Smart Meter Privacy: A Theoretical Framework

Lalitha Sankar, *Member, IEEE,* S. Raj Rajagopalan, Soheil Mohajer, *Member, IEEE,* and H. Vincent Poor, *Fellow, IEEE*

***Abstract*— The solutions offered to-date for end-user privacy in smart meter measurements, a well-known challenge in the smart grid, have been tied to specific technologies such as batteries or assumptions on data usage without quantifying the loss of benefit (utility) that results from any such approach. Using tools from information theory and a hidden Markov model for the measurements, a new framework is presented that abstracts both the privacy and the utility requirements of smart meter data. This leads to a novel privacy-utility tradeoff problem with minimal assumptions that is tractable. For a stationary Gaussian model of the electricity load, it is shown that for a desired mean-square distortion (utility) measure between the measured and revealed data, the optimal privacy-preserving solution i) exploits the presence of high-power but less private appliance spectra as implicit distortion noise, and ii) filters out frequency components with lower power relative to a distortion threshold; this approach appears to encompass most of the proposed privacy approaches.**

***Index Terms*— smart meter, privacy, utility, rate-distortion, inference, leakage.**

I. Introduction

One of the hallmarks of the smart grid is a vastly expanded information collection and monitoring system using smart meters and other new technologies. But the information that is collected and harnessed to create a more efficient grid may potentially be used for other purposes, thereby raising the question of privacy, especially of the residential consumer whose smart meter data is being collected [1], [2].

Privacy of smart meter data has become a popular topic of research. A common proposed model of privacy loss is related to the possibility of inferring appliance usage from load data with the help of load signature libraries. Equivalently, a common proposed solution is the use of energy storage devices (such as a large rechargeable battery) to "flatten" these signatures [3], [4]. Proposals for privacy protection in smart meter data have also used aggregation along dimensions of space (using neighborhood gateways, e.g. [5]) or precision (by noise addition, e.g. [6]). However, these approaches lack a formal model of privacy and thus cannot answer some pertinent questions such as (i) is detection of appliance usage patterns the only means of losing privacy?; (ii) how much privacy is lost in such methods and to what extent do the proposed solutions staunch the loss; and (iii) how much sensitive information can and should be left in the data so that it is still useful? In other words, current approaches to privacy only provide privacy assurances, but cannot provide any guarantees.

We present a formal model for the time-series smart meter data and metrics for utility and privacy of the data. Observing that the meter data is a cumulative load consumed by the appliances that are on in an observation time window, we propose a hidden Markov model for the measurements where the underlying appliance states (on or off) determine the load measurements, which in turn are modeled as real valued correlated Gaussian random variables.

We argue that the model of privacy should be abstract and oblivious of the extraction technology since: i) a technology-specific solution that works now may not provide the same privacy assurance in the future; ii) time series meter data analysis is in its infancy and one can expect that in the future, data from smart meters may be mined to infer personal information in ways that are unknown to us presently [7]. To this end, we choose mutual information as our privacy metric. By the same token, it is likely that consumers may want to share data with third parties in some measured manner to derive some benefits (e.g. energy consumption optimization). Thus, it is essential to guarantee a measure of utility of the revealed meter data. In line with the Gaussian real value model for measurements, we quantify the utility of the distorted data by constraining the mean squared error (distortion energy) between the original and revealed signals.

Our design goal is to provide a framework to accommodate both legitimate objectives, sharing and hiding, in a fair manner without completely sacrificing either. Such an overarching framework that both quantifies privacy and provides a means for measuring the tradeoff between sharing (utility) and hiding (privacy) has not yet been presented. Our privacy focus is to decouple the revealed/collected meter data as much as possible from the personal actions of a consumer. This insight is based on the observation that irregular (intermittent) activity such as kettles or lights turned on manually are much more revealing of personal actions than regular (continuous) activity such as refrigerators or lights on timers. Consequently, our approach to privacy preservation is to distort the data to minimize the presence of intermittent activity in the data. We use the theory of rate distortion to precisely quantify the tradeoff between the utility (mean square distortion) and privacy (information leakage) for our proposed model. We show that the privacy-utility tradeoffs on the total load are achievable using an *interference-aware reverse waterfilling*

The research was supported in part by the National Science Foundation under Grants CCF-10-16671 and CNS-09-05398, in part by the Air Force Office of Scientific Research MURI Grant FA-9550-09-1-0643, and in part by DTRA under Grant HDTRA1-07-1-0037.

L. Sankar and H. V. Poor are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544. Email: {lalitha,poor}@princeton.edu. S. R. Rajagopalan is with HP Labs, Princeton, NJ 08540. Email: raj.rajagopalan@hp.com. S. Mohajer is with the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley. This work was done when he was at Princeton University. Email: {mohajer@eecs.berkeley.edu}

solution, which intuitively translates to suppressing low energy components.

The paper is organized as follows. In Section II, we outline current approaches to smart meter privacy. In Section III, we develop our model, metrics, and the privacy-utility tradeoff framework. We illustrate our results in Section IV and conclude in Section V.

II. Related Work

The advantages and usefulness of smart meters in general is examined in a number of papers; see for example [8] and the references therein. [3] presents a pioneering view of privacy of smart meter information: the authors identify the need for privacy in a home's load signature as being an inference violation (resulting from load signatures of home appliances) rather than an identity violation (i.e. loss of anonymity). Accordingly, they propose home electrical power routing using rechargeable batteries and alternate power sources to moderate the effects of load signatures. They also propose three different privacy metrics: relative entropy, clustering classification, and a correlation/regression metric. However they do not propose any formal utility metrics to quantify the utility-privacy tradeoff.

Recently, [9] proposes additional protection through the use of a trusted escrow service, along with randomized time intervals between the setup of attributable and anonymous data profiles at the smart meter. [2] shows, somewhat surprisingly, that even without *a priori* knowledge of household activities or prior training it is possible to extract complex usage patterns from smart meter data such as residential occupancy and social activities very accurately using off-the-shelf statistical methods. [5] and [2] propose privacy-enhancing designs using neighborhood-level aggregation and cryptographic protocols to communicate with the energy supplier without compromising the privacy of individual homes. However, escrow services and neighborhood gateways support only restricted query types and do not completely solve the problem of trustworthiness. [4] presents a formal state transition diagram-based analysis of the privacy afforded by the rechargeable battery model proposed in [3]. However, [4] does not offer a comparable model of utility to compare the risks of information leakage with the benefits of the information transmitted.

In, [6] the authors present a method of providing *differential privacy* over aggregate queries modeling smart meter measurements as time-series data from multiple sources containing temporal correlations. While their approach has some similarity to ours in terms of time-series data treatment, their method does not seem generalizable to arbitrary query types. On the other hand, [10] introduces the notion of partial information hiding by introducing uncertainty about individual values in a time series by perturbing them. Our method is a more general approach to time series data perturbation that guarantees that the perturbation cannot be eliminated by averaging.

III. Our Contributions

The primary challenge in characterizing the privacy-utility tradeoffs for smart meter data is creating the right abstraction – we need a principled approach that provides quantitative measures of both the amount of information leaked as well as the utility retained, does not rely on any assumptions of data mining algorithms, and provides a basis for a negotiated level of benefit for both consumer and supplier [11]. [4] provides the beginnings of such a model – they assume that in every sampling time instant, the net load is either 0 or 1 power unit represented by the smart meter readings X_k, $k = 1, 2, ...$, are a discrete-time sequence of binary independent and identically distributed values. They model the battery-based filter of [3] as a stochastic transfer function that outputs a binary sequence $\hat{X}_k$ that tells the electricity provider whether the home is drawing power or not at any given moment. The amount of information leaked by the transfer function is defined to be the mutual information rate $I(X; \hat{X})$ between the random variables X and $\hat{X}$. By modeling the battery charging policy as a 2-state stochastic transition machine, they show that there exist battery policies that result in less information leakage than from the deterministic charging policy of [3]. Though [4] does not provide a general utility function to go with the chosen privacy function and the modeling assumptions are extremely simplistic, it nevertheless provides a good starting point for our framework.

In our model, we assume that the load measurements are sampled (at an appropriate frequency) from a smart meter, that they are real-valued, and can be correlated (models the temporal memory of both appliances and human usage patterns). Rather than assume any specific transfer function, we assume an abstract transfer function which maps the input load measurements X into an output sequence $\hat{X}$. As in [4], we assume a mutual information rate as a metric for privacy leakage; however, we allow for the fact that a large space of (unknown to us) inferences can be made from the meter data – we model the inferred data as a random variable Y correlated with the measurement variable X. Thus, the privacy leakage is the mutual information between Y and $\hat{X}$. We also provide an abstract utility function which measures the fidelity of the output sequence $\hat{X}$ by limiting the Euclidean distance (mean square error) between X and $\hat{X}$. Using these abstractions and tools from the theory of rate distortion we are able to meet all our requirements for a general but tractable privacy-utility framework: the privacy and utility requirements provide opposing constraints that expose a spectrum of choices for trading off privacy for utility and vice-versa. adversary too

A. Notation

Before proceeding, we summarize the notation used in the sequel. Random variables (e.g. $H_{k,j}$) are denoted with uppercase letters and their realizations (e.g. $h_{k,j}$) with the corresponding lowercase letters. $\underline{X}$ denotes an n-length vector while bold font $\mathbf{X}$ denotes a matrix. $\mathbf{I}$ denotes the identity matrix. $\mathcal{N}(\mu^n, \mathbf{\Sigma})$ denotes a n-variable real Gaussian distribution with mean μ^n and covariance $\mathbf{\Sigma}$. $\mathbb{E}(\cdot)$ denotes expectation; $(x)^+$ denotes $\max(x, 0)$; $I(\cdot;\cdot)$ denotes mutual information; $h(\cdot)$ denotes differential entropy. Finally, in the sequel we use the term *reverse waterfilling solution* to denote the rate and leakage minimizing source coding solution for Gaussian sources with memory [12, Chap. 4].

B. Model

We write x_t, $t = 1, 2, \ldots, n$, to denote the sampled load measurements from a smart meter. In general, x_t are complex valued corresponding to the real and reactive measurements and are typically vectors for multi-phase systems [13]. For simplicity and ease of presentation, we model the meter measurements as a sequence of real-valued scalars (for example, such a model applies to two-phase 120 V appliances for which one of the two phase components is zero).

At any time, the load measurements are determined by the appliances actively in use. In general, there is a finite time window during which an appliance is used and the meter measurements during that time are correlated with the specific appliance used. Furthermore, the statistics of the load measurements changes as the appliances used change, i.e., the measurement data can be viewed as being generated by a quasi-stationary source.

Let M denote the total number of appliances at a residence; since each appliance can be either on or off in any window of time, we have 2^M possible appliance states. In general, the appliance state at any time is an instantiation of a random process that is highly correlated with the personal habits of members of a household. We denote this state process by $\{S(k)\}$ such that $S(k) \in \left\{0, 1, \ldots, 2^{M-1}\right\}$ is the random state variable in the k^{th} time instant. Associated with this state, is the meter measurement variable $X(k)$ in the same time instant. Formally, we model the joint probability distribution of the states $\underline{S} = \{S(k)\}_{k=1}^{n}$ and measurements $\underline{X} = \{X(k)\}_{k=1}^{n}$ over n time instants as

$$P\left(\underline{S}, \underline{X}\right) = \prod_{k=1}^{n} P(S(k)|S^{k-1}) P\left(X(k)|S(k)\right) \quad (1)$$

$$= \prod_{k=1}^{n} P(S(k)|S_{k-1}) P\left(X(k)|S(k)\right) \quad (2)$$

where (1) results from the fact that conditioned on the state, the measurements are independent of each other and (2) follows from the fact that states are related in a causal and sequential manner, i.e., we have the Markov chain relationship, $S_k - S_{k-1} - S^{k-2}$ for all k.

Remark 1: A rechargeable battery can be viewed as the state $S = 0$ in which no appliance, as viewed by the smart meter, is on.

A hidden Markov model (HMM) (see Fig. 1) such as in (1) and (2) is typically characterized by three parameters: i) the initial state distribution; ii) a state transition matrix; and iii) and a conditional distribution, assumed Gaussian here, which captures the probability density function of a measurement x conditioned on a state s. For a stationary HMM process, the state transition matrix is the same at each time instant, and therefore, the time duration of the different states are the same on average. While in general, the duration of usage of the different appliance states may be different, for simplicity and tractability, we assume the state is held and the underlying probability distribution is stationary in a block of $n>>1$ measurements.

We now present an explicit probability model for the state and the measurements. Our model is based on the following two observations: i) a state S remains unchanged for a continuous period of time, assumed here to be n; ii) in that time,

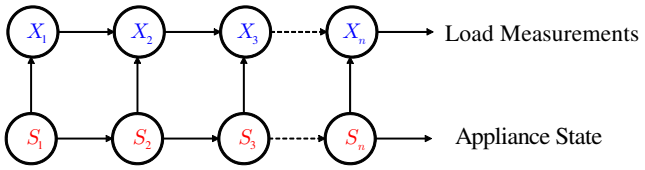

Fig. 1. A hidden Markov model for the meter measurements.

each appliance which is on in that state generates a sequence of random (assumed Gaussian distributed) measurements characteristic of the appliance signature consumption pattern (see for e.g., [14], [13]). The assumption of randomness models the variability in appliance manufacturers and voltage fluctuations. Furthermore, the assumption of normal distribution for total load is a simplification from empirical observations [15] that the power consumption pattern of a typical appliance in the on state is approximately Gaussian.

In general, appliances can be classified as being either on almost all the time such as air conditioners, computers, and refrigerators or as those that come on intermittently such as ovens, toasters, and kettles. Without loss of generality, we consider a state $S = (s_c, s_i)$ in which both a continual appliance s_c and an intermittent appliance s_i are on. Note that s_c and s_i can be viewed as states where only the said appliance is on while the remaining $M-1$ appliances are off (see Fig. 2). Let $G_c^n(s_c)$ and $G_i^n(s_i)$ denote the length n Gaussian distributed time sequences for the states s_c and s_i, respectively. While the transition between states is given by a Markov model, for any state, since a sequence is unique (in autocorrelation and spectral characteristics) for each appliance, we assume that $G_c^n(s_c)$ and $G_i^n(s_i)$ are independent of each other. Note, however, that the entries of each $G_{(\cdot)}^n\left(s_{(\cdot)}\right)$ are correlated due to memory effects. We assume that the length n is chosen such that the memory effects of each state are contained within the sequence. We henceforth model the memory via a length $m_a < n$ for state s_a such that each entry in a window of length n is affected by m_a adjacent entries.

Writing the measurements in (n-length) vector notation, we have

$$X^n(S) = G_c^n(S_c) + G_i^n(S_i) + Z^n \quad (3)$$

where $G_c^n(S_c) \sim \mathcal{N}(\mu_c, \mathbf{R}_{G_c})$ and $G_i^n(S_i) \sim \mathcal{N}(\mu_i, \mathbf{R}_{G_i})$ are independent of each other and independent of the independent and identically distributed (i.i.d.) Gaussian noise vector $Z^n \sim \mathcal{N}\left(0, \sigma^2\mathbf{I}\right)$ and the summation in (3) is a vector (entry-by-entry) summation.

Remark 2: Since the entries of $G_c^n(S_c)$ (resp. $G_i^n(S_i)$) are correlated with each other, in general, each entry of $G_c^n(S_c)$ can be written as a function of its past and future entries and a term independent of them, such as, for example, an autoregression model. For a more general analysis, however, we do not restrict ourselves to any specific correlation model.

Thus, the covariance matrix $\mathbf{R}_X$ has entries $\{\mathbb{E}[X_j X_k]\}_{j,k=1}^{n} \equiv [R_X(j,k)]_{j,k=1}^{n} = [R_X(|j-k| \bmod n)]_{j,k=1}^{n}$ of $X^n(S)$ in (3) is a Toeplitz

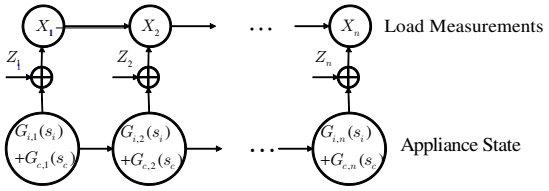

Fig. 2. Meter measurements obtained as a noisy sum of two Gaussian processes, corresponding to the intermittent and continous appliances, respectively over a time window of length n.

matrix with the autocorrelation entries

$$R_X(|j-k|) = \begin{cases} R_{G_c}(|j-k|) + R_{G_i}(|j-k|) + \sigma^2\delta_{|j-k|}, \\ \quad |j-k| = 0, 1, 2, \ldots, m < n, \\ 0, \text{ otherwise;} \end{cases} \tag{4}$$

where $R_{G_{(\cdot)}}(l) = 0, l \in \{m_{(\cdot)}+1, ..., n\}, m = \max(m_c, m_i)$, and δ_l is the Kronecker delta function which is non-zero only for $l = 0$ and 0 otherwise. The circular n-block model for the autocorrelation allows us to use the discrete Fourier transform (DFT) to decompose an n-length correlated sequence into n independent Gaussian measurements subject to the same distortion and leakage constraints, i.e., the Fourier basis is the eigen basis for each of the state sequences. Denoting the DFT matrix by $\mathbf{T}$, we have

$$X_F^n = \mathbf{T}X^n \text{ s.t.} \tag{5}$$

$$\mathbf{F}_X = \mathbb{E}\left[\mathbf{T}X^n (X^n)^\dagger \mathbf{T}^\dagger\right] \tag{6}$$

$$= \mathbf{T}\mathbf{R}_X\mathbf{T}^\dagger \tag{7}$$

i.e., $\mathbf{F}_X$ is a unitary transformation of $\mathbf{R}_X$ and has entries $F_X(k)$, referred to as the power spectral density of the Gaussian process $\{X_n\}$. Thus, the DFT X_F^n of X^n has entries $X_{F,k}$, $k = 1, 2, \ldots, n$, that are independent but not identically distributed Gaussian r.v.'s with variance $F_X(k)$ for the k^{th} entry.

For a unitary transform, the distance-based distortion and mutual information based leakage constraints remain unchanged [12, Chap. 4]. We will use this property in the sequel to determine the minimal leakage fidelity-preserving mapping of the measurement data. The aim of such a privacy-preserving technique is to suppress the signatures of appliances used intermittently as they significantly compromise the end-user privacy relative to the continually running appliances [16].

C. Utility and Privacy Metrics

Since continuous amplitude sources cannot be transmitted losslessly over finite capacity links, a sampled sequence of n load measurements X^n is compressed before transmission. In general, however, even if the sampled measurements were quantized *a priori*, i.e., take values in a discrete alphabet, there may be a need to perturb (distort) the data in some way to guarantee a measure of privacy. However, such a perturbation also needs to maintain a desired level of fidelity.

Intuitively, utility of the perturbed data is high if any function computed on it yields results similar to those from the original data; thus, the utility is highest when there is no perturbation and goes to zero when the perturbed data is completely unrelated to the original. Accordingly, our utility metric is an appropriately chosen average 'distance' *distortion function* between the original and the perturbed data.

Privacy, on the other hand, is maximized when the perturbed data is completely independent of the original. The meter measurements are a result of a specific choice of appliance states which in turn are correlated with the personal habits of the users. Our privacy metric measures the difficulty of inferring private information leaked by the appliance state via the meter measurements. More generally, any desired private information of the data collector's choice, defined as a sequence $\{Y_k\}$ of r.v.'s $Y_k \in \mathcal{Y}$, $-\infty < k < \infty$, which is correlated with the measurement sequence can be inferred from the revealed data. The random sequence $\{Y_k\}$ for all k along with the joint distribution $p_{X^nY^n}$ mathematically captures the space of all inferences that can be made from the measurements. We quantify the resulting privacy loss as a result of revealing perturbed data via the *mutual information* between the two data sequences.

Remark 3: While the space of Y^n sequences can be potentially large, in any time window, one can restrict the analysis to a subset of inferences that are correlated with the appliance state, and correspondingly, with the meter measurements for that window. Since the appliance state determines the meter measurements and also reveals private information of the consumer, we have the following Markov chain: $Y^n - S^n - X^n$, i.e., the joint distribution of (Y^n, S^n, X^n) can be written as

$$P(Y^n, S^n, X^n) = P(S^n) P(Y^n|S^n) P(X^n|S^n). \tag{8}$$

Remark 4: Our model of privacy is between a single user (household) and the electricity provider. It does not consider the leakage possibilities of comparing the perturbed data from two or more different users. On the other hand our model can be extended to address the availability of side-information at the data collector such as income level of the user that may cause further information leakage by incorporating the statistical knowledge of the side information at the meter.

D. Privacy-Preserving Mapping

A smart meter can enable load consumption monitoring at a fine-grained level such as over 15 minute intervals [16, p. 12]; this in turn determines the sampling rate and quantization levels for the meter measurements. The resulting stream of continuous valued discrete data has to be communicated over a finite rate, such as a wireless, link. For efficient transmission, one can exploit the correlations in a window of measurements to compress efficiently. The quality of compression is determined by the fidelity desired of the output, i.e., the utility of the revealed measurements as discussed earlier.

A privacy-preserving mapping also needs to ensure that a minimal amount of information can be inferred about the personal habits. We abstract the resulting problem to one of mapping every meter data sequence to an appropriate sequence that satisfies both the utility (fidelity) and privacy (leakage) constraints. Formally, an (n, M, D, L) code involves an encoder and a decoder described below:

Encoding: In each time window, the meter collects $n \gg 1$ measurements prior to communication. Recall that the corresponding state in this time window is then S_k. The encoding function is then a mapping of the resulting *source sequence* $X^n(S_k) = (X_1\ X_2\ X_3\ \ldots\ X_n)$, for all $k = 1, 2, \ldots, n$, given by

$$F_E : \mathcal{X}^n(\mathcal{S}) \to \mathcal{M} = \left\{1, 2, \ldots, 2^{nR}\right\} \tag{9}$$

where F_E maps the sequence $X^n(S)$ to an index $M \in \mathcal{M}$ which represents a quantized sequence.

Decoding: The decoder (at the data collector) computes an output sequence $\hat{X}^n = \left(\hat{X}_1\ \hat{X}_2\ \hat{X}_3\ \ldots\ \hat{X}_n\right)$, $\hat{X}_k \in \mathbb{R}$, for all k, using the decoding function

$$F_D : \mathcal{M} \to \hat{\mathcal{X}}^n. \tag{10}$$

The encoder is chosen such that the input and output sequences achieve a desired utility given by an average distortion constraint

$$D = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[\left(X_k - \hat{X}_k\right)^2\right] \tag{11}$$

and a constraint on the information leakage about the desired sequence $\{Y_k\}$ from the revealed sequence $\left\{\hat{X}_k\right\}$ is quantified via the leakage function

$$L = \frac{1}{n} I\left(Y^n; \hat{X}^n\right) \tag{12}$$

where $\mathbb{E}[\cdot]$ denotes the expectation over the joint distribution of X^n and $\hat{X}^n$ given by $p_{X\hat{X}}(x^n, \hat{x}^n) = P_{X^n}^n(x^n)\, p_t(\hat{x}^n|x^n)$ where $p_t(\hat{x}^n|x^n)$ is a conditional pdf on $\hat{x}^n$ given x^n. The mean-square error (MSE) distortion function chosen in (11) is typical for Gaussian distributed real-valued data as a measure of the fidelity of the perturbation (encoding). Some examples of the inference sequence Y^n are the known signature sequences for specific appliances which typically leak the most information about the personal habits of a consumer. Thus, Y^n can include the signature sequences for appliances such as kettles, toasters, and appliances which come on at unexpected times or are unusual in usage pattern.

Remark 5: The encoding scheme presented here, is inspired by the theory of rate-distortion in which the focus is on determining the minimal rate at which to compress a data source for a desired fidelity (distortion) level. However, the aim of the encoding here is to guarantee a minimal level of leakage L for a desired fidelity (distortion) D. We formalize this tradeoff below.

E. Utility-Privacy Tradeoff Region

Formally, the utility-privacy tradeoff region $\mathcal{T}$ is defined as follows.

Definition 6: The smart meter utility-privacy tradeoff region $\mathcal{T}$ is the set of all (D, L) pairs for which there exists a coding scheme given by (9) and (10) with parameters $(n, M, D+\epsilon, L+\epsilon)$ satisfying (11) and (12) for n sufficiently large and $\epsilon > 0$.

In classical rate-distortion theory, the constraint is on the number M of encoded (quantized) sequences such that the rate in bits per entry of the sequence is bounded as $M \leq 2^{n(R+\epsilon)}$. The aim then is to determine the infimum of all rates $R(D)$ that is achievable for a desired distortion D. Here, we seek to minimize the average number of bits per entry that is leaked of the correlated sequence Y^n that we wish to hide from the revealed sequence $\hat{X}^n$; while minimizing $R(D)$ is also be desirable from a communication standpoint, minimizing both via the coding scheme may, in general, not be feasible except for specific cases, and therefore, we do not explicitly consider the rate minimization. Formally, the minimal leakage $\lambda(D)$ for a desired fidelity D is defined as follows.

Definition 7: The minimal leakage $\lambda(D)$ achievable for a desired distortion D for a source with memory subject to distortion and leakage constraints in (11) and (12) is given by

$$\lambda(D) = \lim_{n\to\infty} \inf_{p(x^n, y^n)p(\hat{x}^n|x^n)} \frac{1}{n} I\left(Y^n; \hat{X}^n\right). \tag{13}$$

The closure of the set of all achievable distortion-leakage (D, L) pairs is the distortion-leakage region such that the minimal leakage (boundary) is $\lambda(D)$ for any D.

Remark 8: The Markov relationship $Y^n - X^n - \hat{X}^n$ is captured via the set of all distributions in (13) which minimize $\lambda(D)$.

Remark 9: If an additional constraint on minimizing the encoding rate is included, the minimal achievable rate for a desired distortion is given by

$$R(D, L) = \lim_{n\to\infty} \inf_{p(x^n, y^n)p(\hat{x}^n|x^n)} \frac{1}{n} I\left(X^n; \hat{X}^n\right) \tag{14}$$

For $Y_k = X_k$, for all k, i.e., for the case in which the actual measurements need to be private, $\lambda(D) = R(D, L) = R(D)$ where $R(D)$ is the rate-distortion function for the source.

In general, the optimal distribution minimizing the leakage subject to a distortion constraint depends on the joint distribution of the state, measurement, and inference sequences. Modeling this relationship is, in general, not straightforward or known *a priori*. However, since the revealed measurements leak information about the appliance state which in turn can lead to a large set of inferences, we focus directly on the problem of minimizing the leakage of specific states via the revealed data.

F. Privacy-Preserving Spectral Waterfilling

In general, the problem of suppressing specific appliance signatures requires detection of the appliance states at the meter in a given window of time to determine the appliances to suppress. To avoid dependence on any specific appliance detection algorithm, we assume the existence of an external algorithm that can detect (with perfect accuracy) in a given window of time which appliances changed state from off to on or vice versa. One can broadly describe the signal in any such window as a noisy sum of signals from intermittently (more revealing of personal details) and continuously (less revealing) used appliances, with states s_i and s_c, respectively, as given by the model in (3). Thus, we henceforth focus on the problem of suppressing the state s_i relative to the state s_c and the measurement noise.

We consider the state and measurement model in (3) and determine a lower bound on the leakage possible in each window of n measurements. Specifically, we seek to hide the intermittently used appliance by choosing the inference sequence as $Y^n = G_i^n$, and thus, our aim is to minimize the leakage

$$L = \frac{1}{n} I\left(G_i^n; \hat{X}^n\right) \tag{15}$$

in a window of n measurements. Recalling that the DFT is a unitary transformation that preserves Euclidean distance and mutual information, we have

$$\begin{aligned}
L &= \frac{1}{n} I\left(G_{i,F}^n; \hat{X}_F^n\right) && (16)\\
&= \frac{1}{n} h\left(G_{i,F}^n\right) - \frac{1}{n} h\left(G_{F,i}^n | \hat{X}_F^n\right) && (17)\\
&= \frac{1}{n} \textstyle\sum_{k=1}^n \log\left(2\pi e F_{G_i}(k)\right) - \frac{1}{n} h\left(G_{F,i}^n - \hat{X}_F^n | \hat{X}_F^n\right) && (18)\\
&\geq \frac{1}{2n} \textstyle\sum_{k=1}^n \log\left(2\pi e F_{G_i}(k)\right) - \frac{1}{n} h\left(G_{F,i}^n - \hat{X}_F^n\right) && (19)\\
&\geq \frac{1}{2n} \textstyle\sum_{k=1}^n \log\left(2\pi e F_{G_i}(k)\right) - \frac{1}{n} h_G\left(G_{F,i}^n - \hat{X}_F^n\right) && (20)\\
&= \frac{1}{2n} \textstyle\sum_{k=1}^n \log\left(2\pi e F_{G_i}(k)\right) && (21)\\
&\quad - \frac{1}{2n} \textstyle\sum_{k=1}^n \log\left(2\pi e\left(F_{G_c}(k) + \Delta(k) + \sigma^2\right)\right)\\
&= \frac{1}{2n} \textstyle\sum_{k=1}^n \left[\log\left(\frac{F_{G_i}(k)}{\left(F_{G_c}(k) + \Delta(k) + \sigma^2\right)}\right)\right]^+ && (22)
\end{aligned}$$

where (17) follows from the expansion of mutual information, (18) follows from the fact that $\{X_{F,k}\}$ for all k are independent Gaussian distributed r.v.'s, (19) follows from the fact that conditioning does not increase mutual information, (20) follows from the fact that for a fixed variance, Gaussian r.v.'s have the maximal entropy (h_G denotes the entropy of a Gaussian r.v.), i.e., choosing $\hat{X}_F^n$ as independent Gaussian r.v.'s which implies from (11), we have

$$X_F^n = \hat{X}_F^n + Q_F^n \tag{23}$$

where Q_F^n is a sequence of independent Gaussian r.v.'s (intuitively viewed as quantization noise) independent of $\hat{X}_F^n$, (21) follows from (3), (23), and by setting $\Delta(k) \equiv |X_F(k) - \hat{X}_F(k)|^2$ such that $\sum_{m=1}^n \Delta(k) = D$, and finally, (22) follows from the positivity of the mutual information, i.e., $h(G_{F,i}^n | \hat{X}_F^n) < \min(h(G_{F,i}^n - \hat{X}_F^n), h(G_{F,i}^n))$. The optimization in (21) results in the following distortion allocation solution across the frequencies:

$$\begin{aligned}
\Delta(k) = \min\Big(&\left(F_i(k) - F_c(k) - \sigma^2\right)^+,\\
&\left(\lambda - F_c(k) - \sigma^2\right)^+\Big),\ k = 1, 2, \ldots, n
\end{aligned} \tag{24}$$

where the first term in the minimum in (24) comes from the requirement that in (21), for non-negative leakage, the denominator is upper bounded by the numerator and the second term is a result of the optimization in which λ is the Lagrangian variable satisfying the distortion constraint in (11). One may view λ as a *water-level* across the frequencies such that at each frequency only that portion of the spectrum is revealed which is strictly above λ.

The resulting minimal leakage $L \equiv L_\lambda(D)$ in the limit of large n is given by

$$L_\lambda(D) \geq \int_{f: F_i(f) > \lambda > F_c(f) + \sigma^2} \frac{1}{2} \log\left(\frac{F_i(f)}{\lambda}\right) df \tag{25}$$

where f denotes real valued frequencies; the corresponding distortion spectrum is given by

$$\Delta(f) = \begin{cases} 0; & \begin{array}{l} F_i(f) < F_c(f) + \sigma^2; \text{ (or)} \\ \lambda < F_c(f) + \sigma^2 < F_i(f) \end{array} \\ D_1(f); & F_c(f) + \sigma^2 < \lambda < F_i(f) \\ D_2(f); & \lambda > F_i(f) > F_c(f) + \sigma^2 \end{cases} \tag{26}$$

where

$$D_1(f) = \left(\lambda - F_c(f) - \sigma^2\right)^+ \tag{27}$$

$$D_2(f) = \left(F_i(f) - F_c(f) - \sigma^2\right)^+. \tag{28}$$

Note that the term inside the integral in (25) can be viewed as the leakage at each frequency f for a distortion λ. While (25) provides a lower bound on L, the bound can be achieved by using an independent encoding scheme at each frequency subject to an average distortion constraint. In practice, the bound can be approached using techniques such as sub-band coding as used in common audio and image compression formats.

To better understand the solution, we now describe the solution in detail starting from the simplest case of $F_c(f) = \sigma^2 = 0$:

- Case 1: $F_c(f) = 0$, for all f, and $\sigma^2 = 0$ such that $X^n = S_i^n$, i.e., the random sequence in a window of time is a noiseless sequence resulting from having only the appliance S_i in the on state. For this case, since $Y^n = S_i^n$, we wish to reveal X^n subject to a fidelity constraint in (11) and hide X^n subject to a leakage constraint in (12). Let λ_1 denote the water-level for this case. From (24), the solution $\Delta(k) = \lambda_1$ for all k leads to the reverse waterfilling level solution that minimizes the rate for a desired distortion for Gaussian sources with memory. This is because now the expressions for both rate and leakage in (14) and (13) coincide. The privacy-preserving rate-distortion optimal scheme thus reveals only those frequency components with power above the water-level λ. Furthermore, at every frequency only the portion of the signal power which is above the water level λ is preserved by the minimum-rate sequence from which the source can be generated with an average distortion D.
- Case 2: $F_c(f) = 0$, for all f, such that $X^n = S_i^n + Z^n$, i.e., the random sequence in a window of time is a noisy sequence resulting from having only the appliance S_i in the on state. Since measurement noise reduces the fidelity of the appliance signature, we expect that the average leakage to be lesser than that for Case 1. Let λ_2 denote the water-level for this case. The requirement in (24) that $\Delta(f) = \left(\lambda_2 - \sigma^2\right)^+$ implies that for a fixed distortion D, $\lambda_2 > \lambda_1$ for $\sigma^2 > 0$. Furthermore, since $\Delta(f) \leq \left(F_i(f) - \sigma^2\right)^+$, in general, a smaller set of

frequencies, relative to Case 1, are preserved for which the signal power is above the noise power since otherwise the noise suffices to hide the signal. Finally, the average leakage in each preserved frequency, is $\log\left(F_i\left(k\right)/\lambda_2\right) < \log\left(F_i\left(k\right)/\lambda_1\right)$, i.e., the presence of noise can aid in hiding the appliance signature we wish to not reveal.

- Case 3: The observations from Case 2 carry forth to this case also since now S_c^n can also be viewed as noise except with non-identical variances across the frequencies. Thus, only those frequencies are revealed for which $F_i\left(f\right) > \min\left(F_c\left(f\right)+\sigma^2, \lambda\right)$ or $F_c\left(f\right) < F_c\left(f\right)+\sigma^2$. In the latter case, the power of the noise and the continuous appliance signal suffices to suppress the signal to be hidden and therefore, no additional distortion is needed. On the other hand, in the former case, only the signal above the distortion level of $\max\left(F_c\left(f\right)+\sigma^2, \lambda\right)$ is preserved.

Remark 10: For all three cases above, the minimal (compression) rate $R\left(D\right)$ required to achieve a distortion D also results from a water-filling solution except the solution is different in the presence of noise and other appliance signatures is $R\left(D\right) = \int_{f:F_i(f)>\lambda} \log\left(Q\left(f\right)/\lambda\right)/2df$ where for $Q\left(f\right)$ is $F_i\left(f\right)$, $F_i\left(f\right)+\sigma^2$, or $F_i\left(f\right)+F_c\left(f\right)+\sigma^2$ for cases 1, 2, and 3, respectively where λ is chosen such that the distortion spectrum is $\Delta\left(f\right) = Q\left(f\right)$ if $Q\left(f\right) < \lambda$ and $\Delta\left(f\right) = \lambda$, otherwise. Thus, the optimal compression solution does not distinguish between the different signatures or noise in contrast to the reverse water-filling solution which minimizes the leakage and therefore distorts all those frequencies in which the energy (power) of the signal to be suppressed is higher than the water-level.

Remark 11: From (26), we see that at those frequencies in which the power of the state s_i to be suppressed is dominated by the power of the noise and the state s_c, the distortion required is zero. While this suffices for minimizing the leakage, transmitting the data at such frequencies may require additional compression. More generally, this suggests that the combined problem of rate and leakage minimization has to be considered jointly.

Remark 12: While leakage-preserving distortion ensures privacy, the utility in terms of average load consumption is reduced by the distortion level D. However, the knowledge of the distortion level suffices to estimate the average load consumed at the provider end without any loss of privacy.

IV. Illustration

We now illustrate our results with the following examples. Specifically, we model the continuous and intermittent appliance load sequences in (3) as (time-limited) Gauss-Markov processes with an auto-correlation function given by

$$R_{G_{(l)}}\left(k\right) = \begin{cases} P_{(l)}\rho_{(l)}^{-|k|}, & k = 0, \pm1, \pm2, \ldots, m_{(l)}, \\ 0 & k > m_{(l)} \end{cases}, l = i, c \tag{29}$$

where $P_{(l)}$ is the variance, $\rho_{(l)}$ is the correlation coefficient which falls geometrically with increasing difference in measurement indices k, and $m_{(l)}$ is the memory of the l^{th} appliance type, $l = i, c$. The power spectral density (PSD) of this process is given by

$$\mathcal{S}(\omega) = \sum_{k=-\infty}^{\infty} P_{(l)}\rho_{(l)}^{-k}\exp(ik\omega), \quad -\pi \le \omega \le \pi. \tag{30}$$

For the following discussion, we choose the parameters in (29) as follows: $P_i = 12$, $P_c = 8$, $\rho_i = 0.4$, $\rho_c = 0.8$, $m_i = 40$, and $m_c = 120$. These parameters model the observation that the continuously used appliance (state s_c) has a longer memory and a larger correlation coefficient relative to the intermittently used appliance (state s_i); furthermore, while the overall power consumption of state s_c is higher than that of state s_i, the bursty usage pattern of state s_i is incorporated via a larger value for P_i relative to P_c. We choose two different values for the distortion D as 4 and 6.

In Figs. 3 and 4, we plot the PSDs $F_i\left(f\right)$, $F_c\left(f\right)+\sigma^2$, and $F_i\left(f\right)+F_c\left(f\right)+\sigma^2$ of the processes $\{G_i\}$, $\{G_c+Z\}$, and $\{X\}$, respectively, for the parameters described above. Also plotted is the waterlevel λ and the distortion spectrum $\Delta\left(f\right)$. From both figures, we see that the distortion spectrum is zero when the PSD of the noisy continuous process dominates $F_i\left(f\right)$ or the waterlevel λ leading to zero and minimal leakage, respectively, for the two cases. The waterlevel λ determines

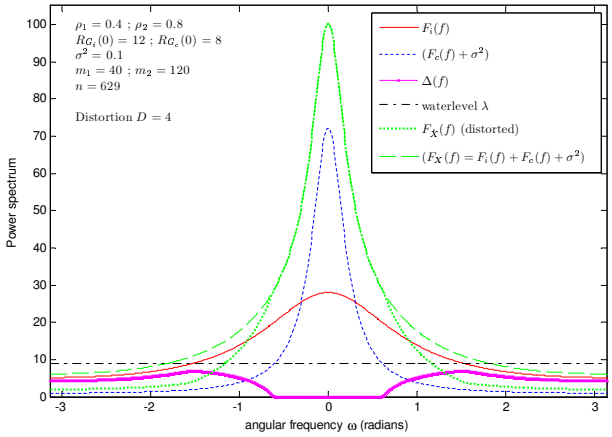

Fig. 3. Signal PSDs, distortion spectrum, and waterlevel λ for $D = 4$.

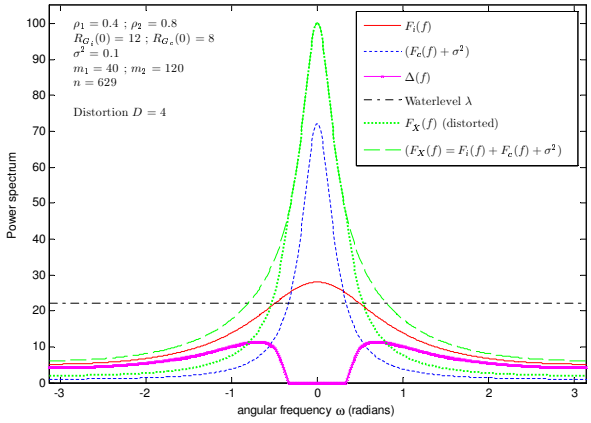

Fig. 4. Signal PSDs, distortion spectrum, and waterlevel λ for $D = 6$.

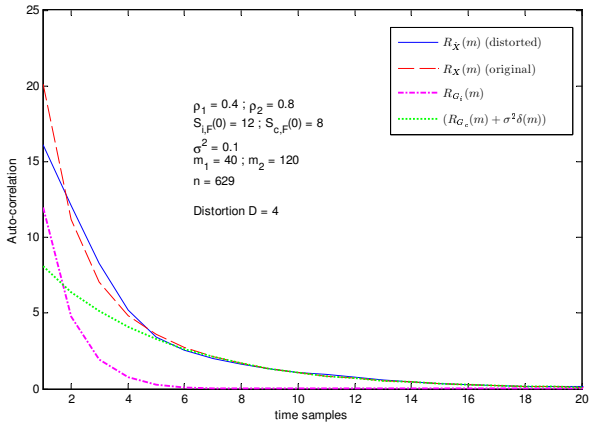

Fig. 5. Time series autocorrelation for the original and distorted signals for $D = 4$.

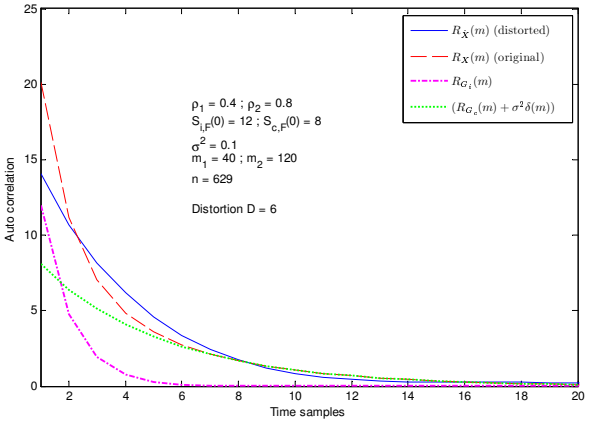

Fig. 6. Time series autocorrelation for the original and distorted signals for $D = 6$.

the distortion level otherwise.

In Figs. 5 (for $D = 4$) and 6 (for $D = 6$), we plot the time series auto-correlation functions $R_X(k)$, $R_{\hat{X}}(k)$, $R_{G_i}(k)$, and $R_{G_c+Z}(k)$ for the processes $\{X\}$, $\{\hat{X}\}$, $\{G_i\}$, and $\{G_c + Z\}$, respectively. We note that the effect of the distortion is captured in a reduction of the variance ($k = 0$ term) of the $\{\hat{X}\}$ process relative to the $\{X\}$ process by D. Furthermore, while the slope of the $\{X\}$ process is dominated by the auto-correlation of the state s_i as observed by comparing the curves for $R_X(k)$ with $R_{G_i}(k)$, the slope of $R_{\hat{X}}(k)$ matches that of $R_{G_c}(k)$. Thus, the signal energy remaining in $\hat{X}$ is dominantly due to the noisy continuous state s_c process.

V. Discussion and Concluding Remarks

Preserving privacy in a measured and flexible way is a paramount societal challenge for smart meter deployment. At the same time, any privacy techniques that dramatically alter the usefulness of smart meter data are not likely to be adopted. The theoretical framework that we have developed here allows us to precisely quantify the utility-privacy tradeoff problem in smart meter data. Given a series of smart meter measurements X, we have revealed a perturbation $\hat{X}$ that allows us to guarantee a measure of both privacy in X and utility in $\hat{X}$. The privacy guarantee comes from the bound on information leakage while the utility guarantee comes from the upper bound on the MSE distance between X and $\hat{X}$.

Our information leakage model of privacy does not depend on any assumptions about the inference mechanism (i.e. the data mining algorithms); instead it presents the least possible (on average) guarantee of information leakage about X, while the utility is preserved in an application-agnostic manner. Our framework is also agnostic about how the perturbation is achieved; for example, it can be achieved using a filter such as a battery or by adding noise or by some novel technique yet to be discovered.

Our model captures the dynamic nature of the appliance states and the smooth continual nature of the measurements via a hidden Markov model and correlated Gaussian measurements, respectively. We have extended classical results from rate distortion theory to obtain tight bounds on the amount of privacy that can be achieved for a given level of utility and vice-versa. We have shown that the critical parameter of choice in the tradeoff is the water level λ, which in turn depends on the distortion bound D that is acceptable. In a practical context, the choice of λ is dictated by the choice of the privacy-utility tradeoff operating point which, in turn, has to be negotiated between the energy provider and consumer.

Our distortion model can be viewed as a filter on the load signal X that suppresses those appliance (intermittent) signatures which reveal the most private information by: i) filtering out all frequencies that have power below a certain threshold (determined directly by λ), and ii) exploiting the presence of continually used appliances which reveal less private information as a pre-existing distortion (noise) at frequencies in which their spectral context is significant. This indirectly exploits the fact that a common household environment has a combination of appliances with various profiles that mask each other and thus having a mixture of appliances is better for privacy in the sense of masking human activity.

Our privacy technique prioritizes the elimination of those characteristics of the load signal that are more correlated with human activity and therefore it is likely to be robust against future data mining algorithms that may be brought to bear on smart meter data. At the same time, our utility constraints guarantee that most of the useful energy consumption information is retained in the revealed load data. This holds out hope that we can reveal significant energy consumption information while at the same time protecting significant personal information in a tunable tradeoff. Finding examples of operating points that correspond to real-world trade-offs would be an interesting avenue for further exploration. Another interesting avenue to explore would be to apply and demonstrate the power of these concepts in a practical context.

References

[1] F. Sultanem, "Using appliance signatures for monitoring residential loads at meter panel level," *IEEE Trans. Power Delivery*, vol. 6, no. 4, pp. 1380 –1385, Oct 1991.

[2] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, "Private memoirs of a smart meter," in *Proc. 2nd ACM Workshop Embedded Sensing Systems for Energy-Efficiency in Building*, New York, NY, USA, 2010, pp. 61–66.

[3] G. Kalogridis, C. Efthymiou, S. Z. Denic, T. A. Lewis, and R. Cepeda, "Privacy for smart meters: Towards undetectable appliance load signatures," in *Proc. IEEE 1st Intl. Conf. Smart Grid Comm.*, Gaithersburg, MD, Oct. 2010, pp. 232–237.

[4] D. Varodayan and A. Khisti, "Smart meter privacy using a rechargeable battery: minimizing the rate of information leakage," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Prague, Czech Republic, 2011.

[5] F. Li, B. Luo, and P. Liu, "Secure information aggregation for smart grids using homomorphic encryption," in *1st IEEE Intl. Conf. Smart Grid Commun.*, Gaithersberg, MD, Oct. 2010, pp. 327–332.

[6] V. Rastogi and S. Nath, "Differentially private aggregation of distributed time-series with transformation and encryption," in *Proc. 2010 Intl. Conf. Data Management*, Indianapolis, Indiana, USA, 2010, pp. 735–746.

[7] *Smart Metering and Privacy: Existing Law and Competing Policies*, Colorado Public Utilities Commission, 2009, http://www.dora.state.co.us/puc/.

[8] G. Deconinck and B. Decroix, "Smart metering tariff schemes combined with distributed energy resources," in *Proc. 4th Intl. Conf. Critical Infrastructures*, Linköping, Sweeden, 2009, pp. 1–8.

[9] C. Efthymiou and G. Kalogridis, "Smart grid privacy via anonymization of smart metering data," in *Proc. IEEE 1st Intl. Conf. Smart Grid Comm.*, Gaithersburg, MD, USA, Oct. 2010, pp. 238–243.

[10] S. Papadimitriou, F. Li, G. Kollios, and P. S. Yu, "Time series compressibility and privacy," in *Proc. 33rd Intl. Conf. Very Large Databases*, Vienna, Austria, 2007, pp. 459–470.

[11] A. Cavoukian, J. Polonetsky, and C. Wolf, "Smartprivacy for the smart grid: embedding privacy into the design of electricity conservation," *Identity in the Information Society*, vol. 3, pp. 275–294, 2010, 10.1007/s12394-010-0046-y. [Online]. Available: http://dx.doi.org/10.1007/s12394-010-0046-y

[12] T. Berger, "Multiterminal source coding," in *Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, 1978.

[13] G. W. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.

[14] H. Y. Lam, G. S. K. Fung, and W. K. Lee, "A novel method to construct taxonomy of electrical appliances based on load signatures," *IEEE Trans. Consumer Electronics*, vol. 53, no. 2, pp. 653–660, May 2007.

[15] M. Marwah, M. Arlitt, G. Lyon, M. Lyons, and C. Hickman, "Unsupervised disaggregation of low frequency power measurements," HP Labs, Tech. Rep., 2010.

[16] *Guidelines for Smart Grid Cyber Security: Vol. 2, Privacy and the Smart Grid*, National Institutes of Standards and Technology, Aug. 2010, http://csrc.nist.gov/publications/.