# A Comprehensive Survey on Energy Theft Detection in Smart Grid

Manya K S
*M.tech Student, Computer Science and Engineering*
*NSS College of Engineering*
Palakkad,Kerala,India
manyaks94@gmail.com

Balagopal N
*Assistant Professor,Computer Science and Engineering*
*NSS College of Engineering*
Palakkad,Kerala,India
balagopaln89@gmail.com

***Abstract*—The smart grid is an interconnected power grid that involves sensors, deployment strategies, smart meters, and real-time data processing. Energy theft is one of the most important issues related to the smart grid implementation. Electricity distribution authorities lose a large amount of income, due to illegal connections or dishonesty of customers for their personal gains. Therefore, it is important to develop efficient and reliable methods to identify illegal users who are committing energy theft. Current energy theft detection schemes are mainly categorized into three groups: classification based, state based and game theory based. Data mining and machine learning technologies are used in classification based methods. Support Vector Machine (SVM), Fuzzy C-means Clustering, Fuzzy logic, Decision tree, Genetic Algorithm, etc. are some of the classification based techniques used to detect electricity theft. State based detection schemes employ specific devices, like wireless sensors and radio-frequency identification tags, to provide a high detection accuracy. In game theory method, the problem of electricity theft detection is formulated as a game between the electric utility and the electricity thief. This paper presents a survey on various energy theft detection methods in the smart grid under these categories.**

***Index Terms*—energy theft, smart grid, smart meter**

## I. Introduction

The smart grid is one of the main research and development direction in today's energy industry. Smart grid technology uses a combination of the digital information technology application and electric power network. Electrical theft is one of the major problem related to the conventional power grid. It leads to huge losses to the utilities in the power sector. Electric theft refers to the intentional and illegal use of electricity. Tapping is one of the major causes of electricity theft. Tampering with the electric meters and billing alterations by the employees are other causes behind the electric theft. In the current scenario, to detect energy losses inspection of customers are carried out based on some predictions. Manual inspection is very time consuming and costly. Thus, to identify electrical theft more efficiently, utilities need to make use of the latest technologies available. To identify and locate the potential areas of electrical theft, information and communication technology can be used in smart grid environment.

The detection and identification of frauds in energy systems have been initially addressed with statistical techniques. A variety of solutions to energy theft have been proposed recently due to the fast development of Advanced Metering Infrastructure (AMI) in the smart grid. Various systems are introduced to detect the theft and reduce the non-operational losses. Classification based, state based and game theory based are some of the important techniques. Classification based approaches take advantage of the detailed energy consumption measurements collected from the smart meters. Under normal condition customers' consumption follow a certain pattern; irregularities in usage pattern can be a sign of some malicious activities. Data mining and machine learning technologies are used to train a classifier based on a sample database, which is then utilized to find abnormal patterns. Support Vector Machine, Fuzzy C-means Clustering, Fuzzy logic, Decision tree, Genetic Algorithm, etc. are some of the classification based techniques used to detect electricity theft. State based detection schemes employ specific devices, like wireless sensors and radio-frequency identification tags, to provide a high detection accuracy. In game theory based methods, the problem of electricity theft detection is formulated as a game between the electricity thief and the electric utility. In the game, the goal of the electricity thief is to steal a predefined amount of electricity while minimizing the likelihood of being detected, while the utility companies want to maximize the probability of detection

The rest of this paper is organized as follows. Section II describes the area of the smart grid. Section III provides a detailed review of the different techniques for theft detection. Section IV summarizes this survey.

## II. Preliminary

### *A. Smart Grid*

The smart grid is a form of an electric power network and it is the next generation power grid. The traditional power grids are generally used to carry power from a few central generators to a large number of customers. It may support all or some of the following operations: electricity generation, electricity transmission, electricity distribution, and electricity control. In short, the digital information technology that allows for two way communication between the utility and its consumers, and the sensing along the transmission lines is what makes the grid smart. A smart grid can be defined as “an interconnected system of information, communication technologies and control systems used to interact with automation, and processes across the entire power sector including electricity generation, transmission, distribution, and the customer. One significant difference between today’s grid and the Smart Grid is two way exchange of information between the consumer and the grid. A smart grid covers the following aspects of the power system:

- The delivery infrastructure, such as circuit breakers, transmission and distribution lines, transformers, smart substations and sensors, etc.
- The end user systems and related distributed energy resources, such as renewable resources, loads, storage, and electrical vehicles, etc.
- The communication networks, such as remote measurement and control networks, inter and intra enterprise networks, and the Internet and Home Area Networks (HAN), etc.
- The management system at various levels of generation and delivery infrastructures, such as transmission and distribution control centers, regional reliability coordination centers, national emergency response centers and smart metering management system, etc.
- The financial and regulatory environment, such as stock and bond markets, government incentives, regulated and the non-regulated rate of return, etc.

One of the most well known common reference models of the smart grid is proposed by the U.S National Institute of Standards and Technology (NIST) in [1]. A conceptual view of the NISTs smart grid reference model is depicted in Fig. 1. The NISTs model is composed of seven domains: generation, transmission, distribution, customers, markets, operations, and service providers. The two-way electrical flows are moving across the top four domains (power generation, transmission, distribution and customer), which are controlled and managed by the bottom three domains (market, operations, and service providers) through communication flows. In addition, three typical customers are listed: Home Area Network (HAN), Building Area Network(BAN) and Industrial Area Network (IAN), where the Advanced Metering Infrastructure (AMI) takes place to monitor and manage the power and information flows through smart meters.

## III. Techniques for Energy Theft Detection

### A. Classification based detection technique

Classification based detection technique is one of the most widely used approaches, which is defined as the load profile classification of electricity consumption of a customer or a group of customers over a period of time. The basic procedure for classification based energy theft detection consists of seven parts: data acquisition, data preprocessing, feature extraction, classifier training and parameter optimization, classification, data postprocessing, and suspected customer list generation.

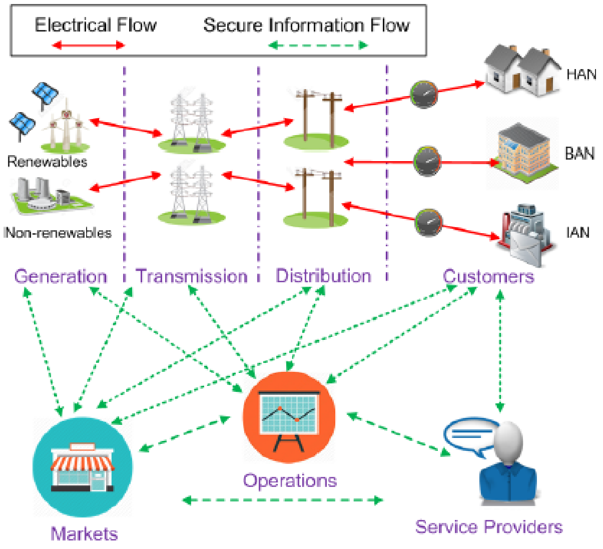

Fig. 1. NIST reference model for smart grid[1]

Support Vector Machines (SVM) are widely utilized to classify the load profiles of customers for detection of energy theft suspects. The main purpose of the SVM algorithm used for classification is to construct an optimal decision function that accurately predicts unseen data into two classes, and minimizes the classification error.

Nagi *el al.* [2] presented a framework to identify and detect non-technical loss activities in the electric utility market. This approach used support vector machine based method for detecting frauds based on irregularities in the customer’s patterns. This method uses historical customer consumption data collected from Tenga Nasional Berhad. Electricity consumption patterns of customers are extracted using data mining and statistical techniques. When a fraud event occurs, then load profiles of customers contain abnormalities. From the meter readings, daily average consumptions features per month are calculated. Features are then normalized and used for training in SVM with gaussian kernal. Support vector machine classifies the customer load profiles for the detection of fraud suspects. Electricity customer consumption data of less than 400 are highly imbalanced out of 260k customers. For the successful identification of fraudulent customers with reduced false positives, it needs to be combined with other parameters. A recall of 0.53 is achieved on the test set.

Nagi *el al.* [3] presented a hybrid approach to find non-technical loss using GA and SVM. The work in [2] is extended by using a GSVM in this work. In SVM classification, load profiles were classified into categories according to their normal and abnormal behaviour. In this technique SVM classifies consumers into 4 different classes: confirmed fraud suspects, unconfirmed fraud suspects, confirm clean suspects, unconfirmed clean suspects. Then, it uses a genetic algorithm in order to globally optimize the hyperparameters of SVM. The hybrid solution of GA-SVM provides a better way of selecting optimal SVM hyper-parameters. This model achieves a test recall of 0.62.

Besides the SVM technique, many other classification methods, such as fuzzy classification and neural networks, are utilized to detect energy theft. Nagi *el al.* [4] used the fuzzy inference system(FIS) to process the list of customers derived from SVM based fault detection method. Fuzzy inference system acts as an intelligent decision making system together with the SVM based detection model in [2] to shortlist customers with high probability of fraud activities and abnormalities. In this method, Fuzzy inference system acts as a post-processing scheme for selecting abnormal customers. The detection rate of the SVM-FIS model is more than [2] and it is more cost effective also. SVM-FIS based fraud detection method has a low hit rate, which indicates a high misclassification of fraudulent consumer detection. A test recall of 0.72 is obtained in this method.

Angelos *et al.* [5] proposed a technique for the classification of electricity consumption profiles. The proposed method consists of two steps. In the first step, a C-means based fuzzy clustering is used to find the similar consumption profiles of customers. The classification process begins when a standard profile is defined. In this step, the classification process calculates the current profile of each client in two steps: First, calculates the current fuzzy membership matrix using current profiles and the centers found out in the clustering process. Then, calculates the Euclidean distance between the current fuzzy membership matrix and the previous one. This value represents the abnormality degree of each client. And, all the clients with values greater than the given values are considered as the abnormal client. This approach was validated with real data, showing satisfactory performance in the identification of irregularities in residential and commercial customers. The major limitation of the method was that it failed to work in situations where fuzzification of features was a difficult task. This method is highly feasible because it requires only five attributes: average consumption, maximum consumption, standard deviation, number of inspections and the average consumption of the residential area. The system achieved consistent performance and has an average precision of 0.745.

Muniz *et al.* [6] describes a combined approach of a neural networks committee and a neuro-fuzzy hierarchical system. The aim is to increase the accuracy of identifying irregular customers, selected from doubtful clients. The proposed method includes two basic modules: filtering and classification. Filtering module contains a combination of five artificial neural networks, aims at improving accuracy. This module is used to extract normal and irregular customers for the classification module. Classification module is developed in the expectation of improving even further detection of abnormalities. But, this improved version also resulted in high false positives. A precision of 0.512 and an accuracy of 0.682 on the test set are obtained.

Costa *et al.* [7] proposed an artificial neural network based method to discover knowledge in databases for classifying the consumers as fraudsters or non-fraudsters. For the fraud detection process, a methodology with data pre-processing and mining using ANNs was applied. A dataset of 22k customers is used for neural networks training. It uses the average consumption of the previous 12 months and other customer features such as location, type of customer, weather, and voltage. On the test set, an accuracy of 0.8717 and a recall of 0.2947 are obtained. Unfortunately, this model was found to be inefficient during an uneven distribution of records. Also, this method obtained low precision which eventually leads to large false positives.

Cody *et al.* [8] used a decision tree machine learning algorithm, M5P to find energy consumption behaviour of consumers for the prediction of future energy consumption and the detection of fraud activities. Real historical data were used in the experiments to generate the decision model. To predict the future energy consumption values more accurately, pre-processing methods such as feature extraction algorithms are used.

To detect the malicious consumers, Jindal et al. [9] using Decision Tree(DT) and SVM classifiers in this paper. The proposed scheme is capable enough to detect the thefts happening anywhere in the power network. To detect the malicious consumers, various features namely number of heavy appliances, number of persons, season, time slot, and temperature are given as input to the DT. The expected electricity consumption for the consumer during a particular time is calculated using DT. This consumption along with other features is provided as input to the SVM classifier which is trained on the collected dataset. This classifier is then used to classify the consumers as normal or fraud based on their features.

Nizar *et al.* [10] proposed a feature selection based method for customer data pre-processing to extract the most relevant information for performing analysis such as classification and clustering. Feature selection is an important step in data mining process to remove irrelevant and redundant features. This will improve the quality of results by providing fast processing, high accuracy. Different feature selection techniques are used in this system. The methods include complete search, best-first search, genetic search and greedy search algorithms for the data. Shape features are derived from the consumption time series considering the impact of lunch times, nights and weekends on the consumption. K-means clustering can be used for clustering similar consumption profiles. In the classification part, a decision tree can be used to predict whether a customer causes non-technical losses or not. An overall test accuracy of 0.997 is obtained.

Ramos *et al.* [11] analyzed the consumption profiles of 5K Brazilian industrial customer profiles. Each consumer profile was represented by ten features including demand billed, installed power, maximum demand etc. In this work, experiments show that SVM slightly performs better than K-nearest neighbors and a neural network. The test accuracies of these methods are 0.09628, 0.9620, 0,9448 respectively.

Sahoo *et al.* [12] proposed a temperature dependent predictive model which uses smart meter data from the distribution transformer to detect electricity theft in an area. This method estimates non-technical losses by subtracting an estimate of the technical losses from the overall losses. It models a temperature dependent model using regression which approximates the technical losses. This work applies the proposed model to a database of 30 customers which the consumption was recorded for six days with meter readings in every 30 minutes for theft levels of 1,2,3,4,6,8 and 10%. The test recalls are 0.2211, 0.7784, 0.9789, 1, 1, 1 and 1 respectively.

Extreme learning machines (ELM) are one hidden layer neural networks in which the weights from the inputs to the hidden layer are randomly set and never updated. Only the weights from the hidden to output layer are learned. Nizar *et al.* [13] applied ELM algorithm to NTL detection in meter readings of 30 minutes, for which a test accuracy of 0.5461 is reported. A self-organizing map (SOM) is a type of unsupervised neural network training algorithm that is used for clustering. Cabral et al. [14] applied SOMs to weekly customer data of 2K customers consisting of meter readings of 15 minutes. This allows clustering customers behavior into fraud or non-fraud. Inspections are only carried out if certain hand-crafted criteria are satisfied including how well a week fits into a cluster and if no contractual changes of the customer have taken place. A test accuracy of 0.9267, a test precision of 0.8526, and test recall of 0.9779 are reported. TABLE 1 shows the comparison results of classification based energy theft detection schemes.

### *B. State based detection technique*

Another common solution for energy-theft detection is state-based detection technique, which uses monitoring state to improve the detection rate. The monitoring state can be derived from wireless sensor networks, RFID. McLaughlin *et al.*[15], [16] proposed an AMI Intrusion Detection System (AMIDS) that uses information fusion to combine the sensors and consumption data from a smart meter to more accurately detect energy theft. AMIDS combines meter audit logs of physical and cyber events with consumption data to more accurately model and detect theft related behaviors.

Ansari *et al.* [17], pointed out that physical attack to smart meters can be extended to a network attack by means of false data injection. They proposed a consumer attack model that is formulated into one type of coin change problem, which minimizes the number of compromised meters without being revealed by maintaining a cumulative load at the aggregation point to which multiple households are connected in todays radial tree-like distribution network. A hybrid detection framework is developed to detect anomalous and malicious activities by incorporating their proposed grid sensor placement algorithm with observability analysis to increase the detection rate.

Cheng et al. [18] proposed a system that implements Radio Frequency IDentification (RFID) technology to help the electricity supply company deal with its ammeter inventory management and prevent energy theft. There are two parts in the proposed system: ammeter inventory management and ammeter verification control. The ammeter inventory management includes an RFID tag on each ammeter, RFID readers, the middleware, and the network with the Enterprise Resource Planning (ERP) system of the electricity supply company. The integrity of the RFID tag can be used to detect energy theft. In addition, the reader acquires the information transmitted from the tag and sends it to the companys ERP system through the network to determine whether it is the approved tag or a different one placed by electricity thieves. Although the RFID technology can be used to detect energy theft, the utility companies have to pay extra cost to install the system.

In order to detect more bad data injection and locate the bad data within a smaller area, Liu *et al.* [19] proposed an Adaptive Partitioning State Estimation (APSE). In this method, the power system is transformed to a weighted undirected graph and divided into several subgraphs. Chi-squares test is used to detect bad data in each subgraph. Since the threshold of a subgraph is expected to be lower than that of the entire system, it will be more sensitive to detect the bad data. The APSE method is based on the proper partitioning graphs and can only detect the bad data on one transmission line. The method to detect multiple bad data should be studied.

### *C. Game theory based detection technique*

Game theory based energy theft detection schemes are proposed recently and provide a new perspective to solve the energy theft issue. Amin *et al.* [20] investigated incentive problems in electricity distribution when customer energy usage is imperfectly observable by the distributor. Paper presented the optimal investment and tariff strategy for the distributor with rational customers from a game theory point of view. They also aimed to provide an effective suggestion to regulator how to decide explicit targets for the allowable losses to remedy the problem of incentive misalignment. The proposed system modeled the energy theft and combat losses as a non-zero sum Stackelberg game with an unregulated distributor, where the distributor acts as leader and the customers act as a follower. Even though the optimal quantity of stolen electricity can be estimated by distributor, smart customers are also possible to avoid detection using sophisticated strategies.

Cardenas *et al.*[21] formulated the problem of electricity theft detection as a game between the distributor and the electricity thief. For the electricity thief, they want to minimize the likelihood of being detected to steal a predefined quantity of electricity. They can achieve it by changing their probability density function of electricity usage during the measurement period. On the other hand, the distributor wants to maximize the probability of energy theft detection and determine the optimal investment incurred by AMIs installation. The Nash equilibrium of the game is found as a probability density function that attackers and defenders must choose in order to send AMI measurements. Game theory based detection provides a new perspective to recognize and address the energy theft.

TABLE I
COMPARISON OF CLASSIFICATION TECHNIQUES

| **Technique used** | **Dataset size** | **Precision** | **Recall** | **Accuracy** |
|---|---|---|---|---|
| SVM[2] | $< 400$ | - | 0.53 | - |
| Genetic SVM[3] | 1171 | - | 0.62 | - |
| SVM and FIS[4] | 100K | - | 0.72 | - |
| Fuzzy clustering[5] | 20K | 0.745 | - | - |
| Neuro-fuzzy system[6] | 20K | 0.512 | - | 0.68 |
| ANN[7] | 22K | 0.65 | 0.2947 | 0.8717 |
| SVM and DT[9] | >500K | - | - | 0.925 |
| SVM, KNN[11] | 5K | - | - | 0.96 |
| NN[11] | 5K | - | - | 0.94 |
| SOM[14] | 2K | 0.8526 | 0.9779 | 0.9267 |

## IV. CONCLUSIONS

This paper presents a survey on energy theft detection in the smart grid. Three categories of theft detection techniques are surveyed, classification based, state based and game theory based. Classification based scheme used machine learning techniques and it has medium accuracy rate and faults positive. The state based method used state monitoring by specific equipment. It has high accuracy rate and low false positive. The price of extra investment required for the monitoring system including device cost, system implementation cost, software cost and operating or training cost. Game theory based energy theft detection schemes are proposed recently and provide a new perspective to solve the energy theft issue. However, how to formulate all players' utility function and potential strategies is still a challenging work. Finally, compare accuracy between classification based methods in TABLE I. Since power theft is a serious problem faced by electricity authority more accurate theft detection system is needed.

## REFERENCES

[1] "Nist framework and roadmap for smart grid interoperability standards," release 1.0, National Institute of Standards and Technology, 2010. [Online]. Available: http://dx.doi.org/10.6028/NIST.SP. 1108r1

[2] J. Nagi, A. M. Mohamad, K. S. Yap and S.K Tiong, "Non-Technical Loss analysis for detection of electricity theft using support vector machines," IEEE 2nd International Power and Energy Conference (PECon 2008), 2008.

[3] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed and A. M. Mohammad, "Detection of abnormalities and electricity theft using genetic Support Vector Machines," 2008 IEEE Region 10 Conference TENCON, 2008.

[4] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed and F. Nagi, "Improving SVM-Based Nontechnical Loss Detection in Power Utility Using the Fuzzy Inference System," IEEE Transactions on Power Delivery, vol. 26, no. 2, pp. 1284-1285, 2011.

[5] E. W. S. dos Angelos, O. R. Saavedra, O. A. Carmona Cortes and A. Nunes de Souza, "Detection and identification of abnormalities in customer consumptions in power distribution systems," IEEE Transactions on Power Delivery, vol. 26, no. 4, pp. 2436-2442, 2011.

[6] C. Muniz, M. M. B. R. Vellasco, R. Tanscheit and K. A. Figueiredo, "Neuro-fuzzy System for Fraud Detection in Electricity Distribution," IFSA/EUSFLAT Conference, pp. 1096-1101, 2009.

[7] B. C. Costa, B. L. Alberto, A. M. Portela, W. Maduro and E. O. Eler, "Fraud detection in electric power distribution networks using an ANN based knowledge-discovery process," International Journal of Artificial Intelligence & Applications, vol. 4, no. 6, 2013.

[8] Christa Cody, Vitaly Ford, Ambareen Siraj, "Decision Tree Learning for Fraud Detection in Consumer Energy Consumption," IEEE 14th International Conference on Machine Learning and Applications, 2015.

[9] Anish Jindal, Amit Dua, Kuljeet Kaur, Mukesh Singh, Neeraj Kumar, and S. Mishra, "Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid," IEEE Transactions on Industrial Informatics, VOL. 12, NO. 3, JUNE 2016.

[10] A. H. Nizar, J. H. Zhao and Z. Y. Dong, "Customer information system data pre-processing with feature selection techniques for non-technical losses prediction in an electricity market," International Conference on Power System Technology (PowerCon 2006), 2006.

[11] C. C. Oba Ramos, A. Nunes de Souza, D. Sinkiti Gastaldello and J. Paulo Papa, "Identification and feature selection of non-technical losses for industrial consumers using the software WEKA," International Conference on Industry Applications, 2012.

[12] S. Sahoo, D. Nikovski, T. Muso and K. Tsuru, "Electricity theft detection using smart meter data," IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2015.

[13] A. H. Nizar, Z. Y. Dong and Y. Wang, "Power Utility Nontechnical Loss Analysis With Extreme Learning Machine Method," IEEE Transactions on Power Systems, vol. 23, issue 3, pp. 946-955, 2008.

[14] J. E. Cabral, J. O. P. Pinto and A. M. A. C. Pinto, "Fraud detection system for high and low voltage electricity consumers based on data mining," Power & Energy Society General Meeting (PES09), 2009.

[15] S. McLaughlin, B. Holbert, S. Zonouz, and R. Berthier, "AMIDS: A multi-sensor energy theft detection framework for advanced metering infrastructures," in Proc. IEEE Third International Conference on Smart Grid Communications (SmartGridComm), 2012, pp. 354-359.

[16] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz, "A multi-sensor energy theft detection framework for advanced metering infrastructures," IEEE Journal on Selected Areas in Communications, vol. 31, no. 7, pp. 1319-1330, 2013.

[17] C. H. Lo and N. Ansari, "CONSUMER: A novel hybrid intrusion detection system for distribution networks in smart grid," IEEE Transactions on Emerging Topics in Computing, vol. 1, no. 1, pp. 33-44, 2013.

[18] B. Khoo and Y. Cheng, "Using RFID for anti-theft in a chinese electrical supply company: A cost-benefit analysis," in Proc. IEEE Wireless Telecommunications Symposium (WTS), 2011, pp. 1-6.

[19] T. Liu, Y. Gu, D. Wang, Y. Gui, and X. Guan, "A novel method to detect bad data injection attack in smart grid," in Proc. IEEE INFOCOM, 2013, pp. 3423-3428.

[20] A. H. Nizar, Z. Y. Dong and Y. Wang, "Power Utility Nontechnical Loss Analysis With Extreme Learning Machine Method," IEEE Transactions on Power Systems, vol. 23, issue 3, pp. 946-955, 2008.

[21] A. A. Cardenas, S. Amin, G. Schwartz, R. Dong, and S. Sastry, "A game theory model for electricity theft detection and privacy-aware control in AMI systems," in Proc. IEEE 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2012, pp. 1830-1837.