Electrical Power and Energy Systems 91 (2017) 230–240

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

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

Detection of energy theft and defective smart meters in smart grids using linear regression

Sook-Chin Yip [a,b,*], KokSheik Wong [c], Wooi-Ping Hew [a], Ming-Tao Gan [b], Raphael C.-W. Phan [b], Su-Wei Tan [b]

[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

Article history:
Received 13 August 2016
Received in revised form 21 November 2016
Accepted 13 April 2017

Keywords:
Energy theft detection
Defective meter detection
Smart Grid
Linear regression
Categorical variable

ABSTRACT

The utility providers are estimated to lose billions of dollars annually due to energy theft. Although the implementation of smart grids offers technical and social advantages, the smart meters deployed in smart grids are susceptible to more attacks and network intrusions by energy thieves as compared to conventional mechanical meters. To mitigate non-technical losses due to electricity thefts and inaccurate smart meters readings, utility providers are leveraging on the energy consumption data collected from the advanced metering infrastructure implemented in smart grids to identify possible defective smart meters and abnormal consumers' consumption patterns. In this paper, we design two linear regression-based algorithms to study consumers' energy utilization behavior and evaluate their *anomaly coefficients* so as to combat energy theft caused by meter tampering and detect defective smart meters. Categorical variables and *detection coefficients* are also introduced in the model to identify the periods and locations of energy frauds as well as faulty smart meters. Simulations are conducted and the results show that the proposed algorithms can successfully detect all the fraudulent consumers and discover faulty smart meters in a neighborhood area network.

1. Introduction

Energy theft, which is also referred to as non-technical loss (NTL) has been a daunting problem for all utility providers (UPs) in the conventional power grid system. NTLs are generally related to energy theft and consumers fraudulent behavior in which there exist a number of methods to deliberately defraud the UPs [1]. NTLs may introduce a series of additional losses, such as reduction in grid reliability and damage to the grid infrastructure. NTLs include meter tampering, meter bypassing, meter switching, tapping on secondary voltages, error in computation of technical losses, defective meters, errors and delay in meter reading and billing, unpaid billing, etc. [2–4]. The latest estimates indicate that UPs suffer from losses up to six billion dollars annually due to energy fraud in the United State alone [5]. In recent years, Smart Grid (SG) is being globally introduced to replace its antiquated predecessor to address some of these issues. One significant feature of SG infrastructure is the replacement of the conventional mechanical meters by smart meters (SMs) in Advanced Metering Infrastructure (AMI).

The introduction of SGs and SMs may contribute to a significant cutback in NTLs by minimizing some types of losses [2,6]. However, the SG, AMI in particular, raises new security risks [5,7–11]. Specifically, AMI can be exploited by the adversaries to perform a number of attacks for manipulating the energy utilization statistics because SMs are vulnerable to more types of attack such as network-borne attacks. In addition, consumers' consumption data may be compromised at three different stages, namely during transmission to UP, while it is being recorded, or after it is stored [12]. Since the conventional methods for mitigating NTLs impose high operational costs (e.g., on-site inspection where extensive deployment of human resources is involved [13,14]), this paper

Abbreviations: NTL, non-technical loss; UP, utility provider; SG, Smart Grid; SM, smart meter; AMI, Advanced Metering Infrastructure; NAN, neighborhood area network; MLR, multiple linear regression; RFID, radio frequency identification; SVM, support vector machine; GA, genetic algorithm; LUD, LU decomposition; DS, distribution station; TL, technical loss; LSE, linear system of equations; TOU, Time-of-Use.

* Corresponding author at: UM Power Energy Dedicated Advanced Centre (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), wong.koksheik@monash.edu (K. Wong), wphew@um.edu.my (W.-P. Hew), mtgan@mmu.edu.my (M.-T. Gan), raphael@mmu.edu.my (R.C.-W. Phan), swtan@mmu.edu.my (S.-W. Tan).

http://dx.doi.org/10.1016/j.ijepes.2017.04.005
0142-0615/

aims to reduce the operational costs of UPs by detecting NTL activities.

In this paper, we propose two linear regression-based algorithms to identify the locations of defective SMs and malicious SMs which are compromised by energy thieves to falsify readings (i.e., data attacks [15]) in the neighborhood area network (NAN). The key idea is to adopt multiple linear regression (MLR) for estimating and evaluating consumers' *anomaly coefficients* based on the reported consumers' energy consumption data. MLR is chosen because it adopts characteristic analysis, which attempts to model the consumers' energy consumption behavior for consideration [16]. Therefore, any anomalies not following the utilization trend may be indicative of energy thefts or metering defects. MLR analysis is especially attractive as it is able to accurately reveal not only the locations of energy thieves and defective SMs, but also the amount of energy theft/loss.

2. Related work

Broadly, energy theft detection techniques, including those that are widely implemented in both conventional power grids and SGs, may be grouped into two categories, namely *state-based detection* and *classification-based detection*.

2.1. State-based detection

This method utilizes monitoring state through mutual inspection [12], wireless sensor networks [15], control units [17], radio frequency identification (RFID) [18] and distribution transformers [19] to identify fraud in power system.

As detailed in [12], Xiao et al. proposed three inspection algorithms to identify malicious SMs in a neighborhood. First, they developed a basic scanning method. Then, they designed a binary tree-based method for inspection when the *malicious SMs* to *honest users* ratio is high, and finally employed an adaptive tree-based method to leverage on the advantages of both the scanning and binary tree algorithms. However, adding an extra meter for each consumer/provider will significantly increase the cost. Meanwhile, the authors in [15] designed an AMI Intrusion Detection System (AMIDS), which utilizes information fusion to combine the consumption and sensors data from a SM to model and identify fraud-related behavior more accurately. In [17], consumers consumption data is compared with the feeder input level. Both individual and aggregated consumption are also compared against the feeder details to detect consumption anomalies. However, their proposal can only detect a small region of electricity theft but not the exact location of fraud. Khoo and Cheng [18] proposed a system that incorporated RFID technology to assist the UPs in ammeter inventory management and mitigate energy theft. Although RFID technology can be implemented to identify electricity theft, UPs have to pay extra cost to install the system. In [19], the author adopted the measure of overall fit of the estimated values to the pseudo feeder bus injection measurements based on consumers' aggregated meter data at the distribution transformers to localize the energy consumption abnormalities. They utilized an analysis of variance to create a list of suspected consumers and estimate the actual consumption based on the state estimation results.

2.2. Classification-based detection

The key idea of this approach is to identify consumers' energy consumption anomalies based on testing datasets consisting of the normal and attack class samples using machine learning [20].

Han et al. [2] designed a NTL fraud detection scheme by using the approximated difference between the actual consumed electricity and billing electricity. On the other hand, Nizar et al. designed a feature selection-based approach to extract features from consumers' behaviors for further analysis [21] to find optimal subsets of features in establishing the load profiles, which describe consumers' energy consumption patterns over a period of time. An attacker model for anomaly detector in meter data management is developed by Mashima and C'ardenas to detect energy theft [22]. In [23], Nagi et al. studied consumers' behaviors and proposed a Remote Meter Abnormality Detection System to detect illegal and abnormal energy consumption trends using meter event logs and remote meter reading. In a different work [24], they proposed a fraud detection framework using Support Vector Machine (SVM). Their proposal chose some suspicious consumers in advance for on-site inspection for fraud based on the abnormal power consumption behavior. SVM is trained to extract features and generate fraud detection model. They also designed a hybrid method for NTL analysis by incorporating Genetic Algorithm (GA) and SVM [25]. Similar to [24], the algorithm selected suspicious consumers for inspection. Then, GA provides an increased convergence and optimized SVM hyper-parameters. Meanwhile, Depuru et al. [26] introduced high performance computing to speed up the energy theft detection through data encoding without compromising the quality of data. The encoded data are then classified to discover the electricity pilfering using SVM and Rule Engine-based algorithms. The authors in [27] shortlisted area with high probability of theft using distribution transformers. Then, their proposal identified the suspicious consumers by observing irregularities of consumption patterns using SVM. The SVM-based energy theft detection schemes [24–27] usually require a large volume of training data with load profiles collected from SMs to extract features from historical data.

Besides, it is crucial to preserve consumers' privacy while detecting energy theft in SGs as detailed in [28,29]. In their paper, Salinas et al. proposed a LU decomposition-based (LUD) algorithm to solve a linear system of equations for consumers' honesty coefficients while ensuring consumers' privacy. However, their proposal is restricted by the dimension of the consumers' energy consumption data (i.e., the data matrix must be a square matrix) due to the characteristic of LUD. In order to meet the dimension requirements, the authors have to change the time granularity. Nevertheless, it might not be practical to *reduce the sampling period or time granularity* indefinitely due to the memory size of SM.

To address some of the limitations of previous work, linear regression-based schemes for identifying energy thefts and defective SMs which are not restricted by the dimension of consumers' power consumption data as well as its time granularity are proposed in this work.

3. Architecture of smart grid in neighborhood area network

Here, we present the electrical and communication network architectures considered in this paper. In AMI, the electrical and communication networks overlay each other and all electrical and communication flows are bidirectional [30]. According to the surveys of SG [9,11], the architecture of SG in a neighborhood area network (NAN) can be illustrated in Fig. 1. Further details on *Electrical network* and *Communication network* will be provided below.

3.1. Electrical network

Similar to the conventional electrical grid system, the power supply of SG in a NAN is usually serviced by the same UP. The UP builds a distribution station (DS), which is also known as fuse