

A Mobile-based Platform for Big Load Profiles Data Analytics in Non-Advanced Metering Infrastructures

Sherin Moussa ^{1,a}

¹ Faculty of Computer and Information Sciences, Information Systems Department, Ain Shams University 11566 Cairo, Egypt

Abstract. With the rapidly increase of electricity demand around the world due to industrialization and urbanization, this turns the availability of precise knowledge about the consumption patterns of consumers to a valuable asset for electricity providers, given the current competitive electricity market. This would allow them to provide satisfactory services in time of load peaks and to control fraud and abuse cases. Despite of this crucial necessity, this is currently very hard to achieve in many developing countries since smart meters or advanced metering infrastructures (AMIs) are not yet settled there to monitor and report energy usages. Whereas the communication and information technologies have widely emerged in such nations, allowing the enormous spread of smart devices among population. In this paper, we present mobile-based BLPDA, a novel platform for big data analytics of consumers' load profiles (LPs) in the absence of AMIs' establishment. The proposed platform utilizes mobile computing in order to collect the consumptions of consumers, build their LPs, and analyze the aggregated usages data. Thus, allowing electricity providers to have better vision for an enhanced decision making process. The experimental results emphasize the effectiveness of our platform as an adequate alternative for AMIs in developing countries with minimal cost.

1 Introduction

A keen attention has been directed in the last decades towards the utilization and optimization of electricity sources, especially with the tremendous increase of demand on the different levels of granularity. Thus, automatic meter reading (AMR) came about in the middle of 80s as an automated mechanism to gather basic electric meter-reading data, instead of depending on the individual collectors of the electricity providers to pass by one-by-one household to get its reading that represents the monthly energy consumption [1]. AMRs tended to be dedicated for collection only on either a monthly or daily basis, without any facilities of control messages or broadcast commands. In 2005, advanced metering infrastructure (AMI) technology has arisen in many countries, evolving from the foundations of AMR [2]. AMIs allow automated, real-time record and send of consumers' consumption data to a central computer in order to analyze and control user consumption.

The analysis of the high velocity, volume and variance of collected AMI data has allowed accurate aggregations for seasonal / long term usage trends, and provided a pattern of the total demands with many associated innovative analytical outcomes, including load profiling, peaks identification, abnormal usages, demand forecasting, electrical thefts, and intelligent pricing tariffs [3-4]. Thus, it allows operators to optimize the energy supply. Due to the lack of satisfactory nation-wide infrastructure and budget to establish their AMIs, many

developing countries are still counting on individual collectors to record consumers' consumptions. Besides missing the vital bundle of advantages formerly discussed, this approach is always subject to the physical presence of the consumer at his premises at the time of the collector's visit and to human errors.

In this study, we present our novel mobile-based Big Load Profiles Data Analyzer (mBLPDA) platform, which is considered to be a cost-effective alternative for AMIs in developing countries. mBLPDA platform uses the widely-spread smart mobile devices to allow each consumer to securely submit his own reading within a predefined duration allowance, associated with some automatically-captured data for the reading to verify the submitted data. The submitted readings are verified then aggregated per consumer to build his load profile (LP) into a central cloud-based server, where further analysis is performed using R language packages. The rest of the paper is organized as follows; section 2 overviews the different studies related to the LPs.

Section 3 discusses our proposed platform, whereas section 4 presents the associated experiments and results to evaluate our proposed platform. Finally, section 5 concludes our study and highlights our future work.

2 Related Work

Many researches have been directed towards the analysis of smart meter readers' data for different purposes. Table

^a Corresponding author: sherinmoussa@cis.asu.edu.eg

1 represents a comprehensive summary for the main studies related to electrical load profiles, clarifying the study area with the applied analytical approaches. Electricity production and consumption were investigated to forecast electricity load consumption in [5] using a genetic algorithm, in [6] using a data-based methodology, and in [7] using support vector machine for regression (SVR) and multilayer perceptron (MLP) for district or single household level. Whereas in [8], a semi-parametric approach based on generalized additive models theory was suggested. In [9], an incremental time series clustering technique using ARIMA time series forecasting model, along with a novel affinity score for determining cluster similarity of time series datasets was proposed. An Incremental Summarization and Pattern Characterization (ISPC) framework was introduced in [10] to incrementally characterize patterns in a data stream of electric meter reader and correlate these across time through incremental learning. Authors in [11-12] proposed a threat model and a predictive usage analytical toolkit for smart meter management applying time series analysis predictive modeling and regression analysis respectively, using data analytics on data collected by AMI in order to identify abnormal consumption trends and possible fraud to reduce electricity theft.

Another methodology for distribution feeder identification was presented in [13] to provide correctness improvement of distribution primary feeder connectivity by identifying the groups of customer meters connected to the same primary distribution feeder. K-means and self-organizing maps (SOMs) were applied in [14] to analyze load profiling in order to provide more customized tariff offers for customers pricing and supply availability. A profile identification, short-term load forecasting, and customer segmentation methodology was proposed in [15] to compute the daily profiles from estimated models. Another cloud-based Dynamic Demand Response (D²R) platform was introduced in [3] to perform curtailment strategy selection, intelligent demand-side management and forecasting, visualize LP patterns, and to relieve peak load. In [16], Fourier transforms and Gaussian time series techniques were investigated to forecast domestic electricity system demand by analyzing LPs. Models were proposed in [17-19] to construct LPs for household electricity in order to determine energy use, forecast usage, predict charge in future energy use, design the power network, and to improve efficiency of energy build structure.

Table 1. Summary of main LPs analytical studies.

Rf No	Objective	Study Area	Techniques
5	Forecast energy consumption	Turkey	Genetic algorithm
8		France	Semi-parametric approach
9		LA, USA	ARIMA time series forecasting model
10		Australia	IPCL and SOMs
7		Ireland	SVR and MLP
6		Finland	SOMs, k-means and hierarchical clustering

14	Customized tariff offers	UK	K-means and SOMs
16	Demand forecasting and peak loads	Ireland	Fourier transforms and Gaussian time series
3		LA, USA	Cloud-based Dynamic Demand Response (D ² R) platform
15		Belgium	K-means, DB scan, and one-step-ahead prediction
18		Romania	K-means and agglomerative clustering
17		Sweden	Matlab Modelling
19		BC, Canada	Conditional demand analysis
32	Analyze energy consumption profiles	Slovenia	Hierarchical clustering, FCM, and PNNs
34		Monte Carlo-type Simulation	Beta Probability Distribution function
35		Southern Ontario, Canada	Auto-regressive time series modelling on Hadoop
36		Public data set	OpenTSDB, Hbase and Hadoop
33		UK	Clustering
13		Improve distribution of primary feeder	Thailand
12	Identify abnormal consumption trends	NY, USA	Time series analysis, predictive modeling, and regression analysis
11		USA	Average Detector, ARMA-GLR, Nonparametric Statistics, and LOF
23	Classify customers' consumption patterns	Norway	PNNs and FCM
25		Ireland	K-means, K-medoid, SOM, SPSS, and Matlab
28		Spain	Frequency-based indices and SOMs
27			SOMs, Matlab, and frequency transform
30		Romania	Modified Follow-The-Leader algorithm, and SOMs
29			Frequency-domain technique
24		Slovenia	PNNs, Wavelet multi-resolution analysis, pattern recognition, and FCM
31			Hierarchal clustering
26		Ireland	Feature extraction and SOMs
20		Reduce energy consumption	Hong Kong
22	Salo, Finland		K-means and SOMs
21	Norway		Demand response analysis

A non-linear algorithm was used in [20] to characterize the growth pattern of household electricity consumption and calculate the power at a certain time to identify the reasons of consumption increase, in addition

to the approaches developed in [21-22] to reduce the total energy consumption in periods of expected peak load.

On the other hand in [23-24], Probabilistic Neural Networks (PNNs) and Fuzzy C Means (FCM) clustering algorithm were used to determine and allocate the typical LPs to consumers. Unsupervised learning based on SOMs have been used in [25-27] to classify, filter and identify customers' consumption patterns in order to learn both their distribution and topology, and segment the demand patterns for electrical customers. Additionally frequency-based indices and hourly LP curve methods were proposed in [28]. In [29-30], authors have modified the Follow-The-Leader algorithm and proposed a frequency-domain approach with SOMs for consumption pattern-based classification of electricity consumers in order to know accurate knowledge of the customers' consumption patterns. Hierarchical clustering was also considered in [31] to determine customers' daily LP in order to cluster the similar units of measured LPs. A hierarchical cluster tree was constructed using the distances between profiles to divide or group each pair of near LPs until coherent clusters were formulated.

In [32], hierarchical clustering and FCM with PNNs were used in order to study the distribution of deviations between the estimated purchased electricity at the power market and the actual consumption. In [33-34], a probabilistic model using beta probability distribution function (beta pdf) and a formulating method for energy LP for domestic buildings were proposed respectively to analyze residential consumer loads. Clustering has been applied based on the proposed scenarios of occupancy patterns. Whereas Hadoop was used in [35] to introduce an auto-regressive time series modelling technique, and in [36] to build a smart meter data analytics approach using OpenTSDB, Hbase in order to store time series data and to compute electricity consumption profiles from residential smart meter data, combined with temperature data and the occupants' daily habits.

Although the previous researches investigated various analytical studies for the LPs and consumption patterns of electricity consumers, however all studies counted on the existence of an AMI to provide the data required for such analysis. Thus, this completely waives all territories having non-AMIs from getting benefits out of their LPs' analytical studies. One main reason for the uniqueness of the proposed research is that it provides an alternative platform for AMIs and individual collectors' visits that require consumers' physical existence to get their meter's readings, allowing developing regions to analyze and control their LPs for better decision making, optimized service in case of peaks, and energy theft detection.

3 Mobile-based BLPDA

In this section, we present our proposed mobile-based platform for big LPs data analytics (mBLPDA). This platform allows non-AMIs environments to track and analyze their energy consumptions for better decision making when AMIs are infeasible. A cloud-based server

is utilized to manage and store the big LPs data, where all connections are made through JSON web services between the different sub-systems. A web-based interface is considered for the administrative services required for the system at the energy provider's side, providing the main analytical results with the appropriate graphical representations. As shown in figure 1, the system architecture of our proposed mBLPDA platform is composed of two main components explained as follows. This architecture can be considered for any energy domain having metering devices, i.e. electricity, gas, water... etc.

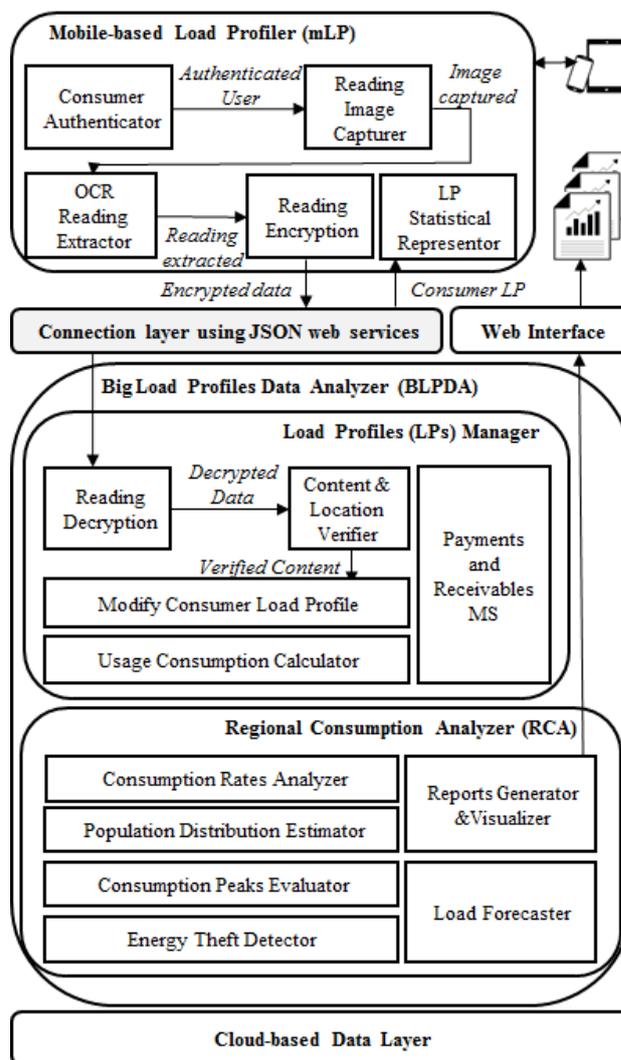


Figure 1. Mobile-based BLPDA Platform System Architecture.

3.1. Mobile-based Load Profiler (mLP)

A mobile-based application has been designed to allow building a secured LP for consumers having non-smart meter devices. This mobile load profiler app (mLP) allows the authenticated users to use their smart mobile devices to capture images with a pre-defined frame size according to the size of the meter device's reading window. An OCR module is integrated to mLP in order to extract the reading digits and the meter device ID out of the captured image. Once the reading digits and ID are

extracted, mLP wraps them together with the date, timestamp, and exact longitude and latitude of capturing the reading image, in addition to the digital signature of the authenticated user. Advanced Encryption Standard (AES) algorithm is then applied to encrypt the wrapped content, which is directly sent to BLPDA server [37]. Thus, readings are secured and associated with the user's authenticated identity during transmission. mLP allows consumers as well to track their consumptions, invoices, and payments on a monthly basis, and to get main analytical insights about their usage patterns.

3.2 Big Load Profiles Data Analyzer (BLPDA)

This represents the core engine of our proposed platform, where big data analytic approaches are investigated to deal with the huge volume of LPs data. It is responsible for the LPs management and the associated analysis per region. BLPDA consists of the following sub-modules.

3.2.1 Load Profiles (LPs) Manager

LPs Manager receives the encrypted readings that are sent simultaneously from the authenticated users through their mLP. The content is first decrypted and verified to confirm that the extracted meter device ID and location match successfully with the registered profile of the authenticated user. The verified data is then accumulated to the consumer's LP according to the date of reading in order to compute his energy consumptions. Hence, LPs are constructed per consumer and distributed based on the geographical location of the meter devices. A payments and receivables management system can be integrated to generate invoices and handle the payment processes.

3.2.2 Regional Consumption Analyzer (RCA)

Considering the LPs accumulated for consumers, RCA uses R language packages to perform two main categories of analysis; per region and per household within a region analytics [38].

- (i) Per region: RCA clusters the huge volume of consumption rates within a region using k-means approach, and forecasts the expected load demand for the next 12 months using time series analysis. Accordingly, usage peaks can be predicted in order to avoid service malfunctioning under load pressures.
- (ii) Per household within a region: RCA estimates an overall approximation for the population distribution within a region using correlation, and detects possible energy theft cases using anomaly detection approach through the new R anomalous package [39-40]. This takes into consideration the LPs' patterns and the local average of consumption per individual.

Analytical reports and graphs are aggregated on a monthly and quarterly basis to visualize the outcomes at the energy provider side as well as the consumer side through mLP. Figure 2 clarifies the system flowchart of mBLPDA platform.

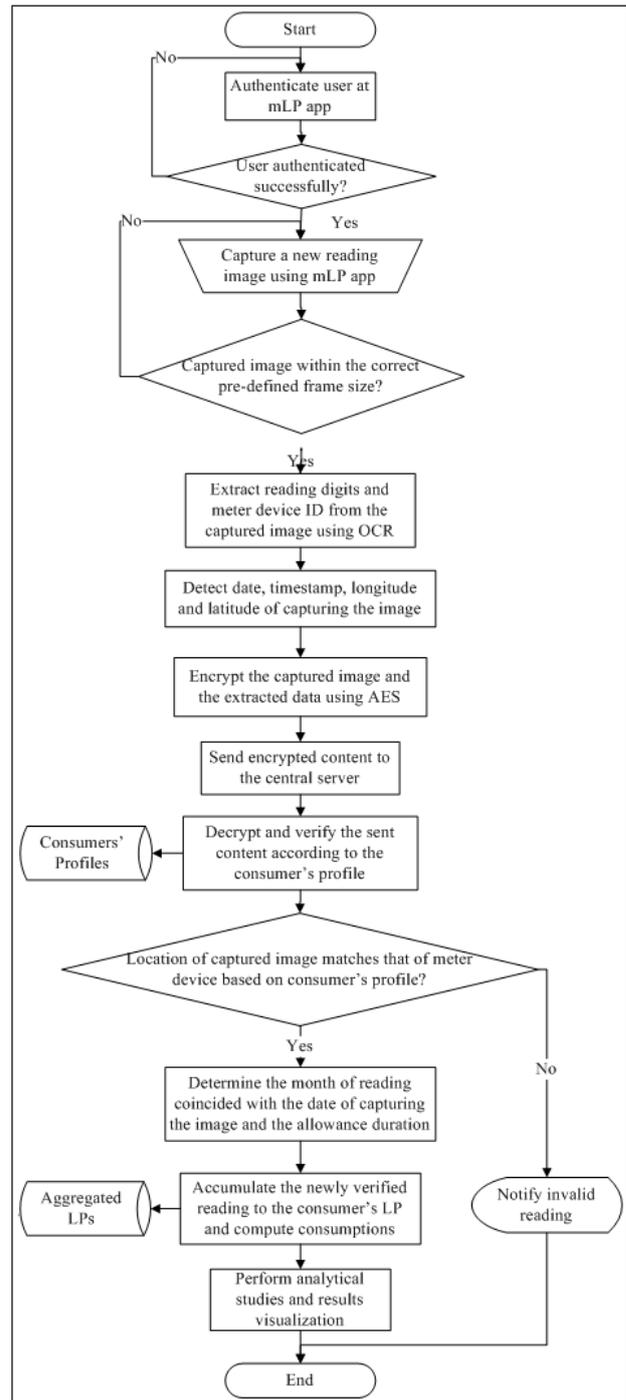


Figure 2. Mobile-based BLPDA Platform System Flowchart.

4 Experiments and Results

In order to evaluate our proposed platform, we operated mBLPDA for 6 months in 4 different regions. 1,648,576 electrical meter readings were collected through mLP Android app. Although mBLPDA architecture is generic for any energy domain having meter devices, we focus in our experiments on the electricity domain in order to pre-define the frame size of capturing to match the local electric meter devices. We have considered one of the most popular electrical meter device models as shown in figure 3, with the edge corners of the adjusted frame size

of capturing. Moreover, a customized OCR module is integrated to mLP to extract the reading digits and meter device ID out of the captured image, having the specific font type and size of the local electric meter devices.

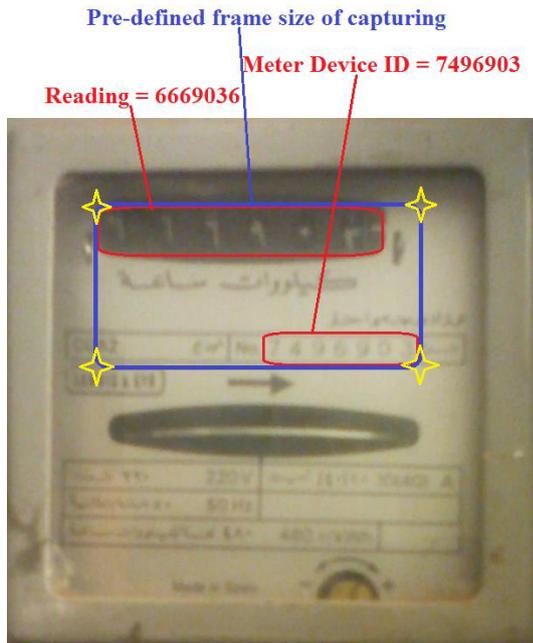


Figure 3. Sample electrical meter device with the pre-defined frame size of capturing.

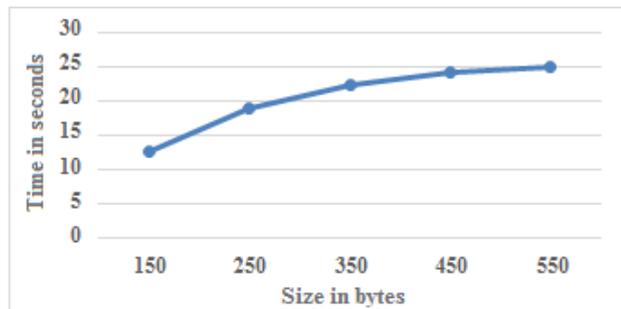


Figure 4. Time in seconds consumed by mLP to process and transmit readings versus the size of content to handle.

Experimental results indicate that mBLPDA platform can provide most of the essential analytical insights as in AMIs. Figure 4 shows the average time in seconds versus the size of content to handle in bytes in order to process a captured reading image by extracting and encrypting the associated content (i.e., reading digits, meter device ID, in addition to the date, timestamp, longitude, and latitude of image capturing), and then to transmit all of these processed data into the central cloud-based server. mLP app takes an average of 12.57 seconds to process and transmit 150 bytes on average, and increases till it reaches around 25.11 seconds for an average size of 550 bytes. Thus, mLP app consumes an overall average time equals to 20.63 seconds for the processing and transmission of one reading. Whereas Figure 5 presents the consumers' LPs collected through mLP app during the pilot duration for 4 the regions, having average consumptions ranged from 4,617,388 KW to 8,886,716.042 KW per month/ region.

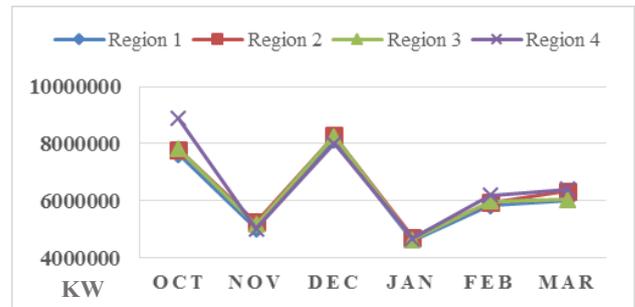


Figure 5. Accumulated LPs using mLP app.

We have used Rand Index (RI) as a measure to evaluate the quality of consumption rates clustering. RI is considered since it equally weighs false positives and false negatives, where both are false, as in the equation below [41]:

$$RI = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

Where TP represents the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. Consumptions rates are clustered into 3 main clusters (i.e., minimal, normal, and high consumption, thus $k=3$). As shown in figure 6, mBLPDA platform has achieved RI values ranged from 0.88 till 0.93, which indicates coherent and high-quality clusters with an overall RI average equals to 0.9. Regarding the evaluation of peaks prediction in mBLPDA platform, we applied F-measure criteria, where it combines both recall and precision parameters to assess the prediction results according to the following equations respectively [42]:

$$F\text{-measure} = 2 \times \left[\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right] \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

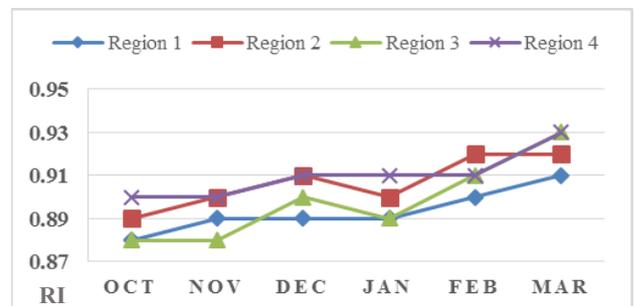


Figure 6. RI results for the quality of consumption rates clustering.

As shown in figure 7, mBLPDA platform successfully achieves F-measure results for consumptions' peaks prediction with a minimum value of 0.71 and a maximum value of 0.95. This indicates a high prediction capability of an overall F-measure value equals to 0.83. We have performed additional experiments as well to investigate anomaly detection at mBLPDA on the household's level, as a preparation for energy theft detection when

accompanied with payments control. A sample anomaly detected on a household's level is presented in figure 8, where the average consumption rate for this consumer during 5 months is 120 KW. The reading recorded for December is 243 KW, which is a way beyond the normal average consumption rate.

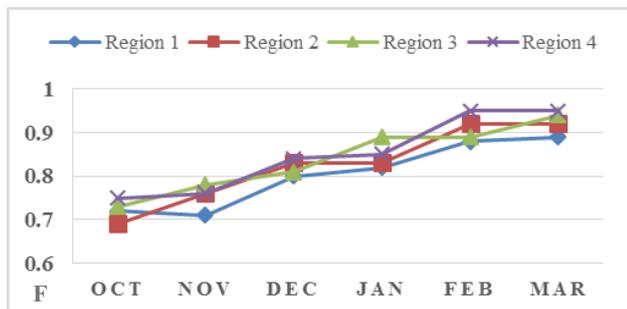


Figure 7. F-measure results for consumption peaks prediction.

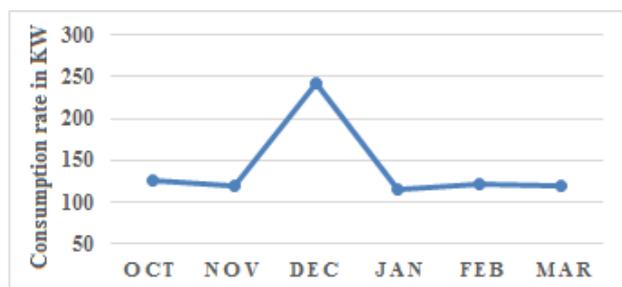


Figure 8. A sample anomaly detected at a household's level.

5 Conclusion and Future Work

Managing and analyzing the energy consumptions of energy providers among the world has become a dramatically increasing demand, where AMIs have supported during the last years to provide useful analytical insights. In this paper, we tackle the problem of building energy LPs in the non-AMIs environments. Many developing countries have not yet settled their AMIs, depending on individual collectors to record readings.

We propose our mobile-based big LPs data analyzer (mBLPDA) platform, where consumers can use mLP smart app to securely submit their own meter reading within an average of 20.63 seconds for the processing and transmission of one reading. Readings are all aggregated on cloud-based server that applies R language packages to analyze consumption rates, forecast load demands and peaks, and to detect energy theft, with an approximated population distribution estimation based on the built LPs. A popular design for local electric meter devices is used. Experiments show that our proposed mBLPDA platform successfully clusters LPs with an average RI equals to 0.9, whereas it predicts peak loads with an average of F-measure equals to 0.83.

As our future work, we investigate considering most of the designs of the local electric meter devices, including other energy meter devices for gas and water, and adding more analytical studies that increase the

visibility of energy providers to support their decision making processes for a better service.

References

1. H. Sui, et al. An AMI system for the deregulated electricity markets, *IEEE Transactions on Industry Applications* **45.6**, 2104-2108 (2009)
2. M. Popa, Data collecting from smart meters in an Advanced Metering Infrastructure, *15th IEEE International Conference on Intelligent Engineering Systems (INES)*, (2011)
3. Y. Simmhan, et al. Cloud-based software platform for big data analytics in smart grids, *Computing in Science & Engineering*, **15.4**, 38-47 (2013)
4. D. Alahakoon, X. Yu, Advanced analytics for harnessing the power of smart meter big data, *IEEE International Workshop on Intelligent Energy Systems (IWIES)*, (2013)
5. M. Çunkaş, U. Taşkıran, Turkey's electricity consumption forecasting using genetic programming, *Energy Sources, Part B: Economics, Planning, and Policy* **6.4**, 406-416 (2011)
6. T. Räsänen, et al. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data, *Applied Energy* **87.11**, 3538-3545 (2010)
7. S. Humeau, et al. Electricity load forecasting for residential customers: Exploiting aggregation and correlation between households, *IEEE Sustainable Internet & ICT for Sustainability (SustainIT)*, (2013)
8. Y. Goude, R. Nedellec, N. Kong, Local short and middle term electricity load forecasting with semi-parametric additive models, *IEEE Transactions on Smart Grid*, **5.1**, 440-446 (2014)
9. Y. Simmhan, M. U. Noor, Scalable prediction of energy consumption using incremental time series clustering, *IEEE International Conference on Big Data*, (2013)
10. D. De Silva, et al. A data mining framework for electricity consumption analysis from meter data, *IEEE Transactions on Industrial Informatics* **7.3**, 399-407 (2011)
11. D. Mashima, A. A. Cárdenas, Evaluating electricity theft detectors in smart grid networks, *Research in Attacks, Intrusions, and Defenses*, Springer Berlin Heidelberg. 210-229 (2012)
12. H. Li, et al. Usage analysis for smart meter management, *8th IEEE International Conference & Expo on Emerging Technologies for a Smarter World (CEWIT)*, (2011)
13. K. Y. Khumchoo, W. Kongprawechnon, Cluster analysis for primary feeder identification using metering data, *6th IEEE International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES)*, (2015)
14. I. Dent, U. Aickelin, T. Rodden, Application of a clustering framework to UK domestic electricity data, arXiv preprint arXiv, 1307.1079 (2013)

15. M. Espinoza, et al. Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series, *IEEE Transactions on Power Systems*, **20.3**, 1622-1630 (2005)
16. F. McLoughlin, A. Duffy, M. Conlon, Evaluation of time series techniques to characterise domestic electricity demand, *Energy* **50**, 120-130 (2013)
17. J. Widén, et al. Constructing load profiles for household electricity and hot water from time-use data—Modelling approach and validation, *Energy and Buildings* **41.7**, 753-768 (2009)
18. E. C. Bobric, G. Cartina, G. Grigoras, Clustering techniques in load profile analysis for distribution stations, *Advances in Electrical and Computer Engineering* **9.1**, 63-66 (2009)
19. K. H. Tiedemann, Using conditional demand analysis to estimate residential energy use and energy savings, *Proceedings of the CDEEE* (2007)
20. T. M. Lai, et al. Electricity consumption in Hong Kong: trend analysis and greenhouse gases emission, *HKIE Transactions*, **21.2**, 81-88 (2014)
21. H. Sæle, O. S. Grande, Demand response from household customers: experiences from a pilot study in Norway, *IEEE Transactions on Smart Grid* **2.1**, 102-109 (2011)
22. T. Räsänen, J. Ruuskanen, M. Kolehmainen., Reducing energy consumption by using self-organizing maps to create more personalized electricity use information, *Applied Energy* **85.9**, 830-840 (2008)
23. D. Gerbec, S. Gašperič, F. Gubina, Allocation of the load profiles to consumers using probabilistic neural networks, *IEEE Transactions on Power Systems*, **20.2**, 548-555 (2005)
24. D. Gerbec, et al. Determining the load profiles of consumers based on fuzzy logic and probability neural networks, *IEEE Proceedings-Generation, Transmission & Distribution* **151.3**, 395-400 (2004)
25. F. McLoughlin, A. Duffy, M. Conlon, A clustering approach to domestic electricity load profile characterisation using smart metering data, *Applied energy*, **141**, 190-199 (2015)
26. C. Beckel, L. Sadamori, S. Santini, Towards automatic classification of private households using electricity consumption data, *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, (2012)
27. S. V. Verdú, et al. Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps, *IEEE Transactions on Power Systems* **21.4**, 1672-1682 (2006)
28. S. V. Verdú, et al. Characterization and identification of electrical customers through the use of self-organizing maps and daily load parameters, *Power Systems Conference and Exposition, IEEE PES*, (2004)
29. E. Carpaneto, et al. Electricity customer classification using frequency-domain load pattern data, *Int. J. of Electrical Power & Energy Systems* **28.1**, 13-20 (2006)
30. G. Chicco, et al. Load pattern-based classification of electricity customers, *IEEE Transactions on Power Systems* **19.2**, 1232-1239 (2004)
31. D. Gerbec, S. Gašperič, F. Gubina, An approach to customers daily load profile determination, *IEEE Power Engineering Society Summer Meeting* **Vol. 1**, (2002)
32. D. Gerbec, S. Gašperič, F. Gubina, Determination and allocation of typical load profiles to the eligible consumers, *IEEE Power Tech Conference Proceedings*, Bologna. **Vol. 1**, (2003)
33. R. Yao, K. Steemers, A method of formulating energy load profile for domestic buildings in the UK, *Energy and Buildings* **37.6**, 663-671 (2005)
34. S. W. Heunis, R. Herman, A probabilistic model for residential consumer loads, *IEEE Transactions on Power Systems* **17.3**, 621-625 (2002)
35. O. Ardakanian, et al. Computing Electricity Consumption Profiles from Household Smart Meter Data, *EDBT/ICDT Workshops*, (2014)
36. S. Prasad, S. B. Avinash, Smart meter data analytics using OpenTSDB and Hadoop, *IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, (2013)
37. J. Daemen, and V. Rijmen, *The design of Rijndael: AES-the advanced encryption standard*, Springer Science & Business Media, (2013)
38. R. Documentation, <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html>
39. R package for detecting unusual time series, <http://robjhyndman.com/hyndsight/anomalous/>
40. R. J. Hyndman, et al. Large-scale unusual time series detection, *Proceedings of International Conference on Data Mining series*, (2015)
41. W. M. Rand, Objective criteria for the evaluation of clustering methods, *Journal of the American Statistical Association* **66 (336)**, 846-850 (1971)
42. D. M. Powers, Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, (2011)