ON METHODS FOR PRIVACY-PRESERVING ENERGY DISAGGREGATION

Ye Wang[†] *Nisarg Raval*[‡] *Prakash Ishwar*[⋆] *Mitsuhiro Hattori*[◦]
Takato Hirano[◦] *Nori Matsuda*[◦] *Rina Shimizu*[◦]

[†]Mitsubishi Electric Research Laboratories, Cambridge, Massachusetts
[‡]Dept. of Computer Science, Duke University, Durham, North Carolina
[⋆]Dept. of Electrical and Computer Engineering, Boston University, Boston, Masschusetts
[◦]Information Technology R&D Center, Mistubishi Electric Corporation, Kamakura, Japan

ABSTRACT

Household energy monitoring via smart-meters motivates the problem of disaggregating the total energy usage signal into the component energy usage and operating patterns of individual appliances. While energy disaggregation enables useful analytics, it also raises privacy concerns because sensitive household information may also be revealed. Our goal is to preserve analytical utility while mitigating privacy concerns by processing the total energy usage signal. We consider processing methods that attempt to remove the contribution of a set of sensitive appliances from the total energy signal. We show that while a simple model-based approach is effective against an adversary making the same model assumptions, it is much less effective against a stronger adversary employing neural networks in an inference attack. We also investigate the performance of employing neural networks to estimate and remove the energy usage of sensitive appliances. The experiments used the publicly available UK-DALE dataset that was collected from actual households.

Index Terms— data privacy, energy disaggregation, factorial hidden Markov model, neural networks, privacy-utility tradeoff

1. INTRODUCTION

Real-time home energy usage monitoring via smart-meters can enable useful analytics for the benefit of both households and power utilities. However, it also raises privacy concerns because sensitive information, such as behavior and occupancy patterns [1] or even the content displayed on a television or computer screen [2], may also be revealed. The general objective in the field of energy disaggregation, also known as non-intrusive (appliance) load monitoring (NILM or NIALM), is to recover the energy usage patterns of individual appliances from the total household energy usage signal. This objective and the techniques used to achieve it are relevant to both the potentially useful analytics and the privacy concerns.

There are various techniques for energy disaggregation, with the data sampling frequency being a major factor determining suitable approaches (see [3] for a recent survey). Exploiting high order harmonics, such as in [4], requires sampling rates on the order of kilohertz to capture these microscopic features. In an approach suitable for the low sampling frequency regime, the Factorial Hidden Markov Model (FHMM) [5] and several of its extensions were applied to energy disaggregation in [6]. Recently, several neural network architectures were successfully applied in [7] to automate feature extraction for energy disaggregation.

In this work, we examine how the total household energy usage signal should be processed to *conceal* the usage patterns of a set of sensitive appliances, while preserving the detection of the other appliances. The related approaches of [8] and [9] apply an HMM and treat a similar problem as an optimization of a statistical privacy-utility tradeoff (related to the general framework of [10]). In both of these works, the utility objective is to minimize the *distortion* with respect to the total household energy signal. However, we argue that the objective of removing the contribution of the sensitive appliance's energy usage from the total household energy signal is more suitable and leads to a simpler approach of equalizing the means and variances across the states of the sensitive appliances.

In our experiments, we show that mean and variance equalization is highly effective against an adversary limited to the FHMM modeling assumptions. However, a concern is whether the model is oversimplified, and hence this approach may fail to suppress features that would allow a more sophisticated adversary to recover sensitive information. We confirm this concern by employing the specific denoising auto-encoder architecture proposed by [7] to effectively attack the mean and variance equalization mechanism. We also investigate applying these neural networks in a privacy mechanism that estimates and removes the energy usage of the sensitive appliances.

2. METHODS

2.1. Data Model

We consider energy usage data collected over T discrete time steps from a household with M appliances. Let $\mathbf{Y} := (Y_1, \ldots, Y_T)$ denote the total household energy usage over time. For each appliance $m \in \{1, \ldots, M\}$, let $\mathbf{X}_m := (X_{m,1}, \ldots, X_{m,T})$ denote its individual energy usage and $\mathbf{S}_m := (S_{m,1}, \ldots, S_{m,T})$ denote its operating states over time. The total household energy usage at time t is

$$Y_t = Z_t + \sum_{m=1}^{M} X_{m,t},$$

where Z_t denotes measurement noise, which could potentially have a non-zero mean representing the energy usage of otherwise unaccounted for appliances.

For our model-based approaches, we adopt additional statistical assumptions. Applying a Gaussian FHMM assumes the following:

- The operating state sequences $\mathbf{S}_1, \ldots, \mathbf{S}_M$ are mutually independent and each is a Markov chain.
- The energy usage $X_{m,t}$ of appliance m at time t depends only on the corresponding state $S_{m,t}$.
- Given that the state $S_{m,t} = s$, the energy usage $X_{m,t}$ is conditionally Gaussian with state-specific mean $\mu_{m,s}$ and variance $\sigma^2_{m,s}$.

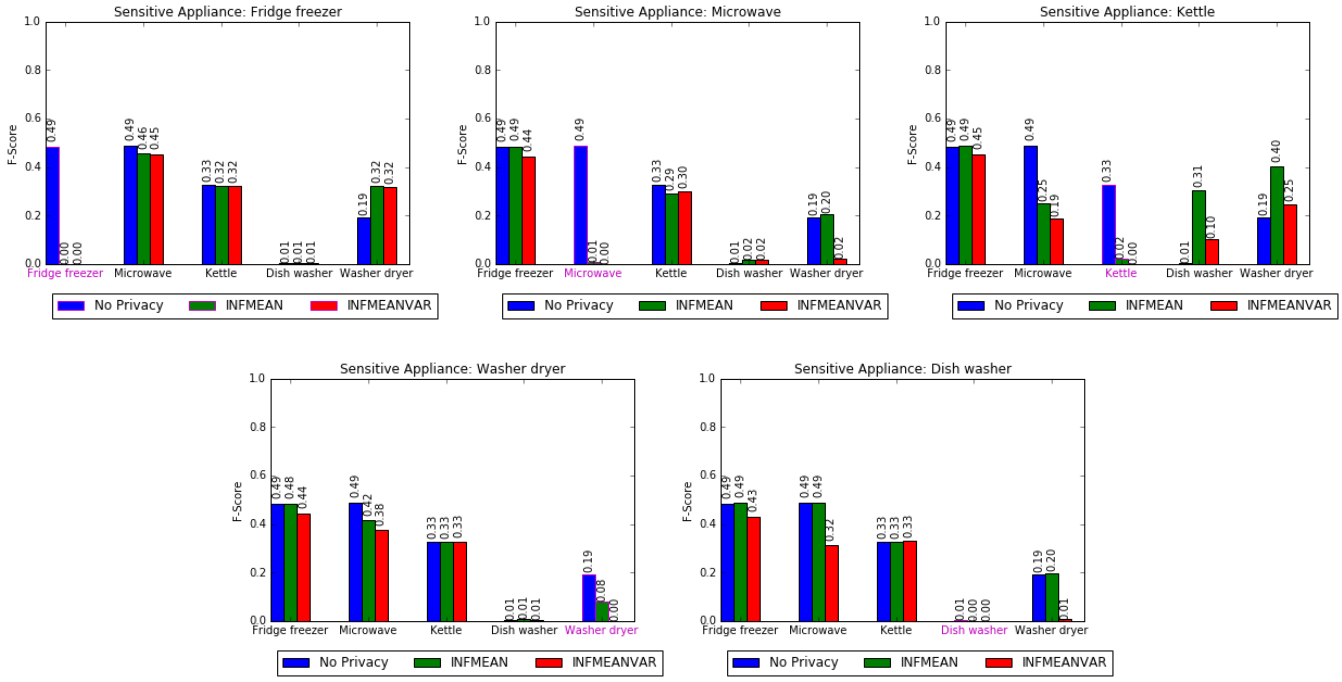

Fig. 1: F-scores for appliance state estimates via the Viterbi algorithm (assuming Gaussian FHMM) applied to the baseline unmodified signal ("No Privacy"), mean equalized signal ("INFMEAN"), and mean and variance equalized signal ("INFMEANVAR"). In each sub-figure, a different appliance is designated as the sensitive appliance.

- The measurement noise Z_t is independent and Gaussian with mean μ_Z and variance σ_Z^2.

Hence, each appliance m is modeled as a mutually independent Gaussian HMM with hidden states $\mathbf{S}_m$ and observations $\mathbf{X}_m$.

2.2. Privacy and Utility Objectives

The goal is to release a modified aggregate energy usage signal $\hat{\mathbf{Y}} := (\hat{Y}_1, \ldots, \hat{Y}_T)$ that enables useful analytics while preserving the privacy of some sensitive information. Specific utility and privacy objectives have a significant impact on the nature of the problem, suitable approaches, and the fundamental privacy-utility tradeoffs that arise. Another practical consideration is what data is directly available to the mechanism that generates $\hat{\mathbf{Y}}$. The achievable privacy-utility tradeoffs are affected by whether both the total energy signal $\mathbf{Y}$ and the states $(\mathbf{S}_1, \ldots, \mathbf{S}_\mathbf{M})$ are available, or only $\mathbf{Y}$ is available.

In general, the household may be concerned with arbitrary sensitive information, such as lifestyle and behavioral details, related to their energy consumption and appliance usage habits. However, for a concrete privacy objective, we consider the specific goal of concealing the *operating states* of a set of k sensitive appliances, which are labeled by the first k indices. The specific privacy objective is to reduce the detectability of the states $(\mathbf{S}_1, \ldots, \mathbf{S}_k)$ from $\hat{\mathbf{Y}}$. For validating a specific system, we will compare the performance of estimating the sensitive states $(\mathbf{S}_1, \ldots, \mathbf{S}_k)$ given the modified signal $\hat{\mathbf{Y}}$ with the baseline estimation performance given the original $\mathbf{Y}$.

For preserving the analytical utility of the modified signal $\hat{\mathbf{Y}}$, a natural approach would be to minimize some distortion metric between $\hat{\mathbf{Y}}$ and $\mathbf{Y}$, if we assume that a signal close to the original should have comparable utility to the unmodified $\mathbf{Y}$, which would implicitly provide optimal utility. However, this distortion minimization objective is not necessarily optimal for all analytical tasks. For example, if the analytics focuses on the non-sensitive appliances (indexed by $k+1, \ldots, M$), it may be more appropriate instead to minimize distortion with respect to

$$\tilde{\mathbf{Y}}_k := \mathbf{Y} - \sum_{m=1}^{k} \mathbf{X}_m,$$

that is, the total energy consumption minus the energy consumed by the sensitive appliances. Note that if the sensitive and non-sensitive appliances were independent, then

$$(\mathbf{S}_{k+1}, \ldots, \mathbf{S}_M, \mathbf{X}_{k+1}, \ldots, \mathbf{X}_M) \leftrightarrow \tilde{\mathbf{Y}}_k \leftrightarrow \mathbf{Y}$$

forms a Markov chain. Hence, the inference of the non-sensitive appliances given $\tilde{\mathbf{Y}}_k$ can only be better than if given $\mathbf{Y}$, since the energy usage of the sensitive appliances can be viewed as noise.

The objective of minimizing some distortion metric between $\hat{\mathbf{Y}}$ and $\tilde{\mathbf{Y}}_k$, such as the mean squared error (MSE) $E[(\hat{\mathbf{Y}} - \tilde{\mathbf{Y}}_k)^\top (\hat{\mathbf{Y}} - \tilde{\mathbf{Y}}_k)]$, is also aligned with the privacy objective of concealing the sensitive appliances, since it equates to the goal of removing their contribution, and hence detectability, of their energy consumption and states. Thus, if the aim is to conceal the sensitive appliances while preserving the detectability of the rest, a suitable approach for both privacy and utility is to attempt to approximate $\tilde{\mathbf{Y}}_k$ with the modified signal $\hat{\mathbf{Y}}$, such as by subtracting out the minimum MSE estimate of $\sum_{m=1}^{k} \mathbf{X}_m$ from $\mathbf{Y}$. In contrast, if the distortion objective was instead measured with respect to $\mathbf{Y}$, such as in [8, 9], then a tradeoff arises between the privacy and utility objectives that can be formulated as an optimization problem, however this is not necessarily the most suitable approach for all analytical objectives.

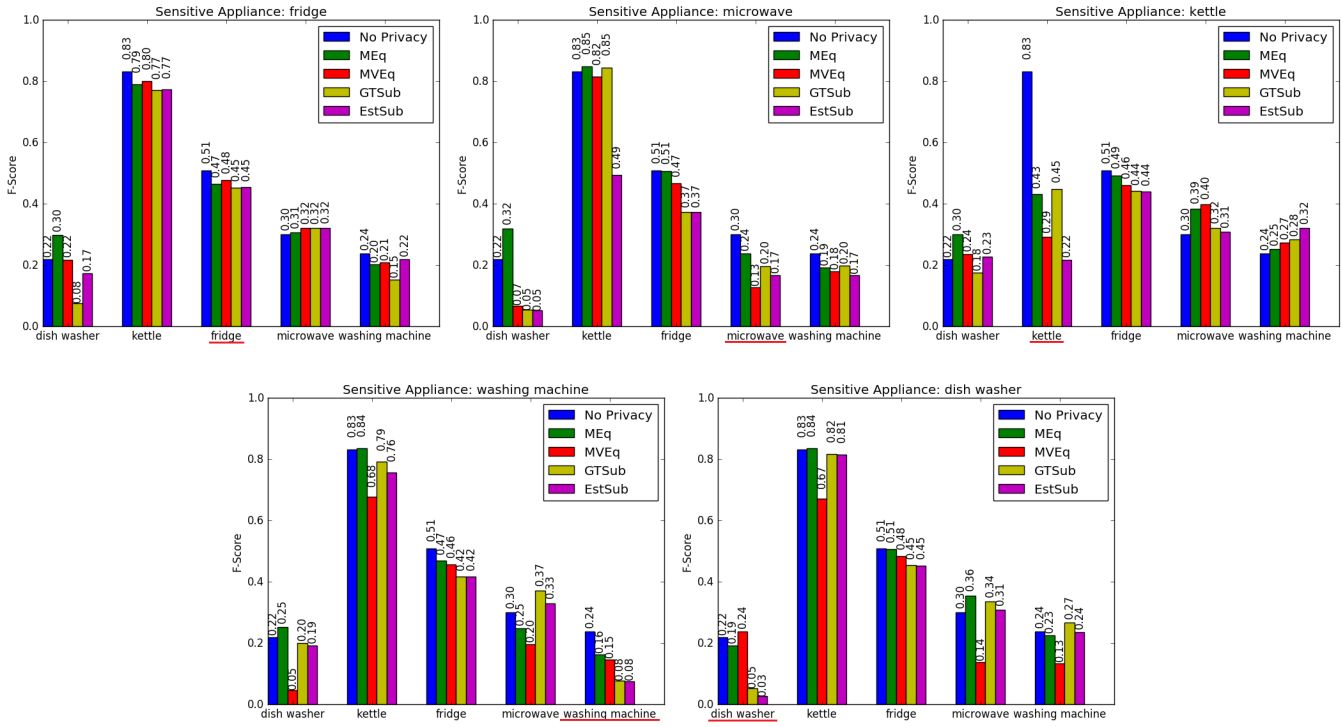

Fig. 2: F-scores for appliance state estimates via the denoising auto-encoder networks applied to baseline unmodified energy signal ("No Privacy"), mean equalized signal ("MEq"), mean and variance equalized signal ("MVEq"), total energy with the ground truth sensitive appliance energy subtracted ("GTSub"), and total energy with the estimated sensitive appliance energy subtracted ("EstSub"). In each subfigure, a different appliance is designated as the sensitive appliance.

2.3. Privacy via Mean and Variance Equalization

If we assume that the Gaussian FHMM is an accurate model for the energy data, then equalizing the mean and variance of the energy usage of the sensitive appliances is a simple approach for concealing the sensitive appliance operating states while preserving the detectability of the other appliances. Specifically, given the total energy signal $\mathbf{Y}$ and the states of the sensitive appliances $(\mathbf{S}_1, \ldots, \mathbf{S}_k)$, the modified signal $\hat{\mathbf{Y}}$ is produced according to

$$\hat{Y}_t := Y_t + \sum_{m=1}^{k} \left(\lambda N_{m,t} - \mu_{m,S_{m,t}} \right), \qquad (1)$$

where $\lambda \in \{0, 1\}$ and $N_{m,t}$ is independent zero-mean Gaussian noise with variance $(\sigma^2_{m,*} - \sigma^2_{m,S_{m,t}})$, where $\sigma^2_{m,*} := \max_{s \in \mathcal{S}_m} \sigma^2_{m,s}$ denotes the maximal variance of the energy usage of appliance m across its operating states. If the states of the sensitive appliances are not available, and the only input to the mechanism is $\mathbf{Y}$, then estimates of these states could first be made from $\mathbf{Y}$ (such as via the Viterbi algorithm) and used instead.

For $\lambda = 0$, the procedure only consists of subtracting out the mean energy usage of each sensitive appliance given their operating states at each time. Note that given its Gaussian distribution, each $\mu_{m,S_{m,t}}$ is the minimum MSE estimate of $X_{m,t}$ given $S_{m,t}$. For $\lambda = 1$, noise is also added in order to equalize the variance of the energy usage of the sensitive appliances across their operating states. In principle, if the Gaussian FHMM assumption is valid, this would result in a modified signal $\hat{\mathbf{Y}}$ that is independent of the states of the sensitive appliances $(\mathbf{S}_1, \ldots, \mathbf{S}_k)$. This follows since $(N_{m,t} + X_{m,t} - \mu_{m,S_{m,t}})$ is zero-mean Gaussian with variance $\sigma^2_{m,*}$, independent of the state $S_{m,t}$.

2.4. Disaggregation with Neural Networks

While the mean and variance equalization privacy mechanism has some theoretical justification, it relies upon modeling assumptions that might not fully capture the complexity of the data. Hence, this simple heuristic might not suppress all potential features linked to the sensitive appliances. Motivated by these concerns, we investigate approaches involving neural networks for both the mean and variance equalization method and other privacy mechanisms.

A denoising autoencoder [11] neural network architecture is employed by [7] for the task of estimating an individual appliance's energy usage $\mathbf{X}_m$ from the total household energy usage $\mathbf{Y}$. The application of a denoising autoencoder is based on the principle that $\mathbf{Y}$ can be viewed as a noisy version of $\mathbf{X}_m$, corrupted by the energy usage of the other appliances. We employ the same network architecture and training procedure as proposed in [7], which takes as input a subsegment of $\mathbf{Y}$ to generate an estimate of $\mathbf{X}_m$ over the same window, and consists of the following layers:

1. Input (nodes $N = w$, window size)
2. 1D convolution (8 filters, size=4, stride=1, linear)
3. Fully connected ($N = 8 * (w - 3)$, ReLU activation)
4. Fully connected ($N = 128$, ReLU activation)
5. Fully connected ($N = 8 * (w - 3)$, ReLU activation)
6. 1D convolution (1 filter, size=4, stride=1, linear)

We additionally train networks for estimating $\mathbf{X}_m$ from the modified energy signal $\hat{\mathbf{Y}}$ generated by various privacy mechanisms. When a network is trained to estimate the same appliance energy usage that the mechanism is attempting to conceal, it would represent an inference attack performed by an adversary. Otherwise, the performance of the network is a benchmark for the utility objective of estimating an appliance while a different one is being concealed by a privacy mechanism. Further, we investigate a privacy mechanism that applies these networks to estimate and remove $\mathbf{X}_m$ from $\mathbf{Y}$ to conceal appliance m.

3. EXPERIMENTAL RESULTS

3.1. UK-DALE Dataset

In our experiments, we used the UK-DALE dataset [12], which consists of electricity usage data collected from five British households. We focused on the $1/6$ Hz data from house 1 of this dataset and five of its major appliances: dishwasher, fridge freezer, kettle, microwave, and washer dryer. The dataset provides the total energy usage signal $\mathbf{Y}$ and the individual appliance energy usage signals $(\mathbf{X}_1, \ldots, \mathbf{X}_5)$. For simplicity, we assume only two operating states for each appliance, representing "off/idle" and "on/active", and we determine the states $(\mathbf{S}_1, \ldots, \mathbf{S}_5)$ by applying a power threshold to $(\mathbf{X}_1, \ldots, \mathbf{X}_5)$. We consider only one sensitive appliance ($k = 1$) at a time, but we rotate each of the five appliances into this role across our experiments.

3.2. FHMM-based Inference and Attacks

We applied a Gaussian FHMM and used the Viterbi algorithm to investigate the effectiveness of mean and variance equalization as a privacy mechanism against an adversary that also assumes a Gaussian FHMM. We tested the mean and variance equalization procedure, for both $\lambda = 0$ and $\lambda = 1$, with each of the five appliances designated in turn as sensitive. When we apply this procedure, we first estimate the states of the sensitive appliance from the total energy signal $\mathbf{Y}$ via the Viterbi algorithm, and then use those state estimates in place of the true states as used in (1).

In Figure 1, we show the F-scores (given by $2(p \cdot r)/(p + r)$, where p and r are precision and recall) for the Viterbi algorithm state estimates generated from the total energy signal $\mathbf{Y}$ ("No privacy"), the modified energy signal produced by mean equalization ($\lambda = 0$, "INFMEAN"), and the modified energy signal produced by mean and variance equalization ($\lambda = 1$, "INFMEANVAR"). Note that for four of the five appliances (except for the dishwasher), the operating states are detected from the original energy usage signal $\mathbf{Y}$ with a non-zero F-score. Mean equalization ($\lambda = 0$) is quite effective, driving the F-score close to zero for each of the appliances (all except the washer dryer) when designated as sensitive. Mean and variance equalization ($\lambda = 1$) drives the F-score to zero for any appliance designated as sensitive, since it removes all distinguishable differences between states given the FHMM. Note that mean and variance equalization often did not significantly degrade the F-scores for the non-sensitive appliances, and in some cases even improved the F-scores. For example, even though the dishwasher was not detectable in the baseline, applying mean and variance equalization for the kettle significantly raised its F-scores above zero.

3.3. Neural Network-based Inference and Attacks

For each appliance m, we trained a baseline auto-encoder network to estimate the appliance energy usage $\mathbf{X}_m$ from $\mathbf{Y}$. Additionally, for each combination of target appliance m and sensitive appliance (indexed as $k = 1$), we trained auto-encoder networks to estimate $\mathbf{X}_m$ from various modified signals $\hat{\mathbf{Y}}$:

- **M-EQ**: mean equalization, $\hat{\mathbf{Y}}$ given by (1) with $\lambda = 0$.
- **MV-EQ**: mean and variance equalization, $\hat{\mathbf{Y}}$ given by (1) with $\lambda = 1$.
- **GT-SUB**: ground truth subtraction, $\hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}_1$, representing a hypothetical, ideal privacy mechanism.

We conducted the following experiments for estimating the energy usage of each appliance $\mathbf{X}_m$ from modified signals $\hat{\mathbf{Y}}$ produced by various privacy mechanisms for concealing appliance $k = 1$:

- *No privacy*: use baseline net to estimate $\mathbf{X}_m$ from input $\mathbf{Y}$.
- *MEq*: use **M-EQ** net ... input $\hat{\mathbf{Y}}$ given by (1) with $\lambda = 0$.
- *MVEq*: use **MV-EQ** net ... input $\hat{\mathbf{Y}}$ given by (1) with $\lambda = 1$.
- *GTSub*: use **GT-SUB** net ... input $\hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}_1$, i.e., testing the hypothetical, ideal privacy mechanism.
- *EstSub*: use **GT-SUB** net ... input $\hat{\mathbf{Y}} = \mathbf{Y} - \hat{\mathbf{X}}_1$, where $\hat{\mathbf{X}}_1$ is estimated from $\mathbf{Y}$ using the baseline net.

From these appliance energy usage estimates, we use power thresholding to determine appliance state estimates.

Figure 2 illustrates the F-scores obtained for the state estimates produced by these experiments. Note that the F-scores for the sensitive appliances in the "MEq" and "MVEq" experiments are generally decreased from the baseline "No privacy" case, but still indicate some detectability. Thus, mean and variance equalization appears less effective against an adversary employing a neural network than one assuming a Gaussian FHMM and applying the Viterbi algorithm. For the "GTSub" experiment, where we subtracted out the ground truth of the sensitive appliance energy usage for a hypothetical, ideal privacy mechanism, we would expect to see very low F-scores for the sensitive appliances. However, this expectation was interestingly unmet (at least for the kettle, fridge, and microwave), indicating that the sensitive appliance can be detected even after its ground truth energy usage has been subtracted out from the total energy signal. Two possible explanations for this observation are: 1) sampling misalignment between the total energy usage and the individual appliance energy usage could prevent a clean subtraction and cause the operation to produce artifacts, or 2) the sensitive appliances could be correlated to other appliances allowing for some detectability even after a clean subtraction. In the "EstSub" experiment, we see that subtracting out the estimated sensitive appliance energy usage achieves similar or better privacy than the ground truth subtraction, which could potentially be explained by the noise in the estimate hindering detection.

4. CONCLUSION

We investigated approaches toward privacy-preserving energy disaggregation that aim to remove the contribution of sensitive appliances from a total household energy usage signal. Standard Gaussian FHMM modeling assumptions suggest a simple heuristic based on mean and variance equalization. This approach is quite effective against an adversary limited to the same assumptions. However, we demonstrate that an adversary employing neural networks can undermine its effectiveness. While simple model assumptions may be sufficient for successful inference, they may be inadequate for ensuring privacy guarantees, since a sophisticated adversary might exploit disregarded features. Applying neural networks directly in the privacy mechanism shows some promise toward addressing complex data, but performance could be improved by future work.

5. REFERENCES

[1] Andrés Molina-Markham, Prashant Shenoy, Kevin Fu, Emmanuel Cecchet, and David Irwin, "Private memoirs of a smart meter," in *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, 2010, BuildSys '10.

[2] Ulrich Greveler, Peter Glösekötterz, Benjamin Justusy, and Dennis Loehr, "Multimedia content identification through smart meter power usage profiles," in *Proceedings of the International Conference on Information and Knowledge Engineering (IKE)*, 2012.

[3] Michael Zeifman and Kurt Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Trans. on Consum. Electron.*, vol. 57, no. 1, pp. 76–84, Feb. 2011.

[4] Christopher Laughman, Kwangduk Lee, Robert Cox, Steven Shaw, Steven Leeb, Les Norford, and Peter Armstrong, "Power signature analysis," *IEEE Power and Energy Magazine*, vol. 1, no. 2, pp. 56–63, 2003.

[5] Zoubin Ghahramani and Michael I. Jordan, "Factorial hidden markov models," *Machine Learning*, vol. 29, no. 2-3, pp. 245–273, 1997.

[6] Hyungsul Kim, Manish Marwah, Martin F. Arlitt, Geoff Lyon, and Jiawei Han, "Unsupervised disaggregation of low frequency power measurements," in *Proceedings of the 2011 SIAM International Conference on Data Mining*, 2011, vol. 11, pp. 747–758.

[7] Jack Kelly and William Knottenbelt, "Neural NILM: Deep neural networks applied to energy disaggregation," in *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, 2015, pp. 55–64.

[8] Lalitha Sankar, S. Raj Rajagopalan, and Soheil Mohajer, "Smart meter privacy: A theoretical framework," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 837–846, 2013.

[9] Murat A. Erdogdu and Nadia Fawaz, "Privacy-utility trade-off under continual observation," in *IEEE International Symposium on Information Theory*, 2015.

[10] Flávio du Pin Calmon and Nadia Fawaz, "Privacy against statistical inference," in *Allerton Conf. on Comm., Ctrl., and Comp.*, 2012, pp. 1401–1408.

[11] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th International ACM Conference on Machine Learning*, 2008, pp. 1096–1103.

[12] Jack Kelly and William Knottenbelt, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," *Scientific Data*, vol. 2, 2015, `https://www.doc.ic.ac.uk/~dk3810/data/`.