Solutions for detection of non-technical losses in the electricity grid: a review

Joaquim L. Viegas[a], Paulo R. Esteves[b], R. Melício[a,c], V. M. F. Mendes[c,d], Susana M. Vieira[a]

[a] *IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1, 1049-001 Lisbon, Portugal*
[b] *Power Data, Portugal*
[c] *Dep.de Física, Escola de Ciências e Tecnologia, Universidade de Évora, Portugal*
[d] *C-MAST, Centre for Mechanical and Aerospace Sciences and Technology, Portugal*

Abstract

This paper is a review of literature with an analysis on a selection of scientific studies for detection of non-technical losses. Non-technical losses occurring in the electric grid at level of transmission or of distribution have negative impact on economies, affecting utilities, paying consumers and states. The paper is concerned with the lines of research pursued, the main techniques used and the limitations on current solutions. Also, a typology for the categorization of solutions for detection of non-technical losses is proposed and the sources and possible attack/vulnerability points are identified. The selected literature covers a wide range of solutions associated with non-technical losses. Of the 103 selected studies, 6 are theoretical, 25 propose hardware solutions and 72 propose non-hardware solutions. Data based classification models and data from consumption with high resolution are respectively required in about 47% and 35% of the reported solutions. Available solutions cover a wide range of cases, with the main limitation found being the lack of an unified solution, which enables the detection of all kinds of non-technical losses.

Keywords: Non-technical losses, Commercial losses, Electricity theft, Electricity fraud, Detection, Modelling, Estimation, Literature review

Email address: joaquim.viegas@tecnico.ulisboa.pt (Joaquim L. Viegas[a])

Contents

1. Introduction

Losses of electric energy in the grid at level of transmission and distribution (T&D) encompass both technical losses and non-technical losses (NTLs) and for the estimation of the later is normally required the estimation of the former [1–4]. Technical losses are naturally occurring losses due to irradiation and as a result of inevitable dissipation of electric energy into the equipment necessary for implementing T&D [1, 4, 5], involving losses in dielectrics and mostly in the conductors by Joule's effect. NTLs are also referred in the literature as commercial losses are non-natural losses associated with the amount of non-billed electricity and billed electricity that is not paid for. The non-billed electricity occurs due to either errors in metering or billing, or non-legitimate behavior of consumers [1, 5, 6]. Non-legitimated behavior, i.e., electric energy use by fraudulent behavior of users, has been detected in association with institutionalized theft, corruption and organized crime [6]. The cost associated with the NTLs has to be covered by the participation of utilities and or of legitimate consumers, i.e., consumers paying bills. Also, if there is public funding for the use of electric energy, as in some states with the aim of enabling supply electricity to non-profitable geography or demography, more public funding is needed. Consequently, NTLs result in augmented costs to utilities, legitimate consumers and states [1, 6, 7].

Table 1: Nomenclature

Nomenclature

NTLs	Non-Technical Losses	T	Period
RQ	Research Question	PES	Power and Energy Society
T&D	Transmission and Distribution	ANN	Artificial Neural Networks
SG	Smart Grid	SVM	Support Vector Machines
SM	Smart Meter	RBS	Rule Based System
AMI	Automatic Metering Interface	DT	Decision Tree
LV	Low Voltage	FS	Feature Selection
MV	Medium Voltage		

The most negative consequences due to the NTLs are found in fragile environments associated either with economies in transition or with developing economies. For example, in Jamaica in 2013, NTLs are up to US$46 million, accounting for 18% of the total fuel bill [1]. In India, yearly losses due to electricity are over 1% of the gross domestic product (GDP) [7, 8]. In general NTLs are prone to have unsustainable consequences in already fragile environments. But although developed economies have less serious consequences regarding NTLs, the existing ones are still considered as needing a convenient treatment. The developed economies of UK and USA have estimated that electricity theft is £173 million [9] and US$6 billion [10] every year, respectively. In Europe, the reduction of NTLs are essential to reap the benefits of currently undergoing deployment of advanced meters [11].

The spreading of the smart grid (SG) concept and the widespread deployment of smart meters (SMs) leads to an extended attack surface to electricity grids and to vulnerability of the software in the meters [12, 13]. A SG intends to enable the intelligent integration and optimization of the whole electricity supply chain, enabling stable distributed generation, efficient transmission and strong engagement of consumers to promote sustainability minded behavior [14–18]. SMs differ from traditional metering systems regarding their advanced communications and processing capabilities, enabling the collection of data regarding high resolution of consumption or consumer services (e.g. automatic efficient control of appliances, demand side management). Cyber attacks can potentially manipulate meter software, delivering fraudulent readings to utilities, disconnecting consumers by remote action and even compromising utility systems operation [12, 13].

Detection of NTLs has been receiving a growing interest both in academia and industry in order to find adequate approaches to face NTLs. Approaches may use statistical analysis to capture the main drivers of fraudulent behavior, customizing adequate methods for software regarding patterns of electric energy usage. Other approaches may use algorithms to analyze the data collected from SMs, enabling the detection of patterns that may indicate the presence of fraudulent behavior. Equipment configurations and grid structures have also been proposed to enable the detection and reduction of NTLs. Althought there are studies that have thoroughly analyzed the issue of NTLs [1, 6], there is no systematic analysis of solutions for the detection of NTLs. In [12, 19] the techniques to detect electricity theft based on smart metering data are analyzed. In [20] an overview of the types of techniques is presented, but in limited way only covering data-based solutions. The authors believe that the growing amount of literature and the wide range of techniques and solutions justify the need for a review on the state-of-the-art for abstraction of the main lines pursued by researchers.

This paper presents a literature review on the topic of the detection of NTLs, giving researchers and utilities an overview of the available methods and requirements to development applications for the detection of NTLs. The paper covers detection techniques for all found vulnerability/attack points that are potential sources of NTLs, analyzing techniques that estimate NTLs at the system level, solutions to identify consumers with a high probability of thieving behavior, and techniques to detect patterns that may implying vulnerabilities in metering equipment. The main contributions of the paper are the following:

- An up-to-date analysis of types and possible attack vectors related to NTLs;
- A review and analysis of studies presenting solutions for detection of NTLs;

- A typology of solutions for the detection of NTLs;
- An analysis of requirements for detection of NTLs;
- An analysis of the limitations of current solutions for detection of NTLs and gaps in current available research.

An analysis of 103 papers is selected to carried out the literature review, using methods inspired by the systematic literature review guidelines [21, 22]. The evolution of the number of studies published, the main journals and conferences involved, the techniques most commonly used, types of data needed and main limitations of currently available solutions for detection of NTLs are presented. The paper is structured as follows: Section 2 presents the method followed in the paper. Section 3 presents an analysis on the types and sources of NTLs. Section 4 presents an analysis of the results on solutions for detection of NTLs. Section 5 presents the conclusions.

2. Method

The main contributions of the paper are derived from gathering, analyzing and summarizing in a systematic way the existing solutions for the detection of NTLs. The methodology used has as guidelines the systematic review guidelines proposed by Kitchenham [21, 22] and the systematic review on energy management systems by Rasool et al. [23].

2.1. Questions

The lines of inquiry for which there is a body of studies large enough to allow capturing the main directions of the proposed solutions are associated with the following research questions (RQs):

- *RQ1: What types of NTLs are considered in literature?*
- *RQ2: Which type of research is conducted on the detection of NTLs?*
- *RQ3: What are the main techniques and data used for the detection of NTLs?*
- *RQ4: What are the limitations of current solutions and future perspectives?*

RQ2, *RQ3* and *RQ4* are main motivations for carrying out the review. An up-to-date response to "*RQ1: What types of NTLs are considered in literature?*" is the main support to address the responses to *RQ2* and *RQ3*.

2.2. Search

The search in the literature review about the detection of non-technical losses is concerned with state-of-art reported in studies appearing since 2000 in the following three databases: ScienceDirect, ACM Digital Library and IEEE Xplore.

2.2.1. Search terms

The search terms are designed to obtain general queries that minimize the chances of missing any relevant study. The queries are built to include studies of NTLs, especial having explicitly text about electricity theft or fraud, the combination of the following terms is used to search in the titles and abstracts of studies in the databases:

1. ”electric” or ”electricity”;
2. ”theft” or ”fraud” or ”non-technical loss” or ”non-technical losses”.

The number of papers resulting from the search queries is presented in Table 2.

Table 2: Results from search queries

Query	Results from query	Date
ScienceDirect	35	26/02/2016
ACM Digital Library	11	26/02/2016
IEEE Xplore	143	26/02/2016

2.3. Criteria

Exclusion and inclusion criteria are used to select the considered studies in the review from the pool of studies resulting from the queries.

2.3.1. Exclusion criteria

- Study not related to non-technical losses in electricity grids;
- Study published before 2000;
- Study presenting the SG, SMs or automatic metering interface (AMI) as a solution to detect and reduce NTLs, without presenting additional details on the detection solution.

2.3.2. Inclusion criteria

- Study provides detection, estimation or prediction of any kind of non-technical losses in the electricity grid;
- Study proposes solution for the detection of any kind of NTLs;
- Study presents a comparison of multiple solutions for the detection of any kind of NTLs;
- Study presents determinant variables and factors on non-technical losses in the electric grid.

2.4. Data collection

The abstract and conclusions of the studies that resulted from the search queries are subjected to a procedure for the identification of the ones not meeting any of the exclusion criteria and meeting at least one of the inclusion criteria. A total of 103 studies resulted from the application of the aforementioned criteria. These studies are the ones subjected to detailed analysis, where the different characteristics of the lines of research are collected in a systematic way. The following attributes are:

- Authors, Title, Year of publication, Journal/conference;
- Source research database;
- Category: Category of solution proposed;
- Type: Type of solution proposed;
- Smart meters: Identifier on the requirement of smart meters for the proposed solution;
- Data: Types of data used and/or required by the proposed solution;
- Consumption/load resolution: Resolution of consumption or load data used and/or required by the solution,
- Techniques: Specific techniques used and/or proposed in the presented solution;
- Real data: Identifier on the use of real data to validate the solution;

The categories, types of solutions, types of data and techniques are specified and analyzed in detail in Section 4.

2.5. Data analysis

The extracted information from the literature review in order to answer the RQs is subjected to a procedure of analysis organized as follows:

1. In 3. the types of NTLs and potential points of attack/vulnerability are identified, pictured and listed in Figure 1 and Table 3, respectively.
2. In 4. the literature review is quantitatively summarized and the distribution per year is pictured in Figure 2, the main publishing journals and conferences are in Table 4.
3. In 4.1. the typology proposed is stated, the lists of the studies organized by categories and types are in Table 5.
4. In 4.2. the techniques for detection of NTLs are analyzed, the main techniques are in Table 7.
5. In 4.3. the requirements for the solutions are analyzed, Table 8 with Table 9 and Table 10 with Table 11 list the requirements of studies considered as *Theoretical study* and *Non-hardware solution*, respectively.
6. In 4.4. limitations on the categories of the solutions in the literature review are summarized.

3. Non-Technical Losses

The total amount of T&D losses, technical and non-technical, amounts to the difference between the total electric energy injected into the T&D and the one associated with the revenue due to the bills paid by customers. While technical losses account for the electric energy dissipated through the equipment necessary to implement the T&D of electricity [1, 6], the NTLs account for the difference between total T&D losses and technical losses. NTLs can be estimated, but exact measurement is not feasible [6]. NTLs have been mostly verified as the result of fraud through hardware tampering, theft by line tapping and unpaid bills. Also, irregularities in the measurement of consumption/billing and collusion with utility employees are considered to be sources of NTLs [1, 6, 24]. The following studies [1, 6, 20, 24] present an overview about sources of NTLs. Also, with the emergence of the SG concept and the emerging global roll-out of meters with advanced communications capabilities, SMs studies are identifying an interest with lines of research on to new potential points of attack/vulnerability [19, 25, 26]. Cyber and data attacks are identified need to be taken into account in solutions for the detection and mitigation of NTLs. A focus on possible attack vectors on metering equipment with advanced communication capabilities [19, 25–27] is a line of recent research that have to be carried out in order to face the menace of NTLs in the context of SGs. Sources of NTLs and attack/vulnerability points are pictured in Figure 1.

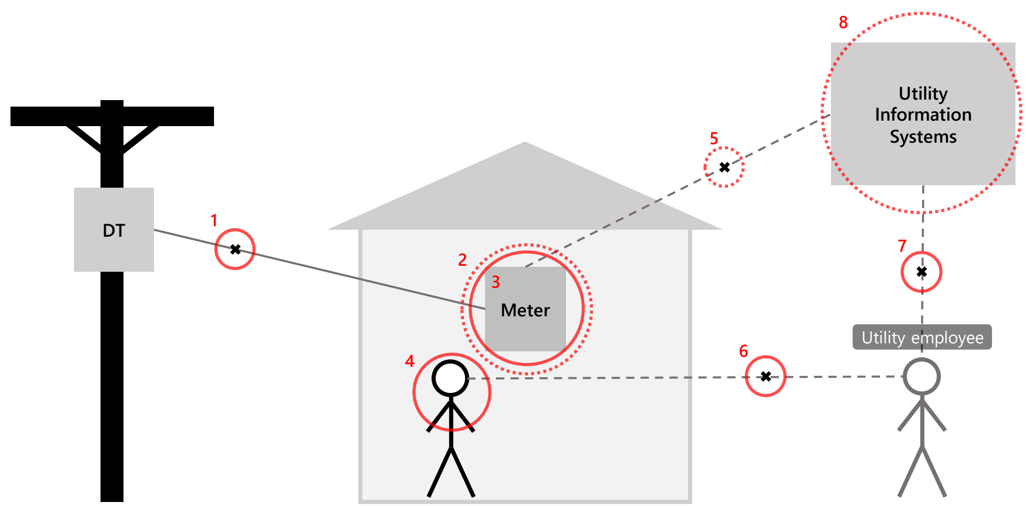

Figure 1: NTLs sources and points of attack/vulnerability.

In Figure 1 the continuous line represents a physical electricity connection and the dashed lines represent channels of communications. 1) points to the distribution line between a medium voltage/low voltage (MV/LV) or low voltage/low voltage (LV/LV) transformer and the household of the electricity customer; 2) points to the meter software; 3) points to the physical meter hardware and components; 4) points to the electricity customer; 5) points to the channel of communications between a meter and the utility; 6) points to the communication and relation between the cufstomer and an utility employee; 7) points to the channel of communication between the employee and the utility and 8) points to the utility information systems.

The capture of the main ideas in the literature review allows a synthesis of imputations derived from acts of the types fraud/theft or fraud/error for the NTLs. A lists of the imputations denominated as zones for NTLs, types and possible sources of attack/vulnerability vectors is presented in Table 3.

Table 3: List of zones, types and possible sources of NTLs

Zone	Type	Sources and attack/vulnerability vectors	Point
Before meter	Fraud/theft	Connecting throw-ups on a distribution feeder [1]	1
Meter	Fraud/theft	Reverse the meter [27]	3
		Disconnect the meter [27]	3
		Bypass meter to remove measurement [27]	3
		Interfering with meter (e.g. by strong magnet) [19]	3
		Compromise meter through remote network exploit [27]	2
		Modify firmware/storage on meter [27]	2
		Steal credential to login to meter [27]	2
		Intercept/alter communications [27]	5
	Fault/error	Inadequacy and inaccuracy of meter reading [20]	3
		Losses due to faulty meter and equipment [20]	3
Billing	Fraud/theft	Non-payment of bills [20]	4
		Arranging billing irregularity help by internal employees [20] (collusion)	6
		Cyber attack to information systems	8
	Fault/error	Inaccurate or erroneous customer electricity billing due to faulty information systems or employer [20]	8
		Inaccurate or erroneous meter reading	7

In table 3 are identified three imputations for sources of NTLs, leading to the type either fraud/theft or fault/error. The last column of the table refers to the numbers in Figure 1 for identification of the points of attack/vulnerability. The characterizations of imputations identified by zones are the following:

1. *Before meter*: Fraud/theft source of NTLs can occur is this zone, such as the illegal tapping of distribution lines and feeders;
2. *Meter*: Fraud/theft source of NTLs can occur in this zone, such as reversing, disconnecting, bypassing and interfering with the meter. Network exploits, software and firmware exploits can also compromise the meters and measurements. Fault/error can occur in this zone, such as

the presence of inaccuracy in the metering equipment and equipment failure;

3. *Billing*: Fraud/theft sources of NTLs can occur in this zone, such as the non-payment of bill, collusion between customer and utility employer to arrange billing reduction and cyber attack to billing systems. Fault/error can occur in this zone, such as inaccurate billing due to faulty system and employer error.

4. Detection of Non-Technical Losses

Data analysis of the attributes is extracted from selected studies published from 2000 to 2016, having the distribution per year pictured in Figure 2.

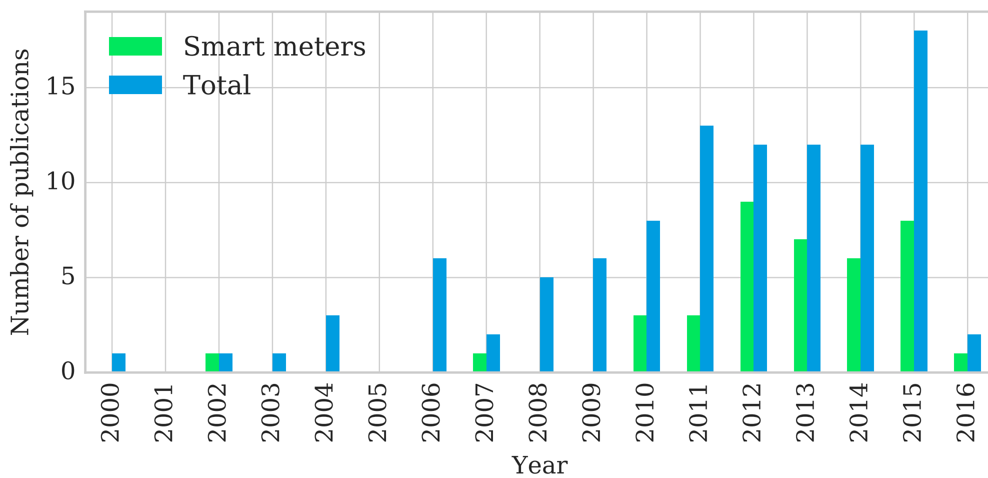

Figure 2: Number of selected studies per year of publication.

In Figure 2 the total number of studies and the ones particularly regarding the research in SMs are presented in blue and in green, respectively. Although with few studies at the beginning from 2000 to 2005, the interest on the detection of NTLs has been more regular from 2006 to 2015. The highest amount of published studies is nineteen in 2015. Also, the development of solutions for detection of NTLs has been growing in popularity in developed countries motivated by policies for higher efficiency and the digitization of the grid, enabling the collection and analysis of data of consumption and asset operations. The studies with solution concerning the use of SMs show that a significant share of the literature has been dedicated to this concern. These type of studies appear mostly from 2010 to 2015. The journals and conferences with at least two selected studies are listed in Table 4.

Table 4: Journals and conferences with at least two selected studies

Journal/conference	#
IEEE PES General Meeting	6
International Journal of Electrical Power & Energy Systems	4
IEEE Transactions on Power Delivery	4
Energy Policy	4
IEEE PES Conference on Innovative Smart Grid Technologies	4
IEEE PES Transmission & Distribution Conference and Exposition: Latin America	4
IEEE Transactions on Power Systems	3
International Conference on the European Energy Market	2
International Conference on Intelligent System Applications to Power Systems	2
IEEE PES Transmission & Distribution Conference and Exposition	2
Utilities Policy	2
International Conference on Power System Technology	2
IEEE International Power and Energy Conference	2
IEEE International Conference on Communications	2
Lecture Notes in Computer Science	2
IEEE International Conference on Smart Grid Communications	2
IEEE Transactions on Smart Grid	2
IEEE PES Conference on Innovative Smart Grid Technologies Europe	2
IEEE PES Power Systems Conference and Exposition	2

In Table 4 are listed the journals and conferences with at least two selected studies, listing 53 of a total of 103 selected studies. A significant number of the studies come from proceedings of conferences related to the IEEE Power & Energy and Society (PES) such as the PES General Meetings, Conferences on Innovative Smart Grid Technologies and T&D Conferences. The journals with more studies published are the International Journal of Electrical Power & Energy Systems, IEEE Transactions on Power Delivery and Energy Policy.

4.1. Typology of studies

Although, initial studies are dominated by the paradigm of analysis of T&D losses and customer subjected to inspection as the staples for the detection of NTLs [6], during the time horizon in observation a wide array of techniques and approaches have been proposed in the literature to detect, estimate and analyze NTLs. Hence, to provide an overview of techniques for the detection of NTLs a typology is proposed with the following three categories:

1. *Theoretical study;*
2. *Hardware solution;*
3. *Non-hardware solution.*

The relation between socio-economic and demographic factors that can help inform policy and decision makers to analyze and reduce the phenomena of NTLs[8, 28, 29] is addressed in studies categorized as *theoretical study*. The installation or implementation of specific equipments, e.g., meters with tempering sensors, RFID equipments, meters with redundancy is addressed and categorized as *hardware solution*, focusing on specific metering hardware, infrastructure and equipment to enable the detection of NTLs. The advance of data processing and communication capabilities have open new lines of research and solutions based on the analysis of consumer data [12, 30] and categorized as *non-hardware solution*, assuming that NTLs result in a deviation from the norm of consumption patterns or other consumer characteristics.

Studies presenting a list or comparison of different NTLs detection techniques only focus on the *non-hardware solution*. In [20] fraud detection techniques are classified as unsupervised, supervised and semi-supervised. In [12] techniques for theft detection are classified as classification-based, state-based and game theory-based. The typology is detailed in the following: a general description of each category is given followed by a more detailed analysis of the types of techniques, the leading studies, advantages, disadvantages and relation to the identified NTLs zones.

Theoretical study: Study on the analysis of factors and variables that may reveal the presence of NTLs in a population or geographical area, tending to focus on analyzing demographic drivers and social aspects related to electricity fraud. This type of study is considered because knowledge of these variables and factors may enable the detection of NTLs in a certain population or area. This category encompasses the following type of technique:

- *Analysis of variables and factors:* The leading studies of this type make use of statistical techniques to find the relationships between socio-demographic, economic, market variables and the amount of theft [31, 32]. In another influential study [8], empirical analysis based on surveys and ethnographic fieldwork is used to understand the main factors related to theft in a region of Tanzania and one of India. A recent study stands out as an example [28], where the author analyzes determinant socio-economic attributes of illegal consumers of electricity through econometric analysis.

  These techniques have the advantage of producing results that can have an high impact, being useful to design policy and make decisions to reduce NTLs. The volume and complexity of the data used is normally easily manageable, has it consists of variables and indicators that aggregate entire regions. The main disadvantage of these studies comes from the scope of the analysis, which focuses on a whole region or country. The studies can estimate and find the drivers of aggregate NTLs but are not adequate to find specific cases of theft and faults in metering or billing.

  Theoretical study is not specifically aimed at any of the zones listed in Table 3, because this category of studies work with data that aggregates all NTLs.

Hardware solution: Study proposing a solution in which the main focus is the characterization and description of equipment that enables the detection or estimation of NTLs. Most studies of this type focus the proposal of metering and sensing hardware. Many of these studies also comprise the ones in the category of *non-hardware solution*, in which a software component is essential for the processing of the data generated by the equipment. This category encompasses the following types of techniques:

- *Metering hardware*: Study presenting solutions specifying metering hardware details and specifications. The selected studies of this type present no apparent common methodology, presenting different ways to design meters or modify the hardware of existing ones to enable easier detection of theft. In [33] and [34], the authors propose a system based on specific processor architectures and algorithms, which enables tampering protection through the detection and communication of intrusion events. RFID tags have been proposed to be used to seal meters and speed up inspections [35]. In Brazil, a metering architecture using two reading points has been tested and concluded as enabling easy theft detection [36].

This type of solutions have the advantage of being able to completely disable some theft options, such as meter reversal and disconnection. The main disadvantage is the significant cost of installing hardware in an high number of households. Metering hardware solutions can only detect NTLs that result from the *meter* zone, as listed in Table 3.

- *Metering infrastructure*: Study presenting solutions based on a set of metering assets and/or sensing hardware, focusing on infrastructure characteristics, such as installation strategies and number of equipments needed based on geographical location. Leading studies proposing solutions of this type focus on the different data collection equipments needed at different locations of the grid (e.g. customers households, distribution transformers and substations) to effectively calculate NTLs and detect their sources [37, 38]. In [26] a comprehensive analysis focused on the threat faced by automated metering infrastructures (AMI) systems is presented, the authors propose different system architectures to combat the different possible attacks.

  These solutions have the advantage of being able to detect any kind of NTLs when the source is in the *meter* or *before meter* zones. The main disadvantage are the high equipment costs associated. While it is not feasible to change overnight the whole architecture of an electrical grid, these studies present important aspects to take into account when utilities take decisions to modernize the electrical networks.

- *Signal generation and processing*: Study presenting solutions that make use of signal generation and processing to detect NTLs. While only a limited number of studies present solutions of this type, they give practical ways to control and detect sources of NTLs. In [7, 39] the use of an harmonic signal generator is proposed, after disconnecting the meters of legal consumers, introducing a signal to the distribution feeders that affects equipments connected to the electricity. In [40] an high frequency signal generator and analysis are used, after disconnecting consumers, to detect illegally connected equipments that contribute to significant loads in distribution.

  This type of solution has the advantage of being able to detect all types of NTLs in the grid. The main disadvantage of these solutions is the dependency on the presence of smart metering systems.

- *Other approaches*: Study that does not fit in any of the aforementioned types. For example, a research group has proposed the use of a light sensor to gather information of public lighting points in Brazil, this is done to make sure the electricity being used for municipal lighting is correctly reported and payed for [41]. Another study uses forensic investigation procedures to find possible cases of collusion and fraud on meters [42].

Non-hardware solution: Study in which the main focus is the characterization and description of a non-hardware solution, i.e. software, which enables the detection or estimation of NTLs. Most studies of this category focus on describing classification techniques that infer the presence of NTLs from electricity consumption or other data. Many of these studies present techniques which require specific hardware requirements for the acquisition of data. This category encompasses the following types of techniques:

- *Classification*: Study presenting techniques that provide predictions for the presence of significant NTLs or theft in a location or consumption end-point. There are many different classification models and algorithms used in the literature, the ones that are more suitable for this application are analyzed in more detail in Section 4.2.

  The studies with highest impact use a support vector-machines (SVM) model to infer, from consumption data and other information about the consumer, the presence of theft or other sources of NTLs. The leading studies supported by SVM propose a method for the electricity provider of peninsular Malaysia that predicts the presence of theft from the monthly consumption data of consumers and risk information, achieving an increases in the detection hit-rate from 3% to 60% [43–47]. Other leading studies that deal with higher resolution data with consumption values collected with time periods equal or under 1 hour have also been developed [20, 30, 48–50]. These studies use load profiling techniques to understand normal customer consumption patterns and classification techniques to predict future behaviors. The studies propose an outlier detection scheme to find irregular consumption points that may be sources of NTLs. Similar techniques are proposed by research groups from Brazil [51–55] and Spain [56–61].

The main advantage of this type of techniques is the low investment cost, utilities have to have computational resources to run these data mining and classification models and make additional use of the data already collected. The main disadvantage is that the presence of sources of NTLs on detection is not guaranteed, being used for enabling a more efficient use of inspection resources. Another possible disadvantage is the data requirements of the techniques (e.g. if high resolution data is required, SMs and remote collection of consumption information are needed). Classification techniques can detect NTLs in all zones listed in Table 3. If these techniques are based on collected consumption data, instead of billing data, is possible to have only ability to detect sources of NTLs located at the zones *before meter* or *meter*, billing subject to manipulations is not detected by this data.

- *Estimation*: Studies presenting techniques that provide an absolute or relative estimate of the amount of NTLs from an area or customer. Leading studies use state estimation to estimate irregularities and errors in customers demand data or use technical loss modeling to estimate aggregated NTLs. In [62] a simple state estimation technique is proposed to estimate the deviation between customers billed and actual electricity consumption. To tackle the new threat of false data injection in the SG context, state estimation has been combined with attacker modeling, enabling the detection of NTLs that would be undetected by traditional methods [63]. To improve the estimation of NTLs, statistical methods are proposed, finding accurate relations between losses and load factors [2, 64]. More recently, spatio-temporal analysis, pattern analysis, generalized additive models and a markov chain model are proposed to estimate NTLs geographically [65].

  State estimation has advantages similar to classification techniques, requiring low investment if the needed data collection assets are already in place. This technique is more precise than classification techniques but require more accurate and complete data on the distribution network loads.

Technical loss modeling has the advantages of having low cost and of requiring data that every utility has, the only disadvantage is that aggregated NTLs are estimated, i.e., there is no ability to detect the specific sources of NTLs. These techniques are similarly to classification and can be applied on all the zones that can be possible sources of NTLs. But in of absence data used for billing is not used may be limited to the zones *before meter* and *meter*.

- *Game theory*: Study presenting detection techniques based on game theory for modeling legitimate consumers, fraudsters and the relationships with the electricity utility. In [66] a comprehensive game theory model to develop and analyze the performance of different classical statistical techniques is proposed for theft detection using smart metering data. The disadvantage of these studies is the need to make strong assumptions about the ways fraud is carried out, only providing estimates on the detection capabilities of techniques under those assumptions. The advantage is that studies provide precise detection capacity estimates under the considered assumptions.

- *Other approaches*: Study based on other less common techniques in the field. Such as [67], where an algorithm to optimally schedule inspections is proposed with the objective of detecting theft.

The performance of classification and estimation techniques is a relevant factor, normally quantified using accuracy or other measures. A comparison between the accuracy of the different solutions is not presented because the studies deal with very different data, coming from different locations, representing different realities and presenting different data types. The data used, in many cases, is synthetic and may not be indicative of the performance in real applications. The only found study where the results presented in multiple studies are compared is [12], reporting on detection rates ranging from 25% to 98% under different conditions and data. No standard validation method is found throughout the analyzed studies.

The validation results of the classification and estimation techniques proposed in the selected studies are always stated as, at least, adequate. A difficulty faced in many of the *non-hardware solution* studies which used real data are the lack of balance in the data, as there are usually many more examples of consumers with no behavior related to NTLs than fraudulent or thieving consumers [43, 45, 57, 68–72].

All selected studies, categorized according to the proposed typology, are listed in Table 5.

Table 5: Categorization of all selected studies

Category Type	Studies
Theoretical study	
Analysis of variables and factors	[8, 28, 29, 31, 32, 73]
Hardware solution	
Metering hardware	[33–36, 74–79]
Metering infrastructure	[26, 37, 38, 80–83]
Signal generation and processing	[7, 39, 40, 84]
Other approaches	[41, 42, 85, 86]
Non-hardware solution	
Classification	[19, 20, 27, 30, 43, 44, 47–61, 68–72, 87–108]
Estimation	[2–4, 13, 62–65, 109–118]
Game theory	[66, 119, 120]
Other approaches	[67]

In Table 5 the majority of studies are in the category of *non-hardware solution*, followed by the types *metering hardware* and *estimation* of the categories *hardware solution* and *non-hardware solution*, respectively. The category of *theoretical study* has the lowest number of studies. The summary of the identified advantages and disadvantages for each type is presented in Table 6.

Table 6: Advantages and disadvantages of the solutions proposed in the studies

Category Type	Advantages	Disadvantages	Examples
Theoretical Study			
Analysis of variables and factors	Generates information valuable for policy and decision making	Unable to detect the specific sources of NTLs	[28, 31, 32]
Hardware solution			
Metering hardware	Can prevent all types of NTLs resulting from the meter	Costs with equipments	[33, 35, 36]
Metering infrastructure	Can prevent all types of NTLs resulting from the electrical network	Costs with equipments	[37, 38]
Signal generation and processing	Can prevent all types of NTLs resulting from the electrical network	Require smart meters	[7, 39, 40]
Non-hardware solution			
Classification	Low cost and use of available resources	Detection is not guaranteed and required data may not be available	[48, 52, 71]
Estimation	State estimation: Low cost and high precision	State estimation: Significant data requirements	[62, 63]
	Technical loss modeling: Low cost	Technical loss modeling: Only estimates aggregated NTLs	[2, 64]
Game theory	Precise estimates of performance	Need to make strong assumptions on fraudulent behavior	[66]

4.2. Techniques

The number of techniques identified in the studies is of 91 and are listed in the Annexes in Table 13 and Table 14. The main specific techniques used in the selected studies that are used in more than two cases are listed in Table 7.

Table 7: Techniques used in at least three selected studies

Technique	#
SVM	16
Load profiling	13
Direct calculation	12
ANN	11
State estimation	8
RBS	7
DT	7
Technical loss modeling	6
FS	4
Optimum-path forest	5
Text mining	3
K-Means	3
Naive Bayes	3

In Table 7 most of the techniques listed are from studies using non-hardware classification solutions, such as SVM, load profiling, artificial neural networks (ANN), rule-based systems (RBS), decision trees (DT), feature selection (FS), optimum-path forest, text-mining, K-Means and naive Bayes. SVM, ANN, RBS, DT and naive Bayes are classification models. These are able to infer a binary indicator or probability of presence of NTLs from a set of inputs. The use of these techniques usually consists on the following phases: 1) processing of data representing past examples of NTLs, 2) fitting of the classification model to the data, 3) evaluation of the performance of the model, and 4) deployment of the model. The inputs can consist in the consumption data of customers, characteristics and other information the utility finds suitable for the task. The following paragraphs analyze and present examples of the most popular techniques used in the proposed classification solutions for the detection of NTLs.

SVM has been one of the leading techniques in classification due to a good performance and ease of adaptation to different applications [71, 93].

In comparison to ANN, SVM are more easily used and tend to result in less overfitting, performing better on data different from the one used for fitting.

The optimum-path forest classifier is a graph-based technique that is less common in the literature and is reported as outperforming SVM and ANN in [51, 52].

DT and RBS present significantly different characteristics than SVM and ANN. The structure of tree and rule-based models is easily interpreted in comparison to the latter, which are close to a black-box approach [30, 58]. These easily understood models have the advantage of being transparent to the operators and being easily adjusted by hand. The disadvantage is a lower performance in comparison with more complex techniques, such as SVM and ANN.

Naive Bayes is a simple probabilistic classifier that presents limitations when dealing with complex data, such has the one usually used for detection of NTLs. This classifier makes strong assumptions on the structure of the data and needs a comprehensive dataset for fitting [27].

FS is normally used together with classification techniques, encompassing the schemes used to find which variables are the most useful to identify the presence of NTLs [48, 54]. The reduction of the number of inputs used in classification can improve the performance and ease interpretation.

Text mining encompasses methods used to transform the data from inspection notes or other text sources in a numerical input format that the classification techniques can use for inference [58, 61].

Load profiling and K-Means are used in various studies categorized as *non-hardware solution*, in some cases as a classification model and in other cases as a method for data pre-processing for use of classification techniques such as SVM. These techniques are used to divide a population of customers or set of electricity patterns in smaller sets that present similar characteristics. The techniques can be used to divide the classification problem in a set of multiple easier problems or to directly identify a group of customers presenting irregular behavior, which may be a strong indicator of NTLs [89]. Finding load profiles representing normal consumption patterns is used compare to new consumption curves, uncovering outliers that can indicate the presence of illicit behavior or faults in metering [48].

State estimation consists on finding the best possible estimate for internal states of the power system using the available measurement data, usually load and current measurements at grid nodes [62, 99]. This type of technique may be used to estimate power flow to customer nodes, enabling the detection of

NTLs when a deviation between estimated and billed electricity consumption is significant. This technique is significantly dependent on the quality of load data and on the presence of automatic collection of measurements from customers households.

Direct calculation is stated when studies propose solutions that need to directly estimate T&D losses, such as calculating the difference between metered electricity flowing out of a distribution feeder and billed energy to the consumers connected [7]. This method is straightforward, but also very dependent on the data available and is not able to pinpoint the source of NTLs. Technical loss modeling is usually used to estimate the amount of technical losses, is used together with direct calculation to infer the amount of NTLs from total T&D losses [64].

4.3. Requirements

The requirements of the solutions presented in the selected studies are analyzed in this section. This analysis mainly focuses on requirements consisting of data, which is essential in the majority of selected studies. For studies presenting an *hardware solution* the requirements are considered to be of material nature, usually presenting unique equipment and hardware needs (e.g. meters with tempering sensors, RFID equipments, meters with redundancy). The manufacturing requirements, infrastructure necessities and financial costs could also be considered in the scope of requirements. Due to the range of possibilities and inability to limit the scope, this section is limited to the categories of *theoretical study* and *non-hardware solution*.

4.3.1. Theoretical study

The number of studies in the category of *theoretical study* per type of data is listed Table 8.

Table 8: Number of studies in *theoretical study* per type of data

Type of data	#
Socio-economic	4
Consumption/load	2
Customer information	2
Load shedding	1
Electricity price	1
Behavior and perceptions	1

Table 8 shows that *theoretical study* is focused on the relation between socio-economic aspects and occurrence of NTLs. For example, in [32] an econometric study of socio-economic factors, consumption and information on consumers and appliances, concludes that: in Brazil low-income urban demography, *favelas*, the illegal behavior is explained not only by low-income, but also by social norms. Data on indicators such as load shedding (related to the amount of interruptions), electricity price, behavior and perceptions is used in one study [31]. Nonetheless, apparent relationships are found between these indicators and the amount of fraud and theft leading to NTLs. The number of studies per combination of types of data are listed in Table 9.

Table 9: Number of studies in *theoretical study* per combination of types of data

Combination of types of data	#
Consumption/load + Customer information + Socio-economic	2
Aggregated non-technical losses + Electricity price + Load shedding	1
Behavior and perceptions + Socio-economic	1
Aggregated non-technical losses + Socio-economic	1

In table 9 the number of studies is lower than the number in the category of *theoretical study*, because [73] focuses on analysis of electricity theft and fraud from a psychology and theory of planned behavior point of view, not using any data. The number of studies is not sufficient to extract additional insights.

4.3.2. Non-hardware solution

The types of data used in the selected studies of the category *non-hardware solution* are listed in Table 10.

Table 10: Number of *non-hardware solution* studies per type of data with at least two uses

Type of data	#
Individual consumption/Load	61
Customer information	13
Load, voltage and current measurements	8
Inspection data	7
Topology	2
Risk information	2
Billing	2
Grid asset information	2
Geo-referenced	2
Demand and load factors	2

Individual consumption or load data is the type of data most commonly used, as many of the solutions presented in the selected studies propose techniques to find consumption patterns assumed to indicate the existence of sources of NTLs.

Customer information, inspection notes, load, voltage and current measurements are also used in a significant number of studies. Load, voltage and current measurements are specially common in studies using state estimation techniques for the estimation of NTLs. The combinations of types of data used in the studies categorized as *non-hardware solution* is listed in Table 11.

Table 11: Number of *non-hardware solution* studies per combination of types of data with at least two uses

Combination of types of data	#
Individual consumption/Load	32
Individual consumption/Load + Customer information	7
Individual consumption/Load + Customer information + Inspection data	5
Grid load, voltage and current measurements	3
Individual consumption/Load + Risk information	2
Individual consumption/Load + Load, voltage and current measurements	2
Grid asset information + Load, voltage and current measurements.	2

Consumption or load data are the only requirement of the majority of solutions. Additional data such as customer information (e.g. type of contract, house type) and inspection notes have been used with the objective of achieving better detection performances. As stated earlier, load, voltage and current measurements are usually used in state estimation based solutions that usually do not include other data types.

As consumption or load data is widely used, its characteristics are further analyzed. The resolution of the data is very relevant for ascertain the adequacy of a solution to the available data and network hardware, as most utilities only have access to monthly billing information on the individual level. The resolution of the consumption/load data used in the studies categorized as *non-hardware solution* is presented in Table 12.

Table 12: Number of studies per resolution of consumption/load data

Resolution	#
High (T not specified)	18
High ($T = 15\ min$)	14
High ($T = 1\ hour$)	2
High ($T = 30\ min$)	1
Monthly	18

In Table 12 T stands for the data collection time period. The resolution is considered *High* if more than one daily measurement is used. The majority of these studies make use of high resolution data, indicating that the availability of such data may be very important to detect the occurrence of NTLs. Nonetheless eighteen studies propose solutions using monthly resolution. Traditionally, customer consumption of electricity has been recorded and billed in a monthly basis, this type of data should be available to all utilities thus making practical the use of solutions that use this data.

4.4. Limitations of available solutions

The analysis of techniques and data requirements of the selected studies unveiled the following main limitations.

***Theoretical study*:**

- Not suitable to identify specific location of the sources o NTLs;
- Less suited to identify the presence of NTLs than to identify the demographic and economic drivers of NTLs.

***Hardware solution*:**

- Unable to detect NTLs resulting from the *billing* zone;
- Solutions consisting in anti-tampering mechanisms are only aimed at the *meter* zone;
- Internet connected meters may result in an widened attack-surface for electricity fraudsters;
- The wide deployment of these solutions requires significant capital expenses.

***Non-hardware solution*:**

- Some solutions assume that NTLs result in a change in consumption information collected from a customer;
- If the solution only analyzes the evolution of the consumption of the customer to infer the presence of NTLs, then is unsuitable to detect sources of NTLs present since the first day of the electrical connection;
- Solutions usually present high levels of false positives in performance evaluation due to the high variability of consumption behaviors;
- Many solutions depend on high resolution consumption data that can be considered a breach of customers privacy;
- Solutions that depend on advanced metering equipment for distributed data collection will have high costs associated if the infrastructure is not already in place.

Throughout the analysis it stands out that no proposed technique is full-proof regarding the detection of all the vulnerability/attack points identified as potential sources of NTLs. The authors believe that research should focus on the development of methods that use multiple solutions in an integrated way. Studies in the category of *Theoretical study* should be used to identify critical locations to prioritize resources. Specific studies in the category of *hardware solution* should be installed to enable the calculation of T&D losses at different locations, log consumption data and detect meter tampering. Studies in the category of *Non-hardware solution* should be used to transform the data of consumers, the data communicated by the meters and T&D losses in actionable information on individuals or zones with high estimates of NTLs and high probabilities of illegal behavior. The only attack/vulnerability points not covered by the aforementioned approach are the cyber attack to the utility information systems and collusion with key employees of the utility. These may only be fought with the use of advanced cyber security measures and good governance, avoiding vulnerabilities which arise from the utility.

5. Conclusions

This review explores the state of the art on detection of NTLs in electricity utilities unveiled by the research reported since 2000 in the following three databases: ScienceDirect, ACM Digital Library and IEEE Xplore. The main focus of the analysis are the solutions proposed, requirements and limitations. A typology to categorize solutions for detection of NTLs is proposed, dividing the solutions in the literature in the categories of *theoretical study*, *hardware solution* and *non-hardware solution*. Studies categorized as: *Theoretical study* deal with the relations between demographics and theft; *Hardware solution* proposes innovative metering equipments and grid structures to detect NTLs; *Non-hardware solution* propose data based methods to identify points of the network with an high probability of being sources of NTLs, or estimate the amount of losses. The selected literature mostly focuses on the category of *non-hardware solution*, 72 of 103 studies are in this category.

The analysis of research unveiled apparent gaps, regarding the category of *theoretical study*, there is a lack of research that analyze the situation in non-developing economies. Developed economies present much lower amounts of NTLs, but the impact in still significant. In the studies in the category of *hardware solution*, there is a lack of analysis on the economic viability of the implementation of the solution proposed, which is important for inferring if the return from reduction of NTLs covers equipment costs. The way these solutions interact with the ones in the category of *non-hardware solution*, in terms of communications, data availability should also be studied more in depth, as these two categories of techniques usually go hand-in-hand. In *non-hardware solution* studies, a standard way to evaluate the techniques is lacking, implying difficulties to compare the solutions proposed in different studies. The lack of public data for this application is very significant regarding this issue, the existence of benchmark examples and standard performance measures should allow for better progress in the field. *Non-hardware solution* techniques are very dependent on data, but there is a lack of analysis on the effect on performance that results from using different types of data, the effect of delays in collection and from different resolutions, this is relevant both to classification and estimation solutions.

In general, authors do not discuss how solutions fit in the structure of utilities. In a horizontally structured electricity industry, the distributors who own grid assets and suppliers who own information on electricity customers are different companies. While there are usually multiple data sharing policies, it would be of interest to understand how the relations between the different involved parties affect the identification of NTLs.

Most studies focus on one type of sources of NTLs, there is a lack of studies that make systematic analysis on the whole range of potential sources. Applications integrated with different solutions that are proposed in current literature to identify NTLs that result from a diversity of sources are not found in the studied literature. Hence, the authors envisage that future research should focus on the development of applications that use multiple solutions in an integrated way.

Acknowledgments

The work of J. L. Viegas was supported by the PhD in Industry Scholarship SFRH/BDE/95414/2013 from FCT and Novabase. S. M. Vieira acknowledges support by Program Investigador FCT (IF/00833/2014) from FCT, co-funded by the European Social Fund (ESF) through the Operational Program Human Potential (POPH). Acknowledgement to FCT, through IDMEC, under LAETA, project UID/EMS/50022/2013.

References

[1] F. B. Lewis, Costly throw-ups': electricity theft and power disruptions, The Electricity Journal 28 (7) (2015) 118–135.

[2] M. E. de Oliveira, A. Padilha-Feltrin, F. J. Candian, Investigation of the relationship between load and loss factors for a Brazilian electric utility, in: 2006 IEEE PES Transmission and Distribution Conference and Exposition: Latin America, 2006.

[3] R. S. Kumar, T. Raghunatha, R. A. Deshpande, Segregation of technical and commercial losses in an 11 kV feeder, in: 7th IEEE GCC Conference and Exhibition, 2013, pp. 76–79.

[4] M. Buevich, A. Jacquiau-Chamski, D. Schnitzer, J. Thacker, T. Escalada, A. Rowe, Short paper : microgrid losses - when the whole is greater than the sum of its parts, in: 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, 2015, pp. 95–98.

[5] P. Antmann, Reducing technical and non-technical losses in the power sector (background paper for the World Bank Group energy sector Strategy), Tech. rep. (2009).

[6] T. B. Smith, Electricity theft: a comparative analysis, Energy Policy 32 (18) (2004) 2067–2076.

[7] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni, Electricity theft: overview, issues, prevention and a smart meter based approach to control theft, Energy Policy 39 (2) (2011) 1007–1015.

[8] T. Winther, Electricity theft as a relational issue: A comparative look at Zanzibar, Tanzania, and the Sunderban Islands, India, Energy for Sustainable Development 16 (1) (2012) 111–119.

[9] IBM, Energy theft: incentives to change, Tech. rep. (2012).

[10] Energy Association of Pennsylvania, Energy theft kills, costs innocent pennsylvanians millions (2007).

[11] Institute of Communication & Computer Systems of the National Technical University of Athen ICCS-NTUA for the European Commission, Study on cost benefit analysis of smart metering systems in EU member states - final report.

[12] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, X. S. Shen, Energy-theft detection issues for advanced metering infrastructure in smart grid, Tsinghua Science and Technology 19 (2) (2014) 105–120.

[13] A. R. Abaide, L. N. Canha, A. Barin, G. Cassel, Assessment of the smart grids applied in reducing the cost of distribution system losses,

7th International Conference on the European Energy Market (EEM 2010) (2010) 1–6.

[14] ETP SmartGrids, European technology platform smart grids: vision and strategy for Europe's electricity networks of the future, Tech. rep. URL http://ec.europa.eu/research/energy/pdf/smartgrids{_}en.pdf

[15] A. Battaglini, J. Lilliestam, A. Haas, A. Patt, Development of SuperSmart grids for a more efficient utilisation of electricity from renewable sources, Journal of Cleaner Production 17 (10) (2009) 911–918.

[16] U.S. Department of Energy, The smart grid: an introduction, Tech. rep. (2008).
URL http://www.oe.energy.gov/SmartGridIntroduction.htm

[17] International Energy Agency, Technology roadmap: smart grids, Tech. rep. (2011).

[18] M. Welsch, M. Howells, M. Bazilian, J. F. DeCarolis, S. Hermann, H. H. Rogner, Modelling elements of smart grids: enhancing the OSeMOSYS (open source energy modelling system) code, Energy 46 (1) (2012) 337–350.

[19] P. Jokar, S. Member, N. Arianpoo, S. Member, V. C. M. Leung, Electricity theft detection in AMI using customers' consumption patterns, IEEE Transactions on Smart Grid 7 (1) (2015) 1–11.

[20] A. H. Nizar, Z. Y. Dong, Identification and detection of electricity customer behaviour irregularities, IEEE PES Power Systems Conference and Exposition (PSCE'09) (2009) 1–10.

[21] B. Kitchenham, S. Charters, Guidelines for performing systematic literature reviews in software engineering, Tech. rep. (2007).

[22] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, S. Linkman, Systematic literature reviews in software engineering - a systematic literature review, Information and Software Technology 51 (1) (2009) 7–15.

[23] G. Rasool, F. Ehsan, M. Shahbaz, A systematic literature review on electricity management systems, Renewable and Sustainable Energy Reviews 49 (2015) 975–989.

[24] J. R. Aguero, Improving the efficiency of power distribution systems through technical and non-technical losses reduction, in: IEEE PES Transmission and Distribution Conference and Exposition, 2012, pp. 1–8.

[25] S. McLaughlin, D. Podkuiko, S. Miadzvezhanka, A. Delozier, P. McDaniel, Multi-vendor penetration testing in the advanced metering infrastructure, in: 26th Annual Computer Security Applications Conference (ACSAC '10), Vol. I, 2010, p. 10.

[26] D. Grochocki, J. H. Huh, R. Berthier, R. Bobba, A. C. Alvaro, W. H. Sanders, AMI threats, intrusion detection requirements and deployment recommendations, in: 2012 IEEE International Conference on Smart Grid Communications (SmartGridComm), 2012, pp. 395–400.

[27] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, S. Zonouz, AMIDS: a multi-sensor energy theft detection framework for advanced metering infrastructures, in: 2012 IEEE International Conference on Smart Grid Communications (SmartGridComm), 2013, pp. 1319–1330.

[28] Ç. Yurtseven, The causes of electricity theft: an econometric analysis of the case of Turkey, Utilities Policy 37 (2015) 70–78.

[29] B. Never, Social norms, trust and control of power theft in Uganda: does bulk metering work for MSEs?, Energy Policy 82 (1) (2015) 197–206.

[30] A. H. Nizar, Z. Y. Dong, J. H. Zhao, P. Zhang, A data mining based NTL analysis method, in: 2007 IEEE Power Engineering Society General Meeting, no. 3, 2007, pp. 1–8.

[31] F. Jamil, On the electricity shortage, price and electricity theft nexus, Energy Policy 54 (2013) 267–272.

[32] L. M. Mimmi, S. Ecer, An econometric study of illegal electricity connections in the urban favelas of Belo Horizonte, Brazil, Energy Policy 38 (9) (2010) 5081–5097.

[33] K. Dineshkumar, R. Prabhu, S. Ramasamy, Development of ARM processor based electricity theft control system using GSM network, in: 2015 International Conference on Circuit, Power and Computing Technologies (ICCPCT), 2015.

[34] S. Ngamchuen, C. Pirak, Smart anti-tampering algorithm design for single phase smart meter applied to AMI systems, in: 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2013, pp. 1–6.

[35] B. Khoo, Y. Cheng, Using RFID for anti-theft in a Chinese electrical supply company: a cost-benefit analysis, in: Wireless Telecommunications Symposium (WTS), 2011.

[36] H. Henriques, A. Barbero, R. Ribeiro, M. Fortes, W. Zanco, O. Xavier, R. Amorim, Development of adapted ammeter for fraud detection in low-voltage installations, Measurement 56 (2014) 1–7.

[37] A. R. Devidas, M. V. Ramesh, Wireless smart grid design for monitoring and optimizing electric transmission in India, in: 4th International Conference on Sensor Technologies and Applications (SENSORCOMM 2010), 2010, pp. 637–640.

[38] P. Kadurek, J. Blom, J. F. G. Cobben, W. L. Kling, Theft detection and smart metering practices and expectations in the Netherlands, IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe) (2010) 1–6.

[39] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni, A conceptual design using harmonics to reduce pilfering of electricity, in: 2010 IEEE PES General Meeting, 2010, pp. 1–7.

[40] A. Pasdar, S. Mirzakuchaki, A solution to remote detecting of illegal electricity usage based on smart metering, IEEE International Workshop on Soft Computing Applications Proceedings (2007) 163–167.

[41] G. M. Soares, A. G. B. Almeida, R. M. Mendes, Detection of street lighting bulbs information to minimize commercial losses, in: 7h International Conference on Sensing Technology (ICST), 2013, pp. 895–900.

[42] R. A. De Faria, K. V. Ono Fonseca, B. Schneider, Sing Kiong Nguang, Collusion and fraud detection on electronic energy meters - a use case of forensics investigation procedures, in: 2014 IEEE Security and Privacy Workshops, 2014, pp. 65–68.

[43] J. Nagi, a.M. Mohammad, K. Yap, S. Tiong, S. Ahmed, Non-Technical Loss analysis for detection of electricity theft using support vector machines, in: 2008 IEEE International Power and Energy Conference, 2008, pp. 907–912.

[44] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, a. M. Mohammad, Detection of abnormalities and electricity theft using genetic support vector machines, in: IEEE Region 10 Annual International Conference (TENCON), 2008, pp. 1–6.

[45] J. Nagi, An intelligent system for detection of non-technical losses in tenaga nasional berhad (TNB) Malaysia low voltage distribution network, Ph.D. thesis (2009).

[46] J. Nagi, K. S. Yap, F. Nagi, S. K. Tiong, S. P. Koh, S. K. Ahmed, NTL detection of electricity theft and abnormalities for large power consumers in TNB malaysia, 2010, pp. 202–206.

[47] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, F. Nagi, Improving SVM-based nontechnical loss detection in power utility using the fuzzy inference system, IEEE Transactions on Power Delivery 26 (2) (2011) 1284–1285.

[48] A. H. Nizar, Z. Y. Dong, J. H. Zhao, Load profiling and data mining techniques in electricity deregulated market, in: 2006 IEEE Power Engineering Society General Meeting, 2006, pp. 1–7.

[49] A. Nizar, J. Zhao, Z. Dong, Customer information system data preprocessing with feature selection techniques for non-technical losses prediction in an electricity market, in: 2006 International Conference on Power System Technology, 2006, pp. 1–7.

[50] A. H. Nizar, Z. Y. Dong, P. Zhang, Detection rules for non technical losses analysis in power utilities, in: IEEE PES General Meeting 2008: Conversion and Delivery of Electrical Energy in the 21st Century, PES, 2008, pp. 1–8.

[51] C. C. O. Ramos, A. N. Souza, J. P. Papa, A. X. Falcão, Fast non-technical losses identification through optimum-path forest, in: IEEE International Conference on Intelligent System Applications to Power Systems, 2009, pp. 1–5.

[52] C. C. O. Ramos, A. N. De Sousa, J. P. Papa, A. X. Falcão, A new approach for nontechnical losses detection based on optimum-path forest, IEEE Transactions on Power Systems 26 (1) (2011) 181–189.

[53] C. C. O. Ramos, A. N. Souza, R. Y. M. Nakamura, J. P. Papa, Electrical consumers data clustering through optimum-path forest, in: 2011 16th International Conference on Intelligent System Applications to Power Systems (ISAP 2011), no. 1, 2011, pp. 1–4.

[54] C. C. Ramos, A. N. Souza, G. Chiachia, A. X. Falcão, J. P. Papa, A novel algorithm for feature selection using harmony search and its application for non-technical losses detection, Computers & Electrical Engineering 37 (6) (2011) 886–894.

[55] E. W. S. Dos Angelos, O. R. Saavedra, O. a. C. Cortés, A. N. De Souza, Detection and identification of abnormalities in customer consumptions in power distribution systems, IEEE Transactions on Power Delivery 26 (4) (2011) 2436–2442.

[56] Í. Monedero, F. Biscarri, C. Léon, J. Biscarri, R. Millán, MIDAS: Detection of non-technical losses in electrical consumption using neural networks and statistical techniques, Lecture Notes in Computer Science 3984 (2006) 725–734.

[57] I. Monedero, F. Biscarri, C. León, J. I. Guerrero, J. Biscarri, R. Millán, Detection of frauds and other non-technical losses in a power utility using Pearson coefficient, Bayesian networks and decision trees, International Journal of Electrical Power & Energy Systems 34 (1) (2012) 90–98.

[58] C. León, F. Biscarri, I. Monedero, J. I. Guerrero, J. Biscarri, R. Millán, Integrated expert system applied to the analysis of non-technical losses in power utilities, Expert Systems with Applications 38 (8) (2011) 10274–10285.

[59] C. León, F. Biscarri, I. Monedero, J. I. Guerrero, J. Biscarri, R. Millán, Variability and trend-based generalized rule induction model to NTL detection in power companies, IEEE Transactions on Power Systems 26 (4) (2011) 1798–1807.

[60] J. I. Guerrero, C. Leon, F. Biscarri, I. Monedero, J. Biscarri, R. Millan, Increasing the efficiency in non-technical losses detection in utility companies, 15th IEEE Mediterranean Electrotechnical Conference (MELOCON 2010) (2010) 136–141.

[61] J. I. Guerrero, C. León, I. Monedero, F. Biscarri, J. Biscarri, Improving knowledge-based systems with statistical techniques, text mining, and neural networks for non-technical loss detection, Knowledge-Based Systems 71 (2014) 376–388.

[62] C. Bandim, J. Alves, J.E.R., J. Pinto, a.V., F. Souza, M. Loureiro, C. Magalhaes, F. Galvez-Durand, Identification of energy theft and tampered meters using a central observer meter: a mathematical approach, 2003 IEEE PES Transmission and Distribution Conference and Exposition 1 (2003) 163–168.

[63] C.-H. Lo, N. Ansari, CONSUMER: a novel hybrid intrusion detection system for distribution networks in smart grid, IEEE Transactions on Emerging Topics in Computing 1 (1) (2013) 33–44.

[64] M. E. de Oliveira, D. F. A. Boson, A. Padilha-Feltrin, A statistical analysis of loss factor to determine the energy losses, in: 2008 IEEE PES Transmission and Distribution Conference and Exposition: Latin America, 2008, pp. 1–6.

[65] L. Faria, J. Melo, A. Padilha-Feltrin, Spatial-temporal estimation for nontechnical losses, IEEE Transactions on Power Delivery 8977 (2015) 1–1.

[66] S. Amin, G. A. Schwartz, Game-theoretic models of electricity theft detection in smart utility networks, IEEE Control Systems Magazine (February 2015).

[67] X. Xia, W. Liang, Y. Xiao, M. Zheng, BCGI: a fast approach to detect malicious meters in neighborhood area smart grid, in: IEEE International Conference on Communications, 2015, pp. 7228–7233.

[68] R. Jiang, H. Tagaris, A. Lachsz, M. Jeffrey, Wavelet based feature extraction and multiple classifiers for electricity fraud detection, in: IEEE PES Transmission and Distribution Conference and Exhibition 2002: Asia Pacific (Volume:3), 2002.

[69] C. Muniz, K. Figueiredo, M. Vellasco, G. Chavez, M. Pacheco, Irregularity detection on low tension electric installations by neural network ensembles, Proceedings of the International Joint Conference on Neural Networks (2009) 2176–2182.

[70] N. Liu, New features for detection of nontechnical losses considering PV installed at customer side, in: 2012 China International Conference on Electricity Distribution, Vol. 1, 2012, pp. 1–4.

[71] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, M. Mohamad, Non-technical loss detection for metered customers in power utility using support vector machines, IEEE Transactions on Power Delivery 25 (2) (2010) 1162–1171.

[72] M. D. Martino, F. Decia, J. Molinelli, A. Fernández, Improving electric fraud detection using class imbalance strategies, in: International Conference on Pattern Recognition Applications and Methods (ICPRAM 2012), 2012, pp. 135–141.

[73] T. Sharma, K. Pandey, D. Punia, J. Rao, Of pilferers and poachers: combating electricity theft in India, Energy Research & Social Science 11 (2016) 40–52.

[74] A. Pyasi, V. Verma, Improvement in electricity distribution efficiency to mitigate pollution, 2008.

[75] A. A. Chauhan, Non-technical losses in power system and monitoring of electricity theft over low-tension poles, in: 2015 Second International Conference on Advances in Computing and Communication Engineering, 2015, pp. 280–284.

[76] A. R. Devidas, M. V. Ramesh, Power theft detection in microgrids, in: International Conference on Smart Cities and Green ICT Systems (SMARTGREENS), 2015, pp. 342–349.

[77] M. U. Hashmi, J. G. Priolkar, Anti-theft energy metering for smart electrical distribution system, in: 2015 International Conference on Industrial Instrumentation and Control (ICIC 2015), 2015, pp. 1424–1428.

[78] R. H. Raju, Design and fabrication of power consumption network to prevent energy pilferage, in: International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), 2015, pp. 21–23.

[79] L. Fucun, G. Hongxia, L. Lijun, W. Zhelong, W. Peng, Anti-theft plug-in metering device and its method based on interlock-delay, in: 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC), 2015, pp. 651–654.

[80] W. Doorduin, H. Mouton, R. Herman, H. Beukes, Feasibility study of electricity theft detection using mobile remote check meters, in: 2004 IEEE Africon, Vol. 1, 2004, pp. 373–376.

[81] R. F. Ghajar, J. Khalife, Brahim Richani, Design and cost analysis of an automatic meter reading system for Electricité du Liban, Utilities Policy 9 (2000) 193–205.

[82] M. C. Evaldt, J. V. C. Dos Santos, R. M. Figueiredo, L. T. Da Silva, M. R. Stracke, Payback analysis in identification and monitoring of commercial losses in distribution networks, in: 9th International Conference on the European Energy Market (EEM), 2012.

[83] V. Paruchuri, S. Dubey, An approach to determine non-technical energy losses in India, in: 14th International Conference on Advanced Communication Technology (ICACT), 2012, pp. 111–115.

[84] A. V. Christopher, G. Swaminathan, M. Subramanian, P. Thangaraj, Distribution line monitoring system for the detection of power theft using power line communication, in: IEEE Conference on Energy Conversion (CENCON), 2014, pp. 55–60.

[85] J. L. Acevedo Parra, E. A. Sáinchez Calderón, Use of the shunts detecting equipment for the identification of illegal power outlets, in: 2006 IEEE PES Transmission and Distribution Conference and Exposition: Latin America (TDC'06), 2006, pp. 3–6.

[86] G. M. Soares, A. G. B. Almeida, R. M. Mendes, E. C. Teixeira, H. A. C. Braga, S. Member, J. G. P. Filho, Performance evaluation of a sensor-based system devised to minimize commercial losses in street lighting networks, in: IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2014.

[87] J. R. Filho, E. M. Gontijo, A. C. Delaíba, E. Mazina, J. E. Cabral, J. O. P. Pinto, Fraud identification in electricity company costumers using decision tree, in: IEEE International Conference on Systems, Man and Cybernetics, Vol. 4, 2004, pp. 3730–3734.

[88] J. Spirić, A. Janjić, Using of fuzzy logic in the struggle with the anauthorized consumption of the electrical energy, in: Regional Conference and Exhibition on Electricity Distribution - Montenegro, 2004.

[89] J. E. Cabral, J. O. P. Pinto, A. M. A. C. Pinto, Fraud detection system for high and low voltage electricity consumers based on data mining, in: 2009 IEEE Power and Energy Society General Meeting, 2009, pp. 1–5.

[90] A. Brun, J. Pinto, A. Pinto, L. Sauer, E. Colman, Fraud detection in electric energy using differential evolution, in: 15th International Conference on Intelligent System Applications to Power Systems, 2009, pp. 1–5.

[91] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni, P. Nelapati, A hybrid neural network model and encoding technique for enhanced classification of energy consumption data, in: 2011 IEEE Power and Energy Society General Meeting, 2011, pp. 1–8.

[92] Z. Marko, N. Hlupiü, D. Basch, Detection of suspicious patterns of energy consumption using neural network trained by generated samples, in: Proceedings of the 33rd International Conference on Information Technology Interfaces (ITI), 2011, pp. 551–556.

[93] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni, Support Vector Machine Based Data Classification for Detection of Electricity Theft, in: IEEE/PES Power Systems Conference and Exposition (PSCE), 2011, pp. 1–8.

[94] D. Mashima, A. a. Cárdenas, Evaluating electricity theft detectors in smart grid networks, Lecture Notes in Computer Science 7462 (2012) 210–229.

[95] Y.-l. Lo, S.-c. Huang, C.-n. Lu, Non-technical loss detection using smart distribution network measurement data, in: IEEE PES Conference on Innovative Smart Grid Technologies, 2012, pp. 1–5.

[96] S. Salinas, M. Li, P. Li, Privacy-preserving energy theft detection in smart grids, in: IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks workshops, Vol. 1, 2012, pp. 605–613.

[97] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni, Enhanced encoding technique for identifying abnormal energy usage pattern, North American Power Symposium (NAPS 2012).

[98] L. A. M. Pereira, L. C. S. Afonso, J. P. Papa, Z. A. Vale, C. C. O. Ramos, D. S. Gastaldello, A. N. Souza, Multilayer perceptron neural networks training through charged system search and its Application for non-technical losses detection, in: IEEE PES Conference on Innovative Smart Grid Technologies, 2013.

[99] S.-C. Huang, Y.-L. Lo, C.-N. Lu, Non-technical loss detection using state estimation and analysis of variance, IEEE Transactions on Power Systems 28 (3) (2013) 2959–2966.

[100] P. Faria, Z. Vale, P. Antunes, A. Souza, Using baseline methods to identify non-technical losses in the context of smart grids, IEEE PES Conference on Innovative Smart Grid Technologies.

[101] J. V. Spirić, S. S. Stanković, M. B. Dočić, T. D. Popović, Using the rough set theory to detect fraud committed by electricity customers, International Journal of Electrical Power & Energy Systems 62 (2014) 727–734.

[102] Z. Wu, T. Zhao, L. He, X. Shen, Smart grid meter analytics for revenue protection, in: International Conference on Power System Technology (POWERCON), 2014, pp. 20–22.

[103] G. Zhou, W. Zhao, X. Lv, F. Jin, W. Yin, A novel load profiling method for detecting abnormalities of electricity customer, in: IEEE PES General Meeting Conference and Exposition, 2014.

[104] P. Faria, Z. Vale, Analysis of consumption data to detect commercial losses using performance evaluation methods in a smart grid, in: IEEE PES Transmission and Distribution Conference and Exposition, 2014, pp. 1–5.

[105] B. Dangar, S. K. Joshi, Electricity theft detection techniques for metered power consumer in GUVNL, Gujarat, India, in: Clemson University Power Systems Conference (PSC), 2015.

[106] D. Rodrigues, C. Ramos, A. Souza, J. Papa, Black hole algorithm for non-technical losses characterization, in: 2015 IEEE Latin American Symposium on Circuits & Systems (LASCAS), 2015, pp. 2–5.

[107] A. Aravkin, M. Wolf, Analytics for understanding customer behavior in the energy and utility industry, IBM Journal of Research and Development 60 (1) (2016) 1–13.

[108] J. V. Spirić, M. B. Dočić, S. S. Stanković, Fraud detection in registered electricity time series, International Journal of Electrical Power & Energy Systems 71 (2015) 42–50.

[109] R. Cruz, C. C. Quintero, F. Pérez, F. Perez, Detecting non-technical losses in radial distribution system transformation point through the real time state estimation method, in: 2006 IEEE PES Transmission and Distribution Conference and Exposition: Latin America, 2006, pp. 1–5.

[110] M. Gemignani, U. S. P. Brazil, C. Tahan, C. Oliveira, Commercial losses estimation through consumers' behavior analysis, CIRED 2009 - 20th International Conference and Exhibition on Electricity Distribution (2009) 8–11.

[111] L. Chen, X. Xu, C. Wang, Research on anti-electricity stealing method based on state estimation, in: IEEE Power Engineering and Automation Conference (PEAM), 2011, pp. 413–416.

[112] S. Weckx, C. Gonzalez, J. Tant, T. D. Rybel, J. Driesen, Parameter identification of unknown radial grids for theft detection, 3rd IEEE PES Conference on Innovative Smart Grid Technologies Europe (2012) 1–6.

[113] A. J. Berrisford, A tale of two transformers: an algorithm for estimating distribution secondary electric parameters using smart meter data, in: Canadian Conference on Electrical and Computer Engineering, 2013.

[114] S. Kaykahie, S. Kowsari Mohaved, A new approach for calculating load and loss factor base on consumer data with fuzzy modelling, in: 22nd International Conference on Electricity Distribution Stockholm, 2013, pp. 10–13.

[115] W. Han, Y. Xiao, NFD: a practical scheme to detect non-technical loss fraud in smart grid, in: 2014 IEEE International Conference Communications (ICC), 2014, pp. 605–609.

[116] S. Sahoo, D. Nikovski, T. Muso, K. Tsuru, Electricity theft detection using smart meter data, in: IEEE PES Conference on Innovative Smart Grid Technologies Conference (ISGT), 2015.

[117] J. A. Porras, H. O. Rivera, F. D. Giraldo, B. S. A. Correa, Identification of non-technical electricity losses in power distribution systems by applying techniques of information analysis and visualization, IEEE Latin America Transactions 13 (3) (2015) 659–664.

[118] S. Salinas, C. Luo, W. Liao, P. Li, State estimation for energy theft detection in microgrids, in: 9th International Conference on Communications and Networking in China, 2014, pp. 96–101.

[119] C. H. Lin, S. J. Chen, C. L. Kuo, J. L. Chen, Non-cooperative game model applied to an advanced metering infrastructure for non-technical loss screening in micro-distribution systems, IEEE Transactions on Smart Grid 5 (5) (2014) 2468–2469.

[120] A. a. Cardenas, S. Amin, G. Schwartz, R. Dong, S. Sastry, A game theory model for electricity theft detection and privacy-aware control in AMI systems, in: 2012 Allerton Conference on Communication, Control, and Computing, 2012, pp. 1830–1837.

Annexes

Data sources and queries

The scientific research databases chosen for the collection of publications are the following:

- *ScienceDirect*;
- *ACM Digital Library*;
- *IEEE Xplore*.

The following three queries are used to obtain relevant publications:

Q1 - *ScienceDirect*:

```
PUB-DATE > 1999
AND
TITLE-ABSTR-KEY (electric OR electricity)
AND
TITLE-ABSTR-KEY (theft OR fraud OR ("non-technical loss") OR
("non-technical losses"))
```

Q2 - *ACM Digital Library*:

```
(acmdlTitle:(electricity electric) AND
acmdlTitle:(theft fraud "non-technical losses"
"non-technical loss"))
OR
(recordAbstract:(electricity electric) AND
recordAbstract:(theft fraud "non-technical losses"
"non-technical loss"))
```

Q3 - *IEEE Xplore:*

```
((("Document Title": "electricity") OR
("Document Title": "electric")) AND
(("Document Title": "theft") OR
("Document Title": "fraud") OR
("Document Title": "non-technical loss") OR
("Document Title": "non-technical losses")))
```

```
OR
((("Abstract": "electricity") OR
("Abstract": "electric")) AND
(("Abstract": "theft") OR
("Abstract": "fraud") OR
("Abstract": "non-technical loss") OR
("Abstract": "non-technical losses")))
```

Table 13: List of all techniques used (part 1 of 2). Techniques are listed and count of selected studies related is in parentheses

Techniques used for detection of NTLS (part 1)			
SVM (Support Vector Machines) (15)	Load profiling (13)	Direct calculation (12)	ANN (Artificial Neural Networks) (11)
State estimation (8)	RBS (Rule Based System) (7)	DT (Decision Trees) (6)	Technical loss modeling (6)
FS (Feature Selection) (4)	Optimum-path forest (4)	Text mining (3)	K-Means (3)
Clustering (3)	Naive Bayes (3)	Predicted Baseline Load (2)	Fuzzy modeling (2)
Anti-tempering system (2)	Econometric analysis (2)	LP (Linear Programming) (2)	Metering hardware (2)
Spectrum signature detection (2)	Fuzzy C-Means (2)	Statistical analysis (2)	ELM (Extreme Learning Machines) (2)
Harmonic generator (2)	Game theory (2)	Average detector (2)	Bayesian Networks (2)
Shewhart (1)	Variability analysis (1)	Forensic investigation procedures (1)	Grid architecture (1)
Genetic Algorithm (1)	Grid identification (1)	Spatial point pattern analysis (1)	Grouping-based inspection (1)
ARMA (1)	CUSUM (1)	Cumulative attestation kernel (1)	Hidden Markov Model (1)
Sensor placement optimization (1)	High performance computing (1)	Signal processing (1)	Honesty coefficient (1)

Table 14: List of all techniques used (part 2 of 2). Techniques are listed and count of selected studies related is in parentheses

Techniques used for detection of NTLS (part 2)			
Specification-based network intrusion detection (1)	Impedance estimation (1)	Statistical process control (1)	Johansen method (1)
Fuzzy logic (1)	Differential Evolution (1)	XMR charts (1)	ANOVA (1)
Attack model (1)	Markov chain (1)	Rough sets (1)	Membership function (1)
Series approximation (1)	Analysis (1)	Shunt sensor (1)	Distance (1)
Socio-technical analysis (1)	SOM (Selft-Organizing Maps) (1)	Spatio-temporal estimation (1)	Adversarial classification (1)
Error correction model (1)	Distributed LU decomposition (1)	Fractional-order Sprott system (1)	Binary Quest tree (1)
Statistical tests (1)	Correlation and distance (1)	Temperature normalization (1)	Embedded Sensing Infrastructure (1)
Theory of planned behavior (1)	Ensemble of models (1)	Wavelet analysis (1)	Privacy preserving (1)
Firefly algorithm (1)	Psychology Theory (1)	DSE (Distributed State Estimation) (1)	Regression (1)
Generalized rule induction model (GRI) (1)	Relational analysis (1)	RFID (1)	Particle Swarm Optimization (1)
Black hole algorithm (1)	Pearson correlation (1)	kNN (K-Nearest Neighbors) (1)	Power-flow model (1)
Charged System Search (1)	Power-flow study (1)	Outlier analysis (1)	Grand Total (189)