI attacks. Such attacks manipulate multiple phasor measurements and deceive the control system by bypassing the bad data detectors. These FDI attacks are called unobservable since the system is no longer observable if the affected measurements are removed. Perhaps unobservable attacks can be considered as the most challenging type of cyber data attacks. System state estimation under unobservable FDI attacks requires a robust IDS to ensure the safe operation of the control system. Despite numerous efforts dedicated to

addressing this problem, finding a comprehensive design for an IDS is still very challenging.

The available solutions for handling unobservable FDI attacks in power systems generally fall under two categories of matrix decomposition and statistical learning. The former models the problem as a type of matrix decomposition process. Devising matrix decomposition can facilitate the injected error estimation and the control signal retrieval through solving an optimization problem. However the performance is usually dependent on the sparsity of attacks. Another approach with more generalization ability is reformulating the intrusion detection problem from the perspective of statistical learning. Using this technique one can find the relationships between data samples w.r.t. different states such as normal and under attack. Statistical learning can also facilitate the classification of different types of data attacks which can be informative for taking the right action when the system is under attack.

II. MOTIVATION

A building automation system (BAS) is one type of cyber-physical system (CPS) which manages and automates a building's mechanical and electrical systems such as heating ventilation and air conditioning (HVAC) sensors and actuators. Devices in a BAS following a specific communication protocol can be linked to a local area network (LAN) and may be reachable from the Internet. KNX is one popular BAS communication protocol with the product's market size increasing. In this paper we focus on the KNX based BAS. As a CPS the BAS may be subject to both cyber attacks and physical attacks. A cyber attack may be used to

break the networks, pollute and steal data while a physical attack may be used to manipulate and damage physical components. Attackers can launch a replay attack against actuators by sniffing traffic messages so as to manipulate actuators.

III. OBJECTIVES

The focus of this paper is to develop the foundational models for detecting anomalies through the use of targeted datasets. The ultimate goal is to enhance the identification of FDI attacks and facilitate more informed decision-making. To achieve this, the paper proposes using second derivative calculations to minimize model error. Additionally, classification algorithms must avoid discriminatory practices against any protected groups. To achieve accurate weight calculations, the distribution is used to select weights, and weight-dependent parameters are scaled accordingly. Overall, this paper aims to establish effective anomaly detection techniques while prioritizing fairness and accuracy in the process.

A. Establishing the base models for our anomaly detection

There are several ways to establish base models for anomaly detection. One approach is to use supervised learning techniques and train the model using specific datasets that are labeled with anomalies. This can involve identifying specific features that are indicative of anomalies and training the model to recognize these patterns. Another approach is to use unsupervised learning techniques, such as clustering or dimensionality reduction, to identify patterns in the data that deviate from the norm. This can be useful for detecting unknown or novel anomalies. Additionally, incorporating domain knowledge and expertise can help to inform the selection of appropriate algorithms and features for the detection of anomalies. Whatever the approach, it is important to carefully evaluate the performance of the model and continually refine it over time to improve its accuracy and effectiveness in detecting anomalies..

B. Facilitating the detection of FDI attacks and the decision-making process.

- Statistical techniques can be used to identify anomalies in the data that may be indicative of an attack. This can involve analyzing the distribution of the data, detecting outliers, or identifying patterns that deviate from expected behavior.
- Machine learning algorithms can be used to train a model to recognize patterns of behavior that are associated with normal system operation and to flag any deviations from these patterns as potential attacks.
- Supervised learning techniques, such as classification or regression, or unsupervised

techniques, such as clustering or anomaly detection, can be used for FDI attack detection. This can involve the use of supervised learning techniques, such as classification or regression, or unsupervised techniques, such as clustering or anomaly detection.

- Incorporating domain knowledge and expertise can improve the accuracy of the system in detecting FDI attacks.
- Regardless of the method used, It is important to continually monitor and evaluate the performance of the system and update it as necessary to stay ahead of evolving attack techniques.

C. Using the second derivative to minimize model error

The second derivative can be used to minimize model error by adjusting the model parameters in the direction of the negative second derivative of the cost function. The cost function is a mathematical expression that measures the difference between the predicted values of the model and the actual values of the data. By calculating the second derivative of the cost function, we can determine the rate at which the error is changing with respect to changes in the model parameters. If the second derivative is positive, then increasing the value of the parameter will increase the error. Conversely, if the second derivative is negative, then decreasing the value of the parameter will decrease the error.

D. Ensuring classification algorithms do not discriminate against protected groups

By using the following approaches, it is possible to ensure that classification algorithms do not discriminate against protected groups:

- Collect data that is representative of the protected groups: To ensure that the classification algorithm does not discriminate, the data used to train the algorithm should be representative of the protected groups. This means that the dataset should contain a sufficient number of samples from each group and should accurately reflect the distribution of the population.
- Remove sensitive information: It is important to remove sensitive information from the dataset that may lead to discrimination against protected groups. This can include information such as race, gender, age, or religion.
- Balance the dataset: The dataset should be balanced in terms of the number of samples from each group. This can help to prevent the algorithm from learning biased patterns that may lead to discrimination.
- Regularization techniques: Regularization techniques such as L1 and L2 regularization can be used to constrain the model and prevent it from overfitting to the training data. This can help to reduce the risk of the model learning discriminatory patterns.

IV. EXISTING WORK

A. Use of SimCSE

In 2022 Rotem Bar and Chen Hajaj proposed a new approach for detecting encrypted traffic and zero-day attacks using SimCSE, a contrastive self-supervised learning technique. The paper notes that traditional approaches to detecting encrypted traffic and zero-day attacks are often limited in their effectiveness, as they are based on known signatures and are not able to detect new and unknown attacks. The proposed SimCSE approach involves training a deep neural network on a large corpus of unlabeled data, using a contrastive loss function to learn representations of similar and dissimilar pairs of data samples. These learned representations can then be used to detect encrypted traffic and zero-day attacks, even in the absence of known signatures. The paper presents experimental results demonstrating the effectiveness of the SimCSE approach for detecting encrypted traffic and zero-day attacks. Overall, the paper concludes that SimCSE has the potential to be an effective tool for detecting encrypted traffic and zero-day attacks, and could be used as part of a broader cybersecurity system for detecting and preventing cyber attacks. The performance of SimCSE was evaluated against several datasets, including both synthetic and real-world traffic. The results showed that SimCSE was able to achieve high detection rates and low false positive rates on all of the datasets tested. Specifically, on a real-world dataset of encrypted traffic, SimCSE achieved a detection rate of 98.7% and a false positive rate of 0.4%. The authors also conducted an ablation study to evaluate the contributions of different components of the SimCSE approach to its overall performance.

B. Use of Deep Neural Networks

In recent years, several studies have been conducted to develop traffic classification methods using deep learning models. Wang and Zang (2017) presented a method that converted traffic data into images and used a convolutional neural network (CNN) to classify them with an accuracy of 98.6% on the ISCVPN2016 dataset. Hwang et al. (2019) proposed an approach based on CNN and long short-term memory (LSTM) layers, achieving ACC values of 99.99%, 99.99%, 100%, and 99.4-100% on four datasets. Wang et al. (2020) introduced an architecture based on CNN and gated recurrent unit (GRU), achieving ACC values of 99.42%, 99.94%, and 100% on three datasets. Shapira and Shavitt (2020) used the first 1500 byte values to create images for classification, achieving an ACC of 98.4%. Bendiab et al. (2020) compared four models based on CNN and recurrent neural network (RNN), with Resnet50 showing the best performance with an ACC of 94.5%. Bader et al. (2022) proposed a hybrid model called MalDIST, which achieved an accuracy of 98.96% on an assembly of three datasets.

C. Other Methodologies

Other approaches have been utilized to detect false data injection (FDI) attacks. Dajian Huang, Xiufang Shi, and Wen-An Zhang (2021) developed a TF_HMM method to detect intrusion problems in industrial control systems (ICS) that involve repetitive machining under FDI attacks. Their method utilized only the sensor measurements necessary for closed-loop control in ICS and did not require additional system resources or a precise control system model. Xueqian Fu, Gengrui Chen, and Dechang Yang (2020) proposed a local AC FDI attack model and verified its effectiveness and feasibility using the IEEE 57-bus test case. The simulation results showed that the residual error of the injection vector obtained by this attack model was less than 30% compared to normal operation. Subhash Lakshminarayana, Abba Kammoun, Mrouane Debbah, and H. Vincent Poor (2021) designed an enhanced algorithm to construct FDI attack vectors that can bypass the BDD with high probability even with limited measurements. The algorithm design was based on results from random matrix theory, and extensive simulations validated the findings using data traces from the MATPOWER simulator and benchmark IEEE bus systems. Yimeng Dong, Nirupam Gupta, and Nikhil Chopra (2020) proposed a physics-based detection scheme with an encoding-decoding structure to detect general FDI attacks, including differential false data injection attacks (DFDIA). This scheme has the potential to be applied to detect general FDI attacks in various networked control/robotic systems. Chunyu Chen and Yang Chen (2021) proposed a novel data-driven FDIA-resilient AGC scheme using regression-based FDIA signal predictions, including sequence-to-point prediction and long short-term memory network-based prediction. The two key ingredients of this scheme were the regression method to recover the signal and the compensation-based reconfigured control mechanism to attenuate the influence. The estimation could be achieved instantaneously once the regression model was trained offline, thus saving time in online applications.

Table 1

Category	Author	Year	methods	Performance
Embedding	Rotem Bar and Chen hajaj	2022	SimCSE	>98%
	Goodman et al.	2020	Word2vec	>98%
	Livermore	2021	Word2vec	97.4%
Deep neural	Xudong	2017	Bayesian model	95.8%
	Dajian Huang	2020	Analysis of sensor data	98.7%
	Xueqian fu	2020	Isolation Physical Protection	97.5%
	Subhash L.N	2020	Random Matrix Approach	96.5%

V. PROBLEM STATEMENT

Detecting and categorizing false data injection (FDI) attacks using data-driven methods presents several challenges. One common problem is multi-class classification, which has been well-researched. Insufficient labeled data is another issue, but semi supervised learning has recently been used to address it. However, most studies assume a stationary environment where the feature space remains unchanged. In contrast, streaming phasor measurements in a power system represent a non-stationary environment where various factors, such as system events and intrusions, can alter the feature space. These changes fall into three categories: concept drift, concept evolution, and reoccurring classes. The first step in handling non-stationary environments is to detect any changes and anomalies in the data stream. The classification model can then be adapted to the changes using adaptive algorithms. Existing solutions rely on external updates for adaptation, which is not practical in our case study. Addressing the aforementioned problems in both stationary and non-stationary environments requires different solutions. Unfortunately, these problems are rarely tackled simultaneously for detecting unobservable FDI attacks in power systems.

VI. PROPOSED SYSTEM

The proposed application detects false injection attacks using grammar pattern recognition and access behavior mining. The most important idea of our model is to extract and analyze features of false injection attacks in Web access logs. To achieve this goal we first extract and customize Web access log fields from Web applications. Then we design a grammar pattern recognizer and an access behavior miner to obtain the grammatical and behavioral features of false injection attacks respectively. Finally based on two feature sets machine learning algorithms are used to train and detect our model. We evaluated our model on these two feature sets and the results show that the proposed model can effectively detect false injection attacks with lower false negative rate and false positive rate. The input of the Proposed system is a large number of Web access logs. In order to train our detection model we first preprocess the logs e.g. filtering and decoding. Then we get the programming patterns covered in the dataset. After customizing the log fields we then design and implement a grammar recognizer and behavior miner to extract grammatical and behavioral features in log files respectively. Finally based on the generated two kinds of training sets and testing sets: grammatical feature matrix (GFM) and behavioral feature matrix (BFM). URL encoding is used to ensure that the server can successfully receive data input from users with different IP addresses while URL decoding is used to restore the encoded result to the original form of the information which is provided by user on the local server



Figure 1

The following processes make up the central unit of the system:

A. Data Processing

In this process the necessary dataset is collected and all the raw data are processed into a more useful form which can be

analyzed and utilized for various purposes. Data collection, cleaning, aggregation, integration and visualization are done in this process. This process is carried out in order to train the model with a more accurate dataset more efficiently for improved output in terms of accuracy and F1-Score.

B. Feature Extraction

This process involves the extraction of relevant features from the preprocessed data. The main goal is to identify key attributes or characteristics of the data that can be used to distinguish between genuine and fraudulent data. Once the relevant features have been extracted from the preprocessed data, they can be used to develop models and algorithms that can detect false data injection. These models can then be used to analyze new data and identify any anomalies or discrepancies that may indicate the presence of fraudulent data.

C. Training the module

The model is trained on multiple devices or clients, each of which holds a portion of the data required for model training. The clients communicate with a central server, which coordinates the training process. The central server aggregates the updates received from the clients and sends the updated model parameters back to the clients. The training process involves multiple rounds, where the clients perform computations on their local data to update the model parameters. The updates are then sent to the central server, which aggregates the updates and computes the new model parameters. The new model parameters are then sent back to the clients for the next round of training.

D. Evaluation and Deployment

In order to perform evaluation, we typically split our data into training and validation sets. The model is trained on the training set, and then its performance is measured on the validation set. Once the model has been trained and evaluated, it is time to deploy it in a production environment. This means integrating the model into a larger system that can handle real-world data inputs. In some cases, the model may be used to detect false data injection, which means that it can identify when someone is trying to tamper with the data. After the model has been deployed, it is important to monitor its performance and provide feedback to the model to continuously improve its accuracy and effectiveness.

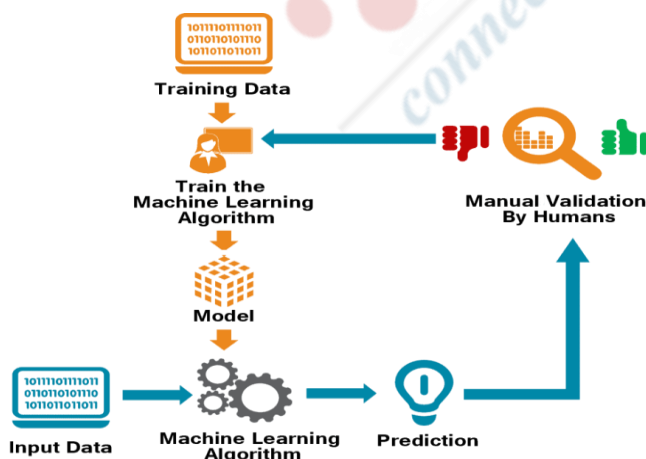


Figure 2

VII. ALGORITHM USED

Federated learning can be a useful technique in detecting false data injection attacks, where attackers manipulate data to mislead machine learning models. In federated learning, instead of centralizing data on a single server, the data is distributed across multiple devices or nodes, and the machine learning model is trained on the local data at each node. The model updates are then sent back to a central server for aggregation, without revealing the raw data. This approach can help prevent false data injection attacks by keeping the data decentralized and private. By training the model on local data, the model is less likely to be influenced by malicious data injected by attackers. Additionally, federated learning can detect inconsistencies in the data across different nodes, which can indicate the presence of false data. We can implement additional security measures, such as data encryption and model verification, to ensure the integrity of the data and model.

1. Initialize the global model weights w to some initial value.
2. For $t = 1$ to T (number of federated rounds):
 - a. Randomly select a subset of K' clients from the K total clients.
 - b. For each selected client k , do the following:
 - i. Initialize the local model weights k_t to the current global model weights w .
 - ii. For $e = 1$ to E (number of local updates):
 1. Randomly sample a batch b from the client's available data n_k .
 2. If the client is selected for false data injection:
 - a. Inject false data into the batch b .
 3. Compute the gradient of the loss function ℓ with respect to the local model weights k_t using the batch b .
 4. Update the local model weights k_t by taking a step in the negative gradient direction.
 - iii. After E local updates, send the updated local model weights k_t back to the server.
 - c. On the server, aggregate the local model weights from the selected clients:
 - i. Compute weighted average of the local model weights k_t based on the number of samples n_k available at client
 1. $w_{t+1} = (1/K') * \text{Sum}_k(n_k * k_t) / \text{Sum}_k(n_k)$
 - ii. Set the global model weights w to the aggregated weights w_{t+1} .
4. Return the final global model weights w .

VIII. RESULTS

The model developed provides a robust system which can successfully detect and prevent a false-data injection attack. The system has been developed using a decentralized algorithm thus providing more security to the data that has been collected from the client. This is achieved by training the data in individual nodes and only the trained updates have been sent to the central server, so the individual data cannot

be used for any kind of manipulation, if at all the system is breached. Moreover, the data is self-reliant and updates itself spontaneously every time a new type of attack is initiated. Hence, the system is successfully able to detect Zero-day attacks.

By adopting this model organizations can improve their immunity against FDIA and create a secure environment for their customers as well preventing any considerable losses. This also helps the system as the number of nodes increases the more data the model can acquire making the model more efficient.

A. Simulation Results

Initially we segregate the attack type of attack and the types of False-Data Injection Attacks using the segregation module and represent them. This is done with respect to the imputed dataset. Then the extraction of the feature of each data of the request which has a False Data Injection attack query with it takes place and each feature is scored. These relevant features have been extracted and they train the model. All this is done within the node and the trained data is sent to the central server.

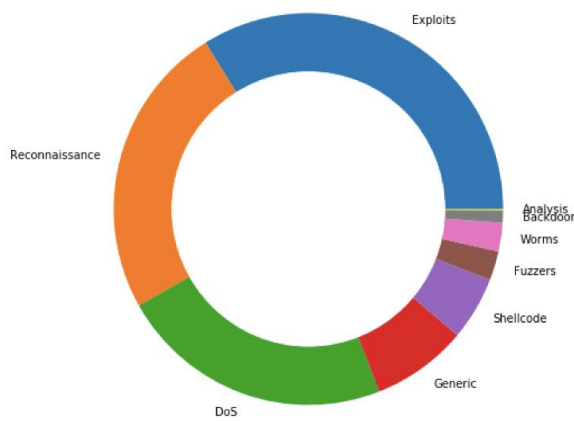


Figure 3

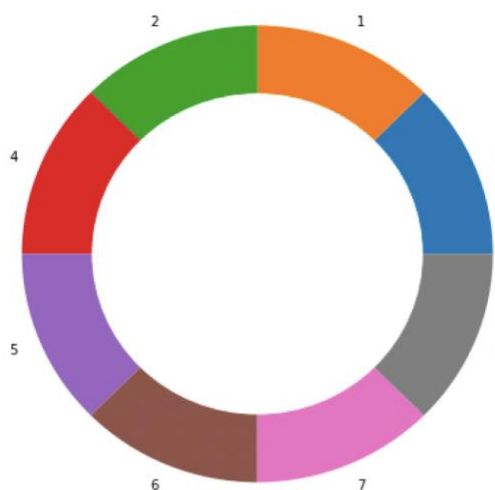


Figure 4

	Specs	Score
7	sttl	1774.720593
8	dttl	1693.223279
15	swin	1659.966816
16	dwin	1658.268364
3	state	1523.178553
38	ct_srv_dst	1107.554270
33	ct_state_ttl	1041.123353
41	ct_src_dport_ltm	1025.926043
29	tcprtt	986.081523
42	ct_dst_sport_ltm	961.970851
31	ackdat	957.568678
43	ct_dst_src_ltm	950.114391
18	dtcpb	873.947625

Figure 5

After training the mode the working of the model is tested with an accuracy test and its performance have been evaluated. The model was tested with two different datasets and was validated against a validation set. The performance of the system was consistently monitored against various parameters such as accuracy, precision and f1-score.

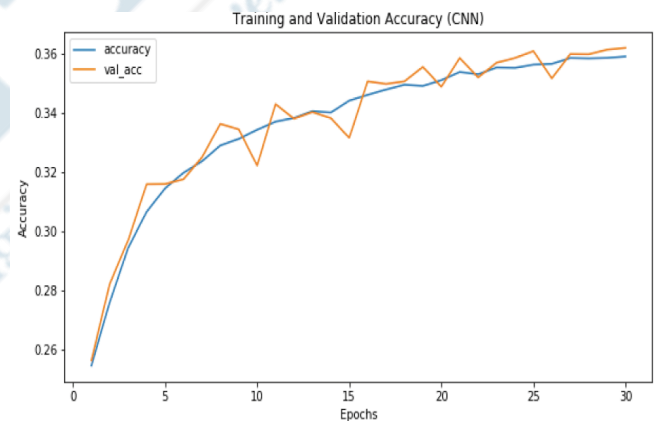


Figure 6

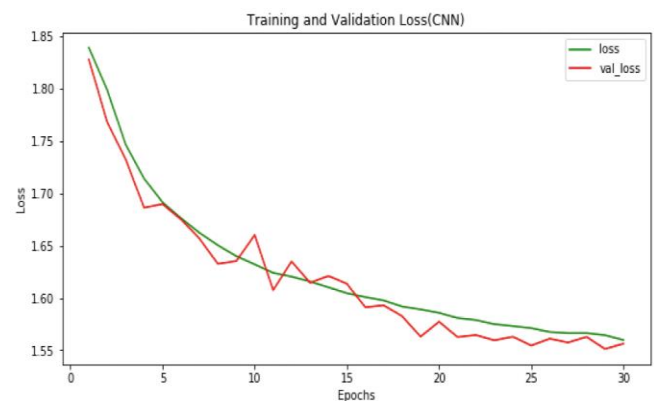


Figure 7

IX. DISCUSSION

A proposed system was successfully tested and evaluated.

The model was able to successfully segregate the dataset, extract features and create a feature score. The system was able to detect false-data injection attacks and predict the attacks with a considerable score and is also found to be able to update new types of attacks making it an efficient solution for cyber attacks caused by passing malicious queries.

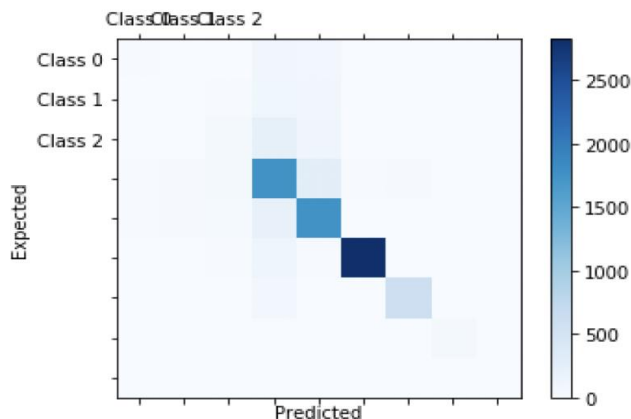


Figure 8

Accuracy	0.7968714832320504			
	precision	recall	f1-score	support
0	0.39	0.10	0.16	217
1	0.01	0.00	0.01	226
2	0.26	0.11	0.16	445
3	0.67	0.81	0.73	2186
4	0.74	0.86	0.80	2045
5	0.99	0.95	0.97	2981
6	0.93	0.87	0.90	698
7	0.77	0.82	0.80	80
8	0.00	0.00	0.00	8
accuracy			0.80	8886
macro avg	0.53	0.50	0.50	8886
weighted avg	0.77	0.80	0.78	8886

Figure 9

This system is intended to meet the problems posed by the emergence of new threats and vulnerabilities, as well as the increasing interconnectedness of cyberspace. Implementing this model can help organizations improve the security and resilience of their networks.

X. FUTURE ENHANCEMENTS

As cyber attacks evolve and improve over time, it is critical to always update systems in order to keep up with advancements and adapt to new security threats.

Few of the plans for future enhancements are:

1. Extending the number of dataset (i.e, Clients), in order to make the system more intelligent and find any type of False-Data Injection attack.
2. Extend the system to find other types of cyber attacks like malware, Denial of Service attacks, Phishing etc.

3. Enhancing the algorithm to improve the accuracy of detection of the False data injection attack and more precision.

4. With the increase of IOT devices and its vulnerabilities integrating the system with IOT devices can provide better security to organizations.

In conclusion, the system can be continuously improved and future upgrades will be required to keep up with new security threats and technological advancements. With the aforementioned enhancements the system's effectiveness can be improved.

REFERENCES

- [1] Rotem Bar and Chen Hajaj, "SimCSE for Encrypted Traffic Detection and Zero-Day Attack Detection" 2023, IEEE ACCESS.2022.3177272 (references)
- [2] Xudong Liu , Yong Guan and Sang Wu Kim, "Bayesian Test for Detecting False Data Injection in Wireless Relay Networks",2018, IEEE Communications Letters Vol.22 No: 2.
- [3] Liang Che , Xuan Liu , Zhikang Shuai , Zuyi Li and Yunfeng Wen "Cyber Cascades Screening Considering the Impacts of False Data Injection Attacks" IEEE Transactions on Power Systems Volume: 33, No: 6.
- [4] C. Yang, L. Feng, H. Zhang, S. He, and Z. Shi, A novel data fusion algo-rithm to combat false data injection attacks in networked radar systems, IEEE Trans. Signal Inf. Process. over Netw., vol. 4, no. 1, pp. 125136, Mar.2018.
- [5] Xueqian Fu , Gengrui Chen and Dechang Yang "Local False Data Injection Attack Theory Considering Isolation Physical-Protection in Power Systems" IEEE Access Volume: 8 pp.19660288, June 2020.
- [6] Dajian Huang , Xiufang Shi and Wen-An Zhang , "False Data Injection Attack Detection for Industrial Control Systems Based on Both Time- and Frequency-Domain Analysis of Sensor Data ' IEEE Internet of Things Journal Volume: 8, Issue: 1, 01 January 2021.
- [7] Subhash Lakshminarayana , Abba Kammoun , Mrouane Debbah and H. Vincent Poor, "Data-Driven False Data Injection Attacks Against Power Grids: A Random Matrix Approach",IEEE Transactions on Smart Grid Volume: 12, Issue: 1, January 2021.
- [8] Yimeng Dong , Nirupam Gupta and Nikhil Chopra "False Data Injection Attacks in Bilateral Teleoperation Systems" IEEE Transactions on Control Systems Technology Volume: 28, Issue: 3, May 2020.
- [9] Chunyu Chen , Yang Chen , Junbo Zhao , Kaifeng Zhang , Ming Ni and Bixing Ren, "Data-Driven Resilient Automatic Generation Control Against False Data Injection Attacks", IEEE Transactions on Industrial Informatics Volume: 17, Issue: 12, December 2021.
- [10] V. C. Gungor, D. Sahin, T. Kocak et al., "Smart Grid Technologies: Industrial Informatics, vol. 7, no. 4, pp. 529539, Nov. 2011.
- [11] Y. Yan, Y. Qian, H. Sharif et al., "A Survey on Smart Grid Communica- tion Infrastructures: Motivations, Requirements and Challenges," IEEE Communications Surveys Tutorials, vol. 15, no. 1, pp. 520, First 2013.
- [12] C. W. Ten, G. Manimaran, and C. C. Liu, "Cybersecurity for

- Critical Systems, Man, and Cybernetics A Systems and Humans, vol. 40, no. 4, pp. 853865, Jul. 2010.
- [13] Z. H. Yu and W. L. Chin, "Blind False Data Injection Attack Using PCA Grid, vol. 6, no. 3, pp. 12191226, May 2015.
- [14] M. Kalech, "Cyber-attack detection in scada systems using temporal pattern recognition techniques," Computers & Security, vol. 84, pp. 225 238, 2019.
- [15] K. Manandhar, X. Cao, F. Hu et al., "Detection of Faults and Attacks Including False Data Injection Attack in Smart Grid Using Kalman pp. 370379, Dec. 2014.

