Data-driven Detection of Electricity Theft Cyber-attacks in PV Generation

M. F. Shaaban, *Senior Member, IEEE*, Usman Tariq, *Member, IEEE*, Muhammad Ismail, *Senior Member, IEEE*, Nouf A. Almadani, and Mohamed Mokhtar, *Member, IEEE*

***Abstract*—Most of the existing research focuses on electricity theft cyber-attacks in the consumption domain. On the contrary, a high penetration level of distributed generators (DGs) may result in increased electricity theft cyber-attacks in the distributed generation domain, which is the focus of this paper. In these attacks, malicious customers can hack into the smart meters monitoring their DG units, which are usually photovoltaic (PV), and manipulate their readings to report higher injected energy to the grid and claim more profit under feed-in tariff programs. This paper proposes a data-driven approach based on machine learning to detect such thefts. We adopt an anomaly detection approach where a theft detection unit (TDU) based on a regression tree model is designed to detect suspicious data. Historical records of solar irradiance, temperature, and smart meter readings are utilized in the training stage of the detector. The probability density function of the error between the actual readings from DG meters and the predicted generation by the regression model is utilized as a metric to detect suspicious data. Several theft scenarios are used to assess the performance of the TDU. Furthermore, a comparison study with other detectors is presented to demonstrate the superiority of the proposed TDU.**

***Index Terms*—Cyber-attacks, electricity theft, machine learning, photo-voltaic, smart grid.**

I. Introduction

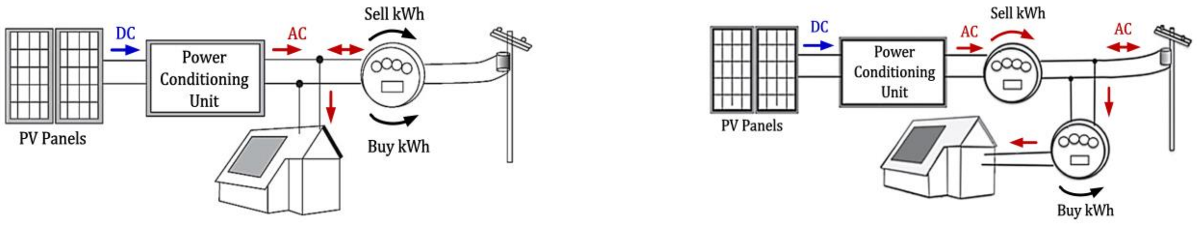
Electricity theft is a vital problem that has resulted in huge financial losses for utility companies in many countries worldwide [1], [2]. According to the Federal Bureau of Investigation (FBI) and International Utilities Revenue Protection Association, energy theft causes financial losses to the electric utility estimated at $6 billion, in the U.S. Electricity theft in the consumption domain is usually committed by conventional tampering of energy meters or rewiring the grid connection. Smart grids keep growing and consequently resulting in new forms of energy theft [3], [4]. Advanced metering infrastructure (AMI) is the backbone of smart grids. It consists mainly of smart energy meters with advanced communication capabilities [5]-[7]. Malicious customers can launch cyber-attacks on these meters to manipulate their reported consumption and hence reduce their bills.

In this context, the smart grid paradigm encourages customers to install their own distributed generation (DG) units to generate energy, sell it back to the grid, and then gain a profit. DG units are usually photovoltaic (PV) panels. Feed-in tariffs (FITs) policy and the net metering system are two approaches adopted by the electric utilities to encourage the customers to invest in renewable energy technologies. In the net metering system, customers feed the excess of the generated solar energy to the grid and receive a reduction on the next bill [8]. Hence, the net metering system requires only one bidirectional meter as shown in Fig. 1(a). On the contrary, FIT is referred to as clean energy cashback, where customers sell all generated energy from PV and get paid for this energy from the grid [9]. The FIT policy is more attractive than the net metering system to encourage customers to produce green energy [10]. The FIT scheme requires two meters. One meter is dedicated to monitoring the energy generated from PV, which is the selling energy to the electric utility, and the other meter is dedicated to monitoring the energy consumed by the customer, as shown in Fig. 1(b). In FIT policy, malicious customers can exploit the electric utility through manipulating the reported energy generation data to claim higher energy generation injected to the grid and hence gain more profits. The weak authentication firmware installed in most of these meters is the main reason for such theft cases where the customers have access to the firmware using the ANSI optical port of these smart meters [12]-[14].

Electricity theft detection has been previously investigated by researchers. However, the scope of most of the previous research work did not consider electricity theft in the distributed generation domain [15], [16], which is considered a pressing problem. The main difference between cyber-attacks applied at the consumption domain or applied at the distributed generation domain is that in the consumption domain, these attacks aim to reduce the consumption bill. On the other hand, the attacks in the distributed generation domain aim to increase the reported generation energy injected to the grid to claim more profits. The work in [15] presented a detector based on the least square error (LSE) and a moving time window to detect the theft in PV panel’s generation. In [16], the authors presented a detector based on Auto Regression Integrated Moving Average (ARIMA) and Kullback-Leibler divergence (KLD) to detect the manipulation in PV readings.

On the contrary, most of the existing research focused mainly on electricity theft detection in the consumption domain [1], [17]-[23]. The work in [17] presented a detector based on an artificial neural network (ANN) to detect suspicious load profiles of customers. They assumed several scenarios of cyber-attacks like assuming the attacker will reduce the consumption by a random amount for each time slot or will reduce the consumption by a fixed number for a specific period. In [18], an electricity theft detector based on random matrix theory (RMT) with cost-effective Distributed Meter Data Management (DMDM) solution was developed. The authors in [1] used some classification and clustering techniques to find the probability of energy theft and suspicious clients were identified by

https://doi.org/10.1109/JSYST.2021.3103272



(a) (b)

Fig. 1. Grid-connected PV systems (a) Net metering scheme and (b) FIT scheme [11].

monitoring abnormalities in consumption patterns. In [19], non-technical losses (NTL) problem was presented, where a fraud detection model (FDM) based on support vector machine (SVM) was developed to extract suspicious customers based on an irregular consumption pattern. In [20], a detection technique based on partially observable Markov decision process (POMDP) and Bollinger bands were presented. In [21], an energy theft detection scheme was proposed using energy privacy preservation in the smart grid network. Furthermore, the combined convolutional neural networks (CNN) technique was utilized to detect abnormal measurements within smart meter data. The authors in [22] presented a comprehensive top-down scheme based on a decision tree (DT) and a SVM. In [23], the authors studied stealth false data injection (FDI) attacks in the consumption domain. It presented a set of restricted Boltzmann machines (RBMs) to detect such theft attacks.

The work in [24] introduced extreme learning machine to detect which busses of a power system are under FDI attacks. Furthermore, a recovery strategy was introduced to address the detected erroneous data. The authors of [25] proposed a machine learning framework based on two-stage approach to identify and locate the cyber-attacks on the control systems in the distribution domain. The work in [26] introduced a federated deep learning algorithm to detect false data in industrial systems. Moreover, the authors of [27] used a feed-forward deep neural algorithm based on wrapper feature extraction unit while authors of [28] used a cloud-based cyber-physical intrusion detection algorithm. The authors of [29] proposed a sequential ensemble detector based on a deep auto-encoder with attention (AEA) to detect various cyberattacks.

This paper focus is detecting electricity theft cyberattacks in the distributed generation domain by developing a theft detection unit (TDU) to detect suspicious data flow to increase the reported injected energy under the FIT programs. More specifically, the proposed TDU tackles the theft incidents caused by cyberattacks to manipulate the PV smart meters readings installed at the customer's premises.

Our early work in [30] focused on developing supervised learning classifiers that are trained (and tested) on both benign and malicious data. Moreover, [30] assumes that a malicious database already exists or can be synthetically created to train the detector However, malicious data may not be known in advance during the training stage of the detector, which is especially true for zero-day attacks, which are the attacks never happened before. Hence, in this work, we aim to develop an unsupervised anomaly detector that is trained only using benign data, which can be collected by the operator during the system normal operation but can be tested on both benign and malicious data. Such an anomaly detector overcomes the challenge associated with the availability of malicious data during the detector's training stage and it stands robust against zero-day attacks. Moreover, malicious data are not known until they are detected to be used in the training stage. However, all malicious data points used in training of the classifier detectors presented in the literature are based on a simulated dataset, i.e., not based on real practical data accurately collected from the system under attack. Therefore, it is better to develop an unsupervised anomaly detector based on benign data only, which tries to capture the characteristics of benign data, and then it can detect any deviation from this benign data. Therefore, our proposed detector compared to our previous work depends on different assumptions and different data availability like the unavailability of malicious data.

The main contributions of this work are summarized as follows:

- We have investigated an anomaly detector. This detector is trained only on benign data. However, it can be tested on both benign and malicious data. Therefore, the developed detector is a general detector that can be used to detect the presence of electricity theft cyber-attack for any PV unit in the system under any cyber-attack functions.
- We investigate the integration of various data sources to develop a machine learning-based electricity theft detection system to detect electricity theft cyber-attacks in solar panels. These data sources include the readings from PV smart meters and meteorological data.
- The proposed anomaly-based TDU presents two sub-stages during training, first a regression stage and an error extraction stage. In the regression sub-stage, the predicted energy generation is specified based on a regression tree. Then, a probabilistic measure is carried out at the error extraction stage to aid in the theft decision.
- In the error extraction sub-stage, the proposed detector relies on the probability density function (PDF) of the error between the reported readings from PV meters and the predicted energy. Thus, the proposed TDU presents not just a classification but a probabilistic measure of the suspiciousness of the malicious data during the test (deployment) stage.
- Simulation results are carried out to evaluate the performance of the proposed TDU and compare its performance against other detectors based on SVM, ARIMA, and LSE models.

The rest of the paper is organized as follows: The problem statement and proposed methodology are explained in Section II. Results and multiple case studies are presented and discussed in Section III. Finally, the conclusions are presented in Section IV.

II. Problem Statement and Proposed Methodology

The objective of this research is to develop a machine learning-based TDU to detect the suspicious data flow reported by the customers following the procedure shown in Fig. 2. In this section, we present the data preparation stage, the training stage, and the theft detection mechanism.

A. Data Preparation

One of the purposes of this work is to integrate data sources in the training process of the machine learning-based detector. These data sources include the readings from PV smart meters and meteorological data (solar irradiance and temperature). Moreover, PV smart meters are mainly affected by the injection from the PV units installed in the downstream, and this assists in confronting the dynamics of the power system. The first and vital step as shown in Fig. 2 is to gather and prepare these data to feed it to the regression model to learn the behavior of the PV panels. Regression is used to find the relationship between one or multiple independent variables called predictors and a single dependent variable called the response or target variable. As shown in Fig. 2, three independent variables will be used as the predictors to predict a single dependent variable or response variable. The three predictors are time, solar irradiance, and ambient temperature while the reading from PV smart meters is considered as the response variable.

In order to create the required data set that includes these readings, historical solar irradiance and temperature data from a weather station in Toronto, Canada are utilized. In this research, we study the behavior of N_C customers with different number of panels and types of panels. To simulate realistic cases, N_{PV} different PV panels types are considered [31] with different capacities and characteristics, as shown in Table I, where $N_{PV} = 11$. To investigate the robustness and generalization ability of the proposed TDU, $N_{PV} - 1$ types will be used for training and a completely different one will be used for testing to prove that our method is general and can work on any panel type, even ones that it was not trained on. These types are randomly assigned to each one of the N_c customers following a discrete uniform distribution. Also, to specify the installed capacities, the number of PV panels installed per customer is randomly selected from N_{panel}^{min} to N_{panel}^{max} panels following the same distribution.

The generated power from the PV panels and hence the output energy for each customer can be calculated using the historical data and the parameters of panel-related characteristics, which represent the readings provided by the PV smart meters. The output power for each customer should be calculated to be used in the training of the proposed TDU in the next step and considered as virtual historical data of the PV smart meter readings. Using the datasheet characteristics of the PV panels, which are shown in Table I, the historical solar irradiance, temperature data, and the relations (1)-(5) [32], the output powers profile of each type of the PV panels can be generated.

$$T_{\mathrm{CELL}} = T_{\mathrm{A}} + \frac{S(T_{\mathrm{NOCT}} - 20)}{0.8\,\mathrm{kW/m^2}} \tag{1}$$

$$I_{\mathrm{PV}} = S[I_{\mathrm{SC}}(1 + K_{\mathrm{I}}(T_{\mathrm{CELL}} - 25))] \tag{2}$$

$$V_{\mathrm{PV}} = V_{\mathrm{OC}}(1 + K_{\mathrm{V}}(T_{\mathrm{CELL}} - 25)) \tag{3}$$

$$FF = \frac{V_{\mathrm{MPP}} I_{\mathrm{MPP}}}{V_{\mathrm{OC}} I_{\mathrm{SC}}} \tag{4}$$

$$P_{\mathrm{PV}} = FF\, V_{\mathrm{PV}}\, I_{\mathrm{PV}}, \tag{5}$$

TABLE I
Characteristics of the 11 PV panels

Type	Max Power (W)	T_{NOCT} (ºC)	I_{MPP} (A)	V_{MPP} (V)	V_{OC} (V)	I_{SC} (A)
1	435	45	5.97	72.9	85.6	6.43
2	245	46	8.11	30.2	37.8	8.63
3	87.5	45	1.78	49.2	61	1.98
4	230	47	6	40.2	50.7	6.7
5	135	45	2.88	47	61.3	3.41
6	240	47	4.86	49.38	59.23	5.44
7	245	47	4.95	49.51	59.45	5.54
8	250	47	5.01	49.91	59.92	5.61
9	255	47	5.09	50.11	60.36	5.70
10	260	47	5.17	50.30	60.36	5.79
11	265	47	5.25	50.48	60.60	5.88

where, T_{CELL} is the cell temperature; T_A is the ambient temperature; S is the solar irradiance; T_{NOCT} is the nominal operating cell temperature at 20°c and 0.8 kW/m^2 irradiance; I_{PV} and V_{PV} are the current and the voltage of the PV module; K_I *and* K_V are the current and the voltage temperature coefficients; I_{MPP} *and* V_{MPP} are the current and the voltage of PV module at maximum power; FF is the fill factor, and P_{PV} is the output power.

The parameters in Table I are not available for the electric utility except for the PV panel capacity. This motivates data-driven approaches to detect electricity theft at the generation side.

The readings of PV smart meter must be normalized to the installed capacity of each customer to avoid any sort of bias and produce to have proper training. Fig. 3(a) shows a sample of the output power waveform on the same day for two customers with type 1 and type 2 PV panels before normalizing, whereas Fig. 3 (b) and (c) shows the normalized output power.

Moreover, all the zero generated power data points at night, where the irradiance is zero, are removed from the data. Detecting a theft during night does not need a smart classifier; in addition, the existence of these data points produced huge bias in the TDU, which should not be a basis for performance evaluation.

Finally, the outcomes from the data preparation stage are the three predictors, readings of PV smart meters, and the installed capacity for each customer, as shown in the top section of Fig. 2.

B. Regression Model Training

The model should learn how to predict the readings of PV smart meters given the three predictors. Hence, the second step is training the regression model with the historical datasets of PV generation, as shown in the left-side of Fig. 2. In this research, historical data of solar irradiance and temperature on an hourly basis are utilized to generate virtual historical data of PV smart meter readings at these conditions as mentioned before. The detector is only trained using benign data as it is difficult to get real malicious data, which makes this anomaly detector more robust against zero-day attacks.

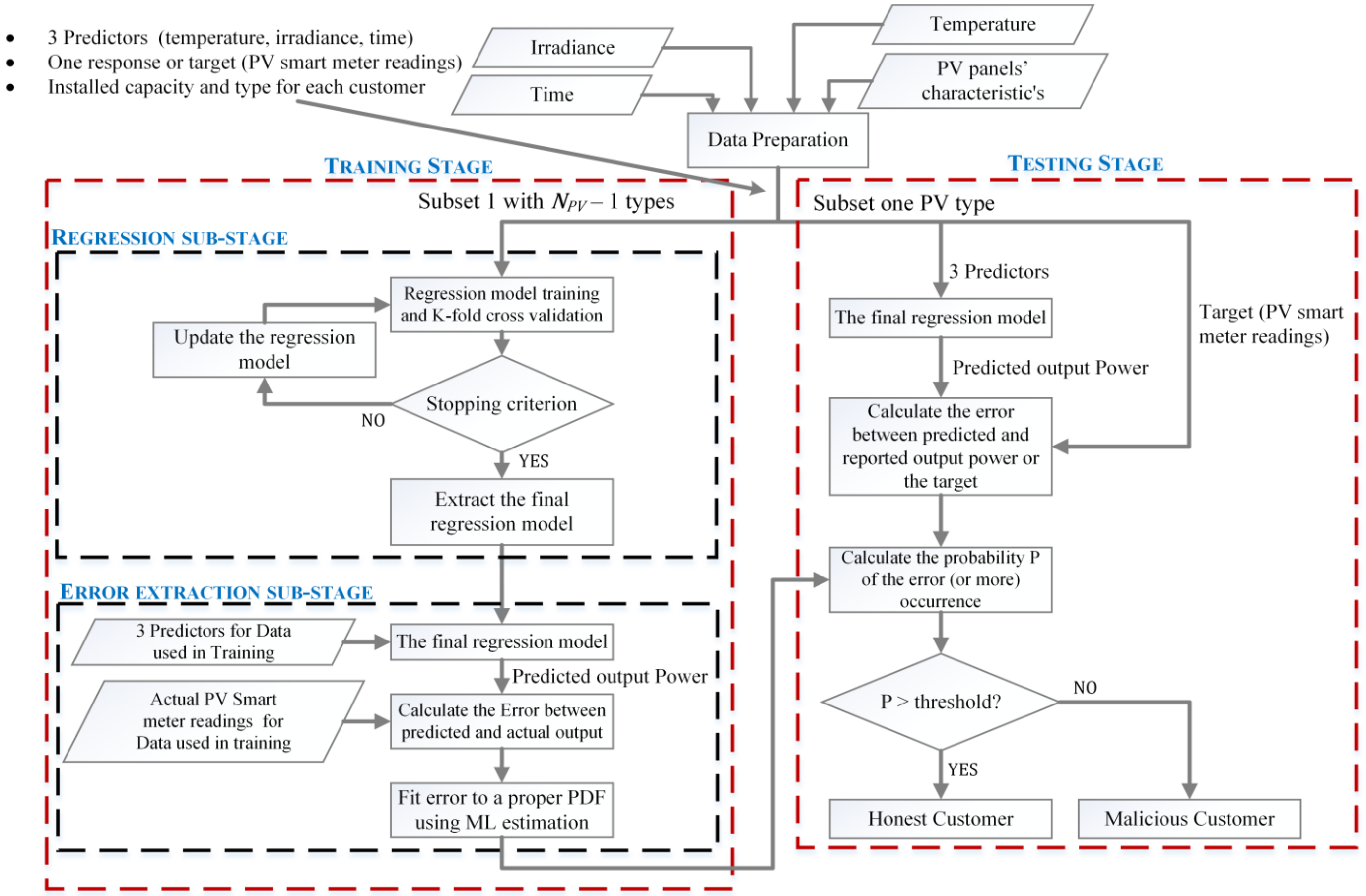
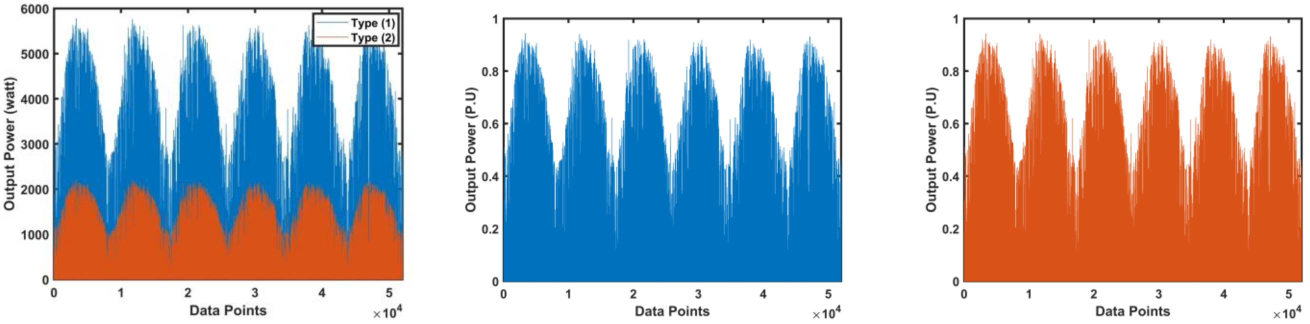

• 3 Predictors (temperature, irradiance, time)
• One response or target (PV smart meter readings)
• Installed capacity and type for each customer
Irradiance
Time
Temperature
PV panels' characteristic's
Data Preparation
TRAINING STAGE
TESTING STAGE
Subset 1 with $N_{PV} - 1$ types
Subset one PV type
REGRESSION SUB-STAGE
Regression model training and K-fold cross validation
Update the regression model
Stopping criterion
NO
YES
Extract the final regression model
ERROR EXTRACTION SUB-STAGE
3 Predictors for Data used in Training
The final regression model
Predicted output Power
Actual PV Smart meter readings for Data used in training
Calculate the Error between predicted and actual output
Fit error to a proper PDF using ML estimation
3 Predictors
The final regression model
Predicted output Power
Target (PV smart meter readings)
Calculate the error between predicted and reported output power or the target
Calculate the probability P of the error (or more) occurrence
P > threshold?
NO
YES
Honest Customer
Malicious Customer

Fig. 2. The Proposed methodology flowchart.



(a) (b) (c)

Fig. 3. (a) Output power of 2 different customers' installed panels, (b) and (c) the normalized output power of the same panels.

The dataset is divided into two subsets. The first subset with the known predictors and known response for the first $\boldsymbol{N_{\mathrm{PV}} - 1}$ PV panel types is used to train different regression models, as shown in Fig. 2. The second subset will not be utilized in the training stage and will be used later in the testing stage to investigate the generalization ability of the proposed TDU.

K-fold cross-validation is used in the training stage of the regression models and will be discussed in the following subsection. Various regression models are investigated during the training stage such as linear regression, SVM regression, and DT regression. The validation error of each model will be utilized to select the best regression model to be used in the TDU. Hence, the model with the least error will be selected for developing the proposed detector.

1) *K-fold Cross-Validation*

K-fold cross-validation is one of the statistical techniques used to evaluate and estimate the misclassification error in machine learning models [33]. This technique is widely used for its simplicity; besides, it can prevent overfitting. In K-fold cross-validation, the dataset assigned for training is divided into K folds. K-1 folds are used to train the model while a single fold is used to test the model. This procedure is repeated multiple times with different folds to ensure that each fold has been used exactly once to validate and test the model. In this research, 5-fold cross-validation method is utilized in the training stage, where the dataset assigned for training has been divided into 5 equal folds. The cross validation is used to compare different models for regression. Then, we choose a model and perform anomaly detection on the data subset not used in this stage.

2) Regression models

Different regression models are utilized in the training stage to select the best model for the proposed TDU. The major models investigated in this work are linear regression, SVM, and DT.

a) Linear Regression

The objective of linear regression is to make the best possible fit regarding the relationship between the predictors and the response variable. A nonlinear relationship can be investigated between the predictors and response variable. However, the relationship between the response variable and the regression model coefficients should be linear. The function representing the response or target as a function of predictors can be written as follows:

$$\hat{P} = \beta^T x + b, \tag{6}$$

where $\hat{P}$ is the predicted response or target which is the predicted output power from PV; $x = [x_1\ x_2\ x_3]^{\mathrm{T}}$ is a vector containing the three predictors which are temperature; irradiance; and time variables; T is the transpose operator; $\beta = [\beta_1\ \beta_2\ \beta_3]$ and b (bias) represents the linear model coefficients to be determined.

The model coefficients are obtained to minimize the LSE between the actual power and the predicted power as expressed in the following equation

$$\min\left(P - \hat{P}\right)^2 = (P - (\beta^T x + b))^2, \tag{7}$$

where P is the actual output power from PV in (5).

Different linear regression techniques are available. For example, stepwise regression systematically adds or removes variables in the linear regression model. The algorithm does this variable selection based upon their statistical significance in explaining the output variable. On the other hand, the robust linear regression is less affected by outliers in the data, whereas basic linear regression gives simply the least square fit, without other intricacies.

b) SVM Regression

In SVM regression, the same function used to predict the response variable in linear regression is utilized. However, an acceptable error is defined in the model, and SVM will determine proper line or hyperplane to fit all data. The objective of SVM regression is to minimize the squared coefficients, not the squared error. The predicted output with an acceptable error ε, the objective function, and the error constraint can be expressed using equations (6) (8), and (9), respectively.

$$\min_{\beta} \left(\frac{1}{2}\ |\beta\, \beta^{\mathrm{T}}|\right) \tag{8}$$

subject to:

$$|P - \hat{P}| \leq \varepsilon, \tag{9}$$

If no function exists to satisfy the previous constraints for all points, a slack variable, ε_n for data point n, can be added to each point. Then, the equations can be written as follow

$$\min_{\beta} \left(\frac{1}{2}\ |\beta\, \beta^{\mathrm{T}}| + C\sum_{n=1}^{N} \varepsilon_n\right) \tag{10}$$

subject to:

$$|P - \hat{P}| \leq \varepsilon + \varepsilon_n, \tag{11}$$

where N is the total number of data points or observations and C is a regularization parameter.

The previous objective function is mathematically simpler to solve in its Lagrange dual formulation. The Lagrange dual formulation is obtained from the original function by defining negative multipliers α_n and γ_n for each data point or observation n. Therefore, the new minimization problem can be described as follows

$$\min_{\alpha,\gamma}\left(\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}(\alpha_n - \gamma_n)(\alpha_m - \gamma_m)x_n x_m^{\mathrm{T}} + \varepsilon\sum_{n=1}^{N}(\alpha_n + \gamma_n) + \sum_{n=1}^{N} P_n(\gamma_n - \alpha_n)\right) \tag{12}$$

subject to:

$$\sum_{n=1}^{N}(\alpha_n - \gamma_n) = 0 \quad \forall n \tag{13}$$

$$0 \leq \alpha_n,\ \gamma_n \ \leq B \quad \forall n, \tag{14}$$

where B is the box constraint.

The modified function used to predict the response variable is described in (15), where $\beta = \sum_{n=1}^{N}(\alpha_n - \gamma_n)\, x_n$.

$$\hat{P} = \beta_0 + \beta\, x^{\mathrm{T}} \tag{15}$$

c) DT Regression

DT learning is considered a predictive modeling approach that is used as one of the supervised machine learning techniques. Both classification and regression problems can be addressed using the DT by splitting the data based on a learned set of parameters. The main idea is that the prediction space is divided into a homogenous subset (non-overlapping regions). Different algorithms are used to construct the decision trees and determine the number of regions. These include approaches such as classification and regression tree (CART) and Iterative Dichotomiser 3 (ID3).

The basic idea in learning a DT for regression is that the prediction space is divided into M regions during training, where M is determined to minimize some error metric. A popular error metric is the mean square error (MSE) between the predicted responses and the actual responses. During testing, the same regions are then used for prediction.

In our case, we have three predictors: temperature, irradiance, and time. We can store this information in the form of a three-dimensional vector for each data point, $x = [x_1\ x_2\ x_3]^{\mathrm{T}}$. Here, T denotes a transpose. If $\hat{P}^m$ is the predicted response or target (i.e. predicted output power from PV) for a point in a region m and $P^{i,m}$ is the actual response (i.e. actual output power from PV) for the ith input point lying in the region m, we seek to form a partition of the three-dimensional space (as we have three predictors in our problem) which minimizes,

$$min\sum_{m}\sum_{i\in m}(P^{i,m} - \hat{P}^m)^2 \tag{16}$$

where i denotes the index for training points and $m \in \{1,2,\ldots,M\}$ is the index of regions.

When training begins, the algorithm for learning the decision tree looks at the entire training set and then chooses the dimension (predictor) j and a split (threshold) s to split the three-dimensional space into two disjoint regions; one region contains the data points whose dimension j has the value less than or equal to s and the other region contains the points whose dimension j has the value greater than s. This choice of dimension j and split s is made so as to minimize,

$$\sum_{i:x_j \leq s} (P^{i,m} - \hat{P}^m)^2 + \sum_{i:x_j > s} (P^{i,n} - \hat{P}^n)^2 \tag{17}$$

where $\hat{P}^m$ denotes the predicted value for the points in region m (for which $x_j \leq s$) and $\hat{P}^n$ denotes the predicted value for the region n (for which $x_j > s$), and $P^{i,m}$ and $P^{i,n}$ are the actual response values for the data points in the two sub-regions. The predicted values for the regions m and n, i.e. $\hat{P}^m$ and $\hat{P}^n$, are essentially the average of the response values of the training points in the two respective regions. For instance for a region q, the predicted value for this region, $\hat{P}^q$, can be given as,

$$\hat{P}^q = \frac{1}{N_q} \sum_{i \in q} P^{i,q} \tag{18}$$

where, N_q is the number of training points in region q and $P^{i,q}$ is the actual response value for training point i in this region.

Each of the two regions, m and n, can then be further split into further sub-regions, so as to minimize the overall training error in (16). If these regions are not further subdivided into other regions, they are known as leaves, otherwise, they are known as nodes, which are then up for further sub-division. This process of training is also known as growing a DT.

Theoretically, one can keep on doing the sub-division process, until each leave contains only one data point or until all the data points in each leave have the same response value. This would result in a zero-training error. However, this may lead to over-fitting and the model may not be able to generalize well on the unseen testing data. Hence, one can use a validation set to decide to which level of sub-division one needs to go, to achieve a good performance on the testing set. This may lead to different types of DTs, such as coarse, medium, and fine DTs. The difference between coarse, medium, and fine DTs is that they respectively have few, medium, and many numbers of leaves that allow low, medium, and high model flexibility.

When we get a test point, we take the grown (trained) DT, follow a sequence of steps, based upon the sequence of splits in the DT, to decide which leaf does the test point falls in. We then assign the response value in that leave to the test point (this is essentially the average of the response values of training points in the leave during training). Here, we have described the CART approach for growing a regression DT. Similar ideas apply to other learning algorithms as well. For further details, the readers are referred to [34].

C. *Theft Detection Mechanism*

The theft detection decision is developed based on the PDF of the error between the predicted output power and the actual one, as shown in left-side of Fig. 2. The distribution of the error is identified to be used as a detection metric. A limit or threshold that indicates the acceptance range will be defined. Therefore, if the prediction error with respect to the customer's data is found to be beyond this limit or threshold, which represents a very small percentage of occurrence and unlikely to happen, then this reported data by the customer is suspicious. The main focus of this research is only on the positive probability indicating that the customer is reporting an injection more than the actual generation from PV.

As aforementioned, the probability of the error between the predicted output power and the PV meter readings will be utilized as a detection metric in the proposed TDU. Hence, the objective of this step is to fit this error extracted from the regression models with a proper PDF. To choose the best PDF, we look into the maximum likelihood for each type. Maximum Log-likelihood estimation is used to find the set of parameters $\hat{\theta}$ that maximizes the probability of occurrence of the data point x_n, as in (19).

$$\hat{\theta} = arg \max_{\theta} \ \sum_{n=1}^{N} log\ [Pr(x_n|\theta)]\, , \tag{19}$$

where $Pr(a)$ is the probability of event a.

Finally, to classify a customer, the reported output power from the smart meter will be compared to the predicted output power. The error between the measured output and the predicted output will be checked using the fitted PDF on the extracted error from the training data and an alarm will be triggered to indicate that theft is detected if the probability of occurrence of this error or higher is below a certain threshold. The larger the error, the smaller its probability to occur, which indicates suspicious data.

III. Results and Discussions

In this section, we present and discuss the results of some case studies to evaluate the performance of the proposed TDU using MATLAB. Various regression models are trained and the root-mean-square-error (RMSE) of each model is determined as shown in Fig. 4 for five-fold cross-validation as discussed earlier to select the most proper model to be used in developing the proposed TDU. 48 points for each customer from $N_c = 400$ customers are utilized to form a benign dataset used in training and testing of the proposed TDU. This benign dataset is divided into two sets. One set is used in the training stage of the proposed unsupervised anomaly detector while the other set is used in the testing stage of the proposed detector. 17712 benign samples are used in the training stage. Whereas, 8928 samples which represent the remaining benign data points and malicious datapoints after applying different cyber-attacks scenarios explained in the next subsection are used to test the performance of the proposed TDU. According to the RMSE, the fine tree model presents the best regression model as illustrated in Fig. 4, which is used for the TDU. Further, the error between the actual and predicted response is fitted with various distribution functions, as shown in Fig. 5. Comparing the likelihood or the log-likelihood (LL) of each distribution function to get the best fit, Beta distribution presents the best (highest) log-likelihood (LL = 37452.5) compared to 35635.9, and 37136.9 for Normal and Weibull distributions, respectively. Therefore, the regression and error extraction sub-stages of the proposed TDU are based on the fine tree algorithm and Beta distribution function, respectively.

A. *Cyber-attacks functions*

One of the challenges facing this research is the lack of data needed to represent malicious customers. In this research, an anomaly detector is developed where benign data is utilized only during the training stage. The detector tries to learn the normal patterns within benign data, and hence, can detect any suspicious deviations from this normal pattern as a sign of malicious behavior. However, in order to test the proposed TDU against honest and malicious customers, cyber-attack functions/scenarios need to be introduced to manipulate the PV smart meter readings in a way that imitates the theft behavior by malicious customers. Hence, a set of cyber-attack functions will be launched on the PV smart meter readings to generate synthetic malicious data to be used during the testing stage of the TDU. The cyber-attack functions claim higher PV energy generation supplied to the electrical grid to gain more profits.

Three cyber-attack functions, illustrated in Table II, are introduced. The first cyber-attack function performs a static percentage attack, in which the malicious customer manipulates the meter reading and reports an increase in the generated energy by a fixed percentage $(\boldsymbol{f_1})$ of the actual generated energy $\boldsymbol{E_{PV}}$ (for example, reporting 110% of the actual generation where $\boldsymbol{f_1 = 10}$%). The second cyber-attack function implements a dynamic percentage attack, in which the malicious customer reports a random increase in the generated energy by a random percentage $(\boldsymbol{f_2(t,d)})$ of the actual generated energy to make it difficult for the electric utility to recognize the abnormality. The third attack function implements a positive shift attack, where a malicious customer reports an increase in the generated energy by adding a fixed value $\boldsymbol{u}$ to the actual generated energy. Thus, the reported generation energy will be $\boldsymbol{u}$ when the actual generation is zero. However, in this case study similar to the previous cases, all the data points during the night are removed to detect the performance of the proposed TDU where detecting a theft during the night does not need a smart classifier. Moreover, the customers may increase the generation by a specific value during the daytime only and then remove this value at the night to avoid detection.

After developing the TDU, a threshold value should be introduced. Selecting the value of threshold will affect the TDU decision since selecting high value may result in detecting the honest customer as a malicious customer and vice versa. The threshold value corresponds to the probability that more than a certain error magnitude (between the predicted and actual power) is likely to be observed in the training dataset as shown in Fig. 6. A threshold corresponding to y% would mean that y% is the probability of observing an error magnitude more than a certain amount e_y. If the probability of the error calculated for a customer falls below y%, then the customer is flagged malicious. The more it falls, the customer becomes more suspicious to be stealing.

To find a proper value for the threshold, the set of data used in the training is fed to the TDU to predict the output power. Then, the predicted output power is compared with the reported readings from PV meter for different cases representing honest customer and static percentage attack with various attack levels f_1 to represent malicious customers. Then, the error of each case, as well as PDF of the error are determined. Finally, the threshold value is investigated to determine its impact on the decision of TDU for pre-known cases. As aforementioned, the training of the TDU is based solely on the benign data. The malicious data is introduced here only for testing the detector's performance. Table III shows the effect of varying the threshold value on the TDU decision. The results in Table III reveal that the best threshold values can be selected are 5% to 10%.

TABLE II
PROPOSED CYBER-ATTACK FUNCTIONS

Attack Type	Representation
Static percentage attack	$E_{\text{reported}} = (1 + f_1)\, E_{\text{PV}}$
Dynamic percentage attack	$E_{\text{reported}} = (1 + f_2(t,d))\, E_{\text{PV}}$
Positive shift attack	$E_{\text{reported}} = \; E_{\text{PV}} + u$

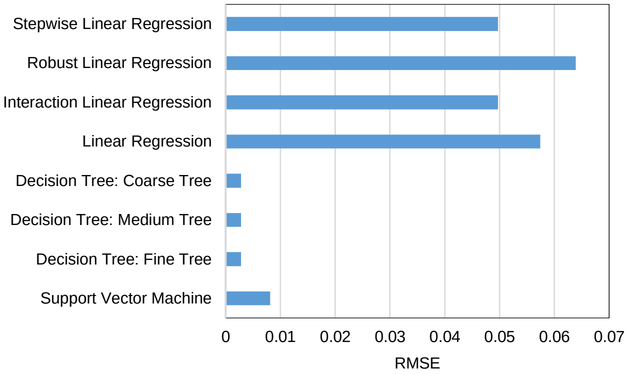


Fig. 4. RMSE of regression models.

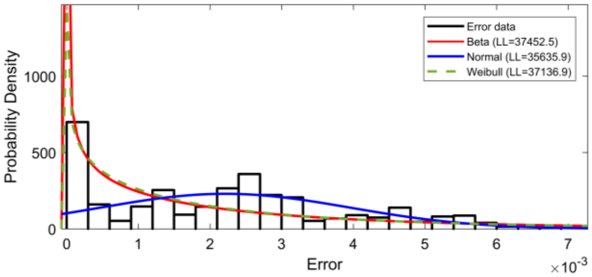


Fig. 5. Fitting of the error PDF

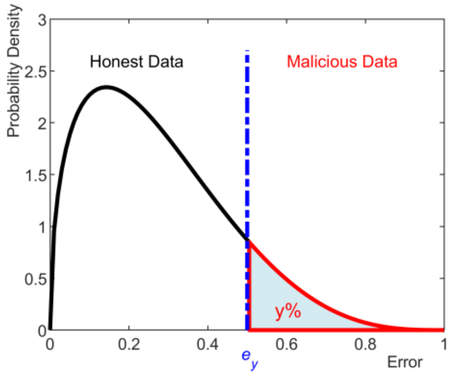


Fig. 6. Fitting of the error PDF

TABLE III
DECISIONS OF TDU BASED ON THRESHOLD VALUE VARIATIONS

		Threshold percentage (y%)				
		20%	15%	10%	5%	1%
Static attack %	0 %	Alarm	Alarm	No Alarm	No Alarm	No Alarm
	2.5 %	Alarm	Alarm	Alarm	Alarm	No Alarm
	5 %	Alarm	Alarm	Alarm	Alarm	Alarm
	10 %	Alarm	Alarm	Alarm	Alarm	Alarm

B. Case 1: Honest Customers

As mentioned in Section III-A, the second subset curtailed from the original dataset will be utilized to evaluate the performance of the proposed TDU. First, the three predictors of the curtailed data points: temperature, irradiance, and time are fed to the proposed TDU to predict the output power. Then, the PV smart meter reading that represents an honest customer is compared with the predicted output power results from the proposed model. Finally, the error is fitted to the Beta distribution and then compared with the selected threshold. Fig. 7(a) illustrates the probability of this error. As shown in Fig. 7(a), the periodic shape is the error for 10 consecutive days excluding the zero generation at night as explained before. It was observed that the error increases for honest data, i.e. the probability decreases, as the PV output increases in the middle of the day. This error is lower when the PV output is lower and it should be zero when the PV output is zero, which was excluded.

Overall, the results of observing this customer for 10 days illustrate that the probability of the error is above the selected threshold, which is set to 5% in this case. Consequently, the TDU will detect this customer as an honest customer.

C. Case 2: Static Percentage Attack

In this case, static percentage attacks with different attack levels are applied to the PV smart meter readings to represent electricity theft by malicious customers. As aforementioned in the previous case, the three predictors will be fed first to the TDU and then the predicted power generated by TDU will be compared with the reported PV smart meter readings after applying cyber-attacks to these readings. Fig. 7(b)-(d) shows the probability of error occurrence with different attack levels. As shown, the error increases at the middle of the day, i.e. probability decreases, when the theft becomes significant. The probabilities below the threshold indicate that this injection is unlikely to happen. Therefore, the proposed TDU will be able to detect such thefts. The higher the amount of theft, the lower the probability falls, which can be used as a measure of the theft attack severity.

D. Case 3: Dynamic Percentage attack

In this scenario, a dynamic percentage attack criterion will be applied to the PV smart meters reading. The dynamic percentage attack's level ($f_2(t,d)$) is generated from the standard uniform distribution on the open interval (0,1). The predicted power from TDU is compared with the manipulated PV smart meter readings. Fig. 7(e) shows the probability of the error in this case. The results reveal the ability of TDU to detect also this malicious data as many data points have a very low probability less than the selected threshold.

E. Case 4: Positive shift attack

In this scenario, a constant value is added to the PV smart meter readings. The attack's level u is selected to be 1% of the peak value of PV smart meter readings. The probability of the error between the predicted power and the manipulated PV smart meter readings in this scenario is illustrated in Fig. 7(f). Unlike the previous cases of using a percentage attack, the probability of the error in this case is very low during low PV generation periods, which is expected and easily detected.

F. Performance Evaluation of the Proposed TDU

To evaluate the performance of the proposed TDU, 100 samples representing honest customers, and 100 samples for malicious customers are fed to the proposed TDU. The malicious dataset is developed based on the three cyber-attack functions. The following performance metrics are evaluated to investigate the performance of the proposed TDU as in (20) - (25).

$$Sensitivity\ (Detection\ Rate) = \frac{TP}{TP + FN} \tag{20}$$

$$Specificity = \frac{TN}{TN + FP} \tag{21}$$

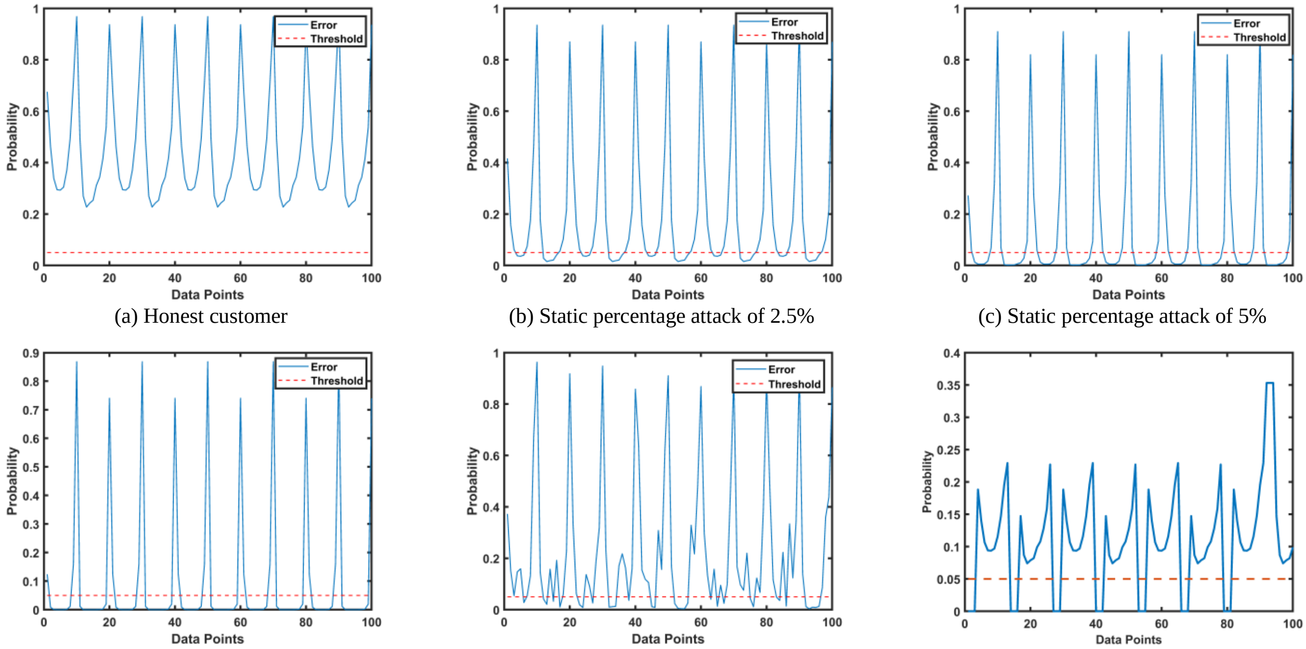
$$Precision = \frac{TP}{TP + FP} \tag{22}$$

$$Negative\ Predictive\ Value\ (NPV) = \frac{TN}{TN + FN} \tag{23}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{24}$$

$$False\ Alarm\ (FA) = 1 - Specificity \tag{25}$$

where TP is true positives, which means the sample is malicious and TDU detects it as malicious; FN is false negatives, which means the sample is malicious and TDU detects it as honest; TN is true negatives, which means the sample is honest and TDU detects it as honest; and FP is false positives, which means the sample is honest, and TDU detects it as malicious. The performance of the proposed TDU is compared with SVM, ARIMA, and LSE detectors to demonstrate the superiority of the proposed TDU. The detectors based on SVM and LSE are trained in a similar manner of our proposed TDU. On the contrary, the training of the detector based on ARIMA model is implemented in a different way. First, the PV smart meter readings for a certain period of the year and the three predictors are used to estimate the ARIMA model parameters. Then, the ARIMA model is used to forecast the PV generation at another period; then, the error is used to make the detection decision. Table IV illustrates confusion/matching matrix for various detectors which provides a summary of the performance of each detector for all samples. Each row represents the actual state of the various samples while each column represents the predicted state of all samples. This matrix is used to determine the proposed indices shown in (20)-(25). The undetected attack samples are static percentage attacks with 2.5% attack level in which the malicious customer manipulates the meter reading and reports an increase in the generated energy by a fixed percentage (2.5%) of the actual generated energy (reporting 102.5% of the actual generation) as the error between the predicted and actual power is so small and thus the probability of this error is high and above the threshold value. Table V illustrates the detection performance for the proposed TDU and the other detectors.



(a) Honest customer (b) Static percentage attack of 2.5% (c) Static percentage attack of 5%

(d) Static percentage attack of 10% (e) Dynamic percentage attack (f) Positive shift attack

Fig. 7. Probability of error occurrence of different case studies: (a) honest customer, (b), (c), (d) Static percentage attacks criterion is applied, f_1= (b) 2.5 %, (c) 5 %, and (d) 10 %, (e) Dynamic percentage attack, and (f) Positive shift attack.

TABLE V
DETECTION PERFORMANCE OF THE PROPOSED TDU IN COMPARISON WITH SVM, ARIMA AND LSE

	Performance Parameters					
Model	**Sens.**	**Spec.**	**Prec.**	**NPV**	**Accur.**	**FA**
TDU	94.5%	88.5%	89.15%	94.15%	91.5%	11.5%
SVM	99%	63%	72.8%	98.44%	81%	37%
ARIMA	100%	67%	75.19%	100%	83.5%	33%
LSE	91.5%	66.5%	73.2%	88.67%	79%	33.5%

The LSE detector has the worst performance in sensitivity, NPV, and accuracy. Also, it has the second-worst performance in the other metrics, as shown in Table V. On the other hand, the detectors based on ARIMA and SVM offer efficient detection of malicious customers with high TP and low FN, which is reflected in their high performance regarding sensitivity and NPV metrics. However, these detectors offer inaccurate performance regarding honest customers with low TN and high FP, compared to the proposed TDU. Thus, the specifity and the precision of the TDU are superior. Overall, the proposed TDU offers the best detection performance compared to SVM, ARIMA, and LSE detectors regarding the most performance metrics like specificity, precision, accuracy, and false alarm. Having various panel types and various forms of malicious samples could confuse the detector in distinguishing between honest and malicious customers. However, the evaluation performance reported in Table V deduces a high detection performance of the proposed TDU when it comes to detection accuracy, false alarm rates, precision, and specificity.

IV. Conclusions

This paper proposes an anomaly detector to detect electricity theft in the distributed generation domain, where this detector is trained only on benign data. Therefore, the anomaly detector is robust against zero-day attacks.

In this work, historical records of solar irradiance and temperature data are utilized to generate virtual historical data of PV smart meter readings to create the required dataset used in the training and testing stages of the proposed TDU. Several regression models are trained using the datasets. The regression model with the least RMSE, which is a DT model, is selected to be used in developing the proposed TDU. Hence, the proposed TDU is developed based on a fine DT to detect the suspicious data reported by the customers. The probability of the error between the actual and predicted response, which is found to follow Beta distribution function, is used by TDU as a detection metric to distinguish between honest and malicious customers by defining a desired threshold value. Furthermore, the robustness of the proposed TDU is evaluated against a new dataset that is not used in the detector's training stage, where a completely different panel type is introduced and different cyber-attack scenarios are applied to benign data to represent the suspicious data reported by malicious customers. Moreover, the performance of the proposed TDU is compared with other detectors based on SVM, ARIMA, and LSE to demonstrate the potential of the proposed TDU. Simulation results revealed that the proposed TDU offers a superior detection performance.

V. References

[1] P. Jokar, N. Arianpoo and V. C. M. Leung, "Electricity Theft Detection in AMI Using Customers' Consumption Patterns," in *IEEE Trans. Smart* Grid, vol. 7, no. 1, pp. 216-226, Jan. 2016.

[2] GTM Research, Utility AMI Analytics for the Smart Grid 2013-2020: Applications, Markets and Strategies, Sept. 2013.

[3] T. R. Sharafeev, O. V. Ju and A. L. Kulikov, "Cyber-Security Problems in Smart Grid Cyber Attacks Detecting Methods and Modelling Attack Scenarios on Electric Power Systems," *2018 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, Moscow, Russia, 2018, pp. 1-6.

[4] Y. Tang, C. Ten and K. P. Schneider, "Inference of Tampered Smart Meters with Validations from Feeder-Level Power Injections," *2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, Chengdu, China, 2019, pp. 2783-2788.

[5] K. Wang, H. Li, S. Maharjan, Y. Zhang and S. Guo, "Green Energy Scheduling for Demand Side Management in the Smart Grid," in *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 2, pp. 596-611, June 2018.

[6] K. Wang et al., "Wireless Big Data Computing in Smart Grid," in *IEEE*

Wireless Commun., vol. 24, no. 2, pp. 58-64, April 2017.

[7] F. Li et al., "Smart Transmission Grid: Vision and Framework," in *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 168-177, Sept. 2010.

[8] S. Seme, K. Sredensek and Z. Praunseis, "Smart grids and net metering for photovoltaic systems," 2017 *International Conference on Modern Electrical and Energy Systems (MEES)*, Kremenchuk, 2017, pp. 188-191.

[9] Cory, Karlynn, Toby Couture, and Claire Kreycik. “Feed-in tariff policy: design, implementation, and RPS policy interactions”, Report No. NREL/TP-6A2-45549. National Renewable Energy Lab (NREL), Golden, CO (United States), 2009.

[10] G. M. Masters, Renewable and Efficient Electric Power Systems, second edition, John Wiley & Sons Inc, 2013.

[11] G. M. Masters, Renewable and efficient electric power systems, Second edition. Hoboken, New Jersey: John Wiley & Sons Inc, 2013.

[12] S. McIntyre, Termineter: Python Smart Meter Testing Framework, Jan. 2018. [Online]. Available: https://tools.kali.org/stress-testing/termineter.

[13] V. B. Krishna, C. A. Gunter and W. H. Sanders, "Evaluating Detectors on Optimal Attack Vectors That Enable Electricity Theft and DER Fraud," in *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 4, pp. 790-805, Aug. 2018.

[14] Electric Sector Failure Scenarios and Impact Analyses Version 3.0, National Electric Sector Cybersecurity Organization Resource, Dec. 2015. [Online]. Available: http://smartgrid.epri.com/doc/NESCOR-15.pdf.

[15] Y. He, G. J. Mendis, J. Wei, "Real-Time Detection of False Data Injection Attacks in Smart Grid: A Deep Learning-Based Intelligent Mechanism," in *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2505-2516, Sept. 2017.

[16] V. B. Krishna, C. A. Gunter and W. H. Sanders, "Evaluating Detectors on Optimal Attack Vectors That Enable Electricity Theft and DER Fraud," in *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 4, pp. 790-805, Aug. 2018.

[17] M. Ismail, M. Shahin, M. F. Shaaban, E. Serpedin and K. Qaraqe, "Efficient detection of electricity theft cyber attacks in AMI networks," 2018 *IEEE Wireless Communications and Networking Conference (WCNC)*, Barcelona, 2018, pp. 1-6.

[18] F. Xiao and Q. Ai, “Electricity theft detection in smart grid using random matrix theory,” *IET Gen., Trans. & Dist.*, vol. 12, no. 2, pp. 371–378, Jan. 2018.

[19] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed and M. Mohamad, "Nontechnical Loss Detection for Metered Customers in Power Utility Using Support Vector Machines," in *IEEE Trans. Power Del.*, vol. 25, no. 2, pp. 1162-1171, April 2010.

[20] Y. Liu and S. Hu, "Cyberthreat Analysis and Detection for Energy Theft in Social Networking of Smart Homes," in *IEEE Trans. Computat. Social Syst.*, vol. 2, no. 4, pp. 148-158, Dec. 2015.

[21] D. Yao, M. Wen, X. Liang, Z. Fu, K. Zhang and B. Yang, "Energy Theft Detection With Energy Privacy Preservation in the Smart Grid," in *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7659-7669, Oct. 2019.

[22] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar and S. Mishra, "Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid," in *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 1005-1016, June 2016.

[23] 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," in *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2505-2516, Sept. 2017.

[24] D. Xue, X. Jing, and H. Liu, ‘‘Detection of false data injection attacks in smart grid utilizing ELM-Based OCON framework,’’ IEEE Access, vol. 7, pp. 31762–31773, 2019.

[25] N. Bhusal, M. Gautam and M. Benidris, "Detection of Cyber Attacks on Voltage Regulation in Distribution Systems Using Machine Learning," in IEEE Access, vol. 9, pp. 40402-40416, 2021.

[26] B. Li, Y. Wu, J. Song, R. Lu, T. Li, and L. Zhao, ‘‘DeepFed: Federated deep learning for intrusion detection in industrial cyber-physical systems,’’ IEEE Trans. Ind. Informat., early access, Sep. 11, 2020.

[27] S. M. Kasongo and Y. Sun, ‘‘A deep learning method with wrapper based feature extraction for wireless intrusion detection system,’’ Comput. Secur., vol. 92, May 2020.

[28] G. Loukas, T. Vuong, R. Heartfield, G. Sakellari, Y. Yoon, and D. Gan, ‘‘Cloud-based cyber-physical intrusion detection for vehicles using deep learning,’’ IEEE Access, vol. 6, pp. 3491–3508, 2018.

[29] A. Takiddin, M. Ismail, U. Zafar and E. Serpedin, "Robust Electricity Theft Detection Against Data Poisoning Attacks in Smart Grids," in IEEE Transactions on Smart Grid, vol. 12, no. 3, pp. 2675-2684, May 2021.

[30] M. Ismail, M. F. Shaaban, M. Naidu and E. Serpedin, "Deep Learning Detection of Electricity Theft Cyber-attacks in Renewable Distributed Generation," in *IEEE Trans. Smart Grid*, doi: 10.1109/TSG.2020.2973681.

[31] “Solar Panels Plus LLC (SPP).” [Online]. Available: http://www.solarpanelsplus.com/. [Accessed: 06-Feb-2020.]

[32] A. T. Umoette, E. A. Ubom, and I. E. Akpan, “Comparative Analysis of Three NOCT-Based Cell Temperature Models,” *Int. J. Syst. Sci. Appl. Math.*, vol. 1, no. 4, p. 69, Dec. 2016.

[33] J. Brownlee, “A Gentle Introduction to k-fold Cross-Validation,” Machine Learning Mastery, May 2018.

[34] E. Alpaydin, “Introduction to machine learning,” *MIT press*, Mar 2020.