Electrical Power and Energy Systems 101 (2018) 189–203

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

An anomaly detection framework for identifying energy theft and defective meters in smart grids

Sook-Chin Yip[a,b,*], Wooi-Nee Tan[b], ChiaKwang Tan[a], Ming-Tao Gan[b], KokSheik Wong[c]

[a] UM Power Energy Dedicated Advanced Center (UMPEDAC), University of Malaya, Malaysia
[b] Faculty of Engineering, Multimedia University, Malaysia
[c] School of Information Technology, Monash University Malaysia, Malaysia

ARTICLE INFO

Keywords:
Anomaly detection
Non-technical losses
Technical losses
Smart grids
AMI
Linear programming

ABSTRACT

Smart meters are progressively deployed to replace its antiquated predecessor to measure and monitor consumers' consumption in smart grids. Although smart meters are equipped with encrypted communication and tamper-detection features, they are likely to be exposed to multiple cyber attacks. These meters may be easily compromised to falsify meter readings, which increases the chances and diversifies the types of energy theft. To thwart energy fraud from smart meters, utility providers are identifying anomalous consumption patterns reported to operation centers by leveraging on consumers' consumption data collected from advanced metering infrastructure. In this paper, we put forward a new anomaly detection framework to evaluate consumers' energy utilization behavior for identifying the localities of potential energy frauds and faulty meters. Metrics known as the *loss factor* and *error term* are introduced to estimate the amount of technical losses and capture the measurement noise, respectively in the distribution lines and transformers. The anomaly detection framework is then enhanced to detect consumers' malfeasance and faulty meters even when there are intermittent cheating and faulty equipment, improving its robustness. Results from both simulations and test rig show that the proposed framework can successfully locate fraudulent consumers and discover faulty smart meters.

1. Introduction

In recent years, the antiquated power grid infrastructure that delivers power to consumers is progressively replaced by a series of digital systems, which is referred to as the smart grids (SGs). This modernized grid greatly assists consumers and utility providers (UPs) to monitor, control and forecast the energy consumption.

Although deploying the SGs has significant technical and social advantages, several security issues arise [1,2]. Specifically, the smart meters (SMs) endowed in the advanced metering infrastructure (AMI) in SGs will change the nature of energy fraud. Instead of making dangerous physical system manipulations, the adversaries could launch more sophisticated cyber attacks [3], which would be detrimental to the power grid (e.g., damage to the power delivery infrastructure, interruption of energy services, etc.) because the SMs are vulnerable to network-borne attacks. According to BBC News, SMs in Spain have been compromised to under-report the energy consumption for reducing the bills [4]. Recent study by Northeast Group also stated that UPs suffer from uncontrolled non-technical losses (NTLs) due to energy theft, which amount to $96 billions globally per annum [5]. In general, the existing NTLs detection schemes are vulnerable to contamination attacks/non-malicious factors and require large sample size for detection analysis, thereby limiting the detection rate [6]. Therefore, an anomaly detection framework that can efficiently detect energy theft attacks against AMI has become significantly imperative for reducing costs and revenue losses incurred due to NTLs.

In this paper, we present two new anomaly detection schemes that adopts *linear programming* (LP) to overcome some of the problems associated with existing NTLs detection schemes. The main contributions to this paper are as follows.

Abbreviations: SGs, smart grids; UPs, utility providers; SMs, smart meters; AMI, advanced metering infrastructure; NTLs, non-technical losses; LP, linear programming; TLs, technical losses; SVM, support vector machine; ELM, Extreme Learning Machine; DT, decision tree; ICT, information and communication technology; AMIDS, AMI intrusion detection system; NILM, non-intrusive load monitoring; LUD, Lower-Upper Decomposition; MDMS, Meter Data Management System; NAN, Neighborhood Area Network; WAN, Wide Area Network; DS, distribution substation; kWh, kilowatt-hours; LSE, linear system of equations; ADF, anomaly detection framework; MLR, multiple linear regression; DER, distributed energy resource; LV, low voltage

* Corresponding author at: UM Power Energy Dedicated Advanced Center (UMPEDAC), Level 4, Wisma R&D University of Malaya, Jalan Pantai Baharu, 59990 Kuala Lumpur, Malaysia.

E-mail addresses: scyip@mmu.edu.my (S.-C. Yip), wntan@mmu.edu.my (W.-N. Tan), cktan@um.edu.my (C. Tan), mtgan@mmu.edu.my (M.-T. Gan), wong.koksheik@monash.edu (K. Wong).

https://doi.org/10.1016/j.ijepes.2018.03.025
Received 26 July 2017; Received in revised form 12 February 2018; Accepted 17 March 2018

Nomenclature

γ	inaccurate meter readings due to faulty smart meters [kW h]
λ	technical losses [kW h]
θ	inaccurate meter readings due to energy theft [kW h]
a_n	anomaly coefficient of consumer n
$a_{t_i,n}$	anomaly coefficient of consumer n at time interval t_i
c_{t_i}	aggregated power supplied by UPs to all the consumers at time interval t_i [kW h]
$c_{t_i}^d$	aggregated power supplied by UPs to all the consumers at time interval t_i on day d [kW h]
d	day
E_{t_i}	error term at time interval t_i [kW h]
$E_{t_i}^d$	error term at time interval t_i on day d [kW h]
g_{t_i}	total energy generated by all the consumers t_i [kW h]
l_{t_i}	loss factor at time interval t_i
$l_{t_i}^d$	loss factor at time interval t_i on day d
$p_{t_i,n}$	energy consumption recorded by the n-th smart meter at time interval t_i [kW h]
$p_{t_i,n}^d$	energy consumption recorded by the n-th smart meter at time interval t_i on day d [kW h]
t_i	time interval [slot]
y_{t_i}	meter discrepancy at time interval t_i [kW h]
$y_{t_i}^d$	meter discrepancy at time interval t_i on day d [kW h]

1. Design two new anomaly detection schemes for detecting energy theft attacks against AMI and locating metering defects in smart grid environment regardless of whether they occur *all the time* or *at varying rates during intermittent periods in a day*;
2. Improve NTLs detection accuracy and reduce false positives by taking the impact of technical losses (TLs) and measurement noise on the detection framework into consideration. Two metrics, referred to as *loss factor* and *error term*, are introduced for capturing the percentage of TLs and amount of measurement noise, respectively, in the service area, and;
3. Investigate and generate a diverse set of NTLs attack functions such that they closely resemble real-world AMI energy frauds/metering defects scenarios.

Our proposed anomaly detection framework realizes a faster, greater flexibility and improved practicality of energy theft/meter irregularities detection based on a small volume of consumers' power consumption data samples regardless of the amount of TLs and the types of consumer. The proposed framework can be extended easily to accommodate more consumers for detection of NTLs.

Our goal is to mitigate NTLs by identifying anomalous consumption patterns within the billing reports transmitted to UPs' operation centers by evaluating consumers' anomaly coefficients. The dataset obtained from a real SG deployment in Ireland is utilized to model normal consumption patterns. As energy theft samples in SGs are non-existent, the evaluation of the proposed detection framework will be performed by simulating energy theft attacks, where consumers' report in the SMs dataset are modified. Besides, an AMI test rig is also constructed in the laboratory to validate the performance and reliability of our proposed anomaly detection framework. Any non-zero anomaly coefficient, which indicates anomaly, may be detected by using the proposed LP models.

The rest of this paper is organized as follows. The related work is discussed in Section 2. Section 3 presents the topological representation of AMI and electrical network in SGs. The background of attack model is described in Section 4. Section 5 discusses the mathematical models for solving the constant and varying anomaly coefficients to detect consumers' malfeasance and defective SMs in SGs. Simulation and experimentation results are discussed in Section 6. Finally, Section 7 concludes this paper.

2. Related work

Energy theft has been a daunting problem for UPs since the early days of energy billing. A variety of energy theft detection schemes have been proposed to reduce the losses incurred due to NTLs. In this work, NTLs detection schemes are classified into two categories: for *conventional power grids* and *smart grids*.

2.1. NTLs detection schemes in conventional power grids

The detection of energy frauds in conventional power grids has been initially addressed with statistical techniques [7]. Subsequently, some of the UPs adopted more advanced techniques which are mostly based on artificial intelligence to solve the issue. One of the popular approaches was to apply support vector machine (SVM) to energy consumption profiles [8,9]. In [8], Nagi et al. proposed a data mining method together with SVM classifier to detect abnormal consumption behaviors using two-year historical consumption data. The long-term trend in energy consumption and computed average daily consumptions of consumers were used to detect malicious consumers. Meanwhile, Nizar et al. [9] investigated the efficiency of SVM technique, Extreme Learning Machine (ELM) and online sequential ELM variant to identify the anomalous consumption trend, which indicates energy fraud based on consumers' load-profile assessments. Several works reported applications of decision tree (DT), fuzzy set and rough set [10–12] to detect NTLs. Nizar et al. [10] adopted Naïve Bayesian and DT to determine the type of data which provides maximum accuracy with reference to NTLs analysis in the electricity distribution sector. In [11], Angelos et al. proposed a fuzzy computational approach for the classification of energy consumption profiles based on C-means-based fuzzy clustering and fuzzy classification. Spiric' et al. [12] utilized the rough set theory to identify electricity fraud committed by energy thieves. Based on the amount of uninvoiced/lost electricity due to fraud, they formed a list of suspicious consumers. However, some of these detection methods are vulnerable to contamination attacks. Specifically, an energy thief can deceive the learning machine to accept a malicious trend as a normal one through granular changes in data and pollution on the dataset. Besides, SVM-based detection approaches typically require long-term monitoring and measurements before theft detection can be performed accurately. The large sample size requirement naturally results in longer detection delay [6].

2.2. NTLs detection schemes in smart grids

With a rapid expansion of information and communication technology (ICT) and growing concerns of cyber attacks, UPs are propelled to invest in AMI [2]. In fact, SMs have been designed to replace conventional analog meters in SGs. SMs eliminate some attack techniques that were common with conventional meters (i.e., meters being disconnected, tilted, reversed, etc.) with alarms [13]. Besides these meter alarms, UPs can identify NTLs with higher accuracy by leveraging the fine granularity of AMI data to correlate events over time and across their entire consumer base with additional information [14]. An AMI intrusion detection system (AMIDS) was presented in [15]. AMIDS collects information of malicious behaviors from three types of information source, namely on-meter anti-tampering sensors, cyber-side network and host-based intrusion detection systems, as well as power measurement-based anomalous consumption detectors, through non-