Cooperative State Estimation for Preserving Privacy of User Behaviors in Smart Grid

Younghun Kim
IBM T. J. Watson Research,
Hawthorne, NY, USA,
kimy@us.ibm.com

Edith C.-H. Ngai
Department of Information Technology,
Uppsala University, Uppsala, Sweden,
edith.ngai@it.uu.se

Mani B. Srivastava
Electrical Engineering Department,
UCLA, Los Angeles, CA, USA,
mbs@ucla.edu

***Abstract*—Smart grid promises a reliable and secure electricity infrastructure to meet the future demand growth. However, the increase of data types and data amount from advanced smart grid introduce new privacy issues, which have to be resolved for customers. This paper presents a cooperative state estimation technique that protects the privacy of users' daily activities. By exploiting the kernel of an electric grid configuration matrix, we develop an error free state estimation technique that can hide the behavioral information of users effectively. The proposed scheme can obfuscate the privacy-prone data without compromising the performance of state estimation. We evaluate our obfuscation scheme using data from 1349 meters in 5 IEEE Electric Test Bus Systems. Our simulation results demonstrate high level of illegibility and resilience of our scheme with an affordable communication overhead.**

I. Introduction

Reliable operation of electric grids requires timely and detailed monitoring of critical electric branches [1], [2]. Smart grid with advanced monitoring technologies can deliver a more reliable and secure electricity infrastructure. Real-time and detailed monitoring of electricity infrastructure, however, comes at a cost of privacy compromise to users' daily activities. Data mining technique can be applied to the energy consumption traces to infer the appliance usage patterns and occupants' activities such as presence, absence, sleep and wake cycle, breakfast activities, and others [6]–[11]. The privacy of consumers is thus considered as a major challenge for the wide adaptation of the smart grid technology. The privacy implications related to demand response enabled system have been widely discussed [12]. Similarly, consumer privacy related guidelines have also been actively debated [13], [14]. In 2010, US California's new smart meter privacy law, the first of its kind, shows strong demands for privacy protection for consumers' energy consumption data [15].

To protect the privacy of consumers, utility providers make use of secure databases with access control. However, secure database alone is not enough for avoiding privacy leaks and data misuse. We have observed privacy leaks through authorized personnel in a medical database [16]. Various privacy aware smart metering techniques have been proposed for different context, such as privacy aware billing [10], secure meter data aggregation [11], and privacy aware home energy management system [9].

This paper aims at addressing privacy issues in grid state estimation. In practice, state estimation needs to be performed in a distributed computing infrastructure, some of which could be hosted by a third party with the following reasons. (1) The scale of electric grids is large and handling every processing in a central place is not a scalable option [2]. (2) A third-party cloud computing infrastructure could be more cost effective than running it by individual utility providers. (3) Deregulated electric markets such as the US electric grids do not have a single central place. Thus multiple parties need to share meter data to estimate the state in wide area grids. Therefore, meter data need to be perturbed from smart meters throughout the estimation process.

Based on the above considerations, we propose a cooperative state vector estimation technique that preserves "the privacy of personal behavior." The key objectives are to make sure that (1) the power consumption measurement is well obfuscated such that the consumers do not fully disclose their private behavioral information in the first place, and (2) the obfuscated data retain the necessary information such that the state vector can be accurately estimated from the perturbed measurement. The contributions of the paper are in two folds.

- *Cooperative State Estimation for Privacy Preservation of Personal Behavior:* We propose a novel privacy preserving state estimation technique to protect the personal behaviors of users. We derive a high dimensional error free obfuscation space by exploiting the structure of the state estimation problem. In addition to state estimation, our formulation allows remote billing from perturbed data.
- *Evaluation with Real Data Sets in US Test Electric Grids Systems:* We evaluate the performance of our data obfuscation scheme with 1349 measurement data sets. We use the data sets as if they are connected to 5 different IEEE Test Systems that are portions of the Middlewestern US Electric Power Grids. We evaluate the illegibility to human inspectors, resilience to automated data mining attackers, and communication overhead.

II. Privacy Preserving State Estimation

A. Static State Estimation in Electric Grid

Let $\mathbf{x} = [x_1, x_2, ..., x_n]^T \in \mathbb{R}^n$ be a state vector that describes the phase angle at each electric branch with the

978-1-4577-1702-4/11/$26.00 ©2011 IEEE

measurement vector $\mathbf{z} = [z_1, z_2, ..., z_m]^T \in \mathbb{R}^m$ where $m > n$. A grid dynamic equation is

$$\mathbf{z} = h(\mathbf{x}) + \mathbf{e}, \tag{1}$$

where $h : \mathbb{R}^n \to \mathbb{R}^m$ is a nonlinear grid dynamics. To reduce the computational complexity [5], [17], a linear approximation is used.

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{e}, \tag{2}$$

where $\mathbf{H} \in \mathbb{R}^{m \times n}$ is a full column rank grid configuration matrix, $\mathbf{e} \in \mathbb{R}^m$ is the measurement noise and unmodelled dynamics whose covariance matrix is $\mathbf{W}^{-1}$.

The best unbiased linear estimation of the state is known in the following form [5],

$$\hat{\mathbf{x}} = (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{z}, \tag{3}$$

where $\hat{\mathbf{x}}$ is the estimated state vector.

B. Our Research Objectives

We focus on the following three objectives to tackle the privacy preserving state estimation problem.

- *Privacy Objective:* We want to perturb the original measurement $z_i(t)$ in a way that the obfuscated measurement $z_{obf,i}(t)$ is far different from $z_i(t)$ and difficult to infer $z_i(t)$ from $z_{obf,i}(t)$.
- *Error Free State Estimation Objective:* The estimated state, $\hat{\mathbf{x}}_{obf}(t)$ from the obfuscated data $\{z_{obf,i}(t) : 1 \leq i \leq m\}$ has to be the same as the estimated state $\hat{\mathbf{x}}(t)$ from the original data $\{z_i(t) : 1 \leq i \leq m\}$.
- *Remote Billing Objective:* To allow automatic billing and dynamic pricing, the accumulated value of the obfuscated data for the billing period T has to be the same as the accumulated consumption of the original data, i.e. $\sum_{t=0}^{T} z_{obf,i}(t) = \sum_{t=0}^{T} z_i(t)$.

III. Proposed Solution

A. State Estimation from Distortion Free Obfuscation

Let $\mathbf{z}_{obf} \stackrel{def}{=} \mathbf{z} + \mathbf{o}$ be an obfuscated measurement vector, where $\mathbf{o} \in Ker(\mathbf{H}^T\mathbf{W})$. With this, we can show that $\hat{\mathbf{x}}_{obf} = \hat{\mathbf{x}}$ even if $\mathbf{z}_{obf} \neq \mathbf{z}$.

Lemma 1. (Degrees of Obfuscatability) *Let $\mathbf{H} \in \mathbb{R}^{m \times n}$ where $m > n$ be a full column rank matrix. The kernel of the linear transformation $((\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W})$, $Ker((\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W})$ has $m - n$ degrees of freedom.*

Proof: This follows from the matrix rank theorem, i.e. rank$((\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W})$ =rank$(\mathbf{H}^T\mathbf{W})$ = rank $(\mathbf{H}^T) = n$. Therefore, dim$(ker((\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W})) = m - n$. ■

In a typical electric grids, $n << m$ [18], which makes the obfuscation space a high dimensional space.

Definition *(Distortion Free Obfuscation Vector Space)* Let $\mathbb{O} = \{o_1, o_2, o_3, ..., o_{m-n}\}$ be a set of basis for $Ker(\mathbf{H}^T\mathbf{W})$. We call $span(\mathbb{O})$ the distortion free obfuscation vector space.

The exact estimated state from the obfuscated measurement thus follows from Eq 3.

Lemma 2. (Distortion Free Obfuscation) *Let $\mathbf{z_{obf}} = \mathbf{z} + \mathbf{o} \in \mathbb{R}^m$ be an obfuscated measurement vector, where $\mathbf{o} \in span(\mathbb{O})$. This obfuscation makes no change in the state estimation problem in Eq. 3.*

Proof: Let $\hat{\mathbf{x}}$ be the estimated state from the original measurement $\mathbf{z}$ and $\hat{\mathbf{x}}_{obf}$ be the estimated state from the obfuscated measurement $\mathbf{z_{obf}}$, then we have

$$\begin{aligned}
\hat{\mathbf{x}}_{obf} &= (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{z_{obf}} \\
&= (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}(\mathbf{z} + \mathbf{o}) \\
&= (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{z} + (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{o} \\
&= (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{z} \\
&= \hat{\mathbf{x}}.
\end{aligned} \tag{4}$$

■

Lemma 2 means that it is possible to change the measurement vector $\mathbf{z}$, which includes behavioral information, without affecting the performance of the state estimation [1].

B. Obfuscation Design Parameters and Choices

Let $span(\mathbb{O}) = \{\eta_1\mathbf{o}_1 + \eta_2\mathbf{o}_2 + ... + \eta_{m-n}\mathbf{o}_{m-n} | \eta_1, \eta_2, ..., \eta_{m-n} \in \mathbb{R}\}$ by definition. Since the basis vectors $\mathbf{o}_j$ are fixed, η_j are the variables that we can choose. Let $\eta_j(t)$ be an obfuscation design parameter at time t, then the obfuscation problem essentially is to choose $\eta_j(t)$. Then we can formulate the three competing objectives as an implicit optimization problem.

$$\begin{array}{ll}
\max_{\eta_i(t)} & \mathbf{Q}(\mathbf{z_{obf}}(t)) \\
subject\ to & \sum_{t=0}^{T} \eta_i(t) = 0 \\
where & \mathbf{z}_{obf}(t) = \mathbf{z}(t) + \mathbf{o}(t) \\
 & T\text{: Billing Interval} \\
 & \mathbf{o}(t) \in span(\mathbb{O}),
\end{array} \tag{5}$$

where $\mathbf{Q}(\cdot)$ is an implicit quality metric of the obfuscation.

Now the problem essentially is to choose $\eta_i(t)$ given the constraints. To solve this problem, we can use an arbitrary random number for the design parameter $\eta_i(t)$ as long as it satisfies $\sum_{t=0}^{T} \eta_i(t) = 0$ constraint. We choose to use, but not limited to, Gaussian, Exponential, Pareto, and Chi squared distributions for the design parameters, and evaluate the obfuscation performance with four different quality metrics such as Pearson's correlation, mutual information, Euclidean distance both in time and frequency domains in the evaluation section.

C. Distributed Infrastructure System Design

There could be many different ways to implement the proposed privacy preserving state estimation system in a distributed infrastructure. To illustrate the idea, we exemplify an implementation of the smart meter network based on a mesh topology in Figure 1. In short, the overall system consists of utility providers, smart meters and computing service providers. The utility providers share the basis set $\mathbf{O}$ with the smart meters. For each billing epoch, a lead meter calculates

[1]Note that the measurement vector $\mathbf{z}$ cannot be computed from the estimated state $\hat{\mathbf{x}}_{obf}$ by using Eq. 2 because it is highly underdetermined.

Procedure 1 Behavioral Privacy Preserving State Estimation

1. Initialization

Compute the basis set $\mathbb{O}$

Send $\mathbb{O}$ to Meters

2. Lead Meter Selection

Select a lead meter

3. Lead: Generate Obfuscation Vector

while Epoch **do**

Pick a random number $\eta_j(t)$ such that $\sum_{t=0}^{T} \eta_j(t) = 0$

Construct an Obfuscation Vector $\mathbf{o}(t) = \sum_{j=1}^{m-n} \eta_j(t)\mathbf{o}_j$

Send each obfuscation entry $o_i(t)$ to each i-th meter

end while

3. Each Meter: Perform Obfuscation

while Epoch **do**

Read Current Measurement $z_i(t)$

Construct an Obfuscated Measurement $z_{obf,i}(t) = z_i(t) + o_i(t)$

Send $z_{obf,i}(t) = e_i z_i(t) + o_i(t)$ to the utility provider

end while

3. State Estimator: Perform Estimation

while Epoch **do**

Construct the obfuscated measurement vector $\mathbf{z}_{obf}(t) = [z_{obf,1}(t), z_{obf,2}(t), ..., z_{obf,m}(t)]^T$

Solve for $\hat{\mathbf{x}}_{obf}(t) = (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{z}_{\mathbf{obf}}(\mathbf{t})$

end while

4. After Epoch Ends

Go to step 2 and continue

a set of obfuscation vector $\mathbf{o}$ [2]. The lead meter generates the obfuscation vector $\mathbf{o}(t) \in span(\mathbb{O})$. The obfuscation vector is a linear combination of the basis vectors with a randomly chosen weights η_j, i.e. $\mathbf{o} = \sum_{j=1}^{m} \eta_j \mathbf{o}_j$. The lead meter splits the vector into m scaler values, and sends each scaler value o_i to the designated meter. Each meter then computes its obfuscated measurement $z_{obf,i} = z_i + o_i$ and sends it to the state estimator. Upon receiving a set of obfuscated measurement, the state estimator can simply construct an obfuscated measurement vector $\mathbf{z}_{obf} = [z_{obf,1}, z_{obf,2}..., z_{obf,m}]^T = [z_1 + o_1, z_2 + o_2..., z_m + o_m]^T = \mathbf{z} + \mathbf{o}$. The obfuscated measurement vector $\mathbf{z}_{obf} = \mathbf{z} + \mathbf{o}$ with $\mathbf{o} \in \mathbb{O}$ makes the state estimation exact by Lemma 2. Procedure 1 describes the data obfuscation process.

IV. Evaluation

To evaluate the performance of our obfuscation scheme, we adopt three quality terms from source code obfuscation: *illegibility, resilience,* and *cost*.

- *Illegibility:* Illegibility measures the level of difficulty of interpreting and mining data to a human inspector. We call the obfuscation illegible if the obscured data can confuse a human inspector.
- *Resilience:* Automated attacks to the obfuscated data can be performed. Resilience defines the level of difficulty for automated deobfuscation attacks. The obfuscation scheme is resilient if an automatic deobfuscation tool cannot recover the original data. We employ four metrics from the literature [9], [19]–[21] to evaluate resilience empirically.
- *Cost:* The major cost is communication overhead.

[2]Directly sharing the configuration matrix $\mathbf{H}$ could make the grids vulnerable to the false data injection attack [17]. Therefore, we do not consider this option in this paper.

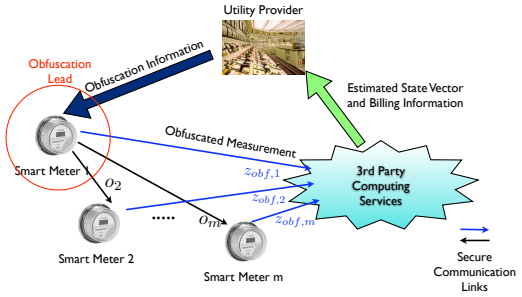


Fig. 1: Mesh-network based System Diagram: For each billing epoch, a lead meter generates the obfuscation vector $\mathbf{o}$ then distributes each obfuscation element o_i to each i-th meter, then each meter obfuscates its measurement to z_{obf}.

TABLE I: IEEE Test Grid Systems

Test System	No. of Meters (m)	No. of States (n)
IEEE 9-bus	27	8
IEEE 14-bus	54	13
IEEE 30-bus	112	29
IEEE 118-bus	490	117
IEEE 300-bus	1122	299

TABLE II: Design Parameter Table

Distribution	Distribution Parameters
Gaussian	STD,$\sigma = 10, 15, 20, 25$
Exponential	Mean,$\mu = 10, 15, 20, 25$
Pareto	Shape variable,$K = 0.75, 1.00, 1.25, 1.5$
Chi Squared	Degrees of Freedom,$D = 10, 15, 20, 25$

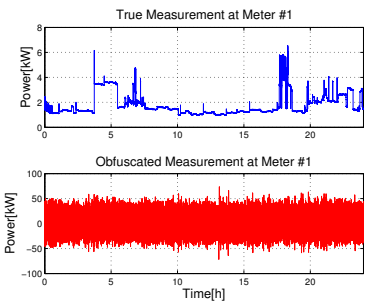
A. Data Set and Evaluation Setting

We evaluate our scheme in IEEE electric grid test systems [18] with real power consumption measurement.

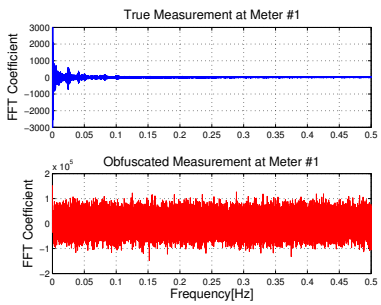
1) Electric Grid Model: The test systems include IEEE 9-bus, 14-bus, 30-bus, 118-bus, and 300-bus systems. Table I summarizes some properties of the test systems. The $\mathbf{H}$ matrix and the basis $\mathbb{O}$ of each test system are computed from MATPOWER[3], a Power System Simulation Tool [18]. The test systems and their parameters are from portions of the US Midwestern electric power grids.

2) Meter Data: We collected data from real-time electric metering systems both in residential and office spaces. We collected 467 residential meter data sets and 882 office meter data sets with the sampling rate of 1Hz. For each test case, we choose a subset of 1349 meter data sets, which includes approximately 116 million data points.

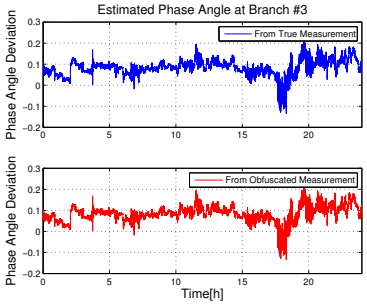
[3]The authors would like to thank Yao Liu for her help in this process.



(a) Original and Obfuscated Data from Meter #1 (Time)



(b) Original and Obfuscated Data from Meter #1 (Freq.)



(c) Estimated State at Branch #3

Fig. 2: Illegibility Test: This figure shows the obfuscated meter data from one meter and the estimated state at one branch to illustrate the efficacy of the obfuscationin the IEEE 30-bus test system. The obfuscation is done with a normal random variable with its standard deviation 20. (a) shows the original power consumption data from a meter in the time domain. (b) is the original and obfuscated data in frequency domain. We see the obfuscated data has almost randomly distributed frequency response whereas the original data has mostly low frequency components. It is difficult to manually inspect the data. (c) shows that the obfuscation does not change the estimated state. The estimated state from the obfuscated data (bottom) is the same as the estimated state from the original data (top).

3) Design Parameters: We evaluate and compare the performance of the obfuscation scheme by varying the distribution of the design parameters. Four probability distributions are selected including Gaussian, Exponential, Pareto, and Chi squared distributions due to their resilience properties. In addition to Gaussian distribution, Pareto distribution is chosen because it is a heavy tailed distribution that makes statistical analysis of randomized signal more difficult [22]. Exponential distribution is known to exhibit the maximum entropy property among all the other random distributions. Chi squared distribution is the sum of a number of independent Gaussian random variables which effectively provides the design parameters higher degrees of freedom by superimposing many independent Gaussian random variables. Table II summarizes the selected random distributions and their design parameters. The design parameters are chosen to guarantee that the obfuscation vectors are sufficiently larger than the consumption data vectors so that the obfuscated data have poor signal-to-noise ratio.

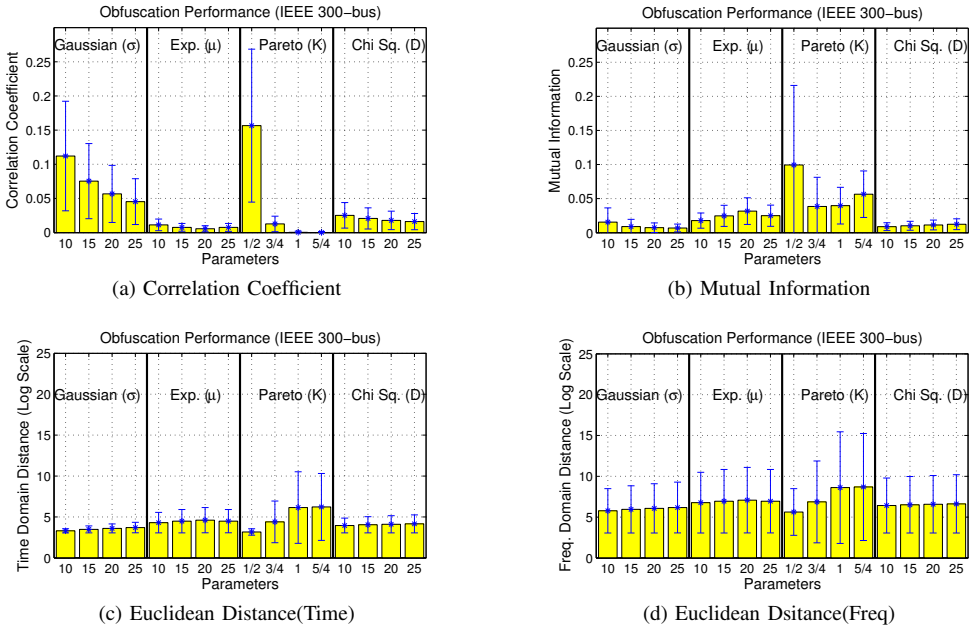
B. Illegibility Test

Illegibility is the level of difficulty for human inspectors to infer any meaningful behavioral information. Figure 2 illustrates the obfuscated power consumption of meter #1 and the estimated state at branch #3 from an instantiation of obfuscation in the IEEE 30-bus system that consists of 112 metering points and 29 states. From the bottom plot of Figure 2a, we see that the obfuscated consumption data look like random variables, while the original consumption data show clear appliance usage patterns. Figure 2b shows the frequency response of the original data and the obfuscated data of meter #1. The original data contain mostly low frequency components, whereas the obfuscated data have almost flat frequency responses which making it difficult to interpret the obfuscated data. Nevertheless, the estimated state from the obfuscated data is exactly the same as that from the original data from Figure 2c. The high level of obscurity from the obfuscated data shown in Figure 2 remains similar regardless of the probability distribution of the design parameters.

C. Resilience Test

To evaluate the resilience of our obfuscation scheme, four performance metrics are used: *Pearson's correlation coefficient*, *Mutual information*, *Euclidean distance in time-domain*, and *Euclidean distance in frequency domain*. They are widely adopted in data mining and pattern recognition for time series data [7], [19]–[21], [23], [24], as well as for evaluating privacy with meter data [9]. By evaluating the performance with different design parameter choices, we show that our obfuscation scheme is resilient to automated attacks with empirical evidence.

Figure 3 shows the error bar plots representing the quality of obfuscation across all pairs of the original and obfuscated measurements. We can see high level of resilience in all of the four evaluation metrics. As the uncertainty of the random variable increases (with greater variance for Gaussian, bigger mean for Exponential, heavier tail for Pareto, and higher degree of freedom for Chi squared distributions), the level of the obscurity becomes higher. In Figure 3a, for example, the correlation between the obfuscated and original data becomes smaller as the variance of the Gaussian random variable increases. The correlation between the two identical signal is 1. From our results, the correlation between the obfuscated and original data are mostly smaller than 0.1, which indicates that it is very hard for matching algorithms to set a proper threshold. Figure 3b shows the changes in mutual information between the original and obfuscated data. The mutual information between the two identical meter data is around 1.6 whereas the mutual information between the original and obfuscated data is mostly less than 0.05. This indicates that the obfuscated data is quite hard from the original data, making it difficult for mutual-information based data mining algorithms to discover related patterns. Similar to the correlation coefficient, it decreases as the uncertainty of the random variables for the design parameters increase. Figure 3c and 3d illustrate the Euclidean distance between



(a) Correlation Coefficient

(b) Mutual Information

(c) Euclidean Distance(Time)

(d) Euclidean Dsitance(Freq)

Fig. 3: Obfuscation performance in IEEE 300-bus system: the performance of the obfuscation is mainly determined by the level of uncertainty of the random variables for the design parameters.

the original and obfuscated data both in time and frequency domains. As we expected, the distance between the original and obfuscated data is more than orders of magnitude, which demonstrates its resilience to the Euclidean distance based data mining algorithms such as clustering.

Figure 4 shows the correlation coefficient and mutual information in the IEEE 9, 14, 30, 118, and 300 test systems varying the standard deviation of the Gaussian random variables. We can see that the obfuscation performance has little dependency on the complexity of the bus system (See table I for the number of meters and state variables), while the standard deviation gives greater impact to the level of obscurity, which implies that the obfuscation design parameters can be independently chosen regardless of the system complexity.

D. Communication Overhead

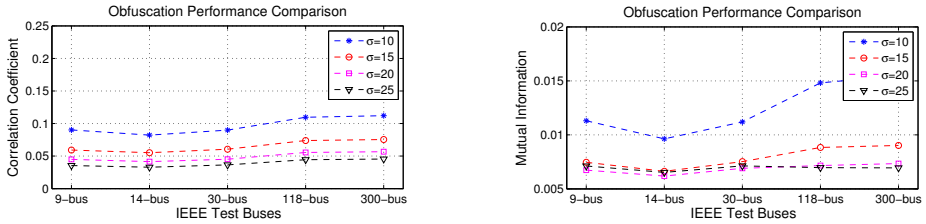
The lead meter sends m obfuscation variables to every meter at every sampling time t. Each meter then sends 1 variable to the utility provider, so that the utility provider receives m variables. Overall the communication overhead for the payload is $2m$. Assuming that 4-byte float data type is used for meter reading, the additional overhead is approximately 4.5 kbytes/sec when running on a IEEE-300 Test Bus System.

V. Related Work

A number of privacy preservation techniques have been investigated in different context for smart grid. Molina-Markham et al. [10] discussed a privacy preserving billing protocol in smart grid. The authors have shown a zero-knowledge billing protocol, in which the utility company does not need raw meter data for billing. Our main goal is different in that we focus on exact state estimation while protecting the privacy of personal behavior. Li et al. [11] proposed a secure in-network aggregation method. The authors use the Paillier crypto-system, an additive homomorphic system, to secure the data aggregation network from network level adversaries. Besides our different focus, state estimation indeed requires both multiplicative and additive homomorphic properties (fully homomorphic), which can only be supported with high-end computing power. In practice, a light-weight implementation that can run in a smart meter platform is still an open question [11], [25], [26]. Kalogridis et al. [9] proposed an energy storage system that protects privacy by managing energy usage within home before metering data are collected. A power mixing algorithm was developed to avoid possible privacy leaks. While we share a similar goal of privacy protection, our approach is solely based on the original system design from the utility company and does not require additional energy management facility at home. On the other hand, our approach can complement the above work to constitute a privacy preserving system with different levels of privacy support. Efthymiou et al. [26] proposed an anonymization technique to protect the identity of high-frequency meter data through an escrow service. Different from consumer identity, our work considers protection towards the "privacy of personal behavior" [14]. Apart from the high-frequency meter data, our work also supports fine-grained monitoring with low-frequency data down to home or smart meter level, while still protecting the privacy of customers' daily activities.

VI. Conclusions and Future Work

We proposed a cooperative state vector estimation scheme that can protect "the privacy of personal behavior" for smart grid customers without compromising the performance of state estimation. Our scheme exploits the kernel of an electric grid configuration matrix and supports error free state estimation based on a distortion free obfuscation vector space. We evaluated the obfuscation performance with 1349 meter data in



(a) Correlation Coefficient (b) Mutual Information

Fig. 4: Performance change over 5 different test systems: the obfuscation performance has little dependency on the complexity of the bus system. This plot shows the correlation and mutual information by changing the variance values in 5 test systems (Lower is better). The change slightly depends on the test systems. However it is not significant compared to the change due to the variance.

5 IEEE Test Electric Bus Systems that are portions of the Middlewestern US electric grids. The empirical evaluation showed that the obfuscation is illegible to human inspectors, and is resilient to automatic attacks that use correlation, mutual information, Euclidean distance in time and frequency domains as similarity metrics.

Although we have provided empirical evidence, more rigorous privacy guarantees remain to be studied in the future. A statistical measure is necessary to quantify how indistinguishable the obfuscated meter data are. Systematic design of the obfuscation vector is another future work. For now, we choose the design parameters such that the obfuscation vector is much larger than the consumption vector to ensure low signal-to-noise ratio. However, electricity consumption may exhibit regular patterns, which can be used for regularization constraints. To design an obfuscation scheme that is resilient to regularization based attacks, we have to understand more about the consumption patterns and possible attack scenarios. Since our approach significantly modifies the meter data. It's necessary to study the impact on bad data detection and isolation [17].

ACKNOWLEDGEMENT

This material is based upon work supported in part by the NSF under award # 0820061 and 0905580, by UCLA's Center for Embedded Networked Sensing, and by the VINNMER program funded by VINNOVA, Sweden. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF, the UCLA, or VINNOVA.

REFERENCES

[1] L. D. Saco, N. Orchard, T. Vorisek, J. Parsons, C. Rochas, A. Z. Morch, V. Lopez, and M. Togeby, "Definition of Smart Metering and Applications and Identification of Benefits," in *European Smart Metring Alliance Project Report*, 2008.

[2] P. Vinter and H. Knudsen, "Using continuous state estimation in grid planning," in *International Conf. on Electricity Distribution*, 2009.

[3] J. Chen and A. Abur, "Placement of PMUs to enable bad data detection in state estimation," *IEEE Trans. on Power Systems*, 2006.

[4] S. Yunting, M. Shiying, W. Lihua, W. Quan, and H. Hailei, "PMU placement based on power system characteristics," in SUPERGEN 2009.

[5] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation and Control*. John Wiley and Sons, 1984.

[6] M. El Mahrsi, S. Vignes, G. Hebrail, and M.-L. Picard, "A data stream model for home device description," in RCIS 2009.

[7] Y. Kim, T. Schmid, Y. Wang, and M. B. Srivastava, "Challenges in resource monitoring for residential spaces," in *BuildSys2009*.

[8] J. Beyea, "The Smart Electricity Grid and Scientific Research," *Science*, vol. 328, no. 5981, 2010.

[9] G. Kalogridis, C. Efthymiou, S. Denic, T. Lewis, and R. Cepeda, "Privacy for smart meters: Towards undetectable appliance load signatures," in *IEEE SmartGridComm*, 2010.

[10] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, , and D. Irvin, "Private memoirs of a smart meter," in *BuildSys*, 2010.

[11] F. Li, B. Luo, and P. Liu, "Secure information aggregation for smart grids using homomorphic encryption," in *IEEE SmartGridComm*, 2010.

[12] M. Lisovich, D. Mulligan, and S. Wicker, "Inferring personal information from demand-response systems," *IEEE Security Privacy*, 2010.

[13] J. Lerner and D. Mulligan, "Taking the 'long view' on the fourth amendment: Stored records and the sanctity of the home," *Stanford Tech. Law Rev.*, vol. 3, 2008.

[14] "Smart grid cybersecurity strategy and requirements," US Nat'l Inst. for Standards and Technology, Tech. Rep., 2009.

[15] "Californiaś new landmark smart meter privacy law," *eMeter*, 2010.

[16] M. Hennessy-Fiske, "Ucla is fined $95,000 for violating patient privacy," *LA Times*, 2010.

[17] Y. Liu, M. K. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," in *ACM CCS*. 2009.

[18] R. D. Zimmerman, C. E. Murillo-Snchez, and R. J. Thomas, "Matpower's extensible optimal power flow architecture," in *IEEE Power and Energy Society General Meeting*, 2009.

[19] S. Mukherjee, Z. Chen, and A. Gangopadhyay, "A privacy-preserving technique for euclidean distance-based mining algorithms using fourier-related transforms," *The VLDB Journal*, vol. 15, no. 4, 2006.

[20] S. Lee, D. Kwon, and S. Lee, "Efficient pattern matching of time series data," in *LNCS Developments in Applied Artificial Intelligence*,2002.

[21] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. PAMI*, 2005.

[22] H. Kargupta and S. Datta, "On the privacy preserving properties of random data perturbation techniques," in *ICDM*, 2003.

[23] W. Qu, X. Li, and Q. Liu, "A time series similar pattern matching algorithm based on singularity event features," 2009

[24] H. Liu, Z. Ni, and J. Li, "Time series similar pattern matching based on empirical mode decomposition," in *ISDA '06*.

[25] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *ACM STOC*. 2009.

[26] C. Efthymiou and G. Kalogridis, "Smart grid privacy via anonymization of smart metering data," in *IEEE SmartGridComm,* 2010.