Electrical Power and Energy Systems 47 (2013) 21–30

Contents lists available at SciVerse ScienceDirect

Electrical Power and Energy Systems

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

High performance computing for detection of electricity theft

Soma Shekara Sreenadh Reddy Depuru, Lingfeng Wang *, Vijay Devabhaktuni, Robert C. Green

EECS Department, University of Toledo, 2801 W. Bancroft St., Toledo, OH 43606-3390, United States

ARTICLE INFO

Article history:
Received 25 June 2012
Received in revised form 7 October 2012
Accepted 9 October 2012
Available online 4 December 2012

Keywords:
Classification
Electricity theft
High performance computing
Parallel programming

ABSTRACT

Transmission and distribution of electricity involve technical as well as Non-Technical Losses (NTLs). Illegal consumption of electricity constitutes a major portion of the NTL at distribution feeder level. Considering the severity and devastating effects of the problem, illegal consumption of electricity has to be detected instantly in real-time. To this end, this paper investigates the possibility and role of High Performance Computing (HPC) algorithms in detection of illegal consumers. This paper designs and implements an encoding procedure to simplify and modify customer energy consumption data for quicker analysis without compromising the quality or uniqueness of the data. This paper parallelizes overall customer classification process. The parallelized algorithms have resulted in appreciable results as displayed in the results section of the paper.

1. Introduction

Losses during generation, transmission, and distribution of electricity are substantial. Distribution of electricity involves both technical losses and NTL. NTL accounts for up to 10–40% of total energy distributed. A major portion of NTL is due to electricity theft (or illegal consumption of electricity). Electricity theft can be defined as using electricity from the utility company without a contract or valid obligation to alter measurement of electricity [1]. Electricity theft is a serious concern for utilities both in developing countries and developed countries. For example, due to electricity theft utilities in India incur losses around $4.5 billion [2] every year. Utilities in USA incur losses about $1–6 billion [3,4]. In Canada, BC Hydro reports $100 million in losses every year [5].

Electricity theft includes tapping energy directly from the distribution feeder, tampering with the energy meter, or using physical methods to evade payment [6,7]. Improper calibration and illegal de-calibration of energy meters during their design [8] can also cause NTL. Several engineered methods are also implemented to manipulate the energy consumption measured by the energy meter. Illegal consumers may use legal electricity from the energy meter for smaller household loads and illegally tapped electricity for heavier loads. On the other hand, billing irregularities are another form of illegal consumption. Fundamentally, methods of electricity consumption are [9]:

- consuming all electricity from grid legally;
- pilfering all electricity from grid illegally; and
- pilfering a portion of household electricity and consuming the rest legally.

In a smart grid environment, smart meters may not be tamperable, but electricity could still be stolen bypassing the smart meter (e.g. tapping energy from the terminals just before they enter the meter). Apart from physical damage to the meter terminals, there may be other techniques used to tamper the smart meter that are yet unknown. An effective way to detect such illegal consumption is by analyzing the energy consumption behavior of all customers on grid. In light of this situation, the authors have previously implemented a procedure that modifies the energy consumption data of customers such that the exclusivity of the data is maintained while simplifying the data. This simplified/encoded data is then classified to detect illegal consumers using both Support Vector Machine (SVM) and Rule Engine based algorithms. This paper reports significant work done in modifying those algorithms such that portions of the algorithms are parallelized to improve the time efficiency. Therefore, parallelized algorithms are implemented to quicken the data loading, data encoding and classification of customers based on their consumption pattern and overall load on the power grid.

In the light of these developments, this paper particularly works on introducing HPC to enhance the performance of previously developed algorithms by speeding up the detection of illegal consumers. The remainder of the paper has been organized as follows: Section 2 completes a brief literature survey, Section 3 analyzes the customer energy consumption patterns; Section 4 illustrates the

* Corresponding author. Tel.: +1 419 530 8154.
E-mail addresses: sdepuru@rockets.utoledo.edu (S.S.S.R. Depuru), lingfeng.wang@utoledo.edu (L. Wang), vijay.devabhaktuni@utoledo.edu (V. Devabhaktuni), robert.c.green@gmail.com (R.C. Green).

0142-0615/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijepes.2012.10.031

data loading and encoding technique; and Section 5 discusses the implementation of parallelized multi-level algorithm, Section 6 discusses the results and Section 7 concludes the paper.

2. Previous work

There has been a substantial and productive research over many years to detect NTL in power grid. In the recent past, several techniques have been proposed and developed for detection of electricity theft. Cabral et al. describe an application that uses rough set, an emergent technique of soft computing for the detection of electricity theft [10]. In another detection technique, total phase current on the feeder line and the total phase current read by all energy meters are collected over a period of time and these two values of the current are compared to estimate the total electricity being lost [11]. Similarly, e-metering systems can be implemented to process data and detect abnormalities in load profiles indicating electricity theft [12].

A smart meter can be used to identify energy consumption of a customer with more details compared to a conventional energy meter. A smart meter based approach for estimation and grouping of illegal consumption has been proposed in [9,13]. Nagi et al. proposed an approach that uses genetic algorithm–support vector machines (GA–SVMs) in detecting electricity theft [14,15]. Nizar et al. investigates the efficiency of the SVM technique, Extreme Learning Machine (ELM) and its online sequential–ELM (OS–ELM) variant, for identification of abnormal load consumption based on a load-profile evaluation [16]. Previous work by the authors proposed a SVM model that identifies illegal consumers and a hybrid evolutionary neural network model that enhances the performance of that SVM model [17,18]. Then, a Rule Engine model was also developed to detect and isolate consumers based on their energy consumption profile. However, in the near future, when these algorithms are operated on data streams from thousands of consumers, millions of data points will have to be analyzed in real-time. Therefore, these algorithms are modified to explore the advantages and performance of HPC.

Over the past decade, several power engineering applications including monitoring, control and optimization applications, power systems analysis and simulations have been performed using HPC to improve their time efficiency. Recently, due to the advent of smart grid, the algorithms used to perform these tasks are becoming more complex and computationally intensive. HPC also aids in real-time analysis, visualization, and intelligent management of the grid [19–21].

3. Analyzing energy consumption patterns

Analyzing customer energy consumption patterns is essential for utility companies to understand the condition of grid and energy consumption behavior of customers. This analysis aids in identifying electricity theft and enhancing power quality. The energy consumption data used in this paper is the same as in [19]. Fig. 1 illustrates the approximate energy consumption of a typical residential customer. Typically, energy consumption will not be uniform over a single day – it varies from time-to-time. Considering these variations, electricity consumption data can be understood as a function of the range of the customer, geographical location, time of the day, season of the year and weather conditions, etc. Here, energy consumption data has been considered to be collected in 15 min intervals, i.e. a total of 96 times a day. Therefore, the number of inputs to SVM model/Rule Engine is 96.

Fig. 2 represents three hypothetical scenarios of energy consumption readings of the same customer for following three cases: Genuine consumption: if the same customer consumes energy required for the entire household legally; Partial illegal consumption: if that customer consumes a portion of required household energy legally and the rest illegally; Complete illegal consumption: if that consumer steals entire portion or most of the household energy illegally.

For implementation and validation of the algorithm, this paper uses energy consumption data of about 20,000 customers. This data follows a specific range of energy consumption over different periods of time for different types of customers. Therefore, energy consumption patterns of these customers are similar to Fig. 1 in terms of the total number of energy readings (96 inputs), but vary in amount/magnitude of energy consumption. This same data has been applied to measure the performance of SVM model and Rule Engine algorithm.

4. Data loading and encoding

Data loading and encoding procedures are parallelized and the implemented procedure is illustrated in this section of the paper. Several parallelizable approaches are also discussed.

4.1. Data loading

Loading the data has been done using one serial and one parallel approach, where energy consumption of multiple customers is loaded at a time. As the methodology presented in this paper requires four data files (training input, training output, testing input, and testing output), the study examines two distinct methods for loading data. In the first, the data is loaded in a typical and sequential fashion. In the second, the data is loaded in a task parallel manner, with each file being loaded simultaneously by separate threads.

4.2. Encoding procedure

To simplify the energy consumption data of customers, the following encoding procedure has been adopted: first, inputs or data points corresponding to a customer (energy consumption reading of that customer's energy meter) with value of zero (zero energy

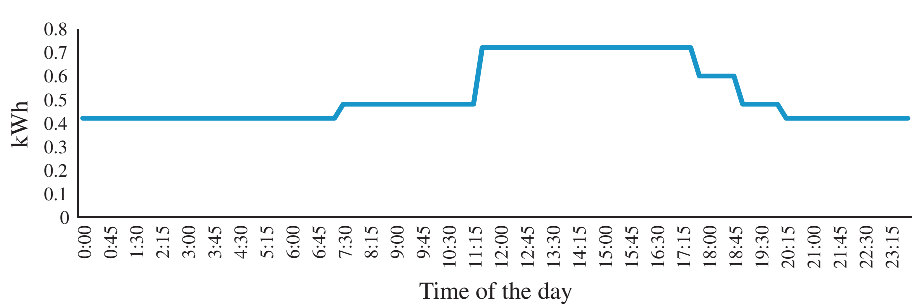

Fig. 1. Approximate energy consumption pattern of a small customer on a weekday.