Information Theoretic Privacy for Smart Meters

Deniz Gündüz*, Jesus Gomez-Vilardebo†, Onur Tan† and H. Vincent Poor,‡

*Dept. of Electrical and Electronic Engineering, Imperial College London, London, UK.
†Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Barcelona, Spain.
‡Department of Electrical Engineering, Princeton University, Princeton, NJ, USA.

***Abstract*—Smart meters (SMs) measure and report energy consumption of individual users to the utility provider at short intervals on the order of minutes. While SM data is used to increase the efficiency in electricity distribution, it also conveys sensitive private data on the energy consumption behaviour of individual customers. In this work, privacy in a smart metering system is studied from an information theoretic perspective in the presence of alternative energy sources and storage units. An alternative energy source provides increased privacy by diversifying the energy source, and the storage device filters the real energy consumption to reduce the leaked data. Connections between this problem and rate-distortion theory is established, and both theoretical and numerical results are presented.**

I. Introduction

Smart grids are advanced electricity distribution networks that exploit digital technology to save energy, increase reliability and reduce the cost both for the customers and the utility providers (UPs). An important aspect of smart grid technology is the advanced control mechanisms that monitor the network closely and enable rapid diagnosis and solutions to problems, and dynamic adaptation to the changes in demand and supply. The essential components that provide such advanced control mechanisms are distributed sensing and measurement devices, such as smart meters (SMs), as well as the communication infrastructure, which establishes a two-way communication network among the SMs and the controllers for real-time information and control.

Compared to conventional electricity meters, SMs measure and report the energy consumption of the user to the UPs much more frequently. This high resolution information on user's energy consumption behaviour provides rapid control and response capability to the UP, which prompted the hasty adoption of SMs worldwide [1]. However, SMs also triggered a growing concern on consumer privacy [2]. It has been repeatedly shown that SM data can be easily analyzed for surveillance purposes by tracking appliance usage patterns, employing non-intrusive appliance load monitors and data mining algorithms [3]–[5]. At the very least, through SM readings it is possible to infer whether a user is at home or not. But, through more advanced pattern recognition techniques energy consumption patterns of individual appliances can be identified with high accuracy even when the SM can read only the aggregated household energy consumption [6]. Thus, even if partially, assuring the privacy of the household's electrical load profile is essential for users.

There is a growing literature on advanced mechanisms to provide privacy to the users of SMs. A major line of research on SM privacy is based on the assumption that the user has access to the SM readings and can manipulate them before forwarding to the UP. Bohli et al. [7] propose sending the aggregated energy consumption of a group of users, [8] proposes noise addition and [9] proposes compression of smart-meter data. The main limitation of these studies is the assumption that the UP depends solely on the SM reading to measure the user's energy consumption profile. However, the UP or other third parties can have other means to keep track of a user's energy consumption directly. The second group of work on the SM privacy problem assume that users are equipped with a certain technology that allows them to store or produce energy, through which they can alter the energy consumption profile observed by the UP. In this framework, the SM readings are not tempered, i.e., the UP can perfectly track the energy it provides to the user over time. The user's goal is to differentiate the appliances' real energy consumption as much as possible from the profile of the energy provided by the UP. While privacy protection using rechargeable batteries (RB) to filter out the real energy consumption is studied in [10]–[12], alternative energy sources (AES) for privacy protection is first proposed in [12], [13].

In our system model, we integrate both an AES and a RB. The energy flow is managed by the energy management unit (EMU). The EMU is responsible for providing the exact amount of power needed by the appliances. The EMU has access to three different energy sources: the power grid, the AES and the RB. At each time instant, the EMU provides the energy required by the appliances from these energy sources, and can also store some extra energy in the battery. We employ stochastic policies at the EMU that decide the amount of power taken from the grid based on the harvested energy, energy demand of the appliances and the state of the RB.

We measure privacy with the amount of leaked information about users' energy consumption to the UP, which is quantified by the mutual information between the users' real energy consumption and the energy provided by the UP. Mutual information has previously been proposed as a measure of privacy in SMs in [14], [15] and [11]. Obviously, with the

This work was partially supported by the Catalan Government under SGR2009SGR1046 and by the Spanish Government under projects TEC2010-17816 and TSI-020400-2011-18.

introduction of an AES, the privacy problem can be resolved in a straightforward manner if the AES is sufficient enough to provide all the required energy by the appliances. However, in general, the energy produced by the AES will be limited. We consider two different settings depending on the nature of the AES. In the first model, we consider a simple AES, such as an energy harvesting (EH) device, without energy storing capabilities. The energy from the AES is thus produced according to the energy generation profile of the underlying energy source, and wasted if, at a given instant, the energy generated surpass the energy required by the appliances and the energy storing capabilities of the RB. For this scenario, we require the EMU to increase both the privacy of the user and the energy efficiency of the system by avoiding wasting energy. We define and characterize the optimal *privacy - wasted power - battery capacity function*. Due to the memory introduced into the system through the battery, analytical expressions for this scenario are elusive, and we use numerical methods to estimate the average wasted power and information leakage rate for various energy management policies.

In the second model, we consider an AES with its own storage unit, which might model an electric vehicle battery that serves as an AES when connected to the household grid, or an independent power grid. Such an AES is assumed to supply the power requested by the appliances as long as the average and peak power constraints are satisfied. To simplify the system model, we eliminate the battery from the EMU and do not allow wasting energy from the AES or from the grid. For this scenario, we characterize the optimal privacy depending on the average and peak power supported by the AES. We provide a single-letter characterization of the *privacy - average power - peak power function* when the input load is an independent and identically distributed (i.i.d.) random variable. For discrete input distributions, we show that the privacy - average power - peak power function can be written as a convex optimization problem with linear constraints. For continuous input distributions, we derive the Shannon lower bound (SLB), and show that it is achieved for exponential input distributions, and for certain average and peak power values for other input load distributions.

We also highlight an equivalence between the privacy problem studied here and the rate - distortion function with a difference distortion measure. This allows us to exploit certain tools from rate - distortion theory to analyze the optimal privacy achievable in a SM system.

The rest of the paper is organized as follows. In Section II, we introduce the system model and define the concept of information leakage rate which will be used to measure the system performance. The privacy - wasted power - battery capacity problem is addressed in Section III and the privacy - average power - peak power problem is addressed in Section IV. Finally, we conclude our work in Section V.

II. System Model and the Privacy Measure

The SM system model considered in this paper is depicted in Fig. 1. We model the energy flow as a discrete time system.

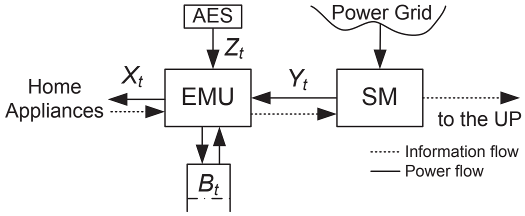

Fig. 1: Smart metering privacy model via an AES.

At each time instant t, the EMU receives the energy demand of the appliances, $X_t \in \mathcal{X}$, called the input load. The EMU provides this energy demand either from the RB, the AES or the power grid. The SM measures the power that is requested from the grid, $Y_t \in \mathcal{Y}$, at each time instant, and reports it to the UP. We call Y_t the output load.

We measure the privacy by the information rate leaked to the UP about the input load. Assuming that the statistical behavior of the energy demand is known by the UP, its initial uncertainty about the real energy consumption can be measured by the entropy rate $\frac{1}{n}H(X^n)$. After the UP observes the output load, this uncertainty is reduced to the equivocation rate

$$\frac{1}{n}H(X^n|Y^n) = \frac{1}{n}H(X^n) - \frac{1}{n}I(X^n;Y^n).$$

Since $H(X^n)$ is a characteristic of the appliances and is assumed to be known, the EMU tries to minimize $I(X^n;Y^n)$ in order to maximize the equivocation. Accordingly, the privacy achieved by an energy management policy is measured by the mutual information rate between the input and output loads. We define the *information leakage rate* as

$$I^{(n)} \triangleq \frac{1}{n}I(X^n;Y^n). \tag{1}$$

The tools available to the EMU to reduce the information leakage rate are directing some of the energy demand to the AES, or filtering the energy demand form the UP using the RB. We consider two different models for the AES. In the first model, studied in Section III, we assume that the AES is an EH device, and harvests a certain amount of energy at each time instant with a certain probability. The statistical behavior of the EH distribution depends on the design of the energy harvester. For example, for a solar energy harvester the average harvested energy can be increased by scaling the size and the efficiency of the solar panel. In this stochastic energy harvesting model, the harvested energy is either directly consumed by the appliances or stored in the RB. Otherwise the harvested energy is lost. The RB is assumed to have finite capacity of $\hat{P}_B$. In this model, the EMU employs energy management policies that decide, at each time instant t, the amount of energy that is provided from the power grid and from the battery, based on the input load up to time t, energy obtained from the EH, the state of the battery and the output load up to the previous time instant.

In the second model studied in Section IV, we consider an AES which has its own RB. In this model, the AES is able to supply all the energy requested by the EMU, as long as the peak and average power constraints are satisfied. The RB at the EMU is not employed and we do not allow the EMU to waste energy neither from the AES nor from the power grid. In this model, the EMU employs energy management policies that decide, at each time instant t, the amount of energy that is provided from the power grid and/or from the AES, based on the input load up to time t, the load obtained from the power grid and the AES up to the previous time instant.

III. The Privacy - Wasted Power - Battery Capacity Function

In this model, we consider the AES as an EH device, and the harvested energy is modeled as a discrete time stochastic process, denoted by $Z^n = Z_1, Z_2, \ldots, Z_n$. Here we assume that Z^n is an i.i.d. sequence with distribution p_Z over $\mathcal{Z} = \{0, 1, \ldots, M\}$. Denoting the battery state at time instant t by B_t, we require X_t, Y_t, Z_t, and B_t to satisfy the following conditions at each time instant t:

$$B_t \leq \hat{P}_B, \tag{2a}$$

$$X_t \leq Y_t + Z_t + B_t - B_{t-1}, \tag{2b}$$

$$0 \leq X_t, Y_t, \tag{2c}$$

which guarantee that the energy stored in the battery is within its capacity, and the energy demand of the appliances is satisfied at each time instant.

Notice that due to the finite capacity of the RB, some of the energy from the grid and from the EH device can be wasted. We measure the *wasted power* after n time instants as

$$P_W^{(n)} \triangleq \frac{1}{n} \sum_{t=1}^{n} (Z_t + Y_t - X_t). \tag{3}$$

At each time instant t, the EMU decides on the amount of energy it gets from the UP and from the battery based on the input load up to time t, X^t, the energy harvested up to the previous time instant, Z^{t-1}, state of the battery B^{t-1}, and the output load Y^{t-1}.

Definition 1: A length-n energy management policy is composed of, possibly random, power allocation functions

$$f_t : \mathcal{X}^t \times \mathcal{Z}^{t-1} \times \mathcal{Y}^{t-1} \times \mathcal{B}^{t-1} \rightarrow \mathcal{Y} \times \mathcal{B},$$

for $t = 1, \ldots, n$, such that X_t, Y_t, Z_t, B_t, satisfy the constrains in (2) at each time instant. The privacy achieved by this policy is given by the *information leakage rate*, $I^{(n)}$, defined in (1), while the *average wasted power* is defined as $\bar{P}_W \triangleq \mathbb{E}\left[P_W^{(n)}\right]$, with $P_W^{(n)}$ defined as in (3). Here the expectation is taken over the probability distributions of the AES, the input and output loads.

Definition 2: An information leakage rate - average wasted power - battery capacity triplet $(I, \bar{P}_W, \hat{P}_B)$ is said to be *achievable* if there exists a sequence of energy management policies, satisfying (2) at each time instant, and $\lim_{n\to\infty} I^{(n)} \leq I$, and $\lim_{n\to\infty} \bar{P}_W^{(n)} \leq \bar{P}_W$.

Definition 3: The information leakage rate - average wasted power - battery capacity *region* is the closure of the set of all achievable triplets $(I, \bar{P}_W, \hat{P}_B)$.

Definition 4: The *privacy - wasted power - battery capacity function*, $\mathcal{I}(\bar{P}_W, \hat{P}_B)$, is the infimum of the information leakage rates such that $(I, \bar{P}_W, \hat{P}_B)$ is in the information leakage rate - average wasted power - battery capacity region.

We restrict our analysis to discrete loads; that is, we assume that there is a minimum unit of energy; and hence, at each time instant t, the input load, harvested energy, battery state and output load are all integer multiples of this energy unit. Over time, we assume that the input load X^n is an i.i.d. sequence with distribution p_X over $\mathcal{X}$. The harvested energy is also modelled as a discrete time stochastic process, where Z^n is also an i.i.d. sequence with distribution p_Z over $\mathcal{Z}$, independent of X.

For this scenario, due to the memory introduced into the system through the battery, a single letter expression for the privacy - wasted power function is elusive. Instead, we focus on a limited set of energy management policies and analyze the achievable privacy - wasted power performance numerically. Note that, energy management policies can be time-varying in general. We consider time-invariant fixed policies in which the transition probabilities and parameters of the energy management policy are fixed throughout the operation. For a fixed policy, the average wasted power $\bar{P}_W$ and the information leakage rate I between the input and the output loads can be estimated numerically. To keep the complexity of possible energy management policies simple, we restrict our attention to those that depend only on the input load, battery state and harvested energy at time t; that is,

$$f_t : \mathcal{X} \times \mathcal{Z} \times \mathcal{B} \rightarrow \mathcal{Y} \times \mathcal{B}$$

for $t = 1, ..., n$. Energy management policies that depend on both the battery state and the previous output load are studied in [11]; however, the authors indicate that they have not found any battery/output conditioned policy that performs better than the optimal policy that acts solely based on the battery state. We have made the same observation in our numerical analysis. Accordingly, to keep our model simple we focus only on battery-conditioned policies in this work.

Due to the discrete time nature of the system, it can be represented by an FSM. The FSM of this system has $|\mathcal{B}|$ states. The management policy specifies the transition probabilities in the FSM. For numerical computation, we sample very long sequences (large n) of X^n, Z^n and Y^n by using the FSM. We then compute the average wasted power by evaluating $P_W^{(n)}$ in (5) for a very long sequence. The weak law of large numbers then ensures that with high probability, $P_W^{(n)} \rightarrow \bar{P}_W$ as $n \rightarrow \infty$. For the computation of the information leakage rate, we use the computation method studied in [16].

For numerical results, we focus on a binary system for its simplicity, as otherwise, the transitions in the state diagram get very complicated and the numerical computation becomes intractable. Consider a system with $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{0, 1\}$. At a given time instant t, the EMU decides stochastically $y = 0$ or

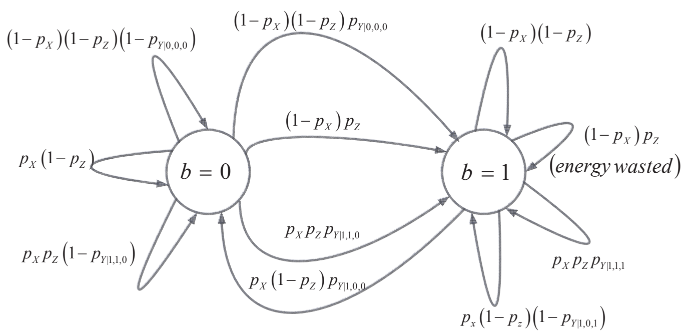

Fig. 2: Finite state diagram for the battery conditioned energy management policy with $s=2$ states. Transition probabilities are also included in the figure.

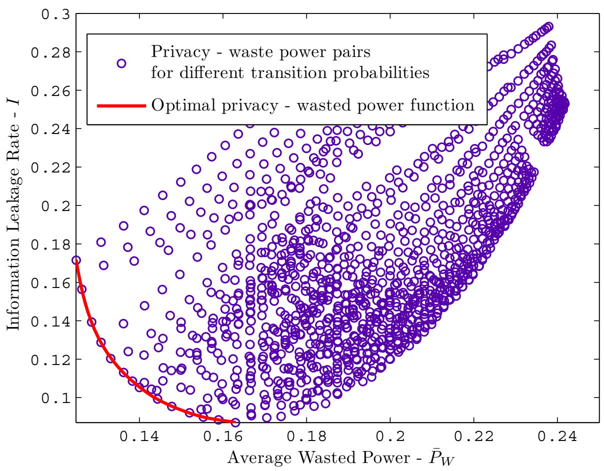

Fig. 3: Information leakage rate, $\mathcal{I}$, versus average wasted energy rate, $\bar{P}_W$, for $p_X=0.5$ and $p_Z=0.5$.

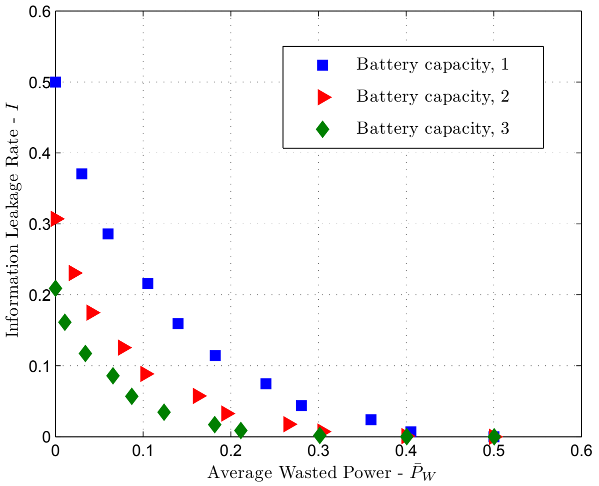

Fig. 4: Information leakage rate, $\mathcal{I}$, versus wasted energy rate, $\bar{P}_W$, for the case of wasting grid energy.

$y=1$ depending on the system state defined by b_{t-1}, x_t and z_t, according to the conditional probabilities $\Pr\{Y=1|X=x, Z=z, B=b\}=p_{Y|x,z,b}$ for all $x\in[0,1]$, $z\in[0,1]$, $b\in[0,\hat{P}_B]$. The EMU can choose any $p_{Y|x,z,b}$ as long as the instantaneous constraints in (2) are satisfied.

In our simulations, we perform an exhaustive search by varying the transition probabilities in Fig. 2 with 0.1 increments. We use $n=10^6$ for the computations. Based on these numerical results we provide various observations and conclusions regarding the optimal operation of the EMU from a joint privacy–energy efficiency perspective.

A. Privacy - Wasted Power Function

First, we consider the binary system described above with a binary battery state, $|\mathcal{B}|=2$. To satisfy, the instantaneous power constraint in (2), we require $p_{Y|1,0,0}=1$. If we, additionally, enforce that no energy can be wasted from the power grid, then $p_{Y|0,1,0}=p_{Y|0,0,1}=p_{Y|1,1,1}=0$. In that case, the possible transitions are depicted in Fig. 2.

In Fig. 3 we characterize the whole trade-off between the privacy and energy efficiency for $p_Z=0.5$ and $p_X=0.5$. Each circle in the figure marks the value of a $\left(I,\bar{P}_W\right)$ pair that can be achieved by assigning different transition probabilities labeled on Fig. 2. The Pareto optimal trade-off curve is the one that is formed by the points on the lower-left corner of the figure, i.e., the points for which I and $\bar{P}_W$ cannot be improved simultaneously. According to the requirements of the system, the operating point can be chosen anywhere on the trade-off curve. For different, values of p_Z, we obtain similar trade-offs; with increasing EH rate p_Z, the minimum information leakage rate decreases while the wasted energy rate increases.

B. Privacy at the expense of wasting grid energy

Next, we study the effect of wasting energy from the grid on the privacy. We consider battery conditioned policies with binary input/output load values, no EH unit, and an RB with capacity of K units. Let RB be fully charged at time instant t, i.e., $b_t=|\mathcal{B}|-1$. Even if the appliances do not consume any energy at time instant $t+1$, i.e., $x_{t+1}=0$, we allow the EMU to demand energy from the UP, i.e., $y_{t+1}=1$, with probability p_W, and $y_{l+1}=0$ with probability $(1-p_W)$. In other words, we allow wasting the grid energy with probability p_W, by which we obscure the information of the UP about the real energy consumption. Fig. 4 illustrates the achievable points on the $\left(I,\bar{P}_W\right)$ trade-off, obtained for an equiprobable input load, $p_X=0.5$, and for increasing RB capacity values, 1, 2, and 3. In this simulation, to keep the simulation time reasonable we find the achievable points, by considering only complementary transition probabilities, such that the sum of the transition probabilities between two states is equal to 1. We can see that the privacy can be significantly improved by wasting more energy, i.e., by increasing p_W. If we increase the RB capacity, as we can see in Fig. 4, both the information leakage rate and the wasted energy rate are improved for the same energy waste probability, p_W.

IV. The Privacy - Average Power - Peak Power Function

In the previous model, the existence of the RB controlled by the EMU introduces memory to the system and prevents us from obtaining single–letter information theoretic results. In this section, we assume that the EMU does not have an RB to filter out the real energy consumption, and hence, at each time instant t, the energy demand of the appliances X_t is provided either by the UP (Y_t), or by the AES (Z_t). In this model, we assume that the AES has its own storage unit, and hence, it is limited by the peak and average power values it can provide, rather than instantaneous constraints as in the previous section. The decision at each time instant is based on the input load up to time t, X^t, as well as the output load and the energy demanded from the AES up to the previous time instant, Y^{t-1} and Z^{t-1}, respectively. In this scenario, at each t, we require X_t, Y_t, Z_t to satisfy

$$0 \leq Z_t \leq \hat{P}_Z, \tag{4a}$$

$$X_t = Y_t + Z_t, \tag{4b}$$

$$0 \leq X_t, Y_t. \tag{4c}$$

We measure the power requested from the AES after n time instants as

$$P_Z^{(n)} = \frac{1}{n} \sum_{t=1}^{n} (X_t - Y_t) \, . \tag{5}$$

Definition 5: A length-n energy management policy is composed of, possibly random, power allocation functions

$$f_t : \mathcal{X}^t \times \mathcal{Z}^{t-1} \times \mathcal{Y}^{t-1} \to \mathcal{Y} \times \mathcal{Z},$$

for $t = 1, \dots, n$, such that X_t, Y_t satisfy the constrains in (4) at each time instant. The privacy achieved by this policy is given by the *information leakage rate* I in (1) while the *average power* requested from the AES is given by $\bar{P}_Z^{(n)} \triangleq \mathbb{E}\left[P_Z^{(n)}\right]$ with $P_Z^{(n)}$ as defined in (5). Here the expectation is taken over the probability distributions of the input and output loads.

Definition 6: An information leakage rate - average power - peak power triplet $(I, \bar{P}_Z, \hat{P}_Z)$ is said to be *achievable* if there exists a sequence of energy management policies of duration n satisfying $\forall t \leq n$ (4) with $\lim_{n\to\infty} I^{(n)} \leq I$, and $\lim_{n\to\infty} \bar{P}_Z^{(n)} \leq \bar{P}_Z$.

Definition 7: The information leakage rate - average power - peak power *region* is the closure of the set of all achievable triplets $(I, \bar{P}_Z, \hat{P}_Z)$.

Definition 8: The *privacy - average power - peak power function*, $\mathcal{I}(\bar{P}_Z, \hat{P}_Z)$, is the infimum of the information leakage rates in the information leakage rate - average power - peak power region.

Our goal is to give a mathematically tractable expression for the privacy–power function, and identify the optimal energy management policy that achieves the highest level of privacy for a given AES. In the next theorem, we show that if input loads are chosen i.i.d. we can characterize the function $\mathcal{I}(\bar{P}_Z, \hat{P}_Z)$ in a single-letter format.

Theorem 1: The privacy - average power - peak power function $\mathcal{I}(\bar{P}_Z, \hat{P}_Z)$ for an i.i.d. input load X with distribution $f_X(x)$ is given by

$$\mathcal{I}(\bar{P}_Z, \hat{P}_Z) = \inf_{\substack{f_{Y|X}(y|x): \mathbb{E}[X-Y] \leq \bar{P}_Z, \\ 0 \leq X-Y \leq \hat{P}_Z}} I(X;Y) \tag{6}$$

Proof: See [13]. ■

Theorem 1 implies that the optimal energy management policy is memoryless; that is, it can be achieved by simply looking at the instantaneous input load, and generating the output load randomly using the optimal conditional probability. This results in a stochastic energy management policy rather than a deterministic one. On the other hand, if the user knew all the future energy demand over a block of n time instants, the same privacy performance could be achieved by a deterministic block-based energy management policy.

We note here the correspondence between the privacy–power function in (6) and the rate-distortion function [17]. The privacy-power function in (6) is indeed a rate-distortion function with the following difference distortion measure:

$$d(x,y) = \begin{cases} x - y & \text{if } 0 \leq x - y \leq \hat{P}_Z, \\ \infty & \text{otherwise.} \end{cases}$$

This correspondence allows us to use various tools from rate-distortion theory to study privacy in a SM system.

We first consider discrete input load distributions. If the input and output alphabets were both discrete, the characterization of the privacy–power function $\mathcal{I}(\bar{P}_Z, \hat{P}_Z)$ in (1) would become a convex optimization problem since the mutual information is a convex function of the conditional probability values, $f_{Y|X}(y_m|x_k)$, for $y_m \in \mathcal{Y}$, $x_k \in \mathcal{X}$, and the constraints are linear. Then, (6) can be solved numerically e.g. using the efficient Blahut-Arimoto (BA) algorithm [17].

The next theorem shows that the output alphabet can be constrained to the input alphabet $\mathcal{Y} = \mathcal{X}$ without loss of optimally. This also implies that for any given discrete input alphabet the optimal output alphabet is also discrete.

Theorem 2: Without loss of optimality the output load alphabet $\mathcal{Y}$ can be constrained to the input load support set, i.e., $\mathcal{Y} = \mathcal{X}$.

Proof: The proof is omitted. It can be found in [18] available online. ■

Theorem 2 implies that the problem in (6) can always be efficiently solved for discrete input distributions.

For a continuous input distribution, the optimal output alphabet is potentially continuous. Consequently, efficient algorithms, such as the BA algorithm, do not yield the optimal solution. In this case, we provide the Shannon lower bound [17] on the privacy - average power - peak power function $\mathcal{I}_{SLB}(\bar{P}_Z, \hat{P}_Z)$, and identify the power region $(\bar{P}_Z, \hat{P}_Z)$ where it is achievable.

We begin by presenting the distribution that maximizes the entropy among those random variables Z with mean $\bar{P}_Z$ and satisfying $0 \leq Z \leq \hat{P}_Z$. From [17, Ch. 11], we know that this distribution is the truncated exponential distribution $Z \sim \mathsf{ExpT}(\bar{P}_Z, \hat{P}_Z)$ with

$$f_Z(z) = \begin{cases} \frac{1}{\lambda_0} e^{-\frac{z}{\lambda_1}}, & 0 \leq z \leq \hat{P}_Z, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

The variables $\lambda_0 \geq 0$ and $\lambda_1 \geq 0$ are chosen to satisfy the constraints

$$\begin{gathered} \int_0^{\infty} f_Z(z) dz = \frac{\lambda_1}{\lambda_0} p = 1, \\ \mathbb{E}[Z] = \lambda_1 - \hat{P}_Z \frac{q}{p} = \bar{P}_Z, \end{gathered} \tag{8}$$

where $q = e^{-\frac{\hat{P}_Z}{\lambda_1}}$ and $p = 1 - q$. This distribution has differential entropy

$$\mathsf{h}\left(\mathsf{ExpT}(\hat{P}_Z, \bar{P}_Z)\right) = \ln(\lambda_0) + \frac{\bar{P}_Z}{\lambda_1},$$

and its Laplace transform $\mathcal{L}f_Z(s) = \mathcal{L}\left(f_Z(z)\right)(s)$ reads

$$\mathcal{L}f_Z(s) = \frac{1}{p} \frac{1 - q e^{-\hat{P}_Z s}}{1 + \lambda_1 s}. \tag{9}$$

We denote by $\delta(x)$ the Dirac delta function and use $f'(x)$ to denote the first order derivative of $f(x)$. In next theorem, we present the SLB on the privacy - average power - peak power function and the achievabiliy region for any piecewise continuous input distribution $f_X(x)$.

Theorem 3: Consider an AES with an average power constraint $\bar{P}_Z$ and a peak power constraint $\hat{P}_Z$. The privacy - average power - peak power function $\mathcal{I}(\bar{P}_Z, \hat{P}_Z)$ for an i.i.d. input load X with differential entropy $\mathsf{h}(X)$ is lower bounded by

$$\mathcal{I}_{SLB}(\bar{P}_Z, \hat{P}_Z) = \mathsf{h}(X) - \ln(\lambda_0) - \frac{\bar{P}_Z}{\lambda_1}, \tag{10}$$

where λ_0 and λ_1 are obtained from (8). For any input distribution $f_X(x)$ continuous on $\mathcal{R}_+$ except for a countable number of jump discontinuities or non-differentiable points $\mathcal{X}_D = \{x_1, ..., x_D\}$, the SLB (10) is achieved for all $\bar{P}_Z$ and $\hat{P}_Z$ satisfying $f_Y(y) \geq 0$ for all $y \in \mathcal{R}_+$ by using the conditional output distribution $f_{Y|X}(y|x) = f_Z(x-y) \frac{f_Y(y)}{f_X(x)}$ where the output distribution is given by

$$f_Y(y) = \sum_{l=0}^{\infty} p q^l g_Y(y - l\hat{P}), \tag{11}$$

and $g_Y(y) = g_{Y_C}(y) + g_{Y_D}(y)$ is a mixture of a continuous and a discrete distribution specified as follows:

$$\begin{aligned} g_{Y_C}(y) &= f_X(y) + \lambda_1 f'_X(y), \; y \in \mathcal{R}_+/\mathcal{X}_D, \\ g_{Y_D}(y) &= \lambda_1 \sum_{i=0}^{D} \Delta_X(x_i) \delta(y - x_i), \; y \in \mathcal{X}_D. \end{aligned}$$

Proof: The SLB in (10) can be obtained from [17]. To find the conditional distribution $f_{Y|X}(y|x)$ that satisfies the SLB with equality [17], we require the random variables $Z = X - Y$ and Y to be independent, and Z to be distributed according to a truncated exponential distribution $Z \sim \mathsf{ExpT}(\bar{P}_Z, \hat{P}_Z)$ with mean $\bar{P}_Z$ and peak value $\hat{P}_Z$. We first obtain the output distribution $f_Y(y)$ from its Laplace transform $\mathcal{L}f_Y(s) = \mathcal{L}(f_Y(y))(s)$. First, recall $\mathcal{L}f_Z(s)$ in (9) and that $\mathcal{L}g_Y(s) = \mathcal{L}\left(g_Y(y)\right)(s)$ is given by

$$\mathcal{L}g_Y(s) = \mathcal{L}f_X(s)\left(1 + \lambda_1 s\right).$$

Then, observe that

$$\begin{aligned} \mathcal{L}f_Y(s) &= \frac{\mathcal{L}f_X(s)}{\mathcal{L}f_Z(s)}, && (12) \\ &= p \frac{\mathcal{L}g_Y(s)}{1 - q e^{-\hat{P}_Z s}}, && (13) \\ &= \sum_{l=0}^{\infty} p q^l \mathcal{L}g_Y(s) e^{-l\hat{P}_Z s}, && (14) \\ &= \sum_{l=0}^{\infty} p q^l \mathcal{L}\left(g_Y(y - l\hat{P}_Z)\right)(s). && (15) \end{aligned}$$

It follows that $f_Y(y)$ is given by (11). The conditional distribution $f_{Y|X}(y|x)$ is obtained using the fact that $f_{X|Y}(x|y) = f_Z(x-y)$. Finally, it can be shown that $\int_0^{\infty} f_Y(y) dy = 1$; and thus, achievability is guaranteed by requiring $f_Y(y) \geq 0$, $\forall y \in \mathcal{R}^+$. ■

Remark 3.1: For an AES with an unlimited peak power constraint $\hat{P}_Z \to \infty$, we have $\lambda_1 \to \bar{P}_Z$, $\lambda_0 \to \bar{P}_Z$, and Z follows an exponential distribution $\mathsf{Exp}(\bar{P}_Z)$. Then, the SLB reduces to

$$\mathcal{I}_{SLB}(\bar{P}_Z) = \mathsf{h}(X) - \ln(e\bar{P}_Z), \tag{16}$$

and the output distribution simplifies to $f_Y(y) = g_Y(y)$.

A. Exponential Distribution

Next we particularize Theorem 3 for an exponential input distribution. Let $X \sim \mathsf{Exp}(m)$, i.e., $f_X(x) = \frac{1}{m} e^{-\frac{x}{m}} u(x)$. From (11) we obtain the output probability distribution as

$$g_{Y_C}(y) = \left(1 - \frac{\lambda_1}{m}\right) \sum_{l=0}^{\infty} p q^l f_X(y - l\hat{P}_Z), \tag{17}$$

$$g_{Y_D}(y) = \frac{\lambda_1}{m} \sum_{l=0}^{\infty} p q^l \delta(y - l\hat{P}_Z). \tag{18}$$

For $Y \in \left\{ l\hat{P}_Z : l = 0, 1, ..., \infty \right\}$, Y follows a discrete geometric distribution, $\mathsf{Geom}(p) = pq^l$. Otherwise, Y follows a mixture of weighted and shifted continuous exponential distributions each with mean m. The SLB achievability condition $f_Y(y) \geq 0$ for all $y \in \mathcal{R}_+$ requires $m \geq \lambda_1$, or equivalently, $\bar{P}_Z \leq \bar{P}_{Z_0}$ with $q_0 = e^{-\frac{\hat{P}_Z}{m}}$, and

$$\bar{P}_{Z_0} = m - \hat{P}_Z \frac{q_0}{1 - q_0}. \tag{19}$$

At $\bar{P}_{Z_0}$ we have

$$\mathcal{I}\left(\bar{P}_{Z_0}, \hat{P}\right) = \mathsf{h}(Y) = \mathsf{h}(\mathsf{Geom}(p_0)). \tag{20}$$

For the exponential input distribution, the SLB is achievable for all possible peak and average power constraints.

For an exponential distribution with mean 1, we depict the privacy - power function in Fig. 5 for different $\hat{P}_Z$ values.

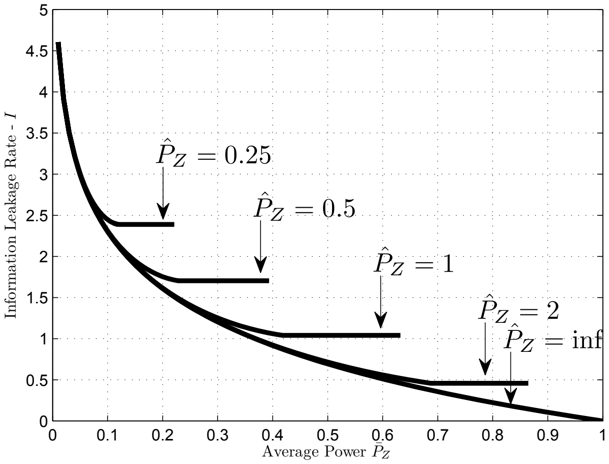

Fig. 5: Privacy-power function for $X \sim \mathsf{Exp}(1)$.

V. Conclusions

We have studied the privacy problem in SM systems for two different yet closely related models. In both cases, we measure the privacy of the system from an information theoretic perspective using the information leakage rate between the input and output loads as the privacy measure. We have shown that the user can hide its energy consumption profile from the UP by employing stochastic energy management policies.

In the first model, we have assumed the availability of an EH device which generates energy at a constant rate in an i.i.d. fashion, and an RB. In this model both the RB and the EH source are used to filter out the real energy consumption of the devices. Note that, due to the finite capacity of the RB and the mismatch between the energy demand of the appliances and the energy generated by the EH device, some of the energy will be wasted. Since the RB can be utilized both to increase the privacy of the user, and to decrease the amount of wasted energy, we have studied the privacy–energy efficiency trade-off. Due to the memory introduced by the RB obtaining closed-form expressions for the information leakage rate is elusive. We have used a numerical method to calculate the information leakage rate - average wasted power trade–off. For the sake of simplicity, we have considered binary input and output loads and focused on battery–dependent energy management policies. We have discussed, the effect of the battery capacity and of wasting grid energy.

In the second part of the paper, we have considered availability of only an AES with a peak and an average power constraint, i.e., no RB. We have characterized the optimal information leakage rate that can be achieved for given average and peak power constraints. We have shown that, for i.i.d. input loads, the privacy–power trade–off has a single-letter expression. In addition, for discrete input alphabets we have shown that the privacy-power function can be evaluated numerically as the solution to a convex optimization problem. For continuous input distributions, we have characterized the Shannon lower bound on the privacy–power function, and provided the closed–form solution for an exponential input load.

References

[1] P. Wunderlich, D. Veit, and S. Sarker, “Adoption of information systems in the electricity sector: The issue of smart metering,” in *Proc. Americas Conference on Information Systems*, Seattle, WA, Aug 2012.

[2] P. McDaniel and S. McLaughlin, “Security and privacy challenges in the smart grid,” *IEEE Security Privacy*, vol. 7, no. 3, pp. 75–77, May-Jun. 2009.

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

[4] 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.

[5] G. Kalogridis and S. Z. Denic, “Data mining and privacy of personal behaviour types in smart grid,” in *Proc. IEEE Int. Conf. Data Mining Wkshp.*, Vancouver, Canada, Dec. 2011.

[6] A. Predunzi, “A neuron nets based procedure for identifying domestic appliances pattern-of-use from energy recordings at meter panel,” in *Proc. IEEE Power Eng. Society Winter Meeting*, New York, USA, Jan. 2002.

[7] J.-M. Bohli, C. Sorge, and O. Ugus, “A privacy model for smart metering,” in *Proc. IEEE Int’l Conf. on Comm.*, Capetown, South Africa, May 2010.

[8] V. Rastogi and S. Nath, “Differentially private aggregation of distributed time-series with transformation and encryption,” in *Proc. Intl. Conf. Data Management*, Indianapolis, Indiana, Oct. 2010.

[9] S. R. Rajagopalan, L. Sankar, S. Mohajer, and H. V. Poor, “Smart meter privacy: A utility-privacy tradeoff framework,” in *Proc. IEEE Int’l Conf. Smart Grid Comm.*, Brussels, Belgium, Oct 2011.

[10] G. Kalogridis, C. Efthymiou, S. Denic, T. A. Lewis, and R. Cepeda, “Privacy for smart meters: Towards undetectable appliance load signatures,” in *Proc. IEEE Int’l Conf. Smart Grid Comm.*, Gaithersburg, MD, Oct. 2010.

[11] D. Varodayan and A. Khisti, “Smart meter privacy using a rechargeable battery: Minimizing the rate of information leakage,” in *Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc.*, Prague, Czech Republic, May 2011.

[12] O. Tan, D. Gündüz, and H. V. Poor, “Smart meter privacy in the presence of energy harvesting and storage devices,” in *Proc. IEEE Int’l Conf. Smart Grid Comm.*, Tainan City, Taiwan, Nov 2012.

[13] D. Gunduz and J. Gomez-Vilardebo, “Smart meter privacy in the presence of an alternative energy source,” in *Proc. IEEE Int’l Conf. on Comm. (ICC) (to appear)*, Budapest, Hungary, June 2013.

[14] G. Kalogridis, R. Cepeda, S. Z. Denic, T. Lewis, and C. Efthymiou, “Elecprivacy: Evaluating the privacy protection of electricity management algorithms,” *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 750–758, Aug. 2011.

[15] L. Sankar, S. R. Rajagopalan, and H. V. Poor, “A theory of utility and privacy of data sources,” in *Proc. IEEE International Symposium on Information Theory*, Austin, TX, June 2010.

[16] D. M. Arnold, H. A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, “Simulation-based computation of information rates for channels with memory,” *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley-Interscience, 1991.

[18] J. Gomez-Vilardebo and D. Gunduz, “Privacy of smart meter systems with an alternative energy source,” in *available at http://bit.ly/10VKUJy*.