Towards Quantitative Evaluation of Privacy Protection Schemes for Electricity Usage Data Sharing

Daisuke Mashima[a], Aidana Serikova[b], Yao Cheng[c], Binbin Chen[a]

[a] *Advanced Digital Sciences Center, Singapore*
[b] *Nazarbayev University, Kazakhstan*
[c] *Institute for Infocomm Research, A*STAR, Singapore*

Abstract

Thanks to the roll-out of smart meters, availability of fine-grained electricity usage data has rapidly grown. Such data has enabled utility companies to perform robust and efficient grid operations. However, at the same time, privacy concerns associated with sharing and disclosure of such data have been raised. In this paper, we first demonstrate the feasibility of estimating privacy-sensitive household attributes based solely on the energy usage data of residential customers. We then discuss a framework to measure privacy gain and evaluate the effectiveness of customer-centric privacy-protection schemes, namely redaction of data irrelevant to services and addition of bounded artificial noise.

Keywords: Privacy, smart meter data, quantitative evaluation

1. Introduction

Thanks to the penetration of smart meters and other types of commodity electricity usage monitoring devices, availability of fine-grained electricity usage data has increased remarkably. Besides utilization by utility companies, for instance demand forecasting and fault/anomaly detection, such data may be shared with third-party service providers either directly from customers (e.g., an energy usage monitoring device may upload data to the service provider's cloud for data analytics, etc.) or via utility companies (e.g., by means of Green Button Connect My Data [1]) to benefit from a variety of services, including energy-saving recommendations, social gaming, and services like demand response.

On the other hand, we are facing a number of new types of privacy risks that were not found in the age prior to the smart grid era. Privacy concerns associated with residential energy usage data have been outlined by National Institute of Standards and Technology (NIST) [2] and include leakage of personally-identifiable information and behavioral information. Moreover, unlike power utility companies that are strictly bound by regulations, other service providers may have the freedom to utilize the collected data for unclaimed purposes and/or share the collected data or analysis results with another party, e.g., advertising or marketing companies, without explicit consent from customers. Therefore, it is not feasible for electricity customers to retain control and awareness over usage of their data once the data are released. Nevertheless, most electricity customers share their data without enough understanding privacy exposure or ways to mitigate such risks [2].

To allow electricity customers to control privacy risks upon sharing electricity usage data with other parties, a framework called customer-centric energy usage management was proposed [3]. This framework can accommodate a variety of data pre-processing schemes applied by customers themselves for privacy protection [4, 5] and is well aligned with policies regarding privacy and data ownership established by utility companies in the US, e.g., [6], as well as European Union [7]. However, they did not show any quantitative evaluation of privacy gains, which can provide electricity customers with meaningful guidelines regarding how much pre-processing is needed to attain the expected level of privacy.

In this paper, we first design mechanisms to estimate privacy-sensitive household information based on household-level energy usage data to highlight potential privacy risks through experiments using real-world energy usage traces [8]. We further discuss a way to measure privacy gains of two privacy-protection mechanisms by means of redaction and artificial noise, which are introduced in the context of the aforementioned customer-centric electricity usage data management [3, 4].

The rest of this paper is organized as follows. In Section 2, we discuss the literature on privacy pertinent to electricity usage data. In Section 3, to educate electricity customers, we demonstrate the feasibility of identifying privacy-sensitive household information with only electricity usage data. In Section 4, we discuss a framework to measure privacy gains and apply it to evaluate the effectiveness of two types of privacy-protection schemes that electricity customers can apply to mitigate privacy risks. We provide supplementary discussion in Section 5 and then conclude the paper in Section 6.

2. Related Work

Kavousian et al. [9] analyzed the determinants of household electricity usage. The results indicated that household characteristics, appliance, electronics stock, and occupants indeed have a large influence on residential electricity usage patterns. An Irish case study [10] also examined the correlation between household/occupant characteristics and electricity usage using a multiple linear regression model. Their results demonstrate that, in addition to household characteristics, household composition and status of the head of household (e.g., age and social class) also have a strong correlation with electricity usage, which has provided a foundation for our investigation.

Beckel et al. [11] used an electricity usage dataset that was collected during a smart meter trial. Along with the electricity usage data, users' responses to a questionnaire before and after the trial are available and include various household characteristics. Based on these ground truth data, the authors demonstrated the feasibility of revealing characteristics from electricity usage data using various classifier models with an overall accuracy of around 70%. This feasibility is further supported by Aderson et al. [12], who demonstrated a concept of energy monitoring for a smart census. Recently, Cong et al. [13] conducted work on discovering missing user attribute labels using smart meter data. In this work, we investigate how much sensitive information can be inferred without any privacy protection, which is based on the feasibility revealed by these efforts. We further introduce extra features to enrich the feature space and apply other data analysis techniques for better accuracy. Moreover, we consider this accuracy as a baseline and evaluate the effectiveness of privacy-protection schemes.

Based on the assumption that the power utility companies fulfill their duty to protect users' electricity usage data as the data custodian, the focus of privacy protection is shifting to data sharing with third-party service providers. In this direction, researchers have proposed customer-centric energy usage management, a privacy protection scheme to enable meaningful data sharing with third parties while preserving users' privacy [3]. We should note that, customer-centric energy usage data management does not aim at privacy protection against utility companies, but against third-party service providers. Thus, it is complementary to, for example, battery-based privacy protection schemes like [14, 15]. Moreover, the framework is orthogonal to privacy protection against attackers targeting smart metering infrastructures, e.g., those summarized in [16]. While [3] implemented privacy protection by means of redaction, there is another work that proposed to add artificial noise before data sharing to mitigate privacy risks [4]. However, to the best of our knowledge, there is no quantitative evaluation regarding how much privacy gain is attained from these protection schemes, which has motivated us to carry out such a study.

3. Estimating Privacy-Sensitive Household Attributes Based on Energy Usage Data

3.1. Residential Energy Usage Dataset

To design and evaluate baseline schemes to estimate privacy-sensitive household attributes, and eventually, to evaluate the effectiveness of privacy-preservation schemes in the next section, we utilize a publicly-available electricity usage dataset collected in the UK, called the Household Electricity Survey (HES) dataset [8]. The primary reason we chose this dataset is that, in addition to electricity usage data with either 10-min or 2-min granularity, this dataset includes various details of each subject household obtained through the survey, which will be discussed later in this section.

Regarding electricity usage data, we used measurements collected at 2-min intervals in 220 households. HES data consist of appliance-level electricity usage data, so we aggregated energy consumption of all appliances in each household to approximate household-level traces. Furthermore, in order to make the data closer to realistic smart meter data, we down-sampled the 2-min interval household-level traces into 10-min intervals. Finally, because the period of data collection differs among households, we normalized the data by using the overall average for each season to remove seasonality.

Table 1: Class definitions for each attribute

Attribute	Class	Definition	# of Samples
Single	1	Single	62
	0	Not Single	158
Occupancy	1	> 2	84
	0	≤ 2	136
Employment_ Status	1	Full-time	123
	0	Otherwise	97
Children	1	With children	72
	0	Without children	148
Social_Grade	1	"A" or "B"	76
	0	Otherwise	144

Among the household details available in the HES dataset, in this study we focused on the following, which are considered to have marketing value and are therefore privacy sensitive: whether the household is occupied by a single person (*Single*), size of household occupancy (*Occupancy*), employment status of a household head (*Employment_Status*), whether a household has any children (*Children*), and the social grade of each household (*Social_Grade*). Class labels were determined based on the data, and their definitions are summarized in Table 1. Namely, Single and Children are defined as boolean (i.e., true or false), Occupancy is set to 1 if the size of occupancy (i.e., the number of residents) is higher than 2 while it is set to 0 otherwise, and Employment_Status is defined as binary regarding whether the household head is a full-time worker or not. In the HES dataset, the social grade has six levels (A, B, C1, C2, D, and E), and we grouped A and B, which correspond to the high social grades, and formed the other group for the rest.

3.2. Designing Baseline Classifiers

This section discusses the design of baseline classifiers that are assumed to be used by a curious (or malicious) third-party energy-data analytics service provider that attempts to reveal the privacy-sensitive data of customers.

We initially defined 114 features derived from the aforementioned energy usage data. Based on our preliminary experiments, features calculated based on weekly data showed better accuracy overall compared to those computed based on monthly data. Thus, the results discussed on this paper are based on the features computed using 1-week-long data. For the experiment in this section, we used the first week of data of each household, resulting in 220 samples. Our initial list of features included basic ones such as the average, variance, and quantiles of electricity usage of each household, as well as features proposed in [11**?**]. In addition, we included features derived from time-series analysis, including autocorrelation, degrees of an autoregressive integrated moving average (ARIMA) model, kurtosis, and skewness, and features based on Fast Fourier Transform (e.g., the most dominant frequency).

Then, we performed feature selection by Random Forest-Recursive Feature Elimination (RF-RFE) [17] for each household attribute to be estimated. This feature selection method provides an importance score for each feature, and according to the score, we first selected 15 features out of the population for each classification. They are summarized in Figure 1. With these features, by using WEKA [18], we applied multiple classifiers that are popularly used, namely AdaBoost, kNN, SVM, Random Forest, Bagging, and BayesNet. Because including all 15 features did not result in the best accuracy, we tweaked the number of features (i.e., selected a different number of features from the top) and conducted experiments for each classifier. As a result, we found that the features highlighted with bold font in Figure 1 provided the best accuracy. Some of the results are shown in Figure 2.

Accuracy in these figures is computed based on the number of correctly-classified samples through a 5-fold cross validation on WEKA. Note here that WEKA's cross-validation implementation applies stratification of data (i.e., the ratio of samples of both classes are roughly the same in all groups). The best classifiers for the household attributes of our interest are summarized in Table 2. Note again that, for the best classifiers, features shown with bold font in Figure 1 are used.

Table 2: Best-performed classifies and accuracy

Household Attribute	Classifier	Accuracy (%)
Single	AdaBoost	79.09
Occupancy	Random Forest	73.18
Employment_Status	Bayes Net	72.72
Children	SVM	75.45
Social_Grade	Random Forest	70.00

As can be seen from the table, privacy-sensitive household attributes can be estimated with over 70% accuracy by using only electricity usage data, and therefore sharing fine-grained electricity usage data should be considered as a serious privacy risk for electricity customers. Comparing our results with those in the literature [11], even though a direct comparison is not completely fair owing to the differences in the datasets and definitions of attributes, our classifiers attained noticeably better performance (over 10% increase) in estimating Social_Grade, while having similar accuracy for Single, Employment_Status, and Children. In the rest of this paper, we assume these classifiers are utilized by curious (or malicious) third-party service providers. The accuracy achieved here (seen in Table 2) is used as the baseline for comparison when we evaluate the effectiveness of privacy-protection schemes.

4. Evaluating Effectiveness of Customer-centric Privacy-protection Schemes

In this section, we evaluate the effectiveness of privacy-protection schemes developed for customer-centric energy usage data management and sharing schemes [3]. In particular, as two data pre-processing techniques that a customer can apply before data sharing, we focus on the redaction of data [3] and the addition of artificial noise [4].

For the experiments in this section, we evaluate the effectiveness of privacy protection in the following way. For the sake of comparison with the baseline discussed in Section 3, we follow a procedure similar to a 5-fold cross validation. Specifically, we randomly form five groups of samples in a stratified manner just as done by WEKA in Section 3.2. For each round, we use four of them for training and the other for testing. The difference from the typical 5-fold cross validation is that, while we use the original electricity usage data for training, for testing we use pre-processed data (see Figure 3). In this way, we can compare the results with those in Table 2. In sum, our experiments emulate a case where a (potentially malicious) service provider has classifiers trained based on original, labeled data collected from a number of customers and attempts to reveal privacy of customers who are submitting either original (Electricity Customer 1 in Figure 4) or pre-processed (Electricity Customer 2 in the same figure) electricity usage data to evaluate the effectiveness of pre-processing for privacy protection.

4.1. Privacy Protection by Redaction

As discussed in [3], hiding some portion of data (e.g., showing only electricity usage during daytime) is considered effective for privacy protection. As can be seen in Figure 1, multiple classifiers rely on consumption during evening as well as night time, which justifies this approach. On the other hand, redacting part of the data is still considered acceptable for many real-world services. For example, services such as demand response, which typically aim at controlling peak-time electricity demand and therefore are particularly interested in consumption during peak times in the afternoon [3].

Single

1) **weekly_day_max**
2) **weekly_total_max**
3) **weekly_total_var**
4) **fft_dominant_freq**
5) **weekly_day_var**
6) **weekly_noon_ percentile_25**
7) **ratio_cons_we_eve_ cons_wd_eve**
8) **weekly_morning_var**
9) **kurtosis**
10) **skewness**
11) ratio_wd_avg_we_avg
12) weekly_noon_max
13) ratio_cons_we_eve_ cons_we_noon
14) ratio_cons_we_noon_ cons_wd_noon
15) weekly_day_ percentile_75

Occupancy

1) **cons_we_day**
2) **weekly_eve_max**
3) **ratio_var_we_avg_ var_wd_avg**
4) **weekly_eve_var**
5) **ratio_cons_we_eve_ cons_wd_eve**
6) **ratio_cons_we_night_ cons_we_day**
7) **weekly_day_var**
8) **ratio_cons_we_eve_ cons_we_noon**
9) **weekly_day_median**
10) weekly_total_var
11) cons_we_eve
12) ratio_daily_night_avg_ day_avg
13) ratio_daily_mor_avg_ noon_avg
14) fft_dominant_freq
15) ratio_cons_we_noon_ cons_we_day

Employment_Status

1) **weekly_eve_var**
2) **ratio_daily_mor_avg_ noon_avg**
3) **ratio_cons_we_mor_ cons_we_noon**
4) **fft_dominant_freq**
5) **ratio_cons_we_noon_ cons_we_day**
6) **ratio_cons_wd_mor_ cons_wd_noon**
7) **ratio_weekly_mor_avg_ noon_avg**
8) **cons_wd_max**
9) **ratio_daily_eve_avg_ noon_avg**
10) ratio_cons_we_eve_ cons_wd_eve
11) ratio_cons_we_night_ cons_we_day
12) ratio_var_we_avg_ var_wd_avg
13) ratio_cons_we_eve_ cons_we_noon
14) var_of_daily_var
15) ratio_cons_wd_max_ cons_wd_min

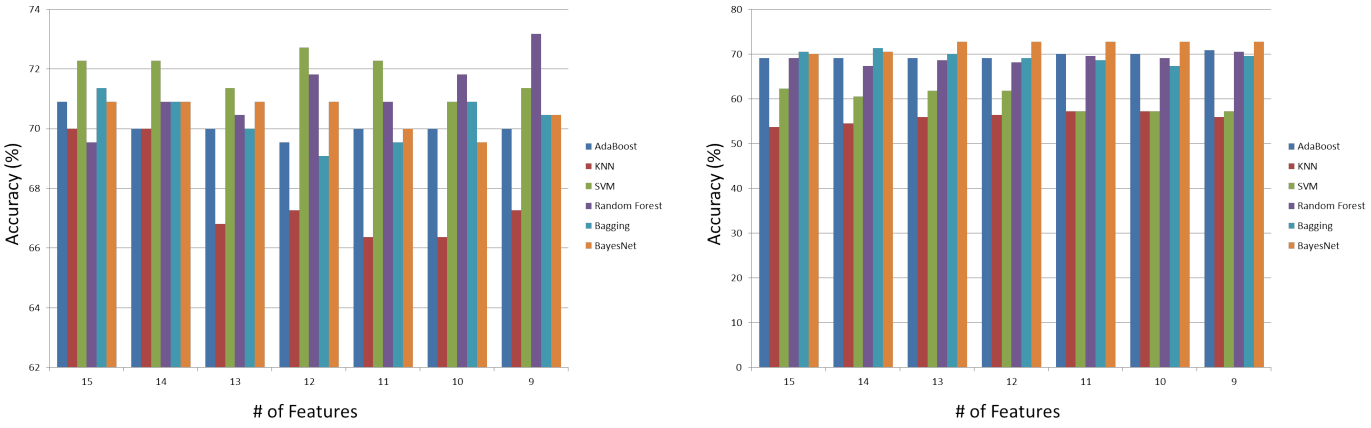
Children

1) **weekly_mor_var**
2) **weekly_day_var**
3) **ratio_cons_we_night_ cons_we_day**
4) **weekly_eve_var**
5) **ratio_cons_we_eve_ cons_wd_eve**
6) **ratio_var_we_avg_ var_wd_avg**
7) **ratio_daily_max_ total_avg**
8) **weekly_total_var**
9) **ratio_cons_we_max_ cons_we_min**
10) **fft_dominant_freq**
11) **ratio_cons_we_noon_ cons_wd_noon**
12) **var_of_daily_var**
13) **ratio_weekly_mor_avg_ noon_avg**
14) ratio_weekly_eve_avg_ noon_avg
15) ratio_cons_wd_mor_ cons_wd_noon

Social_Grade

1) **ratio_weekly_eve_avg_ noon_avg**
2) **ratio_var_we_avg_ var_wd_avg**
3) **ratio_daily_eve_avg_ noon_avg**
4) **ratio_cons_we_noon_ cons_wd_noon**
5) **fft_dominant_freq**
6) **ratio_weekly_mor_avg_ noon_avg**
7) **ratio_cons_wd_night_ cons_wd_day**
8) **ratio_cons_wd_eve_ cons_wd_noon**
9) **ratio_weekly_noon_avg_ total_avg**
10) ratio_cons_wd_mor_ cons_wd_noon
11) weekly_mor_median
12) ratio_daily_mor_avg_ noon_avg
13) ratio_daily_night_avg_ day_avg
14) var_we_avg
15) autocorr_daily

Figure 1: Short-listed features for each household attribute classification. Those highlighted in bold font are features used by the best classifiers.



(a) Occupancy (b) Employment_Status

Figure 2: Accuracy comparison among different classifiers with different numbers of features.

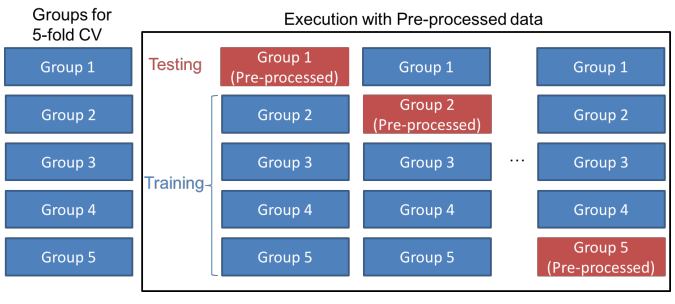


Figure 3: Evaluation of privacy gain following 5-fold cross validation using pre-processed data.

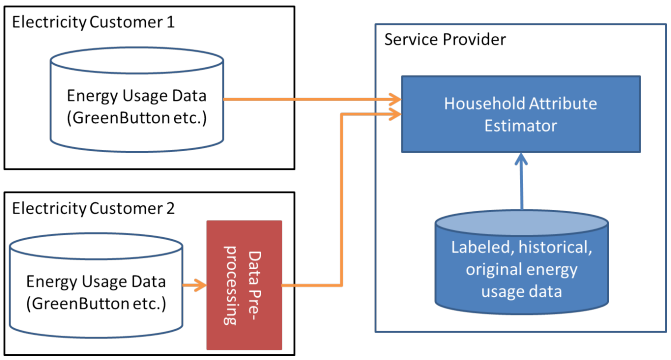


Figure 4: Our model for evaluating privacy gain. Our framework measures privacy gain in terms of differences in estimation accuracy against Electricity Customer 1, who shares original data, and against Electricity Customer 2, who implements customer-centric data pre-processing before data sharing.

We performed two sets of experiments with different degrees of redaction: one redacting electricity usage data except for typical peak hours (10am to 2pm) on each day and the other redacting data except from 6am to 6pm. We

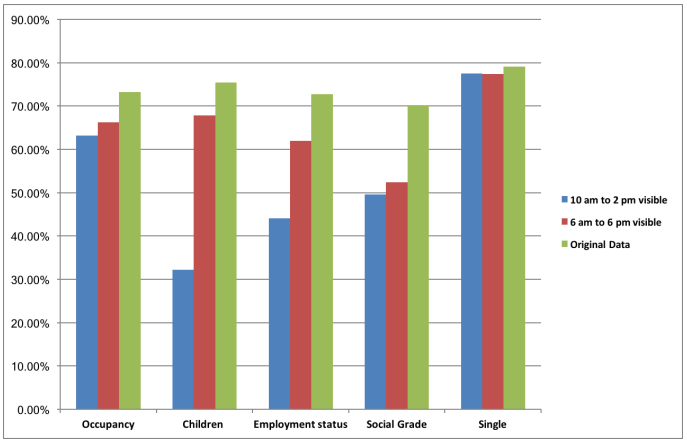


Figure 5: Accuracy of classification using redacted data.

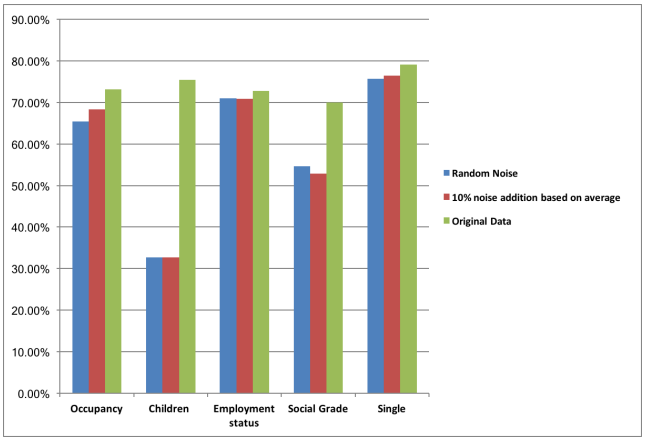


Figure 6: Accuracy of classification using data with artificial noise.

assume that the redacted data are replaced, by the service provider, with the overall average computed based on training data. The results are presented in Figure 5, along with the baseline accuracy from Table 2, which are labeled "Original Data." As can be seen, the accuracy decreases according to the degree of redaction. In particular, accuracy reduction (i.e., privacy gain) is significant for Children, Employment_Status, and Social_Grade.

4.2. Privacy Protection by Artificial Noise

Another privacy-protection strategy is to add an artificial, bounded noise to mask the exact electricity usage. Adding noise would not be preferred for services that require exact data, such as electricity billing and performance evaluation of demand response services. However, a certain amount of noise is considered acceptable for energy-saving recommendation services etc. because approximate data are usually sufficient for many residential customers.

We evaluated the effectiveness of bounded, random noise added on an electricity usage measurement in each time slot. Figure 6 shows the results of experiments with two different types of artificial noise. The first strategy is to add zero-mean, $\pm 10\%$ random noise (i.e., we generated random numbers between 0.9 and 1.1 for each electricity usage measurement and multiplied the factor with the corresponding measurement). The second strategy is slightly more intelligent and adds positive noise when the actual electricity usage of a certain time slot is below the overall average of the household, while adding negative noise otherwise. As can be seen in the figure, we see noticeable decrease in classification accuracy for Children and Social_Grade.

However, compared to the redaction discussed in the previous section, the overall privacy gain by artificial noise seems limited. One plausible reason is that the added noise was to some extent canceled out when computing features based on the sum of the measurements. If we consider further advanced mechanisms to add noise, the impact would be more noticeable. Moreover, the primary motivation for the artificial noise discussed in [4] was to make non-intrusive load monitoring (NILM) or load disaggregation [19, 20] techniques less accurate. In particular, NILM techniques often rely on "load signatures" derived from the energy consumption patterns of each appliance, and noise in electricity usage data makes the signature matching less accurate. Therefore, when the feature set for classification includes those derived based on NILM results (e.g., the usage pattern or frequency of a certain type of appliance), the privacy gain could be more significant.

5. Discussion and Future Research Directions

Based on the results shown in Figures 5 and 6, we can define a privacy-gain metric that summarizes the results for the sake of easier interpretation. For instance, we can calculate the (weighted) average of decreases in accuracy. Alternatively, from the customers' perspective, another metric can be defined in terms of how much information can be correctly identified. The exploration of effective metrics will be part of our future work.

In this study, we assumed that a labeled dataset for training is given. It may be argued that this assumption would be unrealistic because even utility companies do not have customer information other than basic information such as the name of the household head, mailing address, phone number, and billing information. However, there are a non-negligible number of customers who may voluntarily surrender privacy-sensitive information, including those we evaluated in this paper, along with their electricity usage data, through questionnaires requested in exchange for some benefits (e.g., discounts or promotional coupons). By collecting data in such a manner, a service provider could obtain a labeled dataset of a sufficient size in reality.

One limitation of our study is that we did not take into account the adaptation of a data analytics mechanism. A service provider may adjust the feature set and/or

classifiers to better handle pre-processed data (e.g., noisy data or redacted data). In other words, after somehow collecting a sufficient amount of pre-processed data and ground-truth class labels, classifiers could be trained with them. Such a study is part of our future work.

In addition, we simplified the problem into a binary classification for all household attributes of interest. For example, regarding the occupancy size, instead of estimating the actual number, we just paid attention to identifying whether it is an extended family. In general, it is more challenging to estimate exact numbers, as also pointed out in [11]. Although we admit that it is an important part of our future work, the binary information explored in t his paper still has value for marketing and advertisement purposes.

Another direction for future work is to evaluate classifiers that include advanced features such as those derived from non-intrusive load monitoring. It is expected that households with different attributes have different appliance usage patterns. Given the availability of open-source tools such as [20], the derivation of such information becomes feasible.

6. Conclusions

In this paper, we demonstrated the feasibility of estimating privacy-sensitive household attributes that can potentially be abused for unsolicited advertising etc. Based on our experiments using a public dataset, all of five privacy-sensitive attributes considered in this paper can be estimated with over 70% accuracy. We further quantitatively studied the effectiveness of two privacy-protection measures that customers can practically apply before sharing data with potentially malicious third parties, namely redaction and artificial noise.

We hope our contributions shed light on the privacy risks associated with electricity usage data and the quantitative evaluation of privacy-protection schemes not only to counter such risks but also to better educate electricity customers.

Acknowledgment

This research is supported in part by the National Research Foundation, Prime Minister's Office, Singapore under the Energy Programme and administrated by the Energy Market Authority (EP Award No. NRF2014EWT-EIRP002-040) and is also partly supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

References

[1] The green button, [Online]. Available: `http://www.greenbuttondata.org/`, (Date last accessed on Sep. 22, 2017).

[2] NIST Smart Grid, Guidelines for smart grid cyber security: vol. 2, privacy and the smart grid, Guideline, Aug.

[3] G. Lahoti, D. Mashima, W.-P. Chen, Customer-centric energy usage data management and sharing in smart grid systems, in: Proceedings of the first ACM workshop on Smart energy grid security, ACM, 2013, pp. 53–64.

[4] D. Mashima, A. Roy, Privacy preserving disclosure of authenticated energy usage data, in: Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on, IEEE, 2014, pp. 866–871.

[5] D. Mashima, Authenticated down-sampling for privacy-preserving energy usage data sharing, in: Smart Grid Communications (SmartGridComm), 2015 IEEE International Conference on, IEEE, 2015, pp. 605–610.

[6] Privacy policy, [Online]. Available: `https://www.pge.com/en_US/about-pge/company-information/privacy-policy/privacy-policy/privacy-policy.page`, (Date last accessed on Oct. 9, 2017).

[7] Report on the local and regional consequences of the development of smart grids, [Online]. Available: `http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&reference=A7-2014-0019&language=EN`, (Date last accessed on Oct. 9, 2017) (2014).

[8] J. Zimmermann, M. Evans, J. Griggs, N. King, L. Harding, P. Roberts, C. Evans, Household electricity survey: A study of domestic electrical product usage, Intertek Testing & Certification Ltd.

[9] A. Kavousian, R. Rajagopal, M. Fischer, Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior, Energy 55 (2013) 184–194.

[10] F. McLoughlin, A. Duffy, M. Conlon, Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study, Energy and Buildings 48 (2012) 240–248.

[11] C. Beckel, L. Sadamori, T. Staake, S. Santini, Revealing household characteristics from smart meter data, Energy 78 (2014) 397–410.

[12] B. Anderson, S. Lin, A. Newing, A. Bahaj, P. James, Electricity consumption and household characteristics: Implications for census-taking in a smart metered future, Computers, Environment and Urban Systems 63 (2017) 58–67.

[13] Y. Cong, G. Sun, J. Liu, H. Yu, J. Luo, User attribute discovery with missing labels, Pattern Recognition 73 (2018) 33–46.

[14] G. Kalogridis, C. Efthymiou, S. Z. Denic, T. A. Lewis, R. Cepeda, Privacy for smart meters: Towards undetectable appliance load signatures, in: Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on, IEEE, 2010, pp. 232–237.

[15] Z. Zhang, Z. Qin, L. Zhu, J. Weng, K. Ren, Cost-friendly differential privacy for smart meters: exploiting the dual roles of the noise, IEEE Transactions on Smart Grid 8 (2) (2017) 619–626.

[16] M. A. Ferrag, L. A. Maglaras, H. Janicke, J. Jiang, A survey on privacy-preserving schemes for smart grid communications, arXiv preprint arXiv:1611.07722.

[17] P. M. Granitto, C. Furlanello, F. Biasioli, F. Gasperi, Recursive feature elimination with random forest for ptr-ms analysis of agroindustrial products, Chemometrics and Intelligent Laboratory Systems 83 (2) (2006) 83–90.

[18] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, The weka data mining software: an update, ACM SIGKDD Explorations Newsletter 11 (1) (2009) 10–18.

[19] G. W. Hart, Nonintrusive appliance load monitoring, Proceedings of the IEEE 80 (12) (1992) 1870–1891.

[20] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, M. Srivastava, NILMTK: An open source toolkit for non-intrusive load monitoring, in: Proceedings of the 5th international conference on Future energy systems, ACM, 2014, pp. 265–276.