Suspicious Electric Consumption Detection Based on Multi-Profiling Using Live Machine Learning

Thomas Hartmann*, Assaad Moawad*, Francois Fouquet*, Yves Reckinger‡, Tejeddine Mouelhi†,
Jacques Klein*, and Yves Le Traon*
*Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg, {first.last}@uni.lu
†itrust consulting, Luxembourg, {mouelhi}@itrust.lu
‡Creos Luxembourg S.A., Luxembourg, {yves.reckinger}@creos.net

***Abstract*—The transition from today's electricity grid to the so-called smart grid relies heavily on the usage of modern information and communication technology to enable advanced features like two-way communication, an automated control of devices, and automated meter reading. The digital backbone of the smart grid opens the door for advanced collecting, monitoring, and processing of customers' energy consumption data. One promising approach is the automatic detection of suspicious consumption values, *e.g.,* due to physically or digitally manipulated data or damaged devices. However, detecting suspicious values in the amount of meter data is challenging, especially because electric consumption heavily depends on the context. For instance, a customers energy consumption profile may change during vacation or weekends compared to normal working days. In this paper we present an advanced software monitoring and alerting system for suspicious consumption value detection based on live machine learning techniques. Our proposed system continuously learns context-dependent consumption profiles of customers, *e.g.,* daily, weekly, and monthly profiles, classifies them and selects the most appropriate one according to the context, like date and weather. By learning not just one but several profiles per customer and in addition taking context parameters into account, our approach can minimize false alerts (low false positive rate). We evaluate our approach in terms of performance (live detection) and accuracy based on a data set from our partner, Creos Luxembourg S.A., the electricity grid operator in Luxembourg.**

I. Introduction

The vision of the smart grid aims to increase the efficiency and reliability of today's electricity grid [1]. Renewable energies and distributed micro-generations will be seamlessly integrated into the electricity grid, increasing eco-efficiency and sustainability. To tackle the introduced management complexity, advanced new features like two-way communication, an automated control of devices, remotely collecting consumption data from smart meters, and demand time pricing [2] will gradually become the norm. The backbone to realize these new features is the convergence of modern information and communication technology (ICT) with power system engineering [3]. The digital nature of the smart grid also opens the door for advanced collecting, monitoring, and processing of customers' consumption data. One promising possibility coming with this, is the automatic detection of suspicious consumption values. Suspicious consumption values can be due to technical or non-technical losses in the power distribution network, such as electricity theft, cyber-attacks on customers' smart meters or accounting system [4]. Suspicious consumption values are values which are untypical for a certain customer (or group of customers) for a given context, like a typical Monday morning. Non-technical losses are a major concern of any utility company. As an example, in the United States non-technical losses were estimated as between 0.5% and 3.5% of the gross annual revenue [5].

Many approaches based on statistical profiling and machine learning techniques have been suggested in various domains (*e.g.,* [6], [7], [8], [9]) to detect suspicious values. Due to the availability of automated meter reading [3] in smart grids, customers' consumption data can now be regularly collected, stored, and processed. By applying statistical profiling and machine learning techniques for monitoring the electrical consumption of customers, the normal consumption per customer or group of customers can be learned based on profiling. In case a consumption value is too far from the learned profile, an alarm can be created in order to prevent non-technical losses. However, detecting suspicious values in the amount of meter data is challenging, especially because electric consumption highly depends on the context [10], *e.g.,* geographical area, number of residents, temperature, date (vacation, weekday, weekend), type of the heating system, habits of inhabitants, etc. Given the high number of context parameters, it is very difficult to build reliable profiles. For example, a private customer can have a very different consumption profile depending on the weather conditions or during vacation compared to working days. Moreover, context parameters can depend on each other, *e.g.,* the temperature on a winter Sunday afternoon may have different influence on the load profile than the temperature on a summer Monday afternoon. Due to this variability, a single consumption profile per customer, which simply computes an average consumption (similarly to methods for electrical load forecasting), can lead to a significant amount of false positive alarms. Thus, a consumption monitoring system has to take this context variability into account, by always verifying customers' consumption data with respect to a certain context, to avoid a high number of false positive alerts.

In this paper we present a novel software monitoring system for suspicious consumption detection. Unlike most other approaches we use live machine learning techniques to create multiple profiles (*e.g.,* daily, weekly and monthly profiles), specialized per context, instead of a global profile.

The research leading to this publication is supported by the National Research Fund Luxembourg (grant 6816126) and Creos Luxembourg S.A. under the SnT-Creos partnership program.

The great advantage of this technique is, that customer profiles can be continuously updated instead of being created only once in a while. Most importantly, our profiles are context-dependent, *i.e.,* we take context parameters like date, weather conditions, etc. into account. Whenever a value has to be checked, our system selects a set of most relevant profiles based on the context and performs a distance computation. Our monitoring system creates alarms based on the deviation of the analyzed consumption value compared to the selected profile set. To sum up, our monitoring system builds (by learning) multiple customer profiles over time, classifies them and selects the most appropriate ones to decide if a value is suspicious or not. We evaluate our approach in terms of performance (live detection) and accuracy based on a data set from our partner, Creos Luxembourg S.A., the main electricity grid operator in Luxembourg.

The remainder of this paper is structured as follows. Section II describes the context of our work, the main characteristics of a smart grid topology, based on the example of Luxembourg. In Section III we present our live learning based smart grid consumption monitoring system, which we evaluate in Section IV. The related work of this paper is discussed in Section V before the paper concludes in Section VI.

II. Context: Smart Grid Topology and Entities

To present the context of our work, we describe in this section the main characteristics of a typical smart grid topology, based on the smart grid test deployment in Luxembourg. This test deployment contains three different regions, around 300 smart meters, three data concentrators, and a central system. The grid topology in Luxembourg is based on a power line communication [11] (PLC) network and is representative for such smart grid topologies. A more detailed description and analysis of the smart grid topology in Luxembourg can be found in [12]. A major advantage of PLC is that the same media that is used for electric power transmission can be used for establishing the communication network and transmitting data. On the other hand, a major concern with PLC is the amount of electric noise and disturbances that may be encountered, which requires advanced error detection techniques. The main topology devices in the context of this work are:

- *Smart meters* are the cornerstones of the smart grid infrastructure. Installed at customers houses they continuously measure electric consumption and quality of power supply and regularly report these values to utilities for monitoring and billing purposes. In Luxembourg, smart meters send the consumption values every 15 minutes for electricity, respectively every 60 minutes for gas. Another important task of smart meters is load management, as they are able to trigger relays to connect/disconnect specific loads.
- *Data concentrators* collect and store consumption data from a number of associated meters. In regular intervals (several times a day) they send this data, usually via IP connections, to a central control system.
- *Central system* concentrators send their data to a central system where all data are stored, aggregated and analyzed. Because of legal regulations these data must be deleted in regular intervals.

Given the topology structure, our proposed software monitoring and alerting system for suspicious consumption data could be deployed either on a data concentrator level (one instance per data concentrator) or at the central system (one global instance).

III. Suspicious Consumption Value Detection

In this section we describe our approach for suspicious consumption value detection. First, we present an overview of the approach and the involved components. Next, we detail the live or online aspects of our machine learning approach, followed by an explanation of the Gaussian mixture model. Finally, we detail how we learn multiple, context-dependent customer profiles, and how we select and use them for suspicious consumption value detection.

A. Overview: Towards Contextual Learning and Detection

Figure 1 shows a basic overview of our approach. Smart meters continuously report their consumption values, in regular intervals, to utilities for monitoring and billing purposes. In Luxembourg, these intervals are 15 minutes for electricity and 60 minutes for gas. In order to detect suspicious consumption values, this continuous stream of smart meter measurements is first analyzed by a context solver. For every new value, the context solver selects the most appropriate profile for this customer in order to decide what is the normal range of the current measured electricity consumption. For example, for a measurement value on a sunny Monday afternoon (working day) in summer, the context solver will select a profile with comparable context parameters for this customer. The context, for example, can include features like: user type (individual, family, industry, commercial), temporal context (season of the year, month, weekday, holidays), and so forth. The context resolution yields a list of positive and negative profiles. Positive profiles contain information to judge if the measured value is known to be in the normal range, while negative profiles are known to be suspicious values. From these profiles, a decision making step is executed, based on a confidence rate and the probability distribution provided by the profiles. If the measured value is accepted, it is used to train the positive profiles. If it is not accepted, an alert is created. In addition, an interactive validation request is raised to manually validate if the value is indeed suspicious or not. Finally, the negative rated values (suspicious values) are used to train the negative profiles. An initial training period is required to bootstrap the profiles. The update of profiles is a very incremental step and can therefore be performed for every value (as we will show in Section IV). However, it would be also feasible to combine several values instead of checking every value.

B. Live Machine Learning

From observing large sequences of data, machine learning and pattern recognition algorithms can build models that reflect or represent, to a certain degree of accuracy, the domain or the environment on which they are trying to learn from. In real-world environments large sequences of data may not be available in advance, may take too much time to gather, or they can be very expensive in terms of computation power to process in a batch mode. For a reactive system operating in real-time or near real-time is a crucial requirement. In order

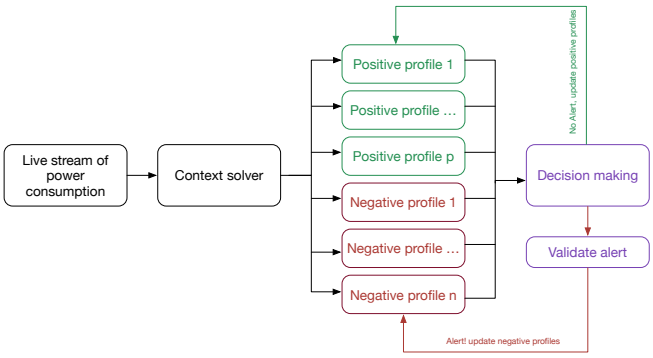

Fig. 1. Suspicious consumption value detection overview

to address this, we use live machine learning algorithms [13] with the following characteristics:

- The algorithms should be able to create or update the models whenever new data arrives (on the fly).
- The computational effort needed for a single update should be minimal and should not depend on the amount of data observed so far.
- The update should only depend on the latest observed value and should not explicitly require access to old data.
- The generated models should be compact and should not grow significantly with the number of observed instances.

C. Gaussian Mixture Model

In this paper we explore modeling power consumption usage by probability density functions (pdf) based on kernel density estimates (KDE). Particularly, we use Gaussian mixture models (GMM), which are known to be a powerful tool in approximating distributions even when their form is not close to Gaussian [14]. A GMM is a probabilistic model that assumes that all data points are derived from a mixture of Gaussian distributions with unknown parameters. Mixture models are basically generalizing k-means clustering.

Definition 1: In a nutshell, A Gaussian mixture model of M components, provides the following probability distribution function of an event x happening:
$p(x) = \sum_{j=1}^{M} w_j K_{\sigma_j}(x)$, with
$K_{\sigma_j}(x) = (2\pi\sigma^2)^{-1/2} \exp^{-(x-x_j)^2/(2\sigma^2)}$.

$K_{\sigma_j}(x)$ is one Gaussian component of the mixture model, with an average of x_j, a standard deviation of σ_j and a weight w_j. These parameters must be learned from the data on the fly. Kristan *et al.*,[15] provides an online learning algorithm called *oKDE* that is able to update the gaussian mixture model in real-time. The implementation of our software monitoring and alerting system is based on this algorithm, which is suitable and fulfils the requirements we presented earlier.

D. Profiling Power Consumption

In order to build consumption profiles in real-time, we feed the measured consumption values from smart meters to our profiler and process them online. In Luxembourg, each smart meter reports its consumption values every 15 minutes for electricity (and 60 minutes for gas). Figure 2 shows an example of measurements from one customer (one smart meter) over a period of 24 hours. The measurements were taken on 31 weekdays (Monday to Friday). The x-axis in the figure represents the time of the day and the y-axis the consumption (active energy consumed) in *Wh*. Every blue point in the figure corresponds to one measurement value. For example, if we take the time 6.00 am, every point along the y-axis belongs to one measurement for one day at 6.00 am. In red, the figure shows the average for every time, *i.e.,* the average value over all days at every measured time (15 minute intervals). Based on these measurements we can create our consumption profiles for this customer by feeding these measures to our profiler (in real-time) and processing them online. For every new value the consumption profile of the customer will be refined (recalculated). Figure 3 shows the profile (the Gaussian mixture model), constructed for the example of Figure 2.

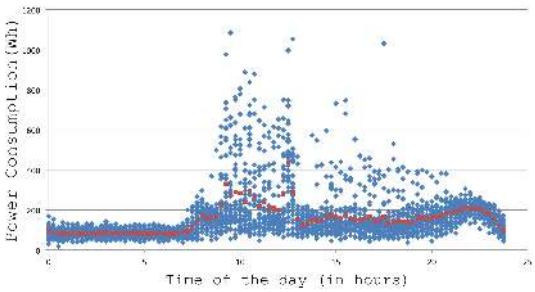

Fig. 2. Power consumption measures (in blue) and average values (in red)

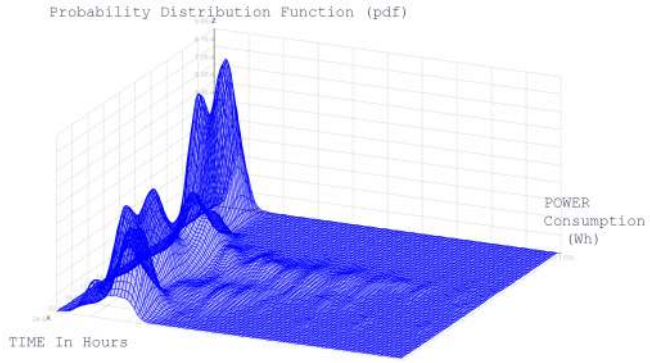

Fig. 3. Probability distribution function (pdf) of the consumption values from Figure 2 built with live machine learning

As an example, the power usage of this user is quite predictable at midnight (varying between 0 and 200 Wh). This is reflected in the profile by a Gaussian Kernel with low variance and we are quite confident (with high probability) that the next midnight measure will be also between 0-200 Wh. However, if we compare this with the consumption at noon, where the user consumes between 0 and 1000 Wh, the profiling has a higher variance, the probability is distributed over a wider range, and thus the prediction is less accurate. In such situations, having a contextual profiling can help to significantly increase the accuracy of the prediction. For instance, during weekends at noon, the consumption may be varying in a less wider range than during weekdays at noon.

IV. Evaluation

In this section we present the evaluation of our proposed consumption monitoring system. We first describe the setup of our evaluation. Next, we validate the efficiency of our approach, followed by a validation of its performance, *i.e.,* accuracy, precision, recall, and F1 score. We compare these results with a traditional, single profile per user approach.

A. Experimental Setup

We evaluated our consumption monitoring system based on real data from the Luxembourg smart grid test deployment. The data are provided by our industrial partner Creos Luxembourg S.A. We analyzed consumption data values from a total of 218 smart meters from three different regions in Luxembourg. For our evaluation we considered the consumption data for a time frame of six weeks for each meter. In Luxembourg, each meter reports its consumption data every 15 minutes, via a data concentrator, to the central system. This results in 804345 consumption values in total.

In order to build multiple profiles we considered three different context parameters in our experiments:

- Customer types: since the electricity consumption for residential and industrial customers differs to a significant extent, we clustered customers in two different groups: residential and industrial customers.
- Business day context: we separated working days from weekends and holidays. The idea behind this is that electricity consumption should highly depend on this context parameter. We further divided a day into 96 time slots (=24x4, corresponding to each of the 15 minute reading intervals).
- Weather condition: we collected the historical temperature readings (with an hourly resolution) from meteolux[1], the official meteorological service provider in Luxembourg, and use it as our third context parameter.

We then evaluated our approach in terms of performance to validate if our system is able to be used in a live monitoring system (live detection) and in terms of accuracy. We run the experiments on a 2.6 GHz Intel Core i7 with 8 GB of RAM.

B. Efficiency: Can We Meet Near Real-Time Expectations?

The core live learning algorithm oKDE to update the profiles of the whole dataset for 218 customers and 803645 consumption values takes 0.22 seconds. This is around 274 nanoseconds per consumption value. It is important to notice, that this step only includes the learning itself, it doesn't contain classification, training, nor decision-making. The whole processing including classification, training, selecting the correct profile, and decision-making takes 1.11 seconds for the complete dataset. This results in an average processing time of 1.37 milliseconds per consumption value. Considering that the interval of consumption reading in Luxembourg is 15 minutes, we are able to process around 656934 consumption readings during one cycle. Assessing the fact that the computation is conducted in a single thread on a classical computer processor (single core on intel i7 processor), we can consider that our approach is fast enough to be used in a live monitoring system. For example, in the case of Luxembourg with approximately 550.000[2] inhabitants, a standard laptop is sufficient to monitor the consumption values of the whole country in live. However, as a threat to validity, our performance and therefore the associated near real-time suspicious value detection can be significantly impacted by network access. To overcome this risk, the proposed computation can be split into geographical specialized nodes, *e.g.,* one monitoring system per region (on a data concentrator level). Then, dedicated databases can distribute data based on their usage probability on each node. Another threat to validity is that the profile selection also can considerably impact the performance. We plan to evaluate this in detail in future work.

C. Effectiveness: Can We Better Detect Suspicious Values?

In order to measure the performance of our multi-context profiling, we compare our approach to a non contextual one, leveraging a single profile per customer approach. We first divide our dataset in two subsets: a training subset of 700.000 and a testing subset of 103.645 consumption values. The latter contains correct, meaning not suspicious consumption values. This is verified by domain experts from Creos.

We then generated two additional sets containing suspicious (false) consumption readings:

- Randomly generated in the interval $[max - 3 * max]$, where max is the maximum electricity consumption value for each customer. We select this interval empirically because it is discriminative, due to the fact that both approaches have 100% accuracy to detect very high deviations. The set evaluates the ability of both approaches (our proposed multi-context profiling and a traditional single profiling per customer approach) to detect suspicious (wrong) consumption values.
- Randomly generated false data between $[max/2 - max]$. This set should test the discriminative power of context-dependent profiling, compared to a single profile per customer. The false data will be seen very real and acceptable if it is not taken within its context.

For our evaluation we use classical metrics, based on:

- True positives (tp): true readings classified as normal.
- True negatives (tn): true readings classified as alert.
- False positives (fp): false readings classified as normal.
- False negatives (fn): false readings classified as alert.

We then calculate from these values, the accuracy, precision, recall, and F1 score [16] for both approaches.

Profiler	True positives (tp)	False Negatives (fn)	Accuracy %
Single	102675	970	99.06%
Multi-Context	102695	950	99.08%

TABLE I. Testing set results from Creos

[1] http://www.meteolux.lu/

[2] http://www.luxembourg.public.lu/fr/societe/population/

Profiler	True negatives (tn)	False Positives (fp)	Accuracy %
Single	92386	11259	89.13 %
Multi-Context	96458	7187	93.06%

TABLE II. TESTING SET RESULTS FROM RANDOMLY GENERATED IN THE INTERVAL [MAX - 3X MAX]

Profiler	True negatives (tn)	False Positives (fp)	Accuracy %
Single	47297	56348	45.63 %
Multi-Context	86582	17063	83.55%

TABLE III. TESTING SET RESULTS FROM RANDOMLY GENERATED IN THE INTERVAL [MAX/2-MAX]

Attribute	Single Profiler	Multi-context profiler
True positives (tp)	102675	102695
True negatives (tn)	139683	183040
False positives (fp)	67607	24250
False negatives (fn)	970	950
Precision	0.602	0.808
Recall	0.99	0.99
Accuracy	0.779	0.918
F1 score	0.749	0.890

TABLE IV. A GLOBAL OVERVEW OF RESULTS

Table I shows the different results of the 3 test sets. Both profiling techniques performed very well on detecting true positives from Creos (more than 99% as shown in table I). However, the single profiler performed worse when it comes to false positives. For the randomly generated sets, the multi-context profiler was able to detect much more accurately that the values do not correspond to the current user, especially when it comes to values that are in the range of [Max/2-Max]. This validates the fact that taking the context into account, a value that might look fine for a single profile because it falls in the acceptable range, it might look suspicious for the multi-context profiler. The only draw back is that the multi-context profilers need much more time and data to bootstrap. In fact, the more contexts and profiles we create per user, the more data are required to reach a stable phase from the learning. Finally, table IV summarizes all tests and provide a single metric (F1 score) to compare both approaches overall. Our multi-context profiler was able to score 18% better than a single profiler.

V. RELATED WORK

Non-technical losses in power grid systems, regardless if they occur due to physically or digitally tampered consumption data or damaged devices, are a major concern of electricity providers. Therefore, it is of no surprise that several different approaches to detect suspicious consumption data have been suggested over the years. Most recent approaches are based on load profiling, becoming possible through automated meter reading. Monedero [17] *et al.,* propose to apply data mining techniques for detection of non-technical losses and present two methodologies, one based on neural networks and one on statistical techniques. They classify customers in two groups, *i)* likely to be affected and *ii)* not affected by non-technical losses. Markoč [18] *et al.,* also suggest to use neural networks to detect suspicious data and show that these can be trained by generated samples instead of real data. In [19] Nagi *et al.,* present an approach for non-technical loss detection based on support vector machines (SVM). Their approach preselects suspicious customers (irregular consumption patterns) to be inspected onsite based on irregularities in their consumption behaviour. In [20] Nizar, Dong, and Wang present a novel method for non-technical loss detection, which also uses data mining techniques for classification. Their approach is built on a novel computational method, called extreme learning machine (ELM). They compare their approach with other classification techniques, such as the support vector machine (SVM) algorithm. The work of Cabral [21] *et al.,* goes in a similarly direction but uses a non-supervised artificial neural network called self-organizing maps (SOM). The work of Espinoza [22] *et al.,* aims at providing a unified framework for electric consumption forecasting and clustering by creating daily profiles of customers. They first generate short-term models that can produce accurate forecasts, extract temperature and seasonal effects and identify the type of customer under scrutiny. Then, they partition the set of time series, using clustering algorithms, based on the customer profiles.

All of these approaches provide good results in terms of accuracy (low false positive rate) and they have in common that they use data mining or machine learning techniques to cluster customers based on their consumption profiles. Whereas these clusters are built off-line and only once in a while our approach applies live machine learning techniques enabling the continuous update and refinement of the customer profiles. This enables our approach to be used in a live monitoring and alarming system instead of using a computational intensive batch process to look for suspicious data. Moreover, our method foresees to build not just one but multiple profiles per customer. In addition learning multiple profiles per customer our method takes context parameters, like temperature and date (*e.g.,* weekday, weekend, vacation) into account. This can help to avoid a high number of false positive alerts. In [6] Espinoza *et al.,* present a very advanced short-term load forecasting approach using kernel-based modeling for nonlinear system identification. Even though their goal is electric load forecasting rather than an alarm system for suspicious consumption values, they also create consumption profiles of customers and most notably, discuss the need to take different context parameters into account.

Since non-technical losses are a major threat for the electrical grid, there are also approaches on the level of smart meters to detect suspicious data. Some intrusion detection systems (IDS) for smart meters apply similar methods to learn the normal behaviour (*e.g.,* sent data) and to check any deviation from it (*e.g.,* Berthier *et al., [23]* and Tabrizi *et al. [24]*). The objective is to be able to detect any attempts to hack into the smart meter, tamper data stored in its memory or manipulate consumption data sent to the central system. However, they are technology dependent and only applicable to a very specific protocol or smart meter OS. Our approach is technology agnostic and is applicable to any protocol or metering system.

VI. Conclusion and Future Work

The transition from today's electricity grid to the so-called smart grid relies heavily on the usage of modern information and communication technology, which also opens the door for advanced collecting, monitoring, and processing of customers' energy consumption data. One promising possibility coming with this, is the automatic detection of suspicious consumption values. Suspicious consumption values can be due to technical losses or non-technical losses in the power distribution network, such as electricity theft, non-payment by customers, and errors in accounting and record-keeping. In this paper we presented an advanced software monitoring and alerting system for suspicious consumption value detection based on live machine learning techniques. Our proposed system continuously learns context-dependent consumption profiles of customers, *e.g.,* daily, weekly, and monthly profiles, classifies them and selects the most appropriate one according to the context, like date and weather. We showed that, by learning not just one but several profiles per customer and in addition taking context parameters into account, our approach can minimize false alerts (low false positive rate). We evaluated our proposed consumption monitoring system based on real data from the smart grid test deployment in Luxembourg in terms of performance and accuracy. We showed that our suspicious value detection is fast enough to be used in a live monitoring system (live detection) and that it is, in many cases, superior in terms of accuracy to other approaches, which mostly use only one profile per customer and do not take context parameters into account. We also showed that even for randomly generated values, where our approach can hardly profit from its multiple, context-dependent profiles, we are comparable (or even slightly better) to other approaches. On the drawback side, our approach comes with an overhead in storage and computation time, due to the usage of multiple context-dependent profiles. The idea behind this approach has been developed in cooperation with Creos Luxembourg S.A. (the main electricity grid operator in Luxembourg) and has been implemented as a prototype monitoring system to detect suspicious consumption values.

In future work we plan to extend the context parameters our system takes into account and integrate more complex ones, like cloud-effects, hours of sunshine, perception, etc. In addition, in our current implementation all context parameters are fixed, meaning that our system is not able to automatically derive or learn new context parameters. In future work we want to explore if it is possible and if it could be helpful to automatically derive new context parameters.

Acknowledgment

The authors would like to thank Robert Graglia from Creos Luxembourg S.A. for his support.

References

[1] S. Amin and B. Wollenberg, "Toward a smart grid: power delivery for the 21st century," *Power and Energy Magazine, IEEE*, vol. 3, no. 5, pp. 34–41, Sept 2005.

[2] P. Samadi, H. Mohsenian-Rad, V. Wong, and R. Schober, "Real-time pricing for demand response based on stochastic approximation," *Smart Grid, IEEE Trans. on*, vol. 5, no. 2, pp. 789–798, March 2014.

[3] H. Farhangi, "The path of the smart grid," *Power and Energy Magazine, IEEE*, vol. 8, no. 1, pp. 18–28, January 2010.

[4] P. Antmann, "Reducing technical and non-technical losses in the power sector," 2009.

[5] T. B. Smith, "Electricity theft: a comparative analysis," *Energy Policy*, vol. 32, no. 18, pp. 2067 – 2076, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0301421503001824

[6] M. Espinoza, J. Suykens, R. Belmans, and B. De Moor, "Electric load forecasting," *Control Systems, IEEE*, vol. 27, no. 5, pp. 43–57, Oct 2007.

[7] M. Espinoza, C. Joye, R. Belmans, and B. De Moor, "Short-term load forecasting, profile identification, and customer segmentation: A methodology based on periodic time series," *Power Systems, IEEE Transactions on*, vol. 20, no. 3, pp. 1622–1630, Aug 2005.

[8] A. Nizar, Z. Dong, and Y. Wang, "Power utility nontechnical loss analysis with extreme learning machine method," *Power Systems, IEEE Transactions on*, vol. 23, no. 3, pp. 946–955, Aug 2008.

[9] J. Cabral, J. Pinto, E. Martins, and A. Pinto, "Fraud detection in high voltage electricity consumers using data mining," in *Transmission and Distribution Conference and Exposition, 2008. T &;D. IEEE/PES*, April 2008, pp. 1–5.

[10] C. Bartusch, M. Odlare, F. Wallin, and L. Wester, "Exploring variance in residential electricity consumption: Household features and building properties," *Applied Energy*, vol. 92, no. 0, pp. 637 – 643, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S030626191100256X

[11] S. Galli, A. Scaglione, and Z. Wang, "For the grid and through the grid: The role of power line communications in the smart grid," *Proc. of the IEEE*, vol. 99, no. 6, pp. 998–1027, June 2011.

[12] T. Hartmann, F. Fouquet, J. Klein, Y. Le Traon, A. Pelov, L. Toutain, and T. Ropitault, "Generating realistic smart grid communication topologies based on real-data," in *Smart Grid Communications (SmartGridComm), 2014 IEEE Int. Conference on*. IEEE, 2014, pp. 428–433.

[13] M. Kristan, D. Skočaj, and A. Leonardis, "Online kernel density estimation for interactive learning," *Image and Vision Computing*, vol. 28, no. 7, pp. 1106–1116, 2010.

[14] M. P. Wand and M. C. Jones, *Kernel smoothing*. Crc Press, 1994.

[15] M. Kristan and A. Leonardis, "Multivariate online kernel density estimation," in *Computer Vision Winter Workshop*, 2010, pp. 77–86.

[16] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," 2011.

[17] Í. Monedero, F. Biscarri, C. León, J. Biscarri, and R. Millán, "Midas: Detection of non-technical losses in electrical consumption using neural networks and statistical techniques," in *Computational Science and Its Applications-ICCSA 2006*. Springer, 2006, pp. 725–734.

[18] Z. Markoc, N. Hlupic, and D. Basch, "Detection of suspicious patterns of energy consumption using neural network trained by generated samples," in *Information Technology Interfaces (ITI), Proceedings of the ITI 2011 33rd Int. Conf. on*. IEEE, 2011, pp. 551–556.

[19] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad, "Nontechnical loss detection for metered customers in power utility using support vector machines," *Power Delivery, IEEE Transactions on*, vol. 25, no. 2, pp. 1162–1171, 2010.

[20] A. Nizar, Z. Dong, and Y. Wang, "Power utility nontechnical loss analysis with extreme learning machine method," *Power Systems, IEEE Transactions on*, vol. 23, no. 3, pp. 946–955, 2008.

[21] J. E. Cabral, J. O. Pinto, E. M. Martins, and A. M. Pinto, "Fraud detection in high voltage electricity consumers using data mining," in *Transmission and Distribution Conference and Exposition, 2008. T&# x00026; D. IEEE/PES*. IEEE, 2008, pp. 1–5.

[22] M. Espinoza, C. Joye, R. Belmans, and B. De Moor, "Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series," *Power Systems, IEEE Transactions on*, vol. 20, no. 3, pp. 1622–1630, 2005.

[23] R. Berthier, W. Sanders, and H. Khurana, "Intrusion detection for advanced metering infrastructures: Requirements and architectural directions," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE Int. Conf. on*, Oct 2010, pp. 350–355.

[24] F. Tabrizi and K. Pattabiraman, "A model-based intrusion detection system for smart meters," in *High-Assurance Systems Engineering (HASE), 2014 IEEE 15th International Symposium on*, Jan 2014, pp. 17–24.