De-pseudonymization of Smart Metering Data: Analysis and Countermeasures

DOI:
10.1109/giots.2018.8534430

Document Version
Accepted author manuscript

Published in:
IEEE Workshop on Industrial Internet of Things Security 2018

De-pseudonymization of Smart Metering Data: Analysis and Countermeasures

Sara Cleemput, Mustafa A. Mustafa, Eduard Marin, and Bart Preneel
imec-COSIC, KU Leuven, Belgium
E-mail: firstname.lastname@esat.kuleuven.be

***Abstract*—Fine-grained metering data threatens users' privacy, as it typically reveals the users' consumption patterns and thereby their behaviour. To address this problem, the use of pseudonyms when sending such fine-grained data has been proposed in the literature. In this paper, we demonstrate experimentally that an adversary who has access to pseudonymized fine-grained data and identifiable billing data can fully de-pseudonymize all users using a simple matching algorithm. Our experiments use real-world metering data collected from ca. 6500 smart meters. As pseudonymization alone is not sufficient to provide privacy, we propose three simple yet effective countermeasures against de-pseudonymization: deliberately not reporting some of the fine-grained metering values, rounding these values before reporting them and regularly changing the pseudonyms. We experimentally demonstrate that our countermeasures considerably improve users' privacy protection without significantly lowering the usefulness of the data. They also do not affect the billing process.**

I. Introduction

The Smart Grid (SG) is an extension of the traditional electrical grid. It supports bidirectional electricity and data flows between its components and entities, with the aim to increase the reliability and efficiency of the grid. One aspect of the SG is that every house will have a Smart Meter (SM) that is capable of measuring and reporting electricity consumption data very frequently, e.g. every 30 minutes. These data can help operators manage the grid more efficiently, and suppliers generate accurate bills for their customers more timely.

However, such fine-grained metering data may pose serious risks on users' privacy [1]. Entities with access to these data (e.g. suppliers) might use non-intrusive load monitoring techniques to infer users' consumption patterns [2]. These patterns can then be used by these entities to infer private information about the users [3], [4] such as their daily schedule, the appliances being used, whether they are at home, when and even which TV channel they are watching [5]. Therefore, such fine-grained metering data is considered highly sensitive. In 2009 the Dutch Senate even rejected a law mandating the use of SMs, based on the right to privacy [6].

Therefore, appropriate measures should be taken to protect users' privacy when processing fine-grained consumption data. One possibility is to have SMs use pseudonyms instead of their real IDs when reporting their fine-grained metering data to the supplier [7]. However, past work has shown that partial de-pseudonymization of the data, i.e. discovering the SM (user) corresponding to a pseudonym, is possible by using statistical measures, as well as additional side-channel information [8]–[11]. Unlike our analysis, these articles assume that the adversary uses complex de-pseudonymization algorithms which are trained with users' fine-grained metering data. In addition, none of these algorithms consider a realistic adversary, i.e. one that has access to both the pseudonymized fine-grained metering data and the monthly aggregate data used for billing.

The main contributions of this paper are twofold:

- It investigates the feasibility of attacks that require only simple techniques to de-pseudonymize users. More specifically, it demonstrates, using a real-world dataset, that an adversary with access to pseudonymized fine-grained data and attributable monthly aggregates, can fully de-pseudonymize users' fine-grained metering data using a simple matching algorithm.
- It proposes and experimentally verifies three simple but effective countermeasures against de-pseudonymization: each SM (i) deliberately omits reporting some of its fine-grained metering data, (ii) reports rounded metering data, or (iii) uses more than one pseudonym per billing period. These countermeasures can all be adopted without any major changes to the smart metering architecture.

The rest of this paper is organised as follows: Section II discusses related work. Section III describes our methodology and the de-pseudonymization process, and proposes three countermeasures. Section IV presents our results which are further discussed in Section V. Section VI concludes the paper.

II. Related Work

Efthymiou and Kalogridis proposed a solution for anonymizing users' metering data [7] based on using two different IDs per SM: (i) the user ID which is used by the SM to report metering data used for billing purposes, and (ii) a high frequency ID, which the supplier cannot link to the user ID and which is used by the SM to report fine-grained metering data. The link between both IDs is known only to a trusted escrow party, such that the supplier is unable to link the pseudonymized metering data to the users.

However, several papers have already shown that partial de-pseudonymization of fine-grained metering data is possible. Jawurek et al. proposed two attack strategies to de-pseudonymize users' metering data: anomaly detection and behaviour pattern matching [8]. Furthermore the authors attempted to link metering data of users stored on two different databases with different pseudonyms. Their algorithm is trained on one of the databases and tested on the other one, and it achieves 83% accuracy in linking the data of the same user in both databases.

However, they are able only to link the two pseudonyms of a user, not to de-pseudonymize the user. Buchmann et al. showed that it is possible to de-pseudonymize users by using machine learning algorithms [9]. They first trained their algorithm using the metering data of known households and extract features for each household. Subsequently, they executed the algorithm on metering data from the same households at a different time period and then tried to find a match between the features extracted during the two time periods. They showed that their algorithm de-pseudonymizes 68% of the 36 households. Tudor et al. proposed an improved version of this algorithm, where they use only five features instead of twelve [10]. They also showed that combinations of different features give different success rates for the de-pseudonymization process. On average, their method outperforms Buchmann's algorithm by 10%. Faisal et al. demonstrated that industrial consumers are easier to re-identify than residential consumers, and concluded that having longer periods of data to train re-identification algorithms is more useful than having high resolution data [11].

A common limitation of the aforementioned work is the assumption that the adversary has access to some of the users' fine-grained metering data, which he uses to train his de-pseudonymization algorithm. Unlike them, we only assume that the adversary has access to users' aggregate monthly consumption data for billing calculations, which is a more realistic assumption. Similar to our paper, Tudor et al. [12] analysed the ability of an adversary to de-pseudonymize users' fine-grained metering data. However, in their analysis SMs report rounded billing values from 1 kWh resolution up to 200 kWh resolution. Unlike Tudor et al. we analyse the case where a supplier obtains users' exact billing data.

III. Methodology

We use real-life consumption data to investigate whether full de-pseudonymization is possible. Next, we propose three countermeasures that improve users' privacy while keeping the usefulness of the data at an acceptable level.

A. Problem Description and Adversarial Model

We study the following use case [7]: for each SM the supplier receives (i) the monthly aggregate, i.e. the overall electricity consumption during that month, coupled to the SM ID, and (ii) all pseudonymized half-hourly consumption data. The latter allows the supplier to create consumption profiles used to purchase electricity on the wholesale market. However, the naive assumption is that the supplier cannot match the pseudonymized data to a specific user, since a priori it does not know which pseudonym corresponds to which SM. Note that only the monthly aggregate is used for billing, hence any modification of the fine-grained metering data does not affect the billing process. This specific set-up will be used in practice for the majority of electricity consumers in the UK [13]. In this paper we assume that the billing period is one month.

Our main goal is to design countermeasures that can be implemented without any major changes to the smart metering architecture and without incurring any substantial overhead, e.g. additional layers of encryption, as this will be computationally heavy for SMs.

The SM itself is considered as a trusted entity, since we assume it is tamper-proof. We consider the supplier as an *honest-but-curious adversary* that follows the protocols correctly, but tries to extract additional information from the different data it receives. Moreover, as the smart metering setting in the UK [13], the supplier has access to both the pseudonymized fine-grained metering data and the attributable monthly aggregate data of all of its consumers.

B. Data Set

Our analysis is based on a real-life dataset, "Electricity Customer Behaviour Trial" [14], that contains 6435 unique users' consumption data, collected at 30-minute intervals from 14th of July 2009 up to 31st of December 2010. To the best of our knowledge this is the largest publicly available data set containing fine-grained electricity consumption data over a period of several months and it has already been used in previous work [15]–[17]. Moreover, as a supplier will usually have some information as to which region the fine-grained consumption data are originating from, the size of the dataset seems sufficiently realistic.

The dataset contains a total of 157,992,996 meter readings. For each reading, the SM ID, the time stamp and the consumption[1] during the 30-minute interval are given. When analysing the data, we found that there are 102,747 SM-month combinations for which all consumption data are present. Since we will use the monthly aggregate to de-pseudonymize the users, we only consider those cases where we have complete data for that user during that month.

C. Experiments

We define the privacy metric as the percentage of users for whom a supplier can match their half-hourly consumption data to their monthly aggregate consumption data and therefore to their unique ID.

1) De-pseudonymization Method: The first step consists of analysing the monthly aggregates from the point of view of the adversary. We start by looking at August 2009, the first full month for which we have measurements. We first check how many monthly aggregates are unique values, i.e. no two users have the same monthly aggregate consumption. For each of the users with a unique monthly aggregate, we know that the adversary can immediately de-pseudonymize them, since exactly one of the sums of half-hourly values will match this aggregate. In the next step, we look at the second month, i.e. September 2009. We consider only those users who have not yet been de-pseudonymized, i.e. users that either had no complete data for August 2009 or a non-unique monthly aggregate for August 2009. Following the same approach, we then try to de-pseudonymize this new set of users. We keep repeating this process until all users are de-pseudonymized. Note that, in contrast to [12] where they use rounded fine-grained data, we use real (unmodified) data.

[1] In our dataset, the resolution of the users' consumption data is 10^{-4} kWh.

Our hypothesis is that the supplier will be able to de-pseudonymize most (if not all) users, as we expect them to have unique monthly aggregates (see Section IV-A for the results). Next we propose three countermeasures against this process.

2) Countermeasure 1 – Missing Data: We propose that SMs omit reporting a certain amount of half-hourly consumption values. For each SM-month combination we randomly discard a certain fraction of the consumption data. The adversary can follow two different strategies to reduce the effectiveness of our countermeasure: replace these omitted values (i) by zero, or (ii) by the average of the two values surrounding the discarded value. For each of these two cases, we assess the improvement in privacy by checking which percentage of the users can still be matched to their half-hourly consumption data. We compare the percentage of successful matches for different percentages of values being omitted.

Since the fraction of discarded data will be small and each SM chooses when to discard data independently, we assume the usefulness of the data will decrease only slightly. We verify this by computing the consumption per half-hourly period, aggregated over all users, and comparing this to the original half-hourly aggregate. Taking into account that most grid management is based on aggregates over a neighbourhood, this is a relevant measure for usefulness. Recall that billing is done using the attributable monthly aggregates, thus our countermeasures, working on the fine-grained data, do not influence the billing process.

3) Countermeasure 2 – Rounded Data: We propose that SMs round all the half-hourly consumption values before reporting them. As before, the improvement in privacy is measured by attempting to match the sets of half-hourly data to the monthly aggregates, and we compare the percentage of successful matches for different rounding thresholds. As with the previous countermeasure, we verified that the effect on the usefulness of the data is small by computing the half-hourly aggregate after rounding the values and compared it with the original one (i.e. with no rounding).

4) Countermeasure 3 – Different Pseudonyms: Our final countermeasure consists of SMs changing their pseudonym after a certain period of time. We assess the improvement in privacy by checking which percentage of the users can still be de-pseudonymized when using a pseudonym that is only valid for one month and half a month, respectively. For this, we check which combinations of two sums, one belonging to the first half of the month and one belonging to the second half of the month, match one of the monthly aggregates.

IV. Results

This section presents the results of the de-pseudonymization process both without and with our proposed countermeasures.

A. Results without Countermeasures

We have investigated the time (i.e. the number of billing cycles) required to de-pseudonymize all users present in the data set, using the method described in Section III-C1. The results are shown in Table I. The rows of this table are the first four months. The second column gives the total number of SMs that have a complete half-hourly dataset for that month. In the third column, the anonymization set is given, i.e. the number of SMs that was not yet de-pseudonymized in any of the previous months. The number of SMs that can be de-pseudonymized during the month in question is given in the fourth column. Finally, the fifth column shows the total percentage of SMs that has already been successfully de-pseudonymized. As can be seen from these results, almost all users (99.89%) are de-pseudonymized after only one month, thus confirming our hypothesis that users have unique monthly aggregates. For every month from month four onwards, the adversary can immediately de-pseudonymize every new user that is added to the dataset.

TABLE I: De-pseudonymization without Countermeasures.

Month	Total nb	Anon. set	De-pseudonymized	% Successful
Aug 09	6282	6282	6275	99.89%
Sep 09	6297	56	51	99.92%
Oct 09	6274	14	11	99.95%
Nov 09	6255	6	6	100 %

B. Results with Missing Data

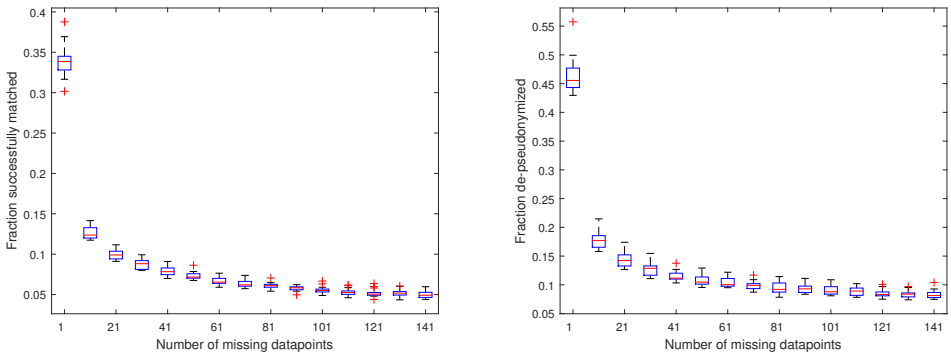
We now describe the de-pseudonymization results when the countermeasure *Missing Data*, detailed in Section III-C2, is implemented. For each month, we considered the set of all SMs for which we have complete data. We first replace a certain amount of the - on average - 1463 half-hourly values in each month by zero, and then attempt to match sets of half-hourly values to monthly aggregates by sorting both the monthly aggregates and the sums. We assume that the smallest sum corresponds to the smallest monthly aggregate etc. Thus, we can compute the fraction of users that was matched (i.e. de-pseudonymized) successfully each month. We finally average our results over all months.

Figure 1a shows the obtained results when replacing the missing values by zero. When omitting only a single data point per SM, on average only 33.95% of the users can be de-pseudonymized. Leaving out 21 datapoints, only 9.95% can be de-pseudonymized. Leaving out 141 datapoints - which corresponds to about 10% of the total number of data points - the adversary can de-pseudonymize less than 5%. Next, we run the same experiment, but instead of replacing the missing datapoints by zero, we replace them by the average of the two values surrounding them. Figure 1b shows the results. As expected, the success rate of the adversary improves as the average value is a better approximation of the missing value.

We also investigated the usefulness of the data. For each number of missing datapoints, we calculate the difference between the new and the original aggregate consumption relative to the original aggregate consumption:

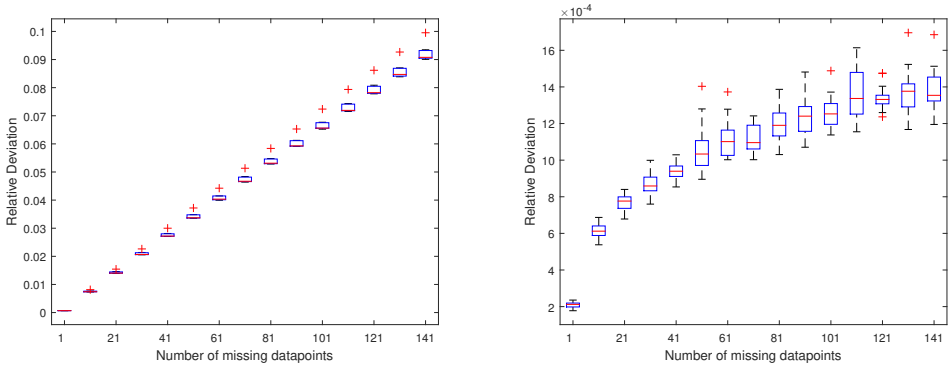
$$deviation_i = \frac{|\sum_t x_{i,t} - \sum_t x'_{i,t}|}{\sum_t x_i}, \tag{1}$$

where $x_{i,t}$ is the consumption of user i at time t and $x'_{i,t}$ is the consumption of user i at time t, with a few



(a) (b)

Fig. 1: Percentage of de-pseudonymized users when the adversary replaces one or more datapoints by (a) zero, or (b) the average of the two values surrounding it.



(a) (b)

Fig. 2: Average relative deviation of the aggregate consumption after replacing some datapoints by (a) zero, and (b) the average of the two values surrounding it vs. the original consumption data.

datapoints replaced by either zero or the average of the values surrounding them. The lower the deviation, the higher the usefulness of the data.

Figure 2a depicts the deviation in function of the number of missing datapoints, when replacing by zero. The relative deviation stays lower than 10%, even when omitting 141 data points. Figure 2b shows this deviation, when replacing by the average of the surrounding values. In this case, the deviation remains extremely small, even when omitting 141 data points. Again this is due to the fact that the average is a better approximation for the missing data point.

C. Results with Rounded Data

We now describe the de-pseudonymization results when the countermeasure *Rounded Data*, detailed in Section III-C3, is implemented. Our approach is similar to the one for the results with missing data, but instead of leaving out data points, we round all data points up to a certain threshold.

Figure 3 shows the percentage of users for which the adversary can still match their half-hourly consumption to their monthly aggregate consumption (i.e. to their ID), in function

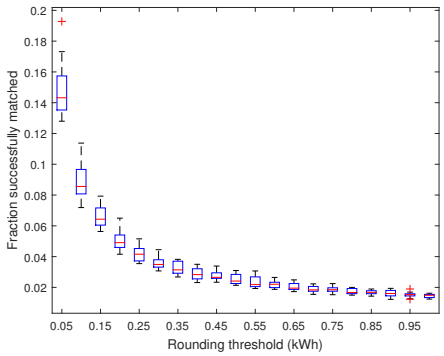


Fig. 3: Percentage of users that can be de-pseudonymized when using different rounding thresholds.

of the rounding threshold. Even with a rounding threshold as small as 0.05 kWh, the adversary can only de-pseudonymize 14.83% of the users. When rounding up to 0.7 kWh, on average less than 2% of the users can be de-pseudonymized.

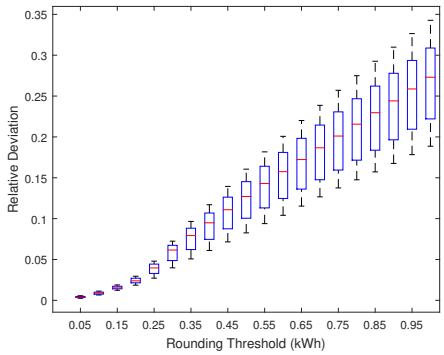


Fig. 4: Average relative deviation of the aggregate consumption after rounding the data vs. before rounding the data.

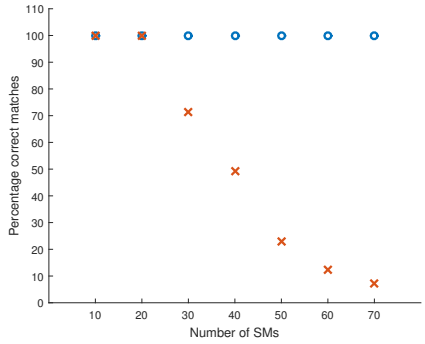


Fig. 5: Percentage of correct matches for different numbers of meters when using two (circle) or four (cross) pseudonyms.

We also investigated the usefulness of the data. Again the deviation of the aggregate is calculated using Equation 1, but this time $x'_{i,t}$ is the rounded consumption of user i at time t. Figure 4 shows a boxplot of the usefulness averaged over all users and all months, for different rounding thresholds. We see that for rounding thresholds lower than 0.1 kWh, the average deviation is less than 1%. Up to 0.25 kWh the average deviation stays under 5%. However, when rounding up to 1 kWh, the deviation is already more than 25%.

D. Results with Different Pseudonyms

We first change the pseudonym every month. Considering the very high level of de-pseudonymization we already achieved after one month in Section IV-A, we expect that this will only improve privacy very minimally. Indeed, when calculating the percentage of successful matches (i.e. the percentage of de-pseudonymized users) for each month, we see that October 2010 is the best month, but the adversary can still de-pseudonymize 99.73% of the users. December 2009 is the worst month, the adversary can de-pseudonymize no less than 99.90% of the users. On average the adversary can de-pseudonymize 99.83% of the users per month.

Next, we describe the de-pseudonymization results when the countermeasure *Different Pseudonyms*, detailed in Section III-C4, is implemented. We change the pseudonym every half-month and define the percentage of successful matches as the number of correct matches, i.e. both the first and the second pseudonym correspond to the meter in question, divided by the total number of matches. When considering only one month (due to the computational complexity of this method), the percentage of correct matches is equal to only 6.34%.

TABLE II: De-pseudonymization Results.

Countermeasure	% De-pseudon.	Deviation
No countermeasure	99.83%	0%
Missing data (1 data point $\rightarrow$ 0)	33.95%	0.07%
Missing data (51 data points $\rightarrow$ 0)	7.84%	3.42%
Missing data (1 data point $\rightarrow$ avg)	46.39%	0.02%
Missing data (51 data points $\rightarrow$ avg)	10.85%	0.09%
Rounding up to 0.05 kWh	14.83%	0.40%
Rounding up to 0.50 kWh	2.50%	12.38%
Two Pseudonyms	6.34%	0%

Finally, we give an illustration of the influence of the number of meters and the number of different pseudonyms within one month. Figure 5 shows the percentage of correct matches, when considering only a small subset of meters. When using only two different pseudonyms per month, we see that all meters can be de-pseudonymized. However, when using four different pseudonyms per month, only with as little as 30 SMs, we can no longer de-pseudonymize all users. With 50 SMs, only 23.04% of the matches are correct.

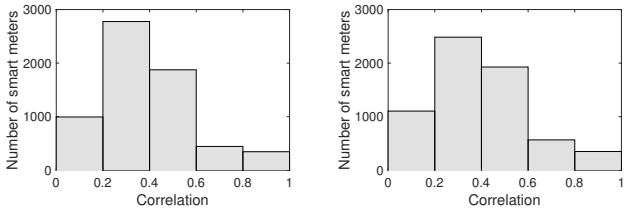
E. Comparison of the Proposed Countermeasures

Table II gives an overview of the results we obtained for the different countermeasures. The best results regarding users' privacy protection are obtained with the rounding countermeasure and a relatively big rounding step (e.g. 0.50 kWh). However, this comes at a cost of degraded data usefulness. The different pseudonyms countermeasure improves the users' privacy protection greatly without affecting the data usefulness. However, its downside is the increased SM complexity as each SM must use at least two different pseudonyms every month. Regarding users' privacy protection, data usefulness and countermeasure simplicity, the missing data countermeasure where SMs omit to send several datapoints per month gives the best trade-off, specially when the supplier replaces these missing datapoints with the average of the two datapoints around them.

V. Discussion

In this paper we have assumed that the electricity price remains the same throughout the day. However, in future a more realistic assumption would be that the price of electricity depends on the time of usage, i.e. there would be multiple tariff periods. In this case the billing would not be based on one monthly aggregate value, but instead on multiple monthly aggregates, one per tariff period. Assuming no countermeasures are being used, this would make the de-pseudonymization even easier, as there would be multiple values to be used for the matching algorithm, rather than just one.

In addition, the accuracy of fine-grained data will be critical in cases when such data is used for real-time Demand



(a) Correlation between the first two weeks vs. the second two weeks

(b) Correlation between the first and third week

Fig. 6: Correlation of users' own consumption data over a specific period of time.

Response (DR). However, our proposed countermeasures introduce a small deviation in the data of only a very small group of users, in each time slot. Normally, a small fraction of users participates in real-time DR programs in a given time slot. Hence, as long as none of the SMs participating in DR programs in a specific time slot is selected to modify its data (for privacy protection reasons), then our countermeasures will not affect the DR programs – each DR-participating user will be reporting accurate data, hence receiving the correct rewards. Also, as accurate aggregates of all SMs (in a region) for each time slot are already available from control SMs placed at strategic places on the distribution grid, an accurate global DR is possible even with some of the SMs not reporting their actual data due to our countermeasures.

In addition, due to the repetitive nature of the users' consumption patterns, it may also be possible to de-pseudonymize some users by looking at their own consumption data over a specific period of time, e.g. by looking at their weekly consumption patterns. This would make our countermeasure less effective, as the adversary may be able to link these consumption patterns to the users even when distinct pseudonyms are used. To verify our hypothesis, we performed some experiments for the cases where a new pseudonym is given to the SMs: (i) every two weeks and (ii) on a per week basis. In other words, we investigate whether giving a new pseudonym every two weeks or every week, respectively, is sufficient to protect the user's privacy or whether it is possible to link different pseudonyms using statistical measures such as the correlation.

Figure 6a shows the correlation between consumption patterns of the first two weeks vs. the second two weeks of November 2009. Similarly, Figure 6b shows the correlation between the first and third week of November 2009. From these figures it is clear that for a non-negligible number of users there is a strong correlation between their consumption patterns in different weeks. From this we can conclude that once a user with a very repetitive consumption pattern has been de-pseudonymized for one particular week, we can de-pseudonymize him in later weeks as well, even if the pseudonym has changed in the mean time. One possible solution would be to give new pseudonyms to users more frequently (e.g. every day or every half-hour), however this will increase the complexity of the system.

VI. Conclusions

We showed that simple pseudonymization does not provide sufficient privacy protection to users. More specifically, adversaries can de-pseudonymize 99.89% of the users after only one month and all users after four months. Based on the obtained results, we presented three practical yet effective countermeasures to increase the users' privacy level. Our results show that all of the three countermeasures yield a significant improvement in privacy, while the loss of data usefulness remains acceptable. With every countermeasure we are able to decrease the percentage of de-pseudonymized users to less than 15%, while keeping the deviation of the half-hourly aggregate below 5%. As future work we plan to investigate the optimal trade-off between privacy gain and loss of data usefulness, the combination of the different countermeasures, as well as their computational complexity.

References

[1] G. Kalogridis, M. Sooriyabandara, Z. Fan, and M. A. Mustafa, "Toward unified security and privacy protection for smart meter networks," *IEEE Systems Journal*, vol. 8, no. 2, pp. 641–654, June 2014.

[2] E. L. Quinn, "Privacy and the new energy infrastracture," *Social Sience Research Networks (SSRN)*, Febuary 2009.

[3] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, december 1992.

[4] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong, "Power signature analysis," *IEEE Power and Energy Magazine*, vol. 1, no. 2, pp. 56–63, 2003.

[5] U. Greveler, P. Glösekötterz, B. Justusy, and D. Loehr, "Multimedia content identification through smart meter power usage profiles," in *Int. Conf. on Information and Knowledge Engineering (IKE)*, 2012, pp. 1–8.

[6] C. Cuijpers and B.-J. Koops, "Het wetsvoorstel slimme meters: een privacytoets op basis van art. 8 EVRM," Universiteit van Tilburg — Centrum voor Recht, Technologie en Samenleving, Onderzoek in opdracht van de Consumentenbond, October 2008.

[7] C. Efthymiou and G. Kalogridis, "Smart grid privacy via anonymization of smart metering data," in *IEEE 1st International Conference on Smart Grid Communications (SmartGridComm)*, October 2010, pp. 238–243.

[8] M. Jawurek, M. Johns, and K. Rieck, "Smart metering de-pseudonymization," in *ACM 27th Annual Computer Security Applications Conference (ACSAC)*, 2011, pp. 227–236.

[9] E. Buchmann, K. Böhm, T. Burghardt, and S. Kessler, "Re-identification of smart meter data," *Personal Ubiquitous Computing*, vol. 17, no. 4, pp. 653–662, 2013.

[10] V. Tudor, M. Almgren, and M. Papatriantafilou, "A study on data de-pseudonymization in the smart grid," in *ACM 8th European Workshop on System Security (EuroSec)*, 2015, pp. 1–6.

[11] M. Faisal, A. A. Cardenas, and D. Mashima, "How the quantity and quality of training data impacts re-identification of smart meter users?" in *IEEE Int. Conf. on Smart Grid Communications*, 2015, pp. 31–36.

[12] V. Tudor, M. Almgren, and M. Papatriantafilou, "Analysis of the impact of data granularity on privacy for the smart grid," in *ACM 12th Workshop on Privacy in the Electronic Society (WPES)*, 2013, pp. 61–70.

[13] Department of energy and climate change, "Smart metering implementation programme, data access and privacy," December 2012. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/43046/7225-gov-resp-sm-data-access-privacy.pdf

[14] Irish social science data archive, "Electricity customer behaviour trial," http://www.ucd.ie/issda/data/commissionforenergyregulationcer/.

[15] F. Laforet, E. Buchmann, and K. Böhm, "Individual privacy constraints on time-series data," *Inf. Syst.*, vol. 54, no. C, pp. 74–91, 2015.

[16] S. Kessler, E. Buchmann, and K. Böhm, "Deploying and evaluating pufferfish privacy for smart meter data," in *IEEE Conf. on Ubiquitous Intelligence and Computing (UIC-ATC-ScalCom)*, 2015, pp. 229–238.

[17] M. A. Mustafa, N. Zhang, G. Kalogridis, and Z. Fan, "DEP2SA: A decentralized efficient privacy-preserving and selective aggregation scheme in advanced metering infrastructure," *IEEE Access*, vol. 3, pp. 2828–2846, 2015.