Frank Englert, Marius Rettberg-Päplow, Sebastian Kössler, Alaa Alhamoud, The An Binh Nguyen,
Doreen Böhnstedt, Ralf Steinmetz
Multimedia Communications Lab
TU Darmstadt
Darmstadt, Germany
{firstname.lastname}@kom.tu-darmstadt.de

***Abstract*—In order to find energy saving potentials, future home energy recommender systems needs a large database of historic energy consumption information from various appliances. Having reference data, those systems could decide whether an appliance is wasting energy or not.**

However, the collection of this reference data degrades the user privacy as energy traces contain sensitive information which allows the exhibition of user behavior. In order to mitigate those privacy implications, we propose a method of sparse data collection. Our proposed solution minimizes the amount of collected reference data by removing energy traces which do not provide new information for the recommender system. Our proposed solution is capable of reducing the collected amount of data by a factor of 2 without lowering the accuracy of the future home energy recommender system.

I. Introduction

Recent studies indicate that residential home inhabitants waste a lot of electricity. According to Fischer 2008 [1], there is a saving potential of 20% of the total electricity consumption in residential homes by providing feedback to the inhabitants. The inhabitants achieve their savings by switching off unnecessary consumers, switching down electrical appliances to the most energy efficient working mode or replacing electrical appliances with more energy efficient alternatives. Currently, an analysis which of those actions are beneficial to save electricity must be carried out manually by the inhabitant.

However, in the future we expect the rise of systems which could apply this analysis automatically. Those systems will most probably need a large stock of reference data from various environments and appliances. In order to find saving potentials, those systems will analyze the electricity consumption in comparison with data from similar appliances or environments. As result, those systems could determine the delta energy between the currently observed appliance and the most energy efficient appliance of the same kind in the reference data set.

For example, the system observed that a refrigerator consumes in average 3 kWh electricity per day. As a singular value, this information is not meaningful. But, if the system has a large reference data set obtained from other refrigerators, it could determine whether this value is too high or appropriate. For example, if a similar refrigerator requires only 1.5 kWh per day, the recommender system could recommend the user to replace its old refrigerator with a new one.

In order to build an extensive set of reference data, monitoring a single environment is not sufficient. Rather than that, as many environments as possible must be observed to build a meaningful set of reference data. Thus, we assume the existence of a collaboratively collected set of energy traces at a centralized location, i.e. a cloud back-end. However, this centralized collection of reference data causes serious privacy implications. Recent studies have shown, that energy traces allow inferring the user's occupancy [2], [3], the activity [4] and the number of residents of a monitored home [5]. While these studies only show a small subset of all possible privacy implications, they clearly imply the importance of methods to reduce privacy implications caused by smart metering.

To mitigate those effects, our proposed solution decides for each observed electricity trace individually whether it contains significant information for the reference data set or not. The trace is only transmitted to the cloud back end, if it contains significant information. In order to make this decision, we compare our data depended approach with a random sampling approach as well as a quota based approach as baseline. Hereby, our data depended selection mechanism could filter electricity traces efficiently. In comparison to the random sampling approach, our data depended electricity trace selection mechanism reduced the amount of shared electricity traces by a factor of two without lowering the accuracy of the home energy recommender system.

This work is structured as follows. First, we briefly discuss related works and methods for measuring and increasing the user privacy. Then, we provide an in depth description of our data selection mechanisms. Building on that, we evaluate and discuss relevant characteristics of our data selection mechanisms. Finally, we conclude this paper by summarizing the most important results.

II. Related Work

There are increasing possibilities to monitor the electricity consumption of buildings with a high level of detail [6], [7]. This monitoring is the foundation for sophisticated energy management and energy saving recommender systems. However, detailed metering of the electricity consumption poses several threats for the user privacy. With an increasing amount of collected data, more and more threats on the user privacy become possible. Even with low time resolution, e.g., 15 minutes sampling interval of whole house metering, it becomes possible to infer occupancy and demographic information on the inhabitants [3], [5], [2]. Higher sample rates even allow to infer present electrical appliances [8], user activities and common user patterns [4], [9]. An illustrative example of those privacy threads is the possibility to infer the TV program from the electricity consumption of a television. As shown by [10], having time stamped secondly electricity metering data of a television, it becomes possible to deduct the exact TV program watched by the user of a television.

In our opinion, relying on cryptographic methods or anonymizing routing protocols (c.f.[11], [12]) for protecting the user privacy is no viable solution to this problem. While those methods are important factors to decouple the personal references from the collected data, cryptographic methods do not reduce the amount of privacy threatening data. It still remains possible for the data collector to re-use the collected data for other purposes. Thus, we will focus on methods to reduce the amount of collected data to the bare minimum.

Three main challenges [13] must be solved in order to enhance the user privacy. First, a solution is needed to make the privacy measurable. Next, good trade-off between data quality and privacy must be found. Last but not least, the user should have the possibility to fine-tune privacy relevant settings in order to maintain a high perceived privacy. In order to enhance the user privacy, measurements should be able to incorporate the achieved level of privacy. Kalogridis [11] expresses three metrics which could be used to quantify the privacy level. Namely, those metrics are either based on the entropy, cluster analysis or regression. However, as those metrics are tailored to whole household energy measurements, they are not applicable in our use case with multiple data streams from different appliances. To increase the user privacy of smart meter installations, many researchers [14], [15], [11] recommend the installation of buffer batteries to average variances in the electricity consumption of a domestic home. By installing buffer batteries between the electrical consumers and the smart meter, the real power draw can be decoupled from the observable power draw. Thus, it becomes possible to smoothen or reshape the electricity consumption such that the privacy implications are reduced. Furthermore, those systems could be used to decrease the electricity bill by shifting the electricity demand to low-price time windows [15]. A more general analysis of methods for finding a good trade-off between privacy and data quality was done by Reinhardt et al. [16]. This work focused on general methods to enhance the user privacy by preprocessing the smart meter data. More specifically, four different privacy enhancing algorithms were used to determine their effect on the user privacy as well as on the data quality.

III. Energy Recommender Systems

Figure 1 shows the architecture of a home energy recommender system. Those systems measure the electricity consumption of all present electrical devices on appliance level. This is achieved either by designated sensor nodes or by using non-intrusive appliance load monitoring approaches[17]. The collected electricity consumption stream is then forwarded to the Recommender Algorithm which compares the observed electricity consumption for each appliance with previously recorded electricity consumptions to find saving potentials. As result, those algorithms could rank the observed electricity consumption with other appliances of the same kind or useful for the same purpose. Depending on the kind of recommendation, different data granularity levels are required. Those are shown in Table I. E.g. for comparing the total electricity consumption with the electricity consumption of friends or similar households, aggregated data with a time resolution in the range of multiple hours can be used. With increasing detail level, finer granularity levels of data are required to derive those recommendations. Usually, home energy recommender systems consist of two parts. Firstly, there are sensors and a gateway node physically installed in each residential home. Secondly, a centralized server ("cloud back-end") collects reference data sets from all connected home installations. In order to keep the reference data set up to date, the collected electricity traces are uploaded to a cloud monitoring. The local part of the system is assumed to be under full control of the user and thus poses no privacy implications. As a consequence, our work aims for limiting the privacy implications caused by the data uploaded to the cloud back end.

TABLE I. Data granularity requirements for different kinds of feedback

Kind of Feedback	Aggregated Data	Time Resolution	Reference Data
Total Consumption	yes	hours. . . days	yes
Device Level Consumption	no	hours. . . days	no
Highlight standby appliances	no	minutes	no
Recommend device replacement	no	minutes	yes
Recommend better operating mode	no	seconds	yes
Recommend device maintenance	no	seconds	yes

One of the most prominent home energy recommender system was implemented by Plugwise [1]. This company offers Zigbee based smart meters which could be plugged between the wall outlet and arbitrary electrical appliances. The main purpose of Plugwise sensors is to provide user feedback about the electricity consumption. However, the Plugwise software is also capable of uploading the observed readings to a cloud back end [2]. Currently, Plugwise is capable of giving feedback to the users which and how much electricity is consumed by present appliances. However, Plugwise announced plans, to build more detailed energy feedback systems based on the collected electricity information in the near future.

IV. Approaches for Data Minimization

In the last section, we described the general working principle of home energy recommender systems. Building upon that, we now describe our contribution to improve the user privacy of those systems. To do so, our proposed solutions extend the Data Uploader with a selection mechanism to decide whether a particular electricity trace from a particular appliance together with its corresponding appliance label should be uploaded or not. This selection mechanism takes an electricity trace and

[1] http://www.plugwise.com

[2] Side-note: Early versions of the Plugwise Source software uploaded all observed electricity readings together with the device names regardless if the user disabled this feature or not. Due to serious privacy concerns in their user community, Plugwise decided to stop this behavior.

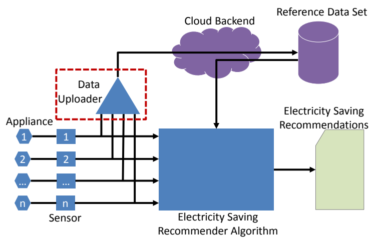

Fig. 1. Architecture of a future home energy recommender system

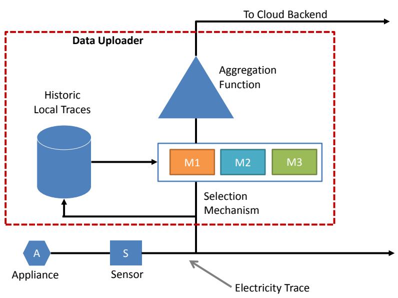

Fig. 2. Extended system architecture with our selection mechanisms.

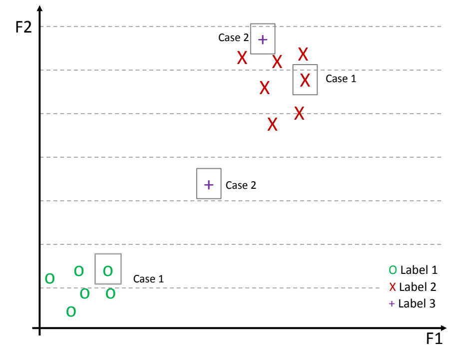

Fig. 3. Exemplary visualization of the Data Depended heuristic for novel information discovery

a locally available subset of the reference data set as input and makes a binary decision if a electricity trace should be uploaded. An electricity trace in this context is a time series of power readings for a particular appliance. This time series starts when the device is switched on and stops when the device is switched off. With other words, the start event of this time series is indicated by a power reading above 0W and the stop event occurs when the power reading of the device under measure goes down to 0W again. For book-keeping purposes, all electricity traces are stored in a local data set with historic electricity traces. It is important to mention, that our enhanced data uploader removes the time relation of the obtained energy traces. Furthermore, it defers the upload process to a hard-coded time window which is the same for all instances. The time relation of electricity traces is most probably not required for energy analysis. Thus, this information should not be collected at all if respecting the user's privacy is required. Having given a high level overview, it is time to describe the core working principle of our solution. In the next subsections, we will describe the implementation of our three data selection mechanisms.

A. Random Sampling

The general idea of the random sampling selection method is rather simple. The decision, whether an electricity trace should be uploaded or not is based on a random variable. Technically spoken, our algorithm draws a number from an equally distributed random number generator. If the value of the number is above a certain threshold, the electricity trace is uploaded to the cloud back end. Otherwise, the electricity trace is dropped from the upload queue.

The dropping probability of this selection method could either be constant or changing over time. This raises the question how to determine a good function to approximate the dropping probability. If the dropping probability is too low, there is no enhancement of the user privacy. On the other hand, if the dropping probability is too high, it affects the accuracy of the energy saving recommender. A method for finding sensible values for the dropping probability is explained in section V. However, as each user might have different privacy preferences, this mechanism allows selecting arbitrary dropping probabilities.

B. Quota Based Sampling

As a core working principle, the quota based sampling mechanism assigns a quota of n available uploads to each appliance. With other words, for each appliance at most n electricity traces are uploaded to the cloud back-end. In contrast to *Random Sampling*, this mechanism has a hard limit of electricity traces which are shared with other users. Our implementation has no sophisticated methods for determining which electricity traces to upload. The algorithm uploads all available electricity traces until the quota for the particular appliance is exhausted.

However, a recent study [18] has shown that electric appliances have a huge set of different electricity consumption patterns. This fact raises the question, if there are more sophisticated approaches to select samples for uploading.

C. Data Depended Sampling

This mechanism tries to determine whether an electricity trace contains novel information with relevance for the reference data set. As the novelty of an electricity trace could not be measured directly, this mechanism tries to estimate the novelty.

This mechanism assumes the availability of labels for each electricity trace. If the system is aware of device names for all monitored appliances, the device names could be used as label. Otherwise, the sensor id might be used as label.

To estimate the novelty of a data sample, the following heuristic is used: First, our algorithm extracts certain features from the electricity trace to analyze as well as from all available historic electricity traces which are stored locally.

Next, it uses the k-Nearest-Neighbors (kNN) algorithm to find k historic traces which are most similar to the current trace. Finally, our algorithm compares the labels of the k historic traces with the label of our current trace. If the labels are equal, the trace contains no novel information. If the labels are different, the trace might contain novel information. Unless stated otherwise, our implementation uses a Euclidean distance metric for finding the nearest neighbors.

This working principle is shown in Figure 3. In this example, we select two neighbors for comparison: $k = 2$. In case 1, all neighbors have the same label. Thus, the electricity trace will most probably contain no novel information and uploading is not required. Things look different in case 2. Here, the nearest neighbors have different labels. The reason for this might be a formerly unobserved behavior of the appliance or the observation of a new appliance.

Currently, our implementation relies on the following features: (1) The total energy of the electricity trace, (2) the length of the electricity trace, (3) the maximum observed power, (4) the average power, (5) a finger print extracted from the inrush power demand.

The supervised machine learning mechanism works best, if a certain amount of historic data from many different appliances is available. Only if a broad spectrum of different features is available in the search space, the algorithm can decide whether a sample is relevant or not. Nevertheless, this mechanism has no possibility to specify a user defined trade-off between data quality and privacy.

V. Evaluation

We evaluate the efficacy of our proposed data minimization approaches in this section. First, we describe general assumption we made, in order to test our mechanisms. Next we describe the concrete evaluation setup. Finally, we demonstrate the effects of our approaches on the user privacy in our particular scenario.

The evaluation of our selection mechanisms is challenging. The reasons for this fact are twofold. First, there is no single measure which could be used to measure the privacy. To work around this issue, we assume a negative correlation between the amount of uploaded data and the user privacy. Thus, from a privacy point of view, it is best to upload as few electricity traces as possible. The second challenge is due to the fact that no ground-truth data for home energy saving recommender systems is available. Thus, the demanded amount of reference data for home energy recommender systems is still unknown and we could not directly measure the impact of our data minimization approach on the accuracy of those systems. In order to address this issue, we use a rather limited approach to generate energy recommendations. First, we use a device identification algorithm [19] to cross-check the label of the electricity trace. If the expected and the classified label match, we search for an instance with the same label and energy consumption lower than the energy consumption of the actual electricity trace instance. If such an instance exists, our algorithm could have made an energy saving recommendation. Next, we count all cases where a recommendation was made and all cases where no recommendation was possible. Building upon that, we define the accuracy of our recommender system as the number of recommendations found divided by the total number of all recommendations:

$$a = \frac{count(r_{possible})}{count(r_{all})} \tag{1}$$

Using this metric, we could indicate whether the energy saving recommendation system is capable of finding energy wastage. However, this metric does not express the height of the saving potential. A reliable set of ground truth data would be required in order to apply such a metric.

A. Evaluation Setup

We used the publicly available Tracebase [19] data set to evaluate our data minimization algorithms. This data set consists of energy traces from different common household appliances which were collected in different residential homes. More precisely, this data set consists of 1,270 power consumption records from 122 different appliances grouped in 31 different classes. Each of those power consumption records contains power measurements for a single appliance with a sample rate of roughly 1Hz over a period of 24h. Those records were collected using Plugwise Circles. Prior to the evaluation, we extracted 8,200 electricity traces from those power consumption records. According to our definition in Section III, each electricity starts when an appliance is switched on and stops when the appliance is switched off again.

B. Trade-off between Data Reduction and Accuracy Loss

The goal of this experiment is to determine the trade-off between data reduction and accuracy. Good selection mechanisms should reduce the number of uploaded electricity traces as much as possible without reducing the recommendation accuracy. We test the effects of our selection mechanisms on the data reduction as well as on the recommendation accuracy in this experiment.

The setup of this experiment is as follows: We use the full set of 8,200 electricity traces as input for our selection mechanisms. The selection mechanism then decides for each input data sample whether it is important or not. Finally, we use all selected electricity traces to build an energy recommender and evaluate its performance as described in Section V-A. As the obtained results may depend on random variables or the order of the input data, we repeated this whole procedure 30 times.

The result of this experiment is shown in Figure 4. The x-axis of the plot shows the resulting data reduction and the y-axis shows the accuracy-loss. To evaluate *Random Sampling* mechanism, we varied the selection probability p from $p = 0.05$ to $p = 0.9$. For high selection probabilities above 60%, this mechanism shows solid results with a low accuracy loss. For lower selection probabilities, the mechanism has a significant decrease of the accuracy. This effect is caused by an inappropriate selection of relevant electricity traces. Obviously, the resulting data reduction directly depends on the selection probability. To evaluate *Quota based Sampling*, we varied the quota q from 20 samples per class to 1,000 samples per class. However, this mechanism performs worse than the *Random*

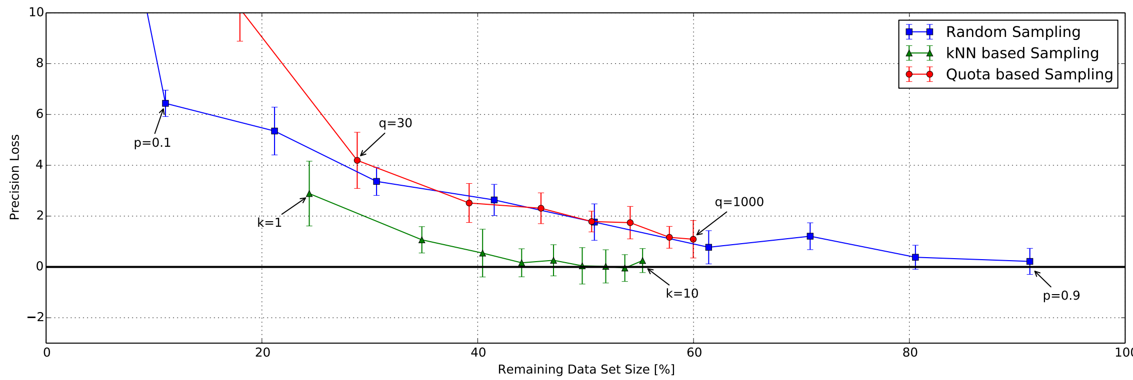

Fig. 4. The Data Depended Sampling provides a better trade-off between data reduction and accuracy loss than Random Sampling or Quota based Sampling.

Sampling for quotas below $q = 30$. For a quota values above $q = 30$, the mechanism converges to $RandomSampling$. For the *Data Depended Sampling* we varied k from 1 to 10. This selection mechanism performs significantly better than random sampling. This mechanism can achieve high data reductions with very small accuracy losses. Furthermore, values of k above 4 were not beneficial at all. For all values of k above 4, there was nearly no clear accuracy-loss observable. However, this mechanism cannot achieve arbitrary low data reductions. The lowest possible remaining data set size of 25% was achieved with $k = 1$.

In this experiment, the *Quota based Sampling* provides no benefits over the *Random Sampling* mechanism. It had equal or even worse results and is slightly more complex to configure. In this experiment, the *Data Depended Sampling* mechanism performed best. It was capable of reducing the resulting data set size by a factor of 2 without significant effects on the accuracy.

C. Time Dependence of Selection Mechanisms

While observing macroscopic effects of our selection mechanisms in the first experiment, we now analyze the dropping probability over time. In terms of increasing the user privacy, the dropping probability of selection mechanisms should increase to 100% over time. When a system is newly installed, it is perfectly ok, if the selection mechanism finds new electricity traces to upload. However, when the system is used for a certain time window and no new appliances are present, the dropping probability should rise to 100%.

In order to evaluate this behavior, we used all electricity traces from our test data set and observed the selection decisions based on the number of previously available reference data points. The results of this experiment are shown in Figure 5. As expected, the *Random Sampling* has a constant dropping probability. The dropping probability of the *Data Depended Sampling* mechanism starts very low but rises quickly with an increasing set of reference data. However, during our experiments, it never reached a point where no new information is discovered. The reason for this may be a limited test data set observed in diverse environments. Nevertheless, the *Data Depended Sampling* mechanism achieved dropping probabilities of up to 80%. In this experiment, the *Quota based*

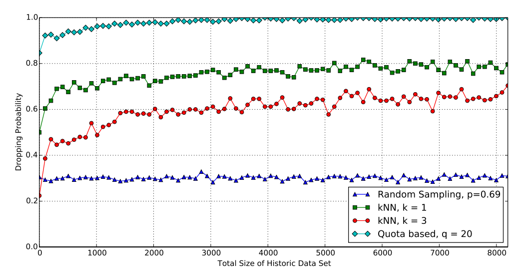

Fig. 5. The dropping probability of our Data Depended selection mechanism increases over time up to 80%.

Sampling mechanism worked best. With a low quota of 20 electricity traces per class, this mechanism could achieve very high dropping probabilities.

VI. Discussion and Future Works

In this section we will discuss our findings and their impact on energy saving recommendation systems. First, we will discuss the appropriate data granularity and then we will review means to enable privacy control for the user. Finally, we discuss how our approaches can be generalized to other problem domains.

In general, the data granularity should be as low as possible. A lower data granularity causes lower information content and thus fewer possibilities for privacy threats. This raises the question on the lowest possible data granularity for energy saving recommender systems. Unfortunately, there is no simple answer to this question. If the recommender system should be able to give appliance level recommendations, the total energy as well as the usage time for each appliance is sufficient in order to generate recommendations. If the recommender system should be able of detecting and recommending appropriate operating modes of an electrical appliance, much more data will be needed.

For the end user, the perceived privacy is very important. Only if the user feels comfortable with the behavior and capabilities of a sensing system, he starts to accept the system. Thus, it is important that the user has the possibility to fine-tune his preferences for privacy protection. As shown in Figure

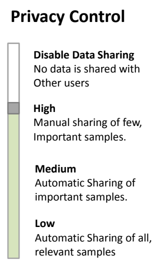

Fig. 6. Mockup user interface for selecting an appropriate personal privacy protection level.

6, we propose four different levels for privacy control:

1) *Disable sharing*: No electricity traces are uploaded to the cloud back-end
2) *High*: No data sample is uploaded automatically. The system selects the most relevant electricity traces. Those traces are presented to the user for manual review and only uploaded if the user acknowledges this.
3) *Medium*: The system selects the most relevant electricity traces and uploads them automatically
4) *Low*: The system selects all relevant electricity traces and uploads them automatically.

We plan to implement this privacy control in our data collection toolbox. The different privacy protection levels could be achieved by using Data Dependent Sampling for the Low privacy protection level. The Data Dependent Sampling with $k = 4$ combines a data reduction by a factor of 2 with nearly no accuracy loss. The Medium privacy protection level could be realized by combining Data Dependent Sampling for the selection of relevant data together with a Quota based selection mechanism for limiting the total amount of transmitted data. The high privacy protection level could be achieved by extending the medium setting with means for user review before the data is uploaded to the cloud back-end.

Finally, we discuss how our selection mechanisms for finding transmission relevant data could be generalized to be applicable in other, similar application domains. Our proposed solution can be used directly in other problem domains where a selection of labeled data should be made. However, most probably our presented relevance metric is not re-usable. A domain specific relevance metric or a more general privacy metric is needed in order evaluate the performance of our selection mechanisms.

VII. Conclusion

In this paper we analyzed different selection mechanisms for transmission relevant, privacy sensitive electricity traces. We compared three different selection mechanisms *Random Sampling*, *Quota based Sampling*, and *Data Dependent Sampling*. These sampling mechanisms can reduce the amount of collected data by a factor of 2 without significant effects on the accuracy of electricity saving recommender systems. Furthermore, we showed options how a user could set up the selection mechanisms in order to match custom privacy requirements. Thus, our work greatly enhances the perceived privacy of home electricity saving recommender systems. This is an important step to foster user acceptance for those systems.

References

[1] C. Fischer, “Feedback on household electricity consumption: a tool for saving energy?” *Energy Efficiency*, vol. 1, no. 1, pp. 79–104, May 2008.

[2] A. Ebadat, G. Bottegal, D. Varagnolo, B. Wahlberg, and K. H. Johansson, “Estimation of building occupancy levels through environmental signals deconvolution,” in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings - BuildSys’13*. New York, New York, USA: ACM Press, 2013, pp. 1–8.

[3] W. Kleiminger, C. Beckel, T. Staake, and S. Santini, “Occupancy Detection from Electricity Consumption Data,” *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings - BuildSys’13*, pp. 1–8, 2013.

[4] A. Alhamoud, F. Ruettiger, A. Reinhardt, F. Englert, D. Burgstahler, D. Böhnstedt, C. Gottron, and R. Steinmetz, “SMARTENERGY . KOM : An Intelligent System for Energy Saving in Smart Home,” in *The 39th IEEE Conference on Local Computer Networks*, 2014.

[5] C. Beckel, L. Sadamori, and S. Santini, “Automatic Socio-Economic Classification of Households Using Electricity Consumption Data Categories and Subject Descriptors,” in *e-Energy*, 2013, pp. 75–86.

[6] X. Jiang, P. Dutta, and D. Culler, “Design and implementation of a high-fidelity AC metering network.” in *Information Processing in Sensor Networks*, 2009, pp. 253 – 264.

[7] D. Phillips, R. Tan, M.-M. Moazzami, G. Xing, J. Chen, and D. Yau, “Supero: A Sensor System for Unsupervised Residential Power Usage Monitoring,” in *PerCom*. San Diego: IEEE, 2013.

[8] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava, “NILMTK : An Open Source Toolkit for Non-intrusive Load Monitoring,” in *e-Energy*, 2014.

[9] M. Milenkovic and O. Amft, “An opportunistic activity-sensing approach to save energy in office buildings,” in *e-Energy*. New York, NY, USA: ACM, pp. 247–258.

[10] U. Greveler, B. Justus, and D. Loehr, “Multimedia content identification through smart meter power usage profiles,” in *Computers, Privacy and Data Protection*, 2012.

[11] G. Kalogridis and C. Efthymiou, “Privacy for smart meters: Towards undetectable appliance load signatures,” in *SmartGridComm*. Gaithersburg, MD: IEEE, 2010, pp. 232–237.

[12] M. Reed, P. Syverson, and D. Goldschlag, “Anonymous connections and onion routing,” *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, pp. 482–494, May 1998.

[13] D. Christin, A. Reinhardt, S. S. Kanhere, and M. Hollick, “A survey on privacy in mobile participatory sensing applications,” *Journal of Systems and Software*, vol. 84, no. 11, pp. 1928–1946, Nov. 2011.

[14] D. Chen, D. Irwin, P. Shenoy, and J. Albrecht, “Combined Heat and Privacy : Preventing Occupancy Detection from Smart Meters,” in *PerCom*, 2014.

[15] L. Yang, X. Chen, J. Zhang, H. V. Poor, and A. Motivation, “Optimal Privacy-Preserving Energy Management for Smart Meters,” in *InfoCom*, 2014.

[16] A. Reinhardt, F. Englert, and D. Christin, “Enhancing User Privacy by Preprocessing Distributed Smart Meter Data,” in *SustainIT 2013*, 2013.

[17] D. Egarter, A. Sobe, and W. Elmenreich, “Evolving Non-Intrusive Load Monitoring,” in *EvoApplications’13*, 2013, pp. 182–191.

[18] J. Kolter and M. Johnson, “REDD: A public data set for energy disaggregation research,” in *ACM Special Interest Group on Knowledge Discovery and Data Mining, workshop on Data Mining Applications in Sustainability*, San Diego, 2011.

[19] A. Reinhardt, P. Baumann, D. Burgstahler, M. Hollick, H. Chonov, M. Werner, and R. Steinmetz, “On the Accuracy of Appliance Identification Based on Distributed Load Metering Data,” in *Sustainable Internet and ICT for Sustainability*, Pisa, 2012, pp. 1–9.