# SMART METER PRIVACY USING A RECHARGEABLE BATTERY: MINIMIZING THE RATE OF INFORMATION LEAKAGE

*David Varodayan**

Hewlett-Packard Laboratories
varodayan@hp.com

*Ashish Khisti*

University of Toronto
akhisti@comm.utoronto.ca

## ABSTRACT

A rechargeable battery may be used to partially protect the privacy of information contained in a household's electrical load profile. We represent the system as a finite state model to make tractable the computation of the rate of information leakage. Specifically, we use a trellis algorithm to estimate the mutual information rate between the battery's input and output loads. We show that stochastic battery policies can leak 26% less information than a so-called best-effort algorithm (that holds the output load constant whenever possible). We finally describe the extension of the technique to more realistic models of the battery system.

***Index Terms*—** information security, data privacy, smart grids, energy storage, batteries

## 1. INTRODUCTION

Deployments of smart electricity meters to residential homes continue unabated around the world. These devices enable utility companies to read their customers' power loads remotely and automatically every few minutes or even more frequently. In contrast, traditional electromechanical meters are typically read just monthly during physical visits by a utility company employee. Smart metering is a critical part of the smart grid vision: the modernization of the electricity infrastructure for improved reliability and energy efficiency [1]. Consider, for example, the expected growth in adoption of plug-in electric vehicles. It is anticipated that their rechargeable batteries (when plugged in at home) could also function collectively as distributed energy storage. These batteries would help smooth out peaks and troughs in the required power supply and, thereby, improve the efficiency of the available generation resources. Smart meters are essential to coordinate the desired charging and discharging of the batteries in real time.

Despite the promise of smart metering, there are risks. A concern, from the customer point-of-view, is loss of privacy [2, 3]. The utility company could employ data mining algorithms to analyze a household's load data and, thus, discover the appliances and usage patterns of the inhabitants [4, 5, 6]. This information could potentially be sold to advertisers or be used for surveillance. One approach to preserving privacy involves anonymization of the data by an intermediary [7], but this merely transfers doubts about trustworthiness from one party to another.

A new (complementary) idea suggests that the household itself partially obscure its load profile using a rechargeable battery (such as the one in a plugged-in electric vehicle) [8]. The setup in Fig. 1 shows the flow of information left-to-right. The battery's input load $X$ is the aggregate load of the appliances. Its output load $Y$ is the combined load of the appliances and the battery, as reported by the smart meter to the utility company. Power, on the other hand, flows right-to-left from the utility company through the battery to the household's appliances. At any time, the battery may perform a combination of the following actions (or none of them) subject to its capacity: relay power directly from the utility to the appliances; store energy from the utility company for future use; deliver previously stored energy to the appliances. In this way, charging and discharging the battery can manipulate the output load $Y$, obscuring some of the information contained in the input load $X$.

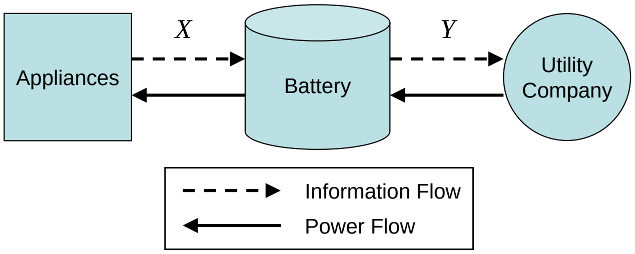

**Fig. 1**. Charging and discharging the battery with power flows can partially decouple the information flows $X$ and $Y$.

In [8], a so-called best-effort algorithm is proposed for charging and discharging the battery. Whenever possible, it holds the output load to its most recent value. This deterministic policy is evaluated with respect to three *ad hoc* metrics of privacy, none of which quantify the amount of information leaked. Moreover, all three metrics measure the privacy of the differential information only, relying on an assumption that the utility's data mining algorithms ignore low frequencies of the load profile. This is not necessarily so, as demonstrated in recent work [9].

In this paper, we take a more principled approach. Section 2 simplifies the battery system as a finite state model and formulates different types of stochastic battery policies. In Section 3, we define the rate of information leakage as the mutual information rate $I(X;Y)$ and describe a method to compute it tractably. Section 4 shows experimentally that stochastic policies are superior to the best-effort policy of [8] in minimizing the rate of information leakage. In Section 5, we discuss extensions of this technique to more realistic models of loads and batteries.

## 2. BATTERY SYSTEM MODEL

Our approach is to model the battery system as simply as possible, so that we can compute the rate of information leakage tractably

*David Varodayan's research was supported, in part, by an ASEE/NSF Corporate Research Postdoctoral Fellowship.

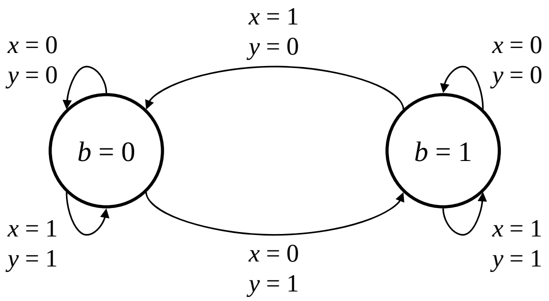

**Fig. 2**. Finite state model depicting the transitions in the binary-load binary-battery model.

and make revealing insights into minimizing it. We discuss extending the technique presented in this paper to more realistic models in Section 5.

### 2.1. Binary Loads and Binary Battery

Consider the input load $X$ to be a binary i.i.d. (independent and identically distributed) sequence $\{x_1, x_2, x_3, \ldots\}$. That is, the appliances consume either 0 or 1 units of power at any discrete time. Let $q_0$ be the probability that $x_t$ is 0 for any time $t$. We also require the output load $Y$ to be a binary sequence $\{y_1, y_2, y_3, \ldots\}$, so the utility provides either 0 or 1 units of power at any time. The rechargeable battery of capacity 1 unit has two states: 0 (discharged) or 1 (charged). Let the binary sequence $\{b_0, b_1, b_2, b_3, \ldots\}$ represent the state of the battery over time, with initial state $b_0 = 0$.

At time $t$, if the battery is in the discharged state ($b_t = 0$), then there are three possible events for time $t + 1$. If the appliances consume 1 unit ($x_{t+1} = 1$), then the battery must remain discharged and the utility provides 1 unit to the appliances ($b_{t+1} = 0, y_{t+1} = 1$). Otherwise, the appliances consume 0 units ($x_{t+1} = 0$) and there is a choice: either the battery remains discharged and the utility provides 0 units ($b_{t+1} = 0, y_{t+1} = 0$), or the battery charges up and the utility provides this 1 charging unit ($b_{t+1} = 1, y_{t+1} = 1$). Likewise, if the battery is in the charged state at time $t$ ($b_t = 1$), there are three other possible events for time $t + 1$. If the appliances consume 0 units ($x_{t+1} = 0$), then the battery must remain charged and the utility provides 0 units ($b_{t+1} = 1, y_{t+1} = 0$). Otherwise, the appliances consume 1 unit ($x_{t+1} = 1$) and there is a choice: either the battery remains charged and the utility provides 1 unit to the appliances ($b_{t+1} = 1, y_{t+1} = 1$), or the battery discharges its power to the appliances and the utility provides 0 units ($b_{t+1} = 0, y_{t+1} = 0$).

All six events are shown as transitions between the two battery states in the finite state model in Fig. 2. For some of the combinations of battery state $b_t$ and input $x_{t+1}$, there is choice about what happens next and, therefore, the battery system requires a policy.

### 2.2. Stochastic Battery Policies

The battery policies are stochastic; that is, every choice is made with a certain probability. We consider two types: battery-conditioned policies $\pi_b$, which depend only on the current battery state $b_t$, and battery/output-conditioned policies $\pi_{by}$, which depend on the current battery state $b_t$ and the current output $y_t$.

The policies $\pi_b(p_0, p_1)$ can be represented by labeling each of the transitions in Fig. 2 with probabilities, as shown in Fig. 3. Recall that $q_0$ is the probability of $x_{t+1}$ being 0. Thus, $p_0$ and $p_1$ are the probabilities of switching battery state $b_{t+1}$ given that there is a choice to do so. The special case $p_0 = p_1 = 0$ produces a policy that never charges nor discharges the battery, whereas setting $p_0 = p_1 = 1$ forces the battery to charge and discharge whenever possible.

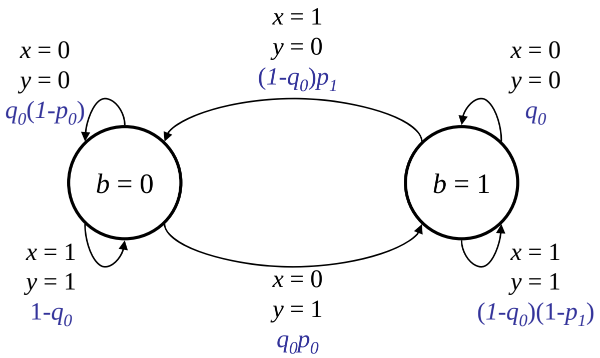

**Fig. 3**. Finite state model depicting battery-conditioned policy $\pi_b(p_0, p_1)$.

The policies $\pi_{by}(p_{00}, p_{01}, p_{10}, p_{11})$ cannot be specified on the graph of Fig. 2. Instead, we augment the state space to represent combinations $(b, y)$ of the battery state and output, and label the transitions with probabilities, as shown in Fig. 4. Here, the $p_{00}$, $p_{01}$, $p_{10}$ and $p_{11}$ are the probabilities of switching output value $y_{t+1}$ given that there is choice to do so. Setting $p_{00} = p_{01} = p_{10} = p_{11} = 0$ forces the battery to keep the output at the same value whenever possible. This special case is, thus, the best-effort algorithm of [8] for our binary-load binary-battery model.

## 3. RATE OF INFORMATION LEAKAGE

We define the rate of information leakage as the mutual information rate $I(X; Y)$. Unlike the three metrics proposed in [8], this definition does not assume that data mining algorithms look for high-frequency signatures. Indeed, this is a principled definition because we make no assumptions about the algorithmic goals or capabilities of the data mining agent at the utility.

Observe that the battery system together with a given policy acts like a communication channel with memory, taking input $X$ and producing output $Y$. Thus, the mutual information rate $I(X; Y)$ can be estimated accurately using a simulation-based technique described in [10]. We briefly summarize the application of the method to our model:

1. Specify the policy $\pi$ and denote the states of the finite state model with an index $s$. For policies $\pi_b$ and $\pi_{by}$, $s$ is equal to $b$ and $(b, y)$, respectively.
2. Sample sequences $\{x_1, \ldots x_n\}$ and $\{y_1, \ldots y_n\}$, where $n$ is large, from the finite state model.
3. Compute $p(y_1, \ldots y_n)$ and $p(x_1, \ldots x_n, y_1, \ldots y_n)$ as described below.
4. Estimate the mutual information rate as

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) \\ &\approx H(X) - \frac{1}{n}\log p(y_1, \ldots y_n) \\ &\qquad + \frac{1}{n}\log p(x_1, \ldots x_n, y_1, \ldots y_n). \end{aligned}$$

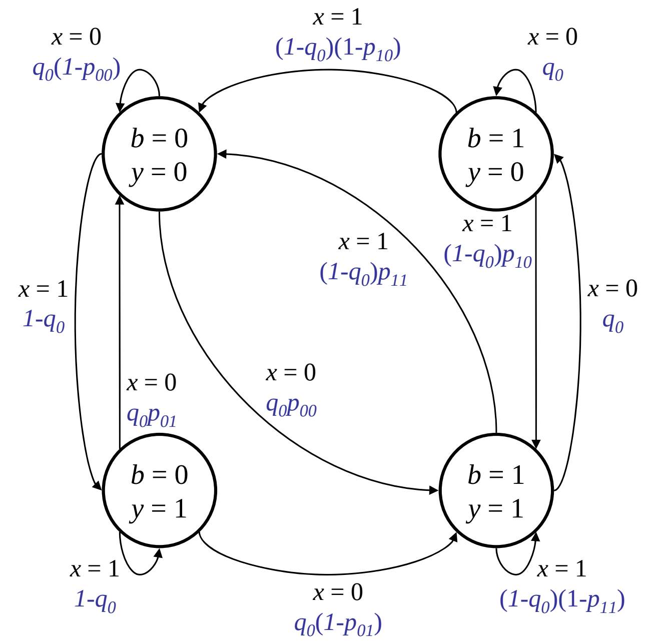

**Fig. 4**. Finite state model depicting battery/output-conditioned policy $\pi_{by}(p_{00}, p_{01}, p_{10}, p_{11})$.

We marginalize the probabilities in Step 3 by converting the finite state model into a trellis with state sequence $\{s_0, \dots s_n\}$ and running the forward pass of the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm [11]. Define the state metrics

$$\begin{aligned}\mu_t(s_t) &= p(s_t, y_1, \dots, y_t)\\ \nu_t(s_t) &= p(s_t, x_1, \dots, x_t, y_1, \dots, y_t).\end{aligned}$$

Then recursively compute

$$\begin{aligned}\mu_{t+1}(s_{t+1}) &= \sum_{x_{t+1}} \sum_{s_t} \mu_t(s_t) p(x_{t+1}, y_{t+1}, s_{t+1}|s_t)\\ \nu_{t+1}(s_{t+1}) &= \sum_{s_t} \nu_t(s_t) p(x_{t+1}, y_{t+1}, s_{t+1}|s_t).\end{aligned}$$

For policies $\pi_b$ and $\pi_{by}$, the nonzero $p(x_{t+1}, y_{t+1}, s_{t+1}|s_t)$ are the transition probabilities labeled on Fig. 3 and 4, respectively. Finally, obtain

$$\begin{aligned}p(y_1, \dots, y_n) &= \sum_{s_n} \mu_n(s_n)\\ p(x_1, \dots, x_n, y_1, \dots, y_n) &= \sum_{s_n} \nu_n(s_n).\end{aligned}$$

## 4. INFORMATION LEAKAGE RESULTS

For different types of input load $X$, we vary the battery policy probabilities by increments of 0.1 and minimize the rate of information leakage $I(X;Y)$. In the computations described in Section 3, we use $n = 10^6$ as recommended by [10].

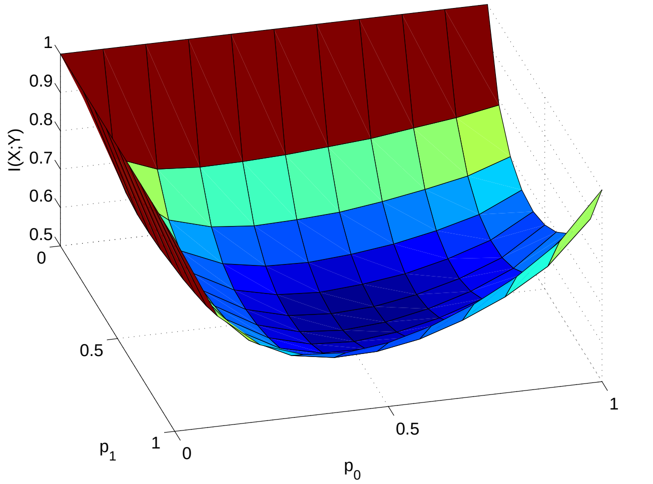

**Fig. 5**. Information leakage rate $I(X;Y)$ of equiprobable input load $X$ under battery-conditioned policies $\pi_b(p_0, p_1)$.

### 4.1. Equiprobable Input Load

When the input load is equiprobable ($q_0 = 0.5$), its entropy rate $H(X) = 1$ bit/symbol.

Fig. 5 plots $I(X;Y)$ for different battery-conditioned policies $\pi_b(p_0, p_1)$. The minimum leakage $I(X;Y)$ of 0.50 bit/symbol, half of $H(X)$, is achieved when $p_0 = p_1 = 0.5$. When $p_0 = p_1 = 0$, $I(X;Y) = 1$ bit/symbol; that is, all the information about $X$ is leaked. Since this policy never charges nor discharges the battery, $y_{t+1} = x_{t+1}$. More interesting is that $p_0 = p_1 = 1$, for which the battery is charged or discharged whenever possible, also leaks $I(X;Y) = 1$ bit/symbol. Inspection of Fig. 2 reveals that $y_{t+1} = x_t$ in this case.

Fig. 6 plots $I(X;Y)$ for battery/output-conditioned policies $\pi_{by}(p_{00}, p_{01}, p_{10}, p_{11})$ for which $p_{00} = p_{11}$ and $p_{01} = p_{10}$. These policies are symmetric with respect to the symbols 0 and 1. The minimum leakage $I(X;Y)$ of 0.50 bit/symbol is achieved when $p_{00} = p_{11} = p_{01} = p_{10} = 0.5$. This policy is the same as $\pi_b(p_0, p_1)$ with $p_0 = p_1 = 0.5$ because the transitions from states $(b = 0, y = 0)$ and $(b = 0, y = 1)$ are identical and so are those from $(b = 1, y = 0)$ and $(b = 1, y = 1)$. The so-called best-effort policy, for which $p_{00} = p_{11} = p_{01} = p_{10} = 0$, has leakage $I(X;Y)$ of 0.68 bit/symbol. Thus, the optimal stochastic policy that does not depend on the output value leaks 26% less information than the best-effort policy that keeps the output equal to its previous value whenever possible.

### 4.2. Biased Input Load

When $q_0 = 0.11$, the input load is biased to value 1 and its entropy rate $H(X) = 0.50$ bit/symbol.

Fig. 7 plots $I(X;Y)$ for different battery-conditioned policies $\pi_b(p_0, p_1)$. The minimum leakage $I(X;Y)$ of 0.23 bit/symbol, less than half of $H(X)$, is achieved when $p_0 = 0.8$ and $p_1 = 0.2$. We did not find any policies appreciably better among battery/output-conditioned policies $\pi_{by}(p_{00}, p_{01}, p_{10}, p_{11})$. Indeed, the best-effort policy, for which $p_{00} = p_{01} = p_{10} = p_{11} = 0$, has leakage $I(X;Y)$ of 0.31 bit/symbol. So, once again, the optimal stochastic policy leaks 26% less information than the best-effort policy.

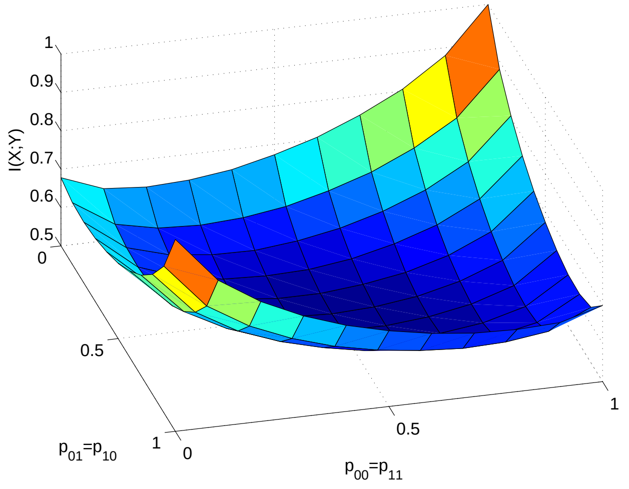

**Fig. 6**. Information leakage rate $I(X;Y)$ of equiprobable input load $X$ under battery/output-conditioned policies $\pi_{by}(p_{00}, p_{01}, p_{10}, p_{11})$ for which $p_{00} = p_{11}$ and $p_{01} = p_{10}$.

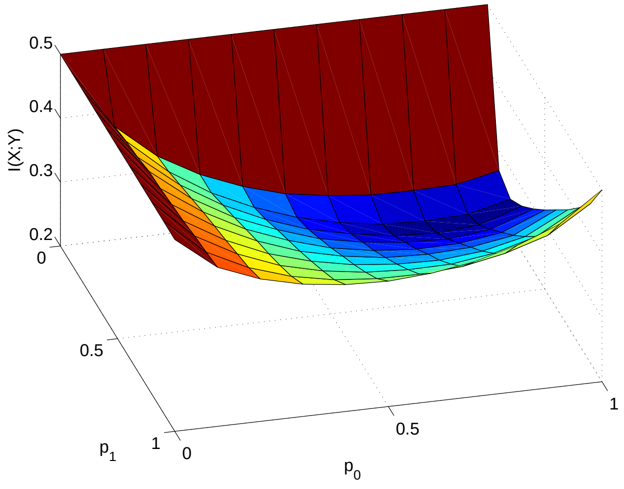

**Fig. 7**. Information leakage rate $I(X;Y)$ of biased ($q_0 = 0.11$) input load $X$ under battery-conditioned policies $\pi_b(p_0, p_1)$.

## 5. EXTENSIONS

We have, so far, limited our investigation to a simple binary-load binary-battery model. We can extend the information leakage rate computation to multilevel loads and batteries by adding more transitions and states, respectively, to the finite state model. Battery policies of greater complexity can be studied by augmenting the state space beyond that shown in Fig. 4. Moreover, including prior input values in the augmented state spaces lets us model Markov (rather than just memoryless) input loads. Time-varying policies and input load statistics are tractable because they can be represented by time-varying trellises, though not finite state models. For continuous loads and batteries, upper and lower bounds on the information leakage rate can be obtained [10].

In addition to data privacy, other properties of battery policies can be considered. We may wish to limit the frequency of charges and discharges to prolong battery life or, if the battery is in an electric vehicle, we may require it to be fully charged at certain times.

## 6. CONCLUSIONS

In this paper, we formulate a rechargeable battery system that partially protects the privacy of a household's electricity load profile as a binary-load binary-battery model. We find that this finite state model makes tractable the computation of the rate of information leakage for different kinds of stochastic battery policies. Specifically, we use a trellis algorithm to estimate the mutual information rate between the battery's input and output loads. Our experimental results show that stochastic battery policies can leak 26% less information than a so-called best-effort policy [8] for both equiprobable and biased input loads. More importantly, our technique for calculating rates of information leakage can be extended to more realistic models, in particular, ones that represent continuous loads and batteries.

## 7. ACKNOWLEDGMENT

The authors are grateful to Pascal Vontobel for valuable discussions.

## 8. REFERENCES

[1] A. Ipakchi and F. Albuyeh, "Grid of the future," *IEEE Power Energy Mag.*, vol. 7, no. 2, pp. 52–62, Mar.-Apr. 2009.

[2] E. L. Quinn, "Privacy and the new energy infrastructure," *Social Science Research Network*, Feb. 2009.

[3] 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.

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

[5] A. Prudenzi, "A neuron nets based procedure for identifying domestic appliances pattern-of-use from energy recordings at meter panel," in *Proc. IEEE Power Eng. Soc. Winter Meeting*, New York, New York, 2002.

[6] 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.

[7] C. Efthymiou and G. Kalogridis, "Smart grid privacy via anonymization of smart metering data," in *Proc. IEEE Smart Grid Commun. Conf.*, Gaithersburg, Maryland, 2010.

[8] 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 Smart Grid Commun. Conf.*, Gaithersburg, Maryland, 2010.

[9] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements," *HP Labs Tech. Report*, 2010.

[10] 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 Trans. Inform. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[11] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error," *IEEE Trans. Inform. Theory*, vol. 20, no. 2, pp. 284–287, Mar. 1974.