Detection of Non-Technical Losses in Smart Distribution Networks: a Review

Anna Fragkioudaki[1], Pedro Cruz-Romero, Antonio Gómez-Expósito, Jesús Biscarri[2], Manuel J. de Tellechea, and Ángel Arcos

[1] Universidad de Sevilla, Spain,
plcruz@us.es
[2] Endesa-Enel, Sevilla, Spain

Abstract. With the advent of smart grids, distribution utilities have initiated a large deployment of smart meters on the premises of the consumers. The enormous amount of data obtained from the consumers and communicated to the utility give new perspectives and possibilities for various analytics-based applications. In this paper the current smart metering-based energy-theft detection schemes are reviewed and discussed according to two main distinctive categories: A) system state-based, and B) artificial intelligence-based.

Keywords: advanced metering infrastructure, electricity theft, smart grid, smart meter.

1 Introduction

As power demand increases in modern societies, the need for an advanced and reliable power grid becomes increasingly imperative. In fact, the traditional power grid, which is still remarkably based on a design already existing for more than 100 years, can no longer satisfy the present-day needs and requirements [1]. The current emergence of smart grids aims to increase the reliability, quality and security of supply, especially in the face of the increased penetration by renewable energy sources in the form of distributed generation [2]. The concept of a smart grid has also come into existence, bringing into the state-of-the-art scope relative advancements in information systems and communication technologies, one of whose cornerstones is the present large-scale deployment in many countries of advanced metering infrastructure (AMI) in order to upgrade the aging energy metering system [3].

One of the principal problems which impacts the efficiency and security of the power distribution networks are the power losses occurring within the process of delivering energy to the consumer. These losses can be decomposed into two categories: i) Technical losses (i.e. losses due to naturally occurring phenomena in the power system, such as power dissipation within transmission lines and transformers), and ii) Non-technical losses (NTL), which can be attributed to the following reasons: a) Actions of utility employees or an operator, such as administrative losses due to accounting errors and record keeping, b) Customer theft, c) Customer non-payment, and d) Theft by non-customers [4].

A critical issue for the distribution utility is that NTL cannot be precisely calculated, only global losses; they are usually estimated as the difference between the total amount of energy fed into the distribution system and the total amount of energy recorded as sold to customers [5]. The excess of unbilled energy is energy that is not scheduled or expected by the utility, thus it can severely affect the power system operation [6]. Critical operational problems that may arise include overloads of generation units and the stressing of network equipment due to congestion and/or over-voltages. These result from the fact that the utility cannot schedule sufficient active and reactive power due to system dynamic uncertainty and insufficient load flow information. Furthermore, these over-loadings can have an impact on the equipment of honest consumers. In extreme cases of excess unplanned load, blackouts and brownouts may also occur. Concerning the distribution utilities, apart from the directly incurred economic losses as a consequence of purchasing energy that is not billed for, maintenance costs also increase due to the aforementioned stressing of the equipment. Hence, NTL deprive utilities from investing in the upgrading of their equipment. Last but not least, the environmental impact of NTL is also considerable due to the increase in CO_2 emissions (the price signal is not considered in the defrauder decisions). A 10% reduction in NTL in India (around 83,000 GWh) would result in 9.2 million tons CO_2 reduction annually [7].

The nature of NTL poses serious challenges to utility companies in detecting dishonest customers. It should also be pointed out that technical losses are correlated with NTL, since the delivery of unbilled energy creates further physical losses on the power system. Thus minimizing NTL contributes to the overall reduction of power losses. ENEL, the Italian electricity utility, was motivated to initiate a large scale roll-out of smart meter-based infrastructure in order to minimize NTL of their distribution network. After the installation of smart meters (SMs) on the consumers' premises, the theft hit-rate raised from 5% to 50% [9]. This massive deployment of SMs is now extended to Spain, being facilitated by Endesa -one of the Spanish distribution utilities- with 6.8 m SMs and 77,000 concentrators having been installed by the end of December 2015. The data provided by the SM devices give a new perspective and unveil numerous possibilities to develop efficient and effective theft detection methods. Research in this respect has recently shown significant progress. As a result, the authors felt motivated to investigate and present in this paper the state-of-the-art in NTL detection methods within the framework of AMI, including artificial-intelligence techniques.

An important part of the implementation of an AMI is the replacement of the legacy mechanical meters with SMs. The bidirectional communication capability of the SM allows remote meter data-reading, recording of higher resolution measurements, as well as outage reporting. SM and AMI data analysis remains a challenging task for several reasons. Support databases with SM data "as is", is infeasible over long time periods due to storage limitations. Those data are processed, depending on the purpose of their use, and compressed before storage. The compression may result in precision reduction of data, which could potentially be useful for future re-analysis. Additionally, real- or nearly real-time data processing can be computationally heavy and resource-consuming. Last but not least, detailed measurements from SMs allow the utility companies to extract the consumer load profile, which is considered sensitive private data and even forbidden by some regulations. There are confidentiality issues arising, with the possibility that such private information can be sold to third-parties such as insurance companies, marketing companies etc. Moreover the consumers may become easier targets for criminals, such as burglars that can infer the victim's daily habits by analyzing their load profile [8].

2 NTL Detection Techniques

There are various ways that the data retrieved from SMs have been analyzed and exploited in order to detect NTL. Existing methods are categorized in this paper in three groups: system state-based, artificial-intelligence-based and game-theory based. In this review we will not consider the last one (see [4, 19] for more information).

2.1 System state-based methods

These methods are based on the coherence of data measured by SMs with respect to the data collected from the network (probably performed on a routine basis by the distribution system operator) and the features of the network (topology and line parameters). Chen et al. present in [10] an electricity anti-theft method based on state estimation (SE) algorithm [11], using redundant data from SM. It is claimed that whichever the technology of stealing may be, the method is applicable. Other advantages of this method include small-scale investment, wide-area and real-time monitoring. It is also suggested that considering that the false voltage, current, or power measurements due to stealing are the bad measurements, then for a limited-size network its status can be estimated with high accuracy, while at the same time localization of the electricity theft point can be achieved. The method was tested on a 10 kV medium voltage (MV) network. However, the authors claim that the theoretical model can be applied on 400 V low voltage (LV) networks as well. A power balancing is initially performed to determine whether there is really need to further investigate a feeder. If the difference between the total of power supply and the sold power exceeds a threshold, then the following methodology is applied. The collected three-phase real-time voltage, current, active and reactive power measurements at the MV/LV transformers are used as inputs to a weighted-least-squares (WLS) three-phase state estimation algorithm. This is applied in order to estimate the loading of the distribution transformer. Note that it was considered that the phases are decoupled in ungrounded MV networks. If the deviation of the estimation from the measurement is greater than a threshold, this suggests the existence of possible electricity theft. No results were presented to validate the performance of the method Another SE approach is developed by Huang et al. in [12, 9] for almost real-time localization of irregular energy consumption and NTL reduction. However, in [9] the SE is complemented with an analysis-of-variance (ANOVA) model, constituting a more detailed two-stage approach. The first-stage includes the implementation of the MV-level SE, as in [12], for load estimation of the MV/LV distribution transformer. This stage aims to identify feeders with tampered or defective meters. Abnormalities within the feeder level in electricity consumption are determined by examining a measure of overall fitting of the estimates to pseudo-measurements on the feeder bus, calculated by aggregated customer data from SM at the MV/LV transformers. Following this phase, ANOVA is performed in order to distinguish suspect customers with abnormal measurements records. A WLS-based, three-phase polar form SE algorithm is implemented to estimate the MV/LV distribution transformer load. This algorithm requires, aside from the network parameters and configuration: a) the hourly LV bus voltage and demand data from SMs at the points of power delivery (aggregated to provide pseudo-measurements at the LV side of the MV/LV transformer), b) outage management system (OMS) data, and c) customer information system (CIS) data in order to examine customer connectivity and construct the feeder framework. When the estimates from the SE are obtained, irregular usage at the distribution transformer level is detected via the examination of the normalized residuals at the point of delivery. Following the identification of bad data, the corresponding LV network is closely investigated by applying ANOVA. To this purpose, for all consumers that belong to that network, their load baseline curves (as estimated by old data dating back a few weeks) are compared with curves obtained by recent SM measurements. The aforementioned method was validated using data from a typical distribution feeder of the Taiwan Power Company. Several NTL cases are demonstrated: a) defective SM with zero reading, b) defective SM with higher reading, and c) electricity theft. The SE was able to identify in every case the bad measurement data. Moreover, when they were replaced by their estimates and the SE was run again, the results were very accurate and close to the actual values. For the ANOVA, two datasets, one of normal and one of fraudulent customers were used, of 8 hours and their baselines for 3 weeks were considered. The model distinguished which was the fraudulent dataset from the 5 ones. Niemira et al. in [13] implement a SE model to detect malicious data attacks by comparing the active and reactive power measurement residuals of a nonlinear SE with those of a linear one (DC model). The main difference of the proposed SE from the traditional ones is that it is designed not only to handle random sensor noise or errors, but also isolated, random bad data. It is assumed that the attacker has partial knowledge of the topology, such as a column of the Jacobian H (which is constant for the DC SE), to prepare an attack measurement vector z. Then, his own measurement and a suitable subset of measurements are modified so as to leave DC residuals unchanged. DC models disregard losses, uneven voltage profiles and reactive power. Thus, if a measurement vector designed for a DC

SE is used with an AC one, there will be an increase in the residuals. Baseline residuals are required for comparison with the current residuals. A 24-bus IEEE network was used to examine the performance of the model. When Monte Carlo noise was added to real measurements, in order to produce data for baseline construction, it was concluded that the active power injection residues of the generators are impacted the most by the attack. Weckx et al. in [14] propose a linearized load flow algorithmic approach using SM data for electricity theft detection via illegal connections, when line lengths are unknown or uncertain. At the same time, basic information of the topology can be extracted and the phase of consumers can be identified in an automatic way. The LV, three-phase, four-wire, radial distribution networks are considered. Active and reactive power, as well as voltage magnitudes, are the required measurements from the SMs to be used in the linear model for the execution of this algorithm:

$$V_{h,k} = V_k^0 + \sum_{\tilde{h}=1}^{N} a_{h,\tilde{h}} P_{\tilde{h},k} + \sum_{\tilde{h}=1}^{N} b_{h,\tilde{h}} Q_{\tilde{h},k} \ , \qquad (1)$$

where k is the time step, h the hth residential consumer and N the total number of houses; V_k^0 is the voltage magnitude at the LV side of the MV/LV transformer, $P_{\tilde{h},k}$ and $Q_{\tilde{h},k}$ the active and reactive power of the consumer $\tilde{h}$ at time step k respectively, and $a_{h,\tilde{h}}$ and $b_{h,\tilde{h}}$ the influence factors of the active and reactive power respectively of consumer $\tilde{h}$ on the voltage magnitude of consumer h.

If there are historical measurements from SMs which are free from fraud, then $a_{h,\tilde{h}}$ and $b_{h,\tilde{h}}$ in 1 can be considered as the unknowns and an ordinary least squares problem is defined. After the influence factors have been determined, then the voltage at each consumer premises can be calculated from (1), using new measurements from SMs that possibly entail electricity theft and can then be compared with the voltage measurement of the SM.

The parameters $a_{h,\tilde{h}}$ and $b_{h,\tilde{h}}$ are also indicators of the relative location of the SM and the phase they are connected to. If the SM h is connected at the same phase as the $\tilde{h}$ the parameter $a_{h,\tilde{h}}$ will be negative since the active power has created a voltage drop. If it is connected to another phase, then the parameter will have a low positive value.

The results of this approach were validated with the simulation of a LV, 4-wire residential feeder in Flanders with 32 customers. The first customer was far away from the substation and the feeder does not have side branches. 1000 steps were required to calculate the influence factors and the identification of the phases was successful. A comparison between the errors of an exact load flow with a 10% uncertainty of cable lengths and the linearized one with unknown cable lengths is also presented; the second case study was found to yield significantly smaller errors (less than 1 V).

In [15] Berrisford describes, within the context of electricity theft detection via SM data, a linear programming optimization method to confirm the network topology by estimating the feeder section impedances, and to provide estimates for the MV/LV transformer LV-side voltage, which in many cases is not measured. The algorithm uses hourly load and voltage measurements. The main idea behind this algorithm is that the voltage of the transformer is equal to the sum of the voltage of any SM and the voltage drop from the transformer to the SM. A set of equations, as many as the SMs, estimating the voltage of the transformer can be formed. The unknown variables in these equations are the line impedances since there are measurements for the SM voltages and active powers. The criterion to obtain the most accurate values for the impedances is the minimization of the transformer voltage variance. This is achieved by using linear programming. When there is theft at a SM, the model exhibits poor convergence. In the simulation, a virtual unknown load with known impedance was added parallel to the irregular SM to represent theft, and the model converged in this case. The method was tested on two transformers of BC Hydro in Canada for hourly measurements of 4 weeks providing promising results. During the first week, for transformer A, the mean standard deviation (MSD) was 0.016% for voltage. The line impedance estimates were consistent for about 4 weeks testing, which implies that the model is accurate. Transformer B had a 0.437 V MSD. This was attributed to the fact that one of the SM had a completely different voltage trend, and it was concluded that it belonged to another transformer. After the SM was removed the MSD was 0.315 V but it was observed that for 2 SM the estimates were not in step with the others. At this point, the virtual load to model theft was added and the MSD decreased to 0.092 V.

Salinas et al. in [16], taking into account customers privacy preserving, propose three distributed algorithms based on peer-to-peer computing in order to calculate customers "honesty coefficients". The distributed LU and QR decompositions are employed to solve a linear system of equations (LSE) while preserving each node's information. For a small network, LU decomposition (LUD) can localize the thieves: unfortunately, the same methodology can prove to be unstable for large networks. For the latter ones, LUD with partial pivoting (LUDP) is implemented, as well as QR decomposition (QRD). The aforementioned methods are applied in cases with constant fraud. In addition to this case, adaptive LUD, LUDP and QUD for scenarios with variable theft activity are also presented.

Those algorithms are intended to be implemented in the SM firmware. An assumption that there is a SM at the concentrator is made, in order to know the overall energy consumption of an area. For a neighborhood with n consumers, let SP be the sampling time, $p_{t_i,j}$ and $\bar{P}_{t_i}$ the recorded energy consumption by the user j at time t_i and the overall consumption recorded at the concentrator level respectively, and k_j the honesty coefficient of the j customer such that $k_j p_{t_i,j}$ gives the actual energy consumption of jth customer for the time period t_i. The sum of all consumers' actual energy at time t_i should be equal to the energy consumption at the concentrator level, thus

$$k_1 p_{t_i,1} + k_2 p_{t_i,2} + \ldots + k_n p_{t_i,n} = \bar{P}_{t_i} \tag{2}$$

The aim is to determine the k_j coefficients. If $k_j = 1$ then the customer is considered honest, if $k_j > 1$ then the recorded energy from that SM is lower than the realized one, and the customer is characterized as fraudulent, and if $0 < k_j < 1$ then the recorded energy is more than the consumed one, suggesting that this SM is defective. With n equations like (i.e. energy measurements for n points in time), a LSE is formed:

$$Pk = \bar{P}\ , \tag{3}$$

where the jth column of P is the recorded energy of the jth SM. Then the data in P are factorized in a lower triangular matrix L and an upper triangular matrix U, so that $P = LU$. A new system is then derived:

$$Ly = \bar{P} \tag{4}$$

$$Uk = y \tag{5}$$

The L, U and y are collaboratively and sequentially calculated by the SMs. For this task, the concentrator has to transmit $\bar{P}_{t_j+1}$ to each SM while each SM calculates only one column of L and y. In order to perform this task it needs the previously calculated columns of these matrices to be transmitted to it by the previous SM. Then backward substitution is used to determine the honesty coefficients k_j. Each SM sends to the previous one the product of one column of U and the calculated honesty coefficient. Additionally, each SM encrypts k_j using the concentrator's public key and the concentrator decrypts all the k_j after the LSE has been solved and the fraudulent SM locations are identified. The LUDP is based on partial pivoting which refers to the exchange of rows of the P matrix in order to arrange all the elements with the greatest absolute value in each column in the diagonal positions. In comparison to LUD, this algorithm requires greater execution time. The QRD algorithm decomposes P into an orthogonal matrix Q $(Q^{-1} = Q^T)$ and an upper triangular matrix R, so that,

$$Rk = Q^T\bar{P} \tag{6}$$

The adaptive LUD, LUDP and QRD algorithms consider variable honesty coefficients. In the area of n consumers, it is assumed that each one may commit fraud with the same probability p. If X is the total number of energy thieves in the area, then X is a random variable with a binomial distribution. When the concentrator decrypts k, it can find the elements that are not equal to 1, denoted as Y, and thus it can calculate the probability of this event happening:

$$P(X = Y) = \binom{n}{Y} p^Y (1-p)^{n-Y} \tag{7}$$

Then, if the customer j commits fraud with different probability p_j, X is a random variable with an expectation $E[X] = \sum_{j=1}^{n} p_j$. By setting a threshold, the concentrator can decide whether a k is valid or not if P is lower than the threshold in which case the SP is reduced, and the process is repeated until the obtained k is the same as the previous one. The performance of the algorithms was verified with simulations where the power measurements were generated using some surveys. LUD performed well with 15 and 30 users with constant honesty coefficients but with 50 users it became unstable, QRD and

LUDP however gave good results with 50 users. With variable coefficients, LUD is stable for 25 users and LUDP, QRD performed well even with 100, 200 and 300 users. Lo and Ansari in [17] deal with false data injection (FDI) attacks by suggesting the combination sum of energy profiles (CONSUMER) attack, involving a number of consumers' SM aiming to achieve a lower consumption record for the attacker and a higher one for the other consumers. The proposed CONSUMER attack model minimizes the number of the violated SMs. This detection technique is based on a grid-sensor placement algorithm that provides increased monitoring to achieve higher hit-rates. In this work, it is assumed that grid operators have complete knowledge of the network topology while radial networks are considered. Let H denote the network configuration matrix, z a set of measurements $z = [P_G, P_1, P_2, \ldots, P_i]^T$ where P_G is the power at the supply point and P_i the power measured by the SM. It is assumed that no irregularities are detected by the traditional bad data detectors. The attacker is considered to have knowledge of H and the state estimation error. With this information, the attacker can build a strategy such that for the normalized residuals applies $\|z_b - H\hat{x}_b\| = \|z - H\hat{x}\| < \delta$, where δ is a pre-determined threshold, z_b and x_b are respectively the measurement and state vectors modified by the attacker. A vector c is designed such that $\hat{x}_b = \hat{x} + c$ and a vector a can be fabricated so that $z_b = z + a = [\bar{P}_G, \bar{P}_1, \bar{P}_2, \ldots, \bar{P}_i]^T \neq 0$ and $a = [a_G, a_1, a_2, \ldots, a_i]^T$, $\sum_{\forall i \in N_{SM}} a_i = a_G = 0$, where N_{SM} is the number of SMs in the examined area. In other words, there are load alterations, and some a_i values will be negative, thus the corresponding SM will exhibit lower energy consumption, and some will be positive by the same overall amount; these will refer to the compromised SMs. The state estimation performed by the grid operators cannot detect the linear alteration of a. The proposed intrusion detection system with power information requires sensor placement across the distribution network. These sensors are of a more simplified design in comparison to SMs, and they belong to the utility. They build a sensor network which is less vulnerable to attacks, as it is designed for grid monitoring. The data of the SMs will be compared with the ones obtained from these sensors. In order to avoid placing sensors in all grid nodes and having an over-determined system, an algorithm that identifies the optimal nodes where sensors should be located is presented. Han et al. in [18] propose a NTL Fraud Detection (NFD) method based solely on data obtained from SMs; no other information of the consumers is required. The criterion used to identify dishonest customers is the difference between the billed energy and the realized consumption. Assuming that technical losses have been estimated by the utility company and excluded, there is also a SM at the distribution transformer recording the overall energy supplied to n customers in a neighborhood. Let E_j denote the energy measured at the distribution transformer and $E_{i,j}$ the actual energy at the ith SM, and $x_{i,j}$ the electricity reported to the utility by the ith SM. By performing energy balance, considering that the technical losses have been calculated and removed, yields

$$E_j = \sum_{i=1}^{n} E_{i,j} \tag{8}$$

If the consumer is honest, then $E_{i,j}/x_{i,j} \approx 1$; for a dishonest customer $|E_{i,j}/x_{i,j} - 1|$ will be very large. For each SM an accuracy coefficient is defined as $a_{i,j} = E_{i,j}/x_{i,j}$. While the reported energy is available, the actual values are not. There is a function for each SM such that $f_i(x_{i,j}) = E_{i,j},\ \ j = 1, 2, \ldots, m$. Based on Taylor approximation $f_i(x) = \sum_{k=m}^{o} a_{k,i} x^k$. By replacing the previous two equations in (8) yields

$$E_j = \sum_{i=1}^{n} \sum_{k=m}^{o} a_{k,i} x_{i,j}^k \tag{9}$$

With m samples of x^k for each SM and E_j known, the accuracy coefficients can be estimated. Simulations were performed to examine the performance of the model.

2.2 Artificial Intelligence-based

Artificial Intelligence-based theft detection techniques are the most popular ones, since they were available to use before the deployment of SMs, and because now they can further advance and improve remarkably within the framework of SMs. These methods usually refer to the classification of the consumers load profile. The aim is to determine irregular patterns in the electricity consumption over time, based on a training dataset that includes normal and irregular cases. The main steps followed in a classification approach are: a) data acquisition, b) data preprocessing, c) feature selection, d) classifier training, e) data-of-interest classification, f) data post-processing, and g) theft-suspects identification.

Nagi et al. in [20] approach the electricity theft detection problem by developing an artificial intelligence technique, namely a support vector machine (SVM). In this method historical consumption data and additional consumers attributes are used to identify irregular consumption profiles that are highly correlated with NTL. The consumers are classified either as "normal" or "fraud" by the SVM model. The consumers' consumption patterns are determined by employing data-mining and statistical analysis tools trying to identify sudden changes in the consumption profiles. Specifically in this paper, the SVM solves a binary classification problem by finding the optimal $f(x) = \text{sgn}(g(x))$, where $g(x)$ is the decision boundary between the two classes, that accurately classifies new data into the two classes while minimizing the classification error. The method of structural risk minimization is exploited. The method was tested using historical data of three Malaysian cities for 265,870 customers and for 25 months. The features that were eventually chosen include: a) 24 daily average energy consumption values for each customer, which correspond to their load profile (estimated as the monthly consumption divided by the number of days between two consecutive measurements), and b) the credit worthiness information CWR (this is produced by the utility's billing system automatically for customers that do not pay their bills) for each customer. The data were normalized, formatted and then used for the training and testing of the SVM model. After collaboration and on-site inspection with Tenaga Nasional Berhad, it was found out that the expected hit rate increased from 3 % to 60 %.

In Nagi et al. [21] the work of [20] was extended, introducing a fuzzy inference system (FIS) in the form of IF-THEN rules. For each customer, an output ranging from 0 to 1 is produced by the FIS. The customers with outputs from 0.5 and higher are considered to have higher probability to be fraudulent. This method seemed to improve the previously 60 % hit rate to 72 %. It is worthy to mention the work of [24], where a method to identify the features that best describe possible illegal consumers is proposed.

Babu et al. in [22] use fuzzy C-Means clustering to categorize consumers based on their consumption patterns. The difference of clustering to classification is mainly that the latter one has a training dataset where the response of the observations is already known and classifies new data. Clustering is the grouping of observations into classes of similar objects. In fuzzy clustering, an observation can belong to more than one class, with a different degree-of-membership. The fraud identification relies on the fuzzy membership function and the normalized Euclidean distances of cluster centers ordered by unitary index score. The highest score represents fraudulent consumers. The method uses five attributes that are considered to describe a consumption pattern. These attributes include: a) the average consumption, b) the maximum consumption, c) the standard deviation of consumption, d) the sum of inspection comments during the last six months, and e) the average consumption of the neighborhood. Data of another twelve months are required for the clustering process. The method was tested with real data from one neighborhood with 57 consumers from India and it achieved a hit rate of 80 %.

Faria et al. in [23] utilize the consumer baseline load calculation methods that have been developed within the context of demand response. For each period of the historical data, the expected consumption is estimated, then this is compared with the realized one and if there is considerable difference the consumer is characterized as a possibly fraudulent one. The baseline types that were used are the following: a) type I, which uses load historical data and may include other data such as weather, and b) type II, which is used for aggregated loads. After the expected energy consumption calculation, statistics regarding the expected and measured consumption are produced and compared. These statistics include whole data average (WDAVG), whole data standard deviation (WDSTD), past data average (PDAVG) and past data standard deviation (PDSTD). Whole data refers to the overall data of the examined consumer, and past data refers to the past data of each calculation period. The performance of the proposed method is demonstrated by a case study.

3 Conclusions

Non-technical losses detection is a hard and challenging issue for the distribution operators. With the massive deployment of SMs, new possibilities to detect electricity theft are opened up. This paper has discussed the challenging issues in energy theft detection and provided some research directions. In addition, NTL detection methods within AMI have been investigated and categorized in three groups. After examination of existing approaches, it can be concluded that each proposal addresses only a few aspects of the multidimensional problem of electricity theft. Therefore, the authors believe that energy-theft detection robust methods of the future will include both system state-based techniques that lie in the Kirchhoff laws applied to low voltage circuits and artificial-based methods that lie in the detection of anomalies in the consumption pattern of consumers. With the assistance of both methods, the weaknesses of each technique, related mainly with lack of information, could be compensated successfully.

Acknowledgment

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7-PEOPLE-2013-ITN) under grant agreement no 607774.

References

1. Lu, R., Liang, X., Li, X., and Shen, X.: An Efficient and Privacy-Preserving Aggregation Scheme for Secure Smart Grid Communications. IEEE Trans. on Parallel and Distributed Systems, 23, 9, 1621–1631 (2012)
2. Kadurek, P., Blom, J., Cobben, J.F., and Kling, W.L.: Theft detection and smart metering practices and expectations in the Netherlands. Innovative Smart Grid Technologies Conference Europe (ISGT Europe), Gothenburg (2010)
3. Jiang, R., Lu, R., Wang, Y., Luo, J., Shen, C., and Shen, X.: Energy-Theft Detection Issues for Advanced Metering Infrastructure. Tsinghua Science and Technology, 19, 2, 105-120 (2014)
4. Amin, S., Schwarz, G.A., Cárdenas, A.A., and Sastry, S.S.: Game-Theoretic Models of Electricity Theft Detection in Smart Utility Networks: Providing New Capabilities with Advanced Metering Infrastructure. Control Systems, IEEE, 35, 1 (2015)
5. Depuru, S.S., Wang L., and Devabhaktuni, V.: Enhanced Encoding Technique for Identying Abnormal Energy Usage Pattern. North American Power Symposium (NAPS), Chanpaign, IL (2012)
6. Smith, T.B.: Electricity theft: a comparative analysis. Energy Policy, 32, 18, 2067-2076 (2004)
7. Depuru, S.S., Wang L., Devabhaktuni, V., and Gudi, N.: Measures and Setbacks for Controlling Electricity Theft. North American Power Symposium (NAPS), Arlington, TX (2010)
8. Salinas, S., Li, M., and Li, P.: Privacy-Preserving Energy Theft Detection in Smart Grids. IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), Seoul (2012).
9. Lu, C.N., Huang, S.C., Lo, Y.L.: Non-Technical Loss Detection using State Estimation and Analysis of Variance. IEEE Transactions on Power Systems, 28, 3, 2959-2966 (2013)
10. Chen, L., Xu, X., and Wang, C.: Research on Anti-electricity Stealing Method Base on State Estimation. Power Engineering and Automation Conference (PEAM), IEEE, Wuhan (2011)
11. Abur, A., Gómez-Expósito, A.: Power System State Estimation: Theory and Implementation. Marcel Dekker, New York (2004)

12. Lo, Y.L., Huang, S.C., and Lu, C.N.: Non-Technical Loss Detection Using Smart Distribution Network Measurement Data. Innovative Smart Grid Technologies - Asia (ISGT Asia), IEEE , Tianjin (2012)
13. Niemira, W., Bobba, R.B., Sauer, P., and Sanders, W.H.: Malicious Data Detection in State Estimation Leveraging System Losses & Estimation of Perturbed Parameters. International Conference on Smart Grid Communications (SmartGridComm), IEEE, Vancouver, BC (2013)
14. Weckx, S., Gonzalez, C., Tant, J., De Rybel, T., and Driesen, J.: Parameter Identification of Unknown Radial Grids for Theft Detection. IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies (ISGT Europe), Berlin (2012)
15. Berrisford, A.J.: A tale of two transformers: An algorithm for estimating distribution secondary electric parameters using smart meter data. 26th Annual IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Regina, SK (2013)
16. Salinas, S., Li, M., and Li, P.: Privacy-Preserving Energy Theft Detection in Smart Grids: A P2P Computing Approach. IEEE Journal on Selected Areas in Communications, 31, 9, 257-267 (2013)
17. Lo, C.H., Ansari, N.: CONSUMER: A Novel Hybrid Intrusion Detection System for Distribution Networks in Smart Grid. IEEE Transactions on Emerging Topics in Computing, 1, 1, 33-44 (2013).
18. Han, W, and Xiao,: NFD: A Practical Scheme to Detect Non-Technical Loss Fraud in Smart Grid. IEEE International Conference on Communications (ICC), Sydney, NSW (2014)
19. Cardenas, A.A., Amin,, Schwartz, G., Dong, R., and Sastry, S.: A Game Theory Model for Electricity Theft Detection and Privacy-Aware Control in AMI Systems, Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL (2012)
20. Nagi, J., Yap, K.S., Tiong, S.K., Ahmed, S.K., and Mohamad, M.: Nontechnical Loss Detection for Metered Customers in Power Utility Using Support Vector Machines, IEEE Transactions on Power Delivery, 25, 2, 1162-1171, (2009)
21. Nagi, J. Yap, K.S., Tiong, S.T., Ahmed, S.K., and Nagi, F.: Improving SVM-Based Nontechnical Loss Detection in Power Utility Using the Fuzzy Inference System. IEEE Transactions on Power Delivery, 26, 1, 1284-1285 (2011)
22. Babu, T.V., Murthy, T.S., and Sivaiah, B.: Detecting Unusual Customer Consumption Profiles in Power Distribution Systems - APSPDCL. IEEE International Conference on Computational Intelligence and Computing Research (ICCIC) , Enathi (2013)
23. Faria, P., Vale, Z., Antunes, P., and Souza, A.: Using Baseline Methods to Identify Non-technical Losses in the Context of Smart Grids. IEEE PES Conference on Innovative Smart Grid Technologies Latin America (ISGT LA), Sao Paulo (2013)
24. Ramos, C.C.O., Papa, J.P., Souza, A.N., Chiachia, G., Falcão, A.X.: What is the Importance of Selecting Features for Non-Technical Losses Identification? International Symposium on Circuits and Systems (ISCAS), Rio de Janeiro (2011)