MALICIOUS NODE DETECTION IN SMART GRID NETWORKS

Faisal Y Al Yahmadi and Muhammad R Ahmed

[1]Marine Engineering Department, Military Technological College,
Muscat, Sultanate of Oman

ABSTRACT

Many countries around the world are implementing smart grids and smart meters. Malicious users that have moderate level of computer knowledge can manipulate smart meters and launch cyber-attacks. This poses cyber threats to network operators and government security. In order to reduce the number of electricity theft cases, companies need to develop preventive and protective methods to minimize the losses from this issue. In this paper, we propose a model based on software that detects malicious nodes in a smart grid network. The model collects data (electricity consumption/electric bill) from the nodes and compares it with previously obtained data. Support Vector Machine (SVM) model is implemented to classify nodes into good or malicious nodes by (high dimensional) giving the statues of 1 for good nodes and status of -1 for malicious (abnormal) nodes. The detection model also displays the network graphically as well as the data table. Moreover, this model displays the detection error in each cycle. It has a very low false alarm rate (2%) and a high detection rate as high as (98%). Future developments can trace the attack origin to eliminate or block the attack source minimizing losses before human control arrives.

KEYWORDS

Smart Grid Networks, Security, Malicious, Attacks, Support Vector Machine.

1. INTRODUCTION

Smart Grid Network (SGN) is an advanced network that merged new technologies and developed infrastructure to prepare the world to overcome the arising challenges expected to be faced in the coming decades. New implementations such as integration of alternative energy sources and decentralized generation will help overcome the growing global power demand expected with the adaptation of Electric vehicles EVs and other smart household appliances. SGN implementation of new technologies allows for two-way stream of both power and data [1]. These implementations will grant the network a greater ability to detect, react and pro-act towards power usage or other businesses. Suspicions power usage patterns by consumers will also be recognised and responded to with the new technology implementation. SGN enables service providers to monitor the behaviour of all stockholders of the electricity. SGN has the capability of enabling the consumer to become an active participant in the network. In order to ensure network economic feasibility and a high quality service with minimum losses, security and safety of supply is prioritised. Some of the benefits that SGN grants beneficiaries are as follows:

- Integration of alternative energy sources
- Decentralized generation
- Reliably electrical supply
- Greener power production
- Active consumer participation
- Better resilience towards grid blackouts

David C. Wyld et al. (Eds): CCSIT, SIPP, PDCTA, AISC, NLPCL, BIGML, NCWMC - 2021
pp. 201-210, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.110716

SGN implementation of information and communication technologies (ICT) allowed the new network to monitor, operate and control the system with added features. These control manners were only available for service providers at the generation phase. However, ICT helped to extend these manners across all SGN phases reaching transmission and distribution phases [2]. Two-way communication enables both service providers and companies to utilize the developed infrastructure for a more efficient grid. Two-way communication also allows consumers to be true active participants with ability to choose new power usage patterns that were not possible with the conventional grid. Moreover, to standardize the new SGN operation, National Institute of Standards and Technology (NIST) proposed an SGN model standardizing SGN architecture as shown in Fig. 1 below.

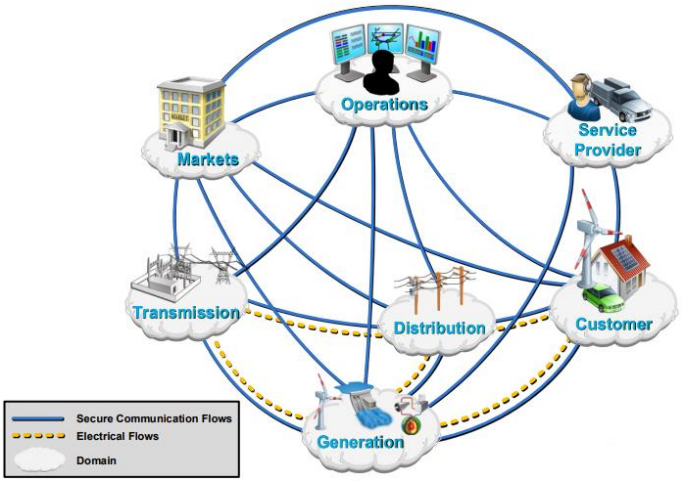

Fig. 1. Smart grid architecture model (SGAM) by [NIST] [3]

The proposed model lists seven domains, which are:

- Generation
- Transmission
- Distribution
- Operations
- Service providers
- Markets
- Consumer.

These above mentioned domains use secure communication in order to operate SGN efficiently. The proposed model also illustrates the electricity path between different domains, which are transmission, distribution, customer and generation, while communication flows across all seven domains.

Security provisioning is a critical necessity for any wired and wireless communication network [4]. Therefore, a machine-learning model will be adopted to detect attacks on SGN. Machine learning technology uses machines learning algorithms to artificially improve their performance as more data is being trained [5]. Machine learning has different techniques and models developed for various applications; one of the uses is solving classification problems. Support Vector Machine (SVM) is a classic machine learning technique which has the ability to classify high dimensional data [6]. This paper aims to develop an algorithm using one of the machine learning techniques, an SVM based model is used and simulated by MATLAB software. The simulation platform was chosen as MATLAB because it has the ability to classify attacked nodes by comparing collected data with average data collected from the same consumer/household. Attack detection revolves around two pillars, which are average electrical power consumption of the consumer (monthly/annually) and average (monthly/annually) electrical bill of the consumer. SGN is developed based on the wireless sensor network (WSN) concept. The data collecting

process starts with nodes representing consumers sending data to a central node and back to the supplier –witch in this case represents the developed algorithm-as shown in Fig. 2 below.

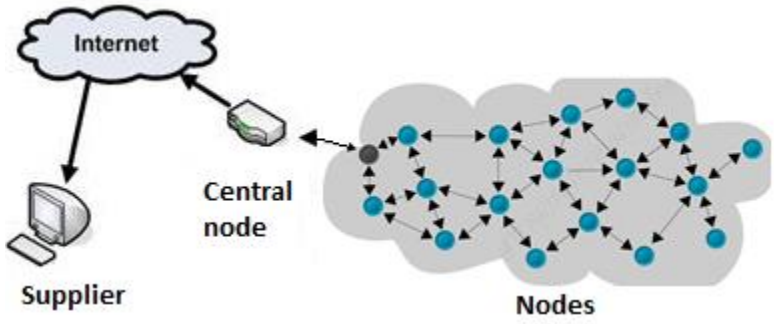

Fig. 2. A conceptual illustration of a generic WSN [7]

Considering the critical importance of SGN security, many algorithms have been developed to ensure secure functionality of SGN.

This paper is organized as follows. Section 2 contains an overview of related works followed by SGN implementations and assumptions in section 3. A detailed methodology is elaborated in section 4 with numerical results of the algorithm in section 5. The paper conclusion is in section 6.

2. RELATED WORKS

Smart grid network introduces enhancements and improved capabilities to the conventional power network making it more complex and vulnerable to different types of malicious attacks. Till today, several works have been done by many researchers to find the best way to detect malicious attacks but very few were focusing on the smart grid malicious attacks. Moreover, no significant importance has been given to finding the malicious attack based on the misbehaviour or abnormal behaviour of the node. Even though some researchers worked based on the misbehaviour but their main focus was to prevent or protect the routing. In the following section, related researchers work will be discussed:

Takiddin et al. in [8] provided answers to three major questions pertaining to the performance of electricity theft detectors in the presence of data poisoning attacks. By proposing a sequential ensemble detector based on a deep autoencoder with attention (AEA), gated recurrent units (GRUs), and feed forward neural networks. The proposed robust detector retains a stable detection performance that is deteriorated only by 1−3% in the presence of strong data poisoning attacks. However, in this method it is normally ensemble performs multiple learners, as a result computation get complicated, which reduce the speed and memory requirements rise.

Zhang et al. in [9] proposed a time series anomaly detection model based on the periodic extraction method of discrete Fourier transform. The detection model determines the sequence position of each element in the period by periodic overlapping mapping, thereby accurately describing the timing relationship between each network message. The experiments demonstrate that the model has the ability to detect cyber attacks such as man-in-the-middle, malicious injection, and Dos in a highly periodic network. The detection model also has a good anomaly detection capability. This model focus on the DoS attacks.

Jiang and Qian in [10] discussed defense mechanisms to either protect the system from attackers in advance or detect the existence of data injection attacks to improve the smart grid security. Focusing on signal processing techniques, this article introduces an adaptive scheme on detection of injected bad data at the control center. Jiang and Qian presented a detection scheme that can self-adaptively detect both non-stealthy and stealthy attacks. The scheme comprises determining two estimates of the state of the monitored system using the state measurement data provided by the remote sensing system at two sequential data collection slots, and determining bad data injection attacks by monitoring the measurement variations and state changes between the two slots. Analysis and simulation results shows that the proposed scheme is efficient in terms of data attack classification and detection accuracy. The research is good to detect data injection attacks.

Zhe et al. in [11] proposed a model based on machine learning to detect smart grid DoS attacks. The model collects network data, then selects features and uses PCA for data dimensionality reduction, and finally uses SVM algorithm for abnormality detection. By testing the SVM, Decision Tree and Naive Bayesian Network classification algorithms on the KDD99 dataset, it is found that the SVM model works best. This method has higher classification detection rate and accuracy, which can effectively improve the security of the smart grid DoS intrusion detection system. This method the data need to go thorough standardization process and in PCA we need to select the principle components otherwise it may miss data features.

Xia et al. in [12] suggest a method to identify all malicious users in a neighbourhood area network. The method uses Group Testing based Heuristic Inspection (GTHI) algorithm, which can estimate the ratio of malicious users on-line, mainly by collecting the information that how many malicious users have been identified during the inspection process. Based upon the ratio of malicious users, the GTHI algorithm adaptively adjusts inspection strategies between an individual inspection strategy and a group testing strategy. The GTHI algorithm outperforms existing methods in some aspects: compared with the BCGI algorithm, it has a wider range of applications; compared with the ATI algorithm, it can locate malicious users within much shorter detection time, regardless of the ratio of malicious users. However, this method does not include the user estimation in the testing phase.

Nandanoori et al. in [13] proposed a Koopman mode decomposition (KMD) based algorithm to detect and identify false data attacks in realtime. The Koopman modes (KMs) are capable of capturing the nonlinear modes of oscillation in the transient dynamics of the power networks and reveal the spatial embedding of both natural and anomalous modes of oscillations in the sensor measurements. The Koopman-based spatio-temporal nonlinear modal analysis is used to filter out the false data injected by an attacker. This algorithm detects the induced attack within 1 second of attack initiation in the presence of load changes in the network. This method normally works only work based on the false data injection.

Patil and Sankpal in [14] proposes an enhanced grid sensor placement (EGSP) algorithm to place grid sensors in the distribution network to monitor and control the smart meters installed in the field. The algorithm provides a simple and efficient way to place grid sensors in the distribution network for monitoring and controlling the smart meters deployed in the distribution network. A simulation model of distribution network has been developed for the analysis of the proposed algorithm. The analytical computation and simulation result shows that the number of grid sensors needed to track all the smart meters connected in the distribution network varies between half the number of SM nodes to equal number of SM nodes depending on how many SM nodes are connected to each EP node. In this method the computation is higher.

He et al. in [15] exploits a deep learning techniques to recognize the behavior features of FDI attacks with the historical measurement data and employ the captured features to detect the FDI

attacks in real-time. The proposed detection mechanism effectively relaxes the assumptions on the potential attack scenarios and achieves high accuracy. Furthermore, an optimization model is proposed to characterize the behavior of one type of FDI attack that compromises the limited number of state measurements of the power system for electricity theft. Method simulation results showed that the detection method can achieve high detection accuracy in the presence of the occasional operation faults. This work well only to predict the potential attack can happen.

The existing literature depicts that the vast majority of present methodologies to find the malicious in smart grid exists are in a general sense based on cryptographic primitives. Typically, in cryptographic solutions, the source uses cryptographic information to create and send additional authentication. As a results the extra information needed and the malicious can be detected based on the additional information data. The other introduced strategies are typically relying upon calculations and high level of training data. However, these methods have high computational overhead, because of every validation requires an immense number of checking to come up with the final decision about the malicious. Therefore, it is essential to develop an effective method to detect the malicious in the smart grid networks.

3. METHODOLOGY

Machine learning has many techniques, Support Vector Machines (SVM) based algorithm is used because of the model ability to classify unreliable data [16]. Which is suitable for high-dimensional data collected from across SGN. Therefore, SVM has been chosen for the proposed solution in this paper.

SVM model categorize the collected data by finding the optimal hyperplane shown in Figure 3 below, which will consist of the largest distance between the two different classes and that distance is called margin [17]. Margin is calculated from the nearest vector to the hyperplane and it must be without interior point as shown in Figure 3 below.

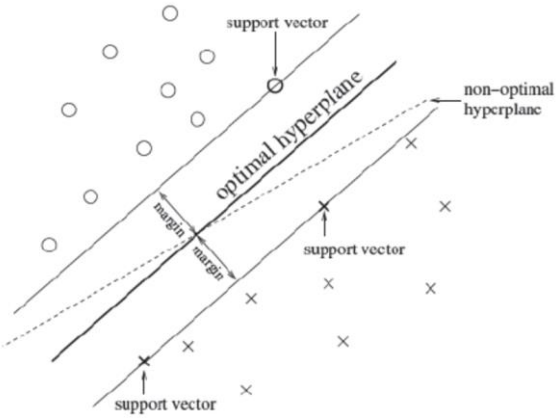

Fig. 3. SVM with its Optimal and non-optimal Hyperplane [18]

The closest point to the hyperplane which will be in contact with the margin parallel lines are called support vectors. Support vectors sets the hyperplane boundary [19]. Figure 3 also shows the two types of data, which are ×'s defining points of a value of 1 and O's defining points of a value of -1. The desired algorithm, a training phase to the system must be conducted offline using a resourceful information source. The training phase uses three Open System Intercommunication9 (OSI) layers, which are a physical layer followed by medium accessed control layer (MAC) ending with a network layer. After training then collecting the desired data,

a data trimming procedure will be implemented on these data sets. Data trimming is a vital step in order to reduce data size which will ultimately allow SVM to process it further. After completing data training and having training sets ready, classification can be done by a linear plane as illustrated in Fig. 4 below.

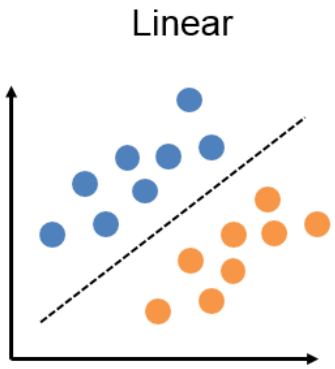

Fig. 4. Linear classification [20]

However, linear classification has limitations when it comes to classifying unreliable data [21]. Therefore, moving the data to a higher dimensional space will allow more functions that were not possible to be applicable such as mapping training sets.

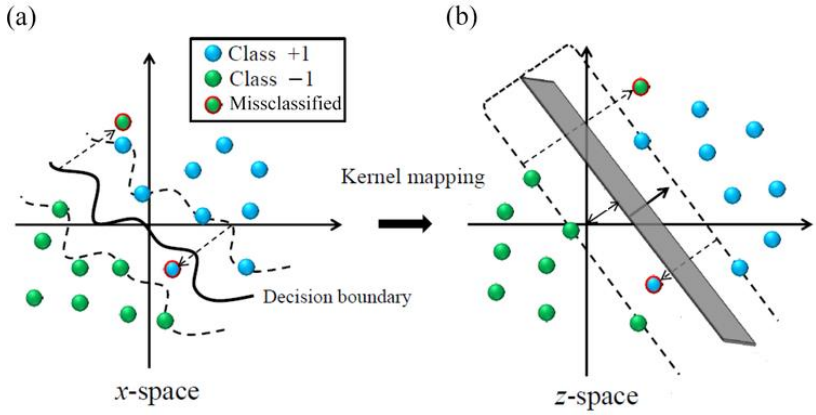

Fig. 5. A problem solved by mapping the training set [22]

As figure 5 shows, a problem that was unsolvable by using linear classification can be classified if training set data moved to a higher dimensional space. After understanding the theoretical part, it is now possible to explain the mathematical calculations behind the SVM method.

Assume that linear separability sample set is (x_i, y_i) with training data sets of:

$$i = 1, \dots, n, x \in R^d, y \in \{+1, -1\}$$

During this research, it's assumed that {1} is the normal and {-1} is the attacked or abnormal. Which leads to the equation of hyperplane classification as follows:

$$w.x + b = 0 \quad \text{---------- (1)}$$

In equation (1), the vector w is a normal vector while b is offset value. Initially we consider that the all the node is good and normal node. The best classifying hyperplane is supported by training data samples. While having this statement in mind, support vectors can be considered as the hyperplane training samples. Moreover, the formulation of the problem will be as follows:

$$\min \emptyset(w) = \frac{1}{2}\|w\|^2 = \frac{1}{2}(w.w)$$
$$subject\ to\ y_i[(w.x_i) + b] - 1 \geq 0, i = 1,2,\dots,n \quad \text{------(2)}$$

Hence, a formulation of the classification function will be as follows:

$$f(x) = sgn\{(w^*.x) + b\} = sgn\{\sum_{i=1}^{n} a_i^* y_i (x_i.x) + b^* \quad \text{--- (3)}$$

And a formulation of the optimal classification function will be as follows:

$$f(x) = sgn\{\sum_{i=1}^{n} a_i^* y_i k(x_i.x) + b^* \quad \text{---------- (4)}$$

The function mentioned above $f(x)$ is kernel function while ai are function multipliers.
In our implementation, the nodes are connected to each other. Specifically, a node connects to a single neighbor node. When all nodes are connected, the optimal hyperplane will be calculated through the previously explained functions and all data from the nodes will be classified into either a normal node or attacked/abnormal node. This process is possible with the use of SVM because of the method ability to classify high-dimensional data.

4. SGN ASSUMPTIONS AND IMPLEMENTATION SCENARIO

In this paper, we considered the following assumptions to implement the methodology:

1- The end used will specify the area of interest. Area of interest has been modelled as a grid Ω of Nx × Ny points scenario. The specified area is given as $A = Nx \times Ny$. Where Nx is the area length in meters (X-Axis) and Ny is the width in meters (Y-Axis) giving the product of the area A.
2- Nodes are sensors that are stationary after deployment (generation of network) and it can be said that nodes are the smart meters that are located in all consumers participating in SGN. Nodes are the communication channel between service provider and consumers and are responsible for collecting and forwarding the monitored data to the central node illustrated previously in Fig. 2.
3- Nodes communicate with Neighbour nodes in a pre-set radio range of (0.25 m2) and to the central node.
4- SVM based algorithm is responsible for classification of nodes.
5- The network is assumed to be synchronized.

The hypothetical scenario was considered from one of the village –AFI- in Al Batinah South Governorate, sultanate of Oman. The Area A in the simulation was set by default to N_x of 500 (m) and a N_y of 500 (m) and The default setting of 75 nodes represents smart meters in households in the shown area above. Average electricity consumption set by default to 30 Kwh. Data collected from electricity provider [23] in the area mention above. The monthly bill is also set to a default 250 Omani Riyals calculated using the online bill calculated provided by the service provider [24].

5. RESULTS

The method was simulated based on the hypothetical scenario considered for the implementation. In order to create the scenario, we have obtained the data about the average electricity consumption of the inheritance of the subscribers from the electricity supplier [23] [24]. The Average electricity bill was set as a base to simulate the network. In our simulation, the basic parameter was set are as follows:

Table 1. Parameters

Parameters (components)	Used values
Number of nodes	75 node
Number of central nodes	1 node
Average electricity consumption	30 Kwh
Average monthly electric bill	250 OMR

In the evaluation process for the effectiveness of the implemented model, we have considered a set of matrices to determine the detection of the attacks.

a) Detection Rate: This is the detection percentage of the attacks based on the total number of attack was performed
b) False positive rate (false alarms): This is the ratio between the number classified as an abnormal node (which is considered as an attacked node) on the total number of normal connections.

The simulation in MATLAB gave us the attack detection accuracy of 98% and the False alarms rate as low as 2% from the total number of attacks. The simulation result is in the figures 7 and 8 shows the generated network and the distribution of the nodes. Fig 8 is the results when we detect the malicious based on the algorithm implemented.

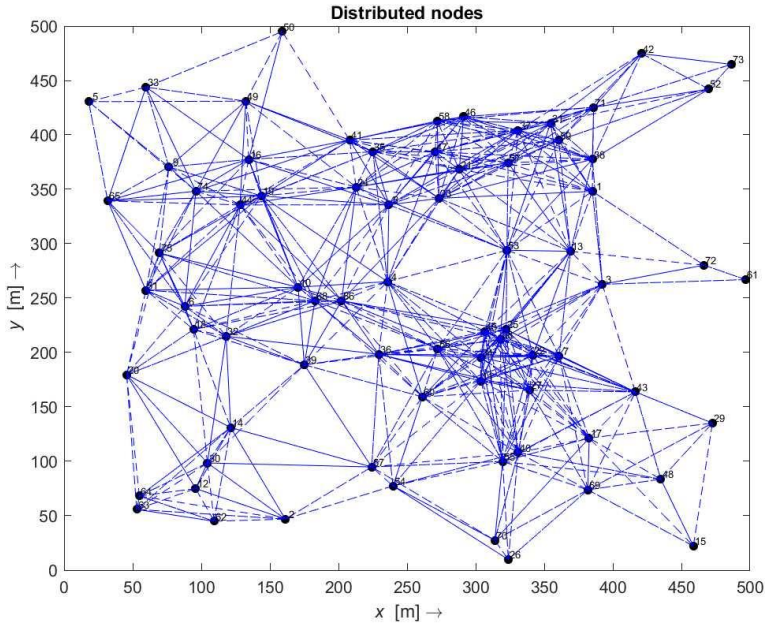

Fig. 7. The SGN Network

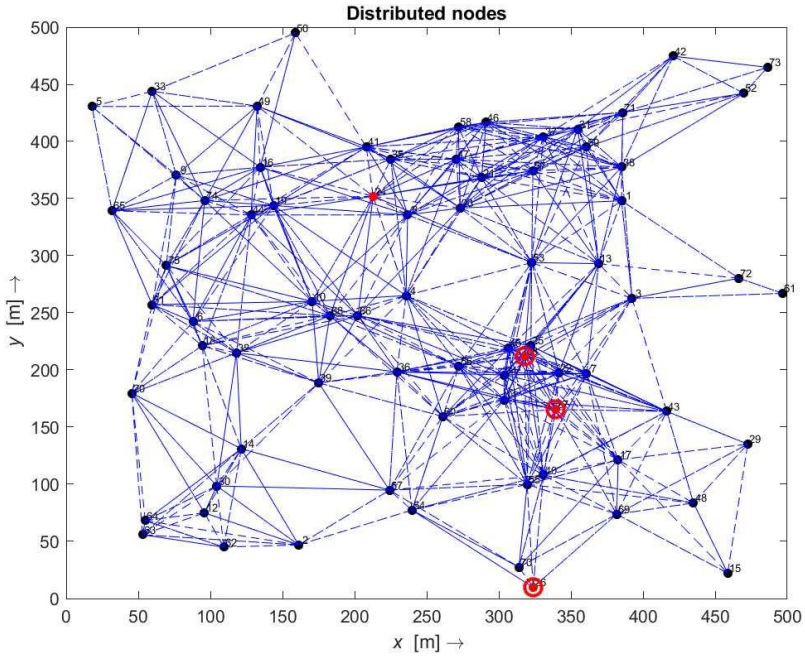

Fig. 8. Detection Malicious Nodes

In the Fig. 8, we can see that there are 75 nodes was distributed in the area and it was randomly checked. Based on the specified parameter the simulation results found 4 malicious nodes. The malicious nodes as circled red.

6. CONCLUSION

Smart Grid Network is backbone infrastructure is the information and communication technology that makes the network vulnerable to malicious attacks. It is essential to detect the attack work on the attack for the uninterrupted and effective supply of the electricity and generate the accurate bill. In this paper, the machine learning approach has been adopted. SVM makes memory efficient and effective for high dimension data. Considering this SVM-based classification framework of machine learning is implemented to detect misbehaving malicious nodes in smart grid networks. The simulation result in MATLAB gave us an effective detection outcome. The result shows us that our detection rate is about 90% and the false positive is only 2%. In future, we would like to simulate the network on a larger scale and implement it at the hardware level.

REFERENCES

[1] J. B. Ekanayake, N. Jenkins, K. Liyanage, J. Wu, and A. Yokoyama, *Smart Grid: Technology and Applications*. John Wiley & Sons, 2012.

[2] F. Skopik and P. Smith, *Smart Grid Security: Innovative Solutions for a Modernized Grid*. Elsevier Science & Technology Books, 2015.

[3] C. Greer *et al.*, "NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 3.0," National Institute of Standards and Technology, NIST SP 1108r3, Oct. 2014. doi: 10.6028/NIST.SP.1108r3.

[4] M. R. Ahmed, S. M. Tahsien, M. Aseeri, and M. S. Kaiser, "Malicious attack detection in underwater wireless sensor network," in *2015 IEEE International Conference on Telecommunications and Photonics (ICTP)*, Dhaka, Bangladesh, Dec. 2015, pp. 1–5, doi: 10.1109/ICTP.2015.7427952.

[5] P. Langley, *Elements of Machine Learning*. Morgan Kaufmann, 1996.

[6] S. Suthaharan, "Support Vector Machine," in *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*, S. Suthaharan, Ed. Boston, MA: Springer US, 2016, p. 1.
[7] "Fig 1: A conceptual illustration of a generic WSN," *ResearchGate*. https://www.researchgate.net/figure/A-conceptual-illustration-of-a-generic-WSN_fig1_301241534 (accessed Jan. 13, 2021).
[8] A. Takiddin, M. Ismail, U. Zafar, and E. Serpedin, "Robust Electricity Theft Detection Against Data Poisoning Attacks in Smart Grids," *IEEE Transactions on Smart Grid*, pp. 1–1, 2020, doi: 10.1109/TSG.2020.3047864.
[9] L. Zhang, X. Shen, F. Zhang, M. Ren, B. Ge, and B. Li, "Anomaly Detection for Power Grid Based on Time Series Model," in *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, Aug. 2019, pp. 188–192, doi: 10.1109/CSE/EUC.2019.00044.
[10] J. Jiang and Y. Qian, "Defense Mechanisms against Data Injection Attacks in Smart Grid Networks," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 76–82, Oct. 2017, doi: 10.1109/MCOM.2017.1700180.
[11] W. Zhe, C. Wei, and L. Chunlin, "DoS attack detection model of smart grid based on machine learning method," in *2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*, Jul. 2020, pp. 735–738, doi: 10.1109/ICPICS50287.2020.9202401.
[12] X. Xia, Y. Xiao, W. Liang, and M. Zheng, "GTHI: A Heuristic Algorithm to Detect Malicious Users in Smart Grids," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 2, pp. 805–816, Apr. 2020, doi: 10.1109/TNSE.2018.2855139.
[13] S. P. Nandanoori, S. Kundu, S. Pal, K. Agarwal, and S. Choudhury, "Model-Agnostic Algorithm for Real-Time Attack Identification in Power Grid using Koopman Modes," in *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, Nov. 2020, pp. 1–6, doi: 10.1109/SmartGridComm47815.2020.9303022.
[14] Y. S. Patil and S. V. Sankpal, "EGSP: Enhanced Grid Sensor Placement Algorithm for Energy Theft Detection in Smart Grids," in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, Mar. 2019, pp. 1–5, doi: 10.1109/I2CT45611.2019.9033759.
[15] Y. He, G. J. Mendis, and J. Wei, "Real-Time Detection of False Data Injection Attacks in Smart Grid: A Deep Learning-Based Intelligent Mechanism," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2505–2516, Sep. 2017, doi: 10.1109/TSG.2017.2703842.
[16] N. Cristianini, J. Shawe-Taylor, and D. of C. S. R. H. J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.
[17] S. S. Keerthi and C.-J. Lin, "Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel," *Neural Computation*, vol. 15, no. 7, pp. 1667–1689, Jul. 2003, doi: 10.1162/089976603321891855.
[18] "Fig. 4. SVM classification with a hyperplane that maximizes the...," *ResearchGate*. https://www.researchgate.net/figure/SVM-classification-with-a-hyperplane-that-maximizes-the-separating-margin-between-the-two_fig3_221926953 (accessed Jan. 13, 2021).
[19] G. C. Calafiore and L. E. Ghaoui, *Optimization Models*. Cambridge University Press, 2014.
[20] J. Sullivan, "Neural Network from Scratch: Perceptron Linear Classifier," *John Sullivan*, Aug. 16, 2017. https://jtsulliv.github.io/perceptron/ (accessed Jan. 13, 2021).
[21] R. Grosse, "Lecture 3: Linear Classification," p. 10.
[22] "Figure 1. Graphical presentation of the support vector machine...," *ResearchGate*. https://www.researchgate.net/figure/Graphical-presentation-of-the-support-vector-machine-classifier-with-a-non-linear-kernel_fig1_299529384 (accessed Jan. 13, 2021).
[23] "Pages - Home." https://mzec.nama.om/en-us/Pages/home.aspx (accessed Mar. 19, 2021).
[24] "Bill Calculator." https://mzec.nama.om/en-us/Pages/billcalculator.aspx (accessed Mar. 19, 2021).