Electrical Power and Energy Systems 95 (2018) 635–643

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

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

Identification of suspicious electricity customers

Josif V. Spirić[a,*], Slobodan S. Stanković[b], Miroslav B. Dočić[b]

[a] I. Strele 6/27, 16000 Leskovac, Serbia
[b] Electric Power Industry of Serbia – Section Leskovac, Stojana Ljubića 16, 16000 Leskovac, Serbia

ARTICLE INFO

Keywords:
Total losses
Rules
Suspicious customer
Energy balance

ABSTRACT

In this paper, customer's invoiced energy time series in the recent past is analyzed in order to detect the cause of increased total electricity losses in the low voltage network. Period from 2003 to 2017 is included in the recent past. This period is organized in five successive decades whose beginnings are shifted for one year. Fuzzy logic is used as a method for determining a set of suspicious electricity customers in the first phase of electricity fraud detecting. At this phase, every customer located in the area of increased total electricity losses is analyzed. For each customer, a time series of invoiced energy are formed. Selected time series data and their relations are used to create fuzzy sets of suspicion. Then, total suspicion value of each customer is determined by using fuzzy logic. Based on the estimated total and technical energy losses in the customer's area (region that is supplied by one or more MV/LV transformer stations) and the balance of total, invoiced and energy of losses, a boundary value of suspicion percentage is determined. All customers, whose percentage suspicion value is greater than the boundary value, are declared suspect. Thus, suspicious customers with their locations which need inspection are obtained. On-site inspection of suspected customers is not performed and is not the subject of this paper.

1. Introduction

Fuzzy logic is used in many processes as a tool for decision-making [2]. Fuzzy sets and fuzzy logic can be applied to identify cases of unnatural behavior of electricity customers [3–8].

This paper deals with the electricity fraud of registered customers. The customers in the 'household' category whose home installations are connected to the low voltage distribution network are considered. Electricity fraud is a major source of non-technical energy losses and a serious problem of power distribution system functioning. To address this problem, several methods have been developed [9–14]. This problem is particularly acute in developing countries, countries in transition and generally in countries with low national income per capita. In addition, on-site inspections of the customers are typically poorly prepared and unsophisticated, and therefore have had limited success in fraud detection. To verify the presence of fraud, real inspection is done by synchronous control of customer's individual electrical devices in the home and measuring devices checking at the connection point. This inspection includes the possibility of entering into the customer's property by the controller. The aim of this paper is to develop an approach that will minimize the need for physical inspections which are currently hampered by various legislative hurdles. This problem is especially acute at the overhead connections of customer's home installation.

The initial data for fuzzy set formation of suspicion are customer's invoiced energy time series. Fuzzy set theory is applied in statistical process control so that most typical limits for monitoring of observed parameters *UCL* (upper control limit), *CL* (center of the process) and *LCL* (lower control limit) cease to be determined lines and move to the fuzzy sets [15–18]. In this way, it comes to fuzzy boundaries and improves the ability of decision-making.

In this paper, monthly, six-monthly and annual data of customer's invoiced energy time series are used directly for fuzzy set formation, which are the basis for fuzzy logical conclusion. Due to the difficulties in conducting a proper and complete on-site inspection, it is necessary to distinguish the customer who is caught stealing and the customer who is suspected of stealing. This paper deals with detection of suspicious customers. To create a list of suspects, it is possible to use the fuzzy set theory.

In the first phase, customers are analyzed individually and the appropriate suspicion level is attributed to each of them. In the second phase, based on the level of total and technical losses in the distribution network and using the energy balance [12,13], a set of suspicious customers who participate in creation of non-technical energy losses is created.

Section 2 provides a literature review of non-technical losses and its chalenges. In Section 3, the basic features of time series with respect to electricity fraud detection are listed. Using fuzzy logic to determine

* Corresponding author.
E-mail address: josif.vspiric@gmail.com (J.V. Spirić).

http://dx.doi.org/10.1016/j.ijepes.2017.09.019
Received 12 June 2017; Received in revised form 7 September 2017; Accepted 18 September 2017

suspicion level of fraud for each customer is described in Section 4. In Section 5, suspicion level is defined according to criteria *A* and *B*, which are descriptors of fraud reasons. The membership function's shapes of fuzzy variables *x* and *y* are determined. Subset of suspicious customer's estimation method in accordance with non-technical losses in low voltage distribution network is presented in Section 6. In Section 7, a specific low voltage distribution network with non-technical losses presence is analyzed and obtained results are presented.

2. Related work

In 1965 L. Zadeh, who is considered to be the founder of fuzzy logic, published the first article of fuzzy sets named "Fuzzy Sets" [1]. In [1], Zadeh says that fuzzy logic is basically a precise logic of imprecision and approximate reasoning.

The level of using fuzzy logic and its ability to solve various problems in many areas is large [2].

In [3] a detailed overview of literature on the detection of non-technical losses is given. The authors point out two basic approaches for detection. The first is based on the balance of total, invoiced and lost energy. The second applies to supervised and unsupervised methods based on the construction of load profile of customers.

Part of the fuzzy logic implementation in energy systems is presented in [4]. For the detection of irregularity for low-voltage customers in [5], a combination of artificial neural network and neural-fuzzy system is applied. The unsupervised algorithm of two steps based on the cluster classification strategy is used to determine suspicious customers of electricity [6].

The Support Vector Machine (SVM) as a general qualification method uses customer load profiles and classifies customer profiles into two categories: normal and fraud. There are several different types of fraud, but research concentrates on abrupt change in customer load profiles [7,28].

In [8], it is noted that the SVM models are more successful than fuzzy models. This could be seen by comparing their AUC (The area under the curve) performance with ROC (Receiver operating characteristic) curves in different NTL proportions. For the optimized fuzzy system, the AUC system is significantly increased for all NTL proportions, but the optimized SVM is still more successful.

The MIDAS project relies on two methodologies for fraud detection. [9]. The first methodology uses neural networks due to problem conditions and works with Kohonen network structure, and second methodology is based on the detection of tolerant values outside the range (outliers).

In [10], selection of representative set of sample patterns is performed first, which is represented by a sample matrix. The following procedure leads to a pattern's frequency in the considered group of customers.

In [11] supervised anomalies detection process is presented with classifiers that are the result of an optimum-path forest (OPF) computation in the feature space induced by a graph. "These kinds of classifiers interpret the classification task as a combinatorial OPF computation from some key samples (prototypes) to the remaining nodes. Each prototype becomes a rout from its optimum-path tree and each node is classified according to its strongly connected prototype that defines a discrete optimal partition (influence region) of the feature space" [11].

Based on customer fraud results, a class of customers with anomalous series is formed and is characterized by corresponding patterns according to their consumption. After the discretization of conditional attributes it is possible to find customers with consumption patterns that are identical to the class of customers with anomalous series patterns. Such customers belong to a boundary region and also, but not certain, may belong to a group of thieves and form a set of suspicious customers [12].

In case of asymmetrical distribution of monthly energy from the customer's time series, the ACL Tukey's control chart is used. According to the values of non-technical losses and energy balance, control limits are set, and customers who violate uniform control limits are declared as suspicious [13].

In [14] detection of fraud and other non-technical losses in distribution companies is based on the use of Pearson's coefficient, Bayesian networks and decision trees. The key of these methods is identification of a customer pattern with a drastic drop in consumption and subsequent stabilization, but a gradual (significant) drop with subsequent stabilization of further consumption is also taken into consideration.

By combining the display of time series on control cards for the purpose of statistical control of the process with the fuzzy theory, the fuzzy control limits are realized and decision-making (15)–(18) is improved.

3. The specifics of registered energy time series with respect to electricity fraud detection

The roles of supplier and customer in energy usage process are different. A customer directly forms energy usage process and can produce assignable (special) causes of variation [19,20]. There are several reasons for reduction of registered energy usage in the period between two measurements [21]:

(a) Illegal use of a part of the supplied energy (fraud);
(b) Housing use changes in terms of stay duration and/or the number of occupants, as well as space function changes;
(c) Partial substitution of electricity with other energy sources;
(d) Failure of one or more measuring systems of three-phase watt-hour meter;
(e) Transit from one to two or more home installation connections to the distribution network;
(f) Active customer (that certainly uses electricity) meter status does not change during several successive readings, so it is designated as zero consumption (0 kWh) in the billing system;
(g) Irregular customer meter reading or inability to do that.

Fraud reason (a) is assignable cause of variation because the customer is implemented intentionally. The reasons referred in (b), (c), (d), (e) and (g) can create wrong image about presence of assignable cause of variation in customer's consumption process and it is one of the key characteristics of registered energy time series. The above-mentioned reasons are affect the stationary of time series. This influence is increased with the series length, since a longer time period provides a greater possibility appearance of one or more cited reasons (b), (c), (e) and (g), which are non-random, but are not the product of energy fraud intention. The reason (e) can be avoided by strict updating of household connections status in the billing system. Customer is responsible for generation of reasons (d), (f) and (g). Supplier may also be held responsible if the same issues are not quickly resolved by sophisticated monitoring of relevant customer data.

Considering the objective of this paper, electricity fraud may be the most significant non-random factor. The point of this paper is to analyze available energy usage and other data about customer by which supplier detects presence of non-random factors in customer's energy usage process. With such conclusion, observed abnormal (suspect) customer's profile is assigned to the customer. In this study, we used available historical data of monthly invoiced energy. A set of arranged energy data per time is energy time series. Based on the energy time series, a good suspicion indicator for fraud can be reduction of monthly energy usage in the last several years [9–14].

4. Using fuzzy logic to determine suspicion level of fraud for each customer

Fuzzy logic is based on the concept of linguistic variables whose