# Information-Theoretic Privacy for Smart Metering Systems with a Rechargeable Battery

Simon Li, Ashish Khisti, and Aditya Mahajan

***Abstract*— Smart-metering systems report electricity usage of a user to the utility provider on almost real-time basis. This could leak private information about the user to the utility provider. In this paper, we investigate the use of a rechargeable battery in order to provide privacy to the user. We assume that the user load sequence is a first-order Markov process, the battery satisfies ideal charge conservation, and that privacy is measured using normalized mutual information (leakage rate) between the user load and the battery output. We study the optimal battery charging policy that minimizes the leakage rate among the class of battery policies that satisfy causality and charge conservation. We propose a series reduction on the original problem and ultimately recast it as a Markov Decision Process (MDP) that can be solved using a dynamic program. In the special case of i.i.d. demand, we explicitly characterize the optimal policy and show that the associated leakage rate can be expressed as a single-letter mutual information expression. In this case, we show that the optimal charging policy admits an intuitive interpretation of preserving a certain invariance property of the state. Interestingly an alternative proof of optimality can be provided that does not rely on the MDP approach, but is based on purely information theoretic reductions.**

***Index Terms*— Information theoretic privacy, smart meters, dynamic programming.**

## I. Introduction

SMART meters are a critical part of modern power distribution systems because they provide fine-grained power consumption measurements to utility providers. These fine-grained measurements improve the efficiency of the power grid by enabling services such as time-of-use pricing and demand response [1]. However, this promise of improved efficiency is accompanied by a risk of privacy loss. It is possible for the utility provider—or an eavesdropper—to infer private information including load taxonomy from the fine-grained measurements provided by smart meters [2]–[4]. Such private information could be exploited by third parties for the purpose of targeted advertisement or surveillance. Traditional techniques in which an intermediary anonymizes the data [5] are also prone to privacy loss to an eavesdropper. One possible solution is to partially obscure the load profile by using a rechargeable battery [6]. As the cost of rechargeable batteries decreases (for example, due to proliferation of electric vehicles), using them for improving privacy is becoming economically viable.

In a smart metering system with a rechargeable battery, the energy consumed from the power grid may either be less than the user's demand—the rest being supplied by the battery; or may be more than the user's demand—the excess being stored in the battery. A rechargeable battery provides privacy because the power consumed from the grid (rather than the user's demand) gets reported to the electricity utility (and potentially observed by an eavesdropper). In this paper, we focus on the mutual information between the user's demand and consumption (i.e., the information leakage) as the privacy metric. Mutual information is a widely used metric in the literature on information theoretic security, as it is often analytically tractable and provides a fundamental bound on the probability of detecting the true load sequence from the observation [7]. Our objective is to identify a battery management policy (which determine how much energy to store or discharge from the battery) to minimize the information leakage rate.

We briefly review the relevant literature. The use of a rechargeable battery for providing user privacy has been studied in several recent works, e.g., [6] and [8]-[11]. Most of the existing literature has focused on evaluating the information leakage rate of specific battery management policies. These include the "best-effort" policy [6], which tries to maintain a constant consumption level, whenever possible; and battery conditioned stochastic charging policies [8], in which the conditional distribution on the current consumption depends only on the current battery state (or on the current battery state and the current demand). In [6], the information leakage rate was estimated using Monte-Carlo simulations; in [8], it was calculated using the BCJR algorithm [12]. The methodology of [8] was extended by [9] to include models with energy harvesting and allowing for a certain amount of energy waste. Bounds on the performance of the best-effort policy and hide-and-store policy for models with energy harvesting and infinite battery capacity were obtained in [10]. The performance of the best effort algorithm for an individual privacy metric was considered in [11]. None of these papers address the question of choosing the optimal battery management policy.

Manuscript received August 26, 2016; revised September 13, 2017; accepted January 25, 2018. Date of publication February 23, 2018; date of current version April 19, 2018. This paper was presented in part at the 2015 Signal Processing for Advanced Wireless Communications, in part at the 2016 International Zurich Seminar, and in part at the 2016 American Control Conference.

S. Li was with the Department of Electrical Engineering and Computer Science, University of Toronto, Toronto, ON M5S3G4, Canada (e-mail: yb.li90@gmail.com).

A. Khisti is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S3G4, Canada (e-mail: akhisti@comm.utoronto.ca).

A. Mahajan is with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC H3A2A7, Canada (e-mail: aditya.mahajan@mcgill.ca).

Communicated by H. Permuter, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2018.2809005

Rate-distortion type approaches have also been used to study privacy-utility trade-off [13]–[15]. These models allow the user to report a distorted version of the load to the utility provider, subject to a certain average distortion constraint. Our setup differs from these works as we impose a constraint on the *instantaneous* energy stored in the battery due to its limited capacity. Both our techniques and the qualitative nature of the results are different from these papers.

Our contributions are two-fold. First, when the demand is Markov, we show that the minimum information leakage rate and optimal battery management policies can be obtained by solving an appropriate dynamic program. These results are similar in spirit to the dynamic programs obtained to compute capacity of channels with memory [16]–[18]; however, the specific details are different due to the constraint on the battery state. Second, when the demand is i.i.d., we obtain a single letter characterization of the minimum information leakage rate; this expression also gives the optimal battery management policy. We prove the single letter expression in two steps. On the achievability side we propose a class of policies with a specific structure that enables a considerable simplification of the leakage-rate expression. We find a policy that minimizes the leakage-rate within this restricted class. On the converse side, we obtain lower bounds on the minimal leakage rate and show that these lower bound match the performance of the best structured policy. We provide two proofs. One is based on the dynamic program and the other is based purely on information theoretic arguments.

After the present work was completed, we became aware of [19], where a similar dynamic programming framework is presented for the infinite horizon case. However, no explicit solutions of the dynamic program are derived in [19]. To the best of our knowledge, the present paper is the first work that provides an explicit characterization of the optimal leakage rate and the associated policy for i.i.d. demand.

### A. Notation

Random variables are denoted by uppercase letters ($X$, $Y$, etc.), their realization by corresponding lowercase letters ($x$, $y$, etc.), and their state space by corresponding script letters ($\mathcal{X}$, $\mathcal{Y}$, etc.). $\mathcal{P}_X$ denotes the space of probability distributions on $\mathcal{X}$; $\mathcal{P}_{X|Y}$ denotes the space of condition distributions from $\mathcal{Y}$ to $\mathcal{X}$. $X_a^b$ is a short hand for $(X_a, X_{a+1}, \dots, X_b)$ and $X^b = X_1^b$. For a set $\mathcal{A}$, $\mathbb{1}_{\mathcal{A}}(x)$ denotes the indicator function of the set that equals 1 if $x \in \mathcal{A}$ and zero otherwise. If $\mathcal{A}$ is a singleton set $\{a\}$, we use $\mathbb{1}_a(x)$ instead of $\mathbb{1}_{\{a\}}(x)$.

Given random variables $(X, Y)$ with joint distribution $P_{X,Y}(x, y) = P_X(x)q(y|x)$, $H(X)$ and $H(P_X)$ denote the entropy of $X$, $H(Y|X)$ and $H(q|P_X)$ denote conditional entropy of $Y$ given $X$ and $I(X; Y)$ and $I(q; P_X)$ denote the mutual information between $X$ and $Y$.

## II. Problem Formulation and Main Results

### A. Model and Problem Formulation

Consider a smart metering system as shown in Fig. 1. At each time, the energy consumed from the power grid must equal the user's demand plus the additional energy that is either stored in or drawn from the battery. Let $\{X_t\}_{t\geq 1}$, $X_t \in \mathcal{X}$, denote the user's demand; $\{Y_t\}_{t\geq 1}$, $Y_t \in \mathcal{Y}$, denote the energy drawn from the grid; and $\{S_t\}_{t\geq 1}$, $S_t \in \mathcal{S}$, denote the energy stored in the battery. All alphabets are finite. For convenience, we assume $\mathcal{X} := \{0, 1, \dots, m_x\}$, $\mathcal{Y} := \{0, 1, \dots, m_y\}$, and $\mathcal{S} := \{0, 1, \dots, m_s\}$. Here $m_s$ corresponds to the size of the battery. We note that such a restriction is for simplicity of presentation; the results generalize even when $\mathcal{X}$ and $\mathcal{Y}$ are not necessarily contiguous intervals or integer valued. To guarantee that user's demand is always satisfied, we assume $m_x \leq m_y$ or that $\mathcal{X} \subseteq \mathcal{Y}$ holds more generally.

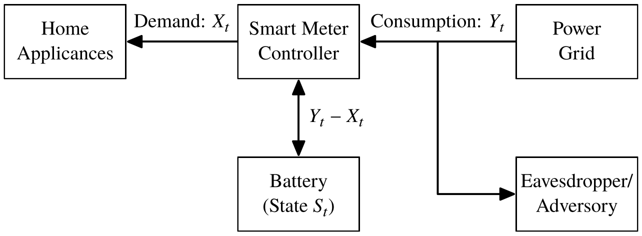

Fig. 1. A smart metering system.

The demand $\{X_t\}_{t\geq 1}$ is a first-order time-homogeneous Markov chain[1] with transition probability $Q$. We assume that $Q$ is irreducible and aperiodic. The initial state $X_1$ is distributed according to probability mass function $P_{X_1}$. The initial charge $S_1$ of the battery is independent of $\{X_t\}_{t\geq 1}$ and distributed according to probability mass function $P_{S_1}$.

The battery is assumed to be ideal and has no conversion losses or other inefficiencies. Therefore, the following conservation equation must be satisfied at all times:

$$S_{t+1} = S_t + Y_t - X_t. \tag{1}$$

Given the history of demand, battery charge, and consumption, a randomized *battery charging policy* $\mathbf{q} = (q_1, q_2, \dots)$ determines the energy consumed from the grid. In particular, given the histories $(x^t, s^t, y^{t-1})$ of demand, battery charge, and consumption at time $t$, the probability that current consumption $Y_t$ equals $y$ is $q_t(y \mid x^t, s^t, y^{t-1})$. For a randomized charging policy to be feasible, it must satisfy the conservation equation (1). So, given the current power demand and battery charge $(x_t, s_t)$, the feasible values of grid consumption are defined by

$$\mathcal{Y}_\circ(s_t - x_t) = \{y \in \mathcal{Y} : s_t - x_t + y \in \mathcal{S}\}. \tag{2}$$

Thus, we require that

$$q_t(\mathcal{Y}_\circ(s_t - x_t) \mid x^t, s^t, y^{t-1}) := \sum_{y \in \mathcal{Y}_\circ(s_t - x_t)} q_t(y \mid x^t, s^t, y^{t-1}) = 1.$$

The set of all such feasible policies is denoted by $\mathcal{Q}_A$.[2] Note that while the charging policy $q_t(\cdot)$ can be a function of the entire history, the support of $q_t(\cdot)$ only depends on the present

[1] In practice, the energy demand is periodic rather than time homogeneous. We are assuming that the total demand may be split into a periodic predictable component and a time-homogeneous stochastic component. In this paper, we ignore the predictable component because it does not affect privacy.

[2] With a slight abuse of notation, we use $\mathcal{Q}_A$ to denote the battery policy for both the infinite and finite-horizon problems

value of $x_t$ and $s_t$ through the difference $s_t - x_t$. This is emphasized in the definition in (2).

The quality of a charging policy depends on the amount of information leaked under that policy. There are different notions of privacy; in this paper, we use mutual information as a measure of privacy. Intuitively speaking, given random variables $(Y, Z)$, the mutual information $I(Y; Z)$ measures the decrease in the uncertainty about $Y$ given by $Z$ (or vice-versa). Therefore, given a policy $\mathbf{q}$, the information about $(X^T, S_1)$ leaked to the utility provider or eavesdropper is captured by $I^{\mathbf{q}}(X^T, S_1; Y^T)$, where the mutual information is evaluated according to the joint probability distribution on $(X^T, S^T, Y^T)$ induced by the distribution $\mathbf{q}$ as follows:

$$\begin{aligned}&\mathbb{P}^{\mathbf{q}}(S^T = s^T, X^T = x^T, Y^T = y^T)\\ &= P_{S_1}(s_1)P_{X_1}(x_1)q_1(y_1 \mid x_1, s_1)\prod_{t=2}^{T}\Big[\mathbb{1}_{s_t}\{s_{t-1} - x_{t-1} + y_{t-1}\}\\ &\qquad \times Q(x_t|x_{t-1})q_t(y_t \mid x^t, s^t, y^{t-1})\Big].\end{aligned}$$

We use information leakage *rate* as a measure of the quality of a charging policy. For a finite planning horizon, the information leakage rate of a policy $\mathbf{q} = (q_1, \dots, q_T) \in \mathcal{Q}_A$ is given by

$$L_T(\mathbf{q}) := \frac{1}{T} I^{\mathbf{q}}(X^T, S_1; Y^T), \tag{3}$$

while for an infinite horizon, the worst-case information leakage rate of a policy $\mathbf{q} = (q_1, q_2, \dots) \in \mathcal{Q}_A$ is given by

$$L_\infty(\mathbf{q}) := \limsup_{T\to\infty} L_T(\mathbf{q}). \tag{4}$$

We are interested in the following optimization problems:

*Problem A: Given the alphabet $\mathcal{X}$ of the demand, the initial distribution $P_{X_1}$ and the transistion matrix $Q$ of the demand process, the alphabet $\mathcal{S}$ of the battery, the initial distribution $P_{S_1}$ of the battery state, and the alphabet $\mathcal{Y}$ of the consumption:*

1) *For a finite planning horizon $T$, find a battery charging policy $\mathbf{q} = (q_1, \dots, q_T) \in \mathcal{Q}_A$ that minimizes the leakage rate $L_T(\mathbf{q})$ given by* (3).
2) *For an infinite planning horizon, find a battery charging policy $\mathbf{q} = (q_1, q_2, \dots) \in \mathcal{Q}_A$ that minimizes the leakage rate $L_\infty(\mathbf{q})$ given by* (4).

The above optimization problem is difficult because we have to optimize a multi-letter mutual information expression over the class of history dependent probability distributions. In the spirit of results for feedback capacity of channels with memory [16]–[18], we show that the above optimization problem can be reformulated as a Markov decision process where the state and action spaces are conditional probability distributions. Thus, the optimal policy and the optimal leakage rate can be computed by solving an appropriate dynamic program. We then provide an explicit solution of the dynamic program for the case of i.i.d. demand.

*Remark 1:* We note that the class of policies we consider in $\mathcal{Q}_A$ are randomized policies i.e., the output at any given time is governed by the conditional distribution $q_t(y_t \mid x^t, s^t, y^{t-1})$. The class of *deterministic policies* where $y_t = h_t(x^t, s^t, y^{t-1})$ is a deterministic function of the past inputs, state and outputs is a special case of randomized policies where $q_t(\cdot)$ is an atomic distribution. As will be apparent from out results, deterministic policies do not suffice to minimize the leakage rate and hence we focus on the class of randomized policies.

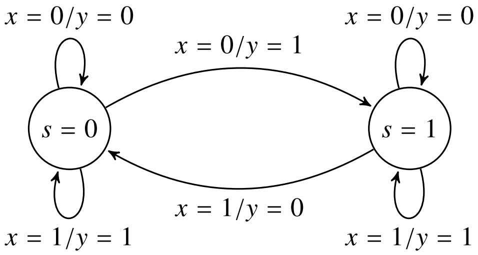

Fig. 2. Binary System model. The battery can be either in $s = 0$ or $s = 1$. The set of feasible transitions from each state are shown in the figure.

### *B. Example: Binary Model*

We illustrate the special case when $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0, 1\}$ in Fig. 2. The input, output, as well as the state, are all binary valued. When the battery is in state $s_t = 0$, there are three possible transitions. If the input $x_t = 1$ then we must select $y_t = 1$ and the state changes to $s_{t+1} = 0$. If instead $x_t = 0$, then there are two possibilities. We can select $y_t = 0$ and have $s_{t+1} = 0$ or we can select $y_t = 1$ and have $s_{t+1} = 1$. In a similar fashion there are three possible transitions from the state $s_t = 1$ as shown in Fig. 2. We will assume that the demand (input) sequence is sampled i.i.d. from an equiprobable distribution, i.e., $P_X(0) = P_X(1) = \frac{1}{2}$.

Consider a simple policy that sets $y_t = x_t$ and ignores the battery state. It is clear that such a policy will lead to maximum leakage $L_T = 1$. Another feasible policy is to set $y_t = 1 - s_t$. Thus whenever $s_t = 0$, we will set $y_t = 1$ regardless of the value of $x_t$, and likewise $s_t = 1$ will result in $y_t = 0$. It turns out that the leakage rate for this policy also approaches 1. To see this note that the eavesdropper having access to $y^T$ also in turn knows $s^T$. Using the battery update equation (1) the sequence $x^{T-1}$ is thus revealed to the eavesdropper, resulting in a leakage rate of at least $1 - 1/T$.

In [8], a probabilistic battery charging policy is introduced that only depends on the current state and input i.e., $q_t(y_t|x^t, s^t) \triangleq q_t(y_t|x_t, s_t)$. Furthermore the policy makes equiprobable decisions between the feasible transitions i.e.,

$$q_t(y_t = 0|x_t, s_t) = q(y_t = 1|x_t, s_t) = 1/2, \quad \text{if } x_t = s_t \tag{5}$$

and $q_t(y_t|x_t, s_t) = \mathbb{1}_{x_t}(y_t)$ otherwise. The leakage rate for this policy was numerically evaluated in [8] using the BCJR algorithm and it was shown numerically that $L_\infty = 0.5$. Such numerical techniques seem necessary in general even for the class of memoryless policies and i.i.d. inputs, as the presence of the battery adds memory into the system.

As a consequence of our main result it follows that the above policy admits a single-letter expression for the leakage rate[3] $L_\infty = I(S^* - X; X)$, thus circumventing the need for numerical techniques. Furthermore it also follows that this leakage rate is indeed the minimum possible one among the class of all feasible policies. Thus it is not necessary for the battery system to use more complex policies that take into account the entire history. We note that a similar result was shown in [20] for the case of finite horizon policies. However the proof in [20] is specific to the binary model. In the present paper we provide a complete single-letter solution to the case of general i.i.d. demand, and a dynamic programming method for the case of first-order Markovian demands, as discussed next.

### *C. Main Results for Markovian Demand*

We identify two structural simplifications for the battery charging policies. First, we show (see Proposition 1 in Section III-A) that there is no loss of optimality in restricting attention to charging policies of the form

$$q_t(y_t|x_t, s_t, y^{t-1}). \tag{6}$$

The intuition is that under such a policy, observing $y^t$ gives partial information only about $(x_t, s_t)$ rather than about the whole history $(x^t, s^t)$.

Next, we identify a sufficient statistic for $y^{t-1}$ when the charging policies are of the form (6). To do so, we use an approach inspired by [16]–[18] and formulate a sequential optimalization problem with partial observations where the state at time $t$ is $(X_t, S_t)$, the observation is $Y_{t-1}$ and the action takes values in a set $\mathcal{A}$ given by

$$\mathcal{A} = \big\{a \in \mathcal{P}_{Y|X,S} : a(\mathcal{Y}_\circ(s-x) \mid x, s) = 1, \\ \forall (x, s) \in \mathcal{X} \times \mathcal{S}\big\}. \tag{7}$$

The set $\mathcal{A}$ is convex and compact.[4]

Based on action $a_t \in \mathcal{A}$, the next observation $Y_t$ is chosen according to the conditional distribution $a_t(\cdot|x_t, s_t)$, the state evolves according to (1), and the system incurs a per-step cost given by

$$\log \frac{a_t(y_t|x_t, s_t)}{\mathbb{P}(Y_t = y_t|Y^{t-1} = y^{t-1})}.$$

We show that Problem A is equivalent to minimizing the total expected cost for the above sequential optimization problem (see Proposition 2 in Sec. III-B), which, in turn, is similar to a partially observable Markov decision process (POMDP) (but with some differences; see Sec. III-C) and may be solved using dynamic programming. For that matter, given a policy $\mathbf{q}$ and any realization $y^{t-1}$ of $Y^{t-1}$, define the belief state $\pi_t \in \mathcal{P}_{X,S}$ as follows: for all $x \in \mathcal{X}$, $s \in \mathcal{S}$,

$$\pi_t(x, s) = \mathbb{P}^{\mathbf{q}}(X_t = x, S_t = s|Y^{t-1} = y^{t-1}). \tag{8}$$

Then, we show (see Theorems 1 and 2 below) that there is no loss of optimality in restricting attention to charging policies of the form

$$q_t(y_t|x_t, s_t, \pi_t). \tag{9}$$

Such a charging policy is Markovian in the belief state $\pi_t$ and the optimal policies of such form can be searched using a dynamic program.

To succinctly write the dynamic program, for any $a \in \mathcal{A}$, we define the Bellman operator $\mathscr{B}_a : [\mathcal{P}_{X,S} \to \mathbb{R}] \to [\mathcal{P}_{X,S} \to \mathbb{R}]$ as follows: for any $V : \mathcal{P}_{X,S} \to \mathbb{R}$ and any $\pi \in \mathcal{P}_{X,S}$,

$$[\mathscr{B}_a V](\pi) = I(a; \pi) \\ + \sum_{\substack{x\in\mathcal{X}, s\in\mathcal{S}, \\ y\in\mathcal{Y}}} \pi(x, s)a(y \mid x, s)V(\varphi(\pi, y, a)) \tag{10}$$

where the function $\varphi$ is a non-linear filtering equation defined in Sec. III-C.

Our main results are the following:

*Theorem 1: We have the following for Problem A with a finite planning horizon $T$:*

1) Value functions*: Iteratively define* value functions $V_t : \mathcal{P}_{X,S} \to \mathbb{R}$ *as follows. For any* $\pi \in \mathcal{P}_{X,S}$, $V_{T+1}(\pi) = 0$, *and for* $t \in \{T, \dots, 1\}$,

$$V_t(\pi) = \min_{a\in\mathcal{A}} [\mathscr{B}_a V_{t+1}](\pi), \quad \forall \pi \in \mathcal{P}_{X,S}. \tag{11}$$

*Then,* $V_t(\pi)$ *is continuous and concave in* $\pi$.

2) Optimal policy*: Let* $f_t^*(\pi)$ *denote the arg min of the right hand side of* (11). *Then, optimal policy* $\mathbf{q}^* = (q_1^*, \dots, q_T^*)$ *is given by*

$$q_t^*(y_t \mid x_t, s_t, \pi_t) = a_t(y_t \mid x_t, s_t), \text{ where } a_t = f_t^*(\pi_t).$$

*Thus, there is no loss of optimality in restricting attention to charging policies of the form* (9).

3) Optimal leakage rate*: The optimal (finite horizon) leakage rate is given by* $V_1(\pi_1)/T$, *where* $\pi_1(x, s) = P_{X_1}(x)P_{S_1}(s)$. □

See Section III for proof.

*Theorem 2: We have the following for Problem A with an infinite planning horizon:*

1) Value function*: Consider the following fixed point equation*

$$J + v(\pi) = \min_{a\in\mathcal{A}} [\mathscr{B}_a v](\pi), \quad \forall \pi \in \mathcal{P}_{X,S}, \tag{12}$$

*where* $J \in \mathbb{R}$ *is a constant and* $v : \mathcal{P}_{X,S} \to \mathbb{R}$.

2) Optimal policy*: Suppose there exists* $(J, v)$ *that satisfy* (12). *Let* $f^*(\pi)$ *denote the arg min of the right hand side of* (12). *Then, the time-homogeneous optimal policy* $\mathbf{q}^* = (q^*, q^*, \dots)$ *given by*

$$q^*(y_t \mid x_t, s_t, \pi_t) = a(y_t \mid x_t, s_t), \text{ where } a = f^*(\pi)$$

*is optimal. Thus, there is no loss of optimality in restricting attention to charging time-homogeneous policies of the form* (9).

3) Optimal leakage rate*: The optimal (infinite horizon) leakage rate is given by* $J$. □

[3] The random variable $S^*$ is an equiprobable binary valued random variable, independent of $X$. See Sec. II-E.

[4] If $a_1, a_2 \in \mathcal{A}$, then any linear combination $a'$ of them, where $a' = \lambda a_1 + (1-\lambda)a_2$ with $\lambda \in (0, 1)$, also satisfies $a'(\mathcal{Y}_\circ(s - x \mid x, s) = 1$ for all $(x, s) \in \mathcal{X} \times \mathcal{S}$ and, therefore, belongs to $\mathcal{A}$. Hence $\mathcal{A}$ is convex. Moreover, it is easy to see that $\mathcal{A}$ is closed and $\mathcal{P}_{Y|X,S}$ is compact. Therefore, $\mathcal{A}$ is also compact.

*Proof:* Given the result of Theorem 1, the result of Theorem 2 follows from standard dynamic programming arguments. See, for example, [21, Th. 5.2.4 and eq. (5.2.18)]. □

There are various conditions that guarantee the existence of a $(J, v)$ that satisfies (12). Most of these conditions require the ergodicity of the process $\{\Pi_t\}_{t\geq 1}$ for every stationary Markov policy $f: \mathcal{P}_{X,S} \to \mathcal{A}$. We refer the reader to [22, Ch. 3] and [23, Ch. 10] for more details.

The dynamic program above resembles the dynamic program for partially observable Markov decision processes (POMDP) (see [24, Ch. 5]) with hidden state $(X_t, S_t)$, observation $Y_t$, and action $A_t$. However, in contrast to POMDPs, the expected per-step cost $I(a; \pi)$ is not linear in $\pi$. Nonetheless, one could use computational techniques from POMDPs to approximately solve the dynamic programs of Theorems 1 and 2. See Section III-D for a brief discussion.

*Remark 2:* Although the above results assume that the demand is a first-order Markov chain, they extend naturally to the case when the demand is higher-order Markov. In particular, suppose the demand $\{X_t\}_{t\geq 1}$ is a $k$-th order Markov chain. Then, we can define another process $\{\tilde{X}_t\}_{t\geq 1}$ where $\tilde{X}_t = (X_{t-k+1}, \dots, X_t)$, and use $\tilde{X}_t$ in Theorems 1 and 2.

### *D. Main Result for i.i.d. Demand*

Assume the following:

*Assumption 1:* The demand $\{X_t\}_{t\geq 1}$ is i.i.d. with probability distribution $P_X$.

We provide an explicit characterization of optimal policy and optimal leakage rate for this case.

Define an auxiliary state variable $W_t = S_t - X_t$ that takes values in $\mathcal{W} = \{s - x : s \in \mathcal{S}, x \in \mathcal{X}\}$. For any $w \in \mathcal{W}$, define:

$$\mathcal{D}(w) = \{(x, s) \in \mathcal{X} \times \mathcal{S} : s - x = w\}. \tag{13}$$

Then, we have the following.

*Theorem 3: Define*

$$J^* = \min_{\theta\in\mathcal{P}_S} I(S - X; X) = \min_{\theta\in\mathcal{P}_S} \big\{H(S - X) - H(S)\big\} \tag{14}$$

*where $X$ and $S$ are independent with $X \sim P_X$ and $S \sim \theta$. Let $\theta^*$ denote the arg min in* (14)*. Define $\xi^*(w) = \sum_{(x,s)\in\mathcal{D}(w)} P_X(x)\theta^*(s)$. Then, under Assumption 1*

1) *$J^*$ is the optimal (infinite horizon) leakage rate.*
2) *Define $b^* \in \mathcal{P}_{Y|W}$ as follows:*

$$b^*(y|w) = \begin{cases} P_X(y)\frac{\theta^*(y+w)}{\xi^*(w)} & \text{if } y \in \mathcal{X} \cap \mathcal{Y}_\circ(w) \\ 0 & \text{otherwise }. \end{cases} \tag{15}$$

*We call $b^*$ as a structured policy with respect to $(\theta^*, \xi^*)$. Then, the memoryless charging policy $\mathbf{q}^* = (q_1^*, q_2^*, \dots)$ given by*

$$q_t^*(y \mid x, s, \pi_t) = b^*(y \mid s - x) \tag{16}$$

*is optimal and achieves the optimal (infinite horizon) leakage rate.* □

Note that the optimal charging policy is *time-invariant* and *memoryless*, i.e., the distribution on $Y_t$ does not depend on $\pi_t$ (and, therefore on $y^{t-1}$).

The proof, which is presented in Section IV, is based on the standard arguments of showing achievability and a converse. On the achievability side we show that the policy in (15) belongs to a class of policies that satisfies a certain invariance property. Using this property the multi-letter mutual information expression can be reduced into a single-letter expression. For the converse we provide two proofs. The first is based on a simplification of the dynamic program of Theorem 2. The second is based on purely probabilistic and information theoretic arguments.

### *E. Binary Model (Revisited)*

We revisit the binary model in Section II-B, where the demand has equiprobable distribution, i.e., $P_X(x) = \frac{1}{2}$ for $x \in \{0, 1\}$. Consider $\theta(0) = p$, $\theta(1) = 1 - p$. Let $W = S - X$. Then,

$$\mathbb{P}(W = w) = \begin{cases} \frac{1}{2}\, p, & \text{if } w = -1 \\ \frac{1}{2}, & \text{if } w = 0 \\ \frac{1}{2}(1 - p), & \text{if } w = 1. \end{cases} \tag{17}$$

Thus,

$$I(W; X) = H(W) - H(S) = 1 - \tfrac{1}{2}\, h(p)$$

where $h(p)$ is the binary entropy function:

$$h(p) = -p \log p - (1 - p) \log(1 - p).$$

Thus, the value of $p$ that minimizes $I(W; X)$ is $p^* = \frac{1}{2}$ and the optimal leakage rate is $I(W; X) = \frac{1}{2}$.

The corresponding $\xi^*$ is obtained by substituting $p = \frac{1}{2}$ in (17). For ease of notation, we denote $b^*(\cdot|w)$ as $[b^*(0|w), b^*(1|w)]$. Then,

$$b^*(\cdot| - 1) = [0, 1], \quad b^*(\cdot|0) = [\tfrac{1}{2}, \tfrac{1}{2}], \quad b^*(\cdot| + 1) = [1, 0].$$

It can be shown that this strategy is the same as (5), which was proposed in [8]. This yields an analytical proof of the result in [8].

### *F. Salient Features of the Result for i.i.d. Demand*

Theorem 3 shows that even if consumption could take larger values than the demand, i.e., $\mathcal{Y} \supset \mathcal{X}$, under the optimal policy, $Y_t$ takes values only in $\mathcal{X}$. This agrees with the intuition that a consumption larger that $m_x$ reveals that the battery has low charge and that the power demand is high. In extreme cases, a large consumption may completely reveal the battery and power usage thereby increasing the information leakage.

We now show some other properties of the optimal policy.

*Property 1:* The mutual information $I(S - X; X)$ is strictly convex in the distribution $\theta$ and, therefore, $\theta^* \in int(\mathcal{P}_S)$.

See Appendix I for proof.

As a consequence, the optimal $\theta^*$ in (14) may be obtained using the Blahut-Arimoto algorithm [25], [26].

*Property 2:* Under the battery charging policy specified in Theorem 3, the power consumption $\{Y_t\}_{t\geq 1}$ is i.i.d. with marginal distribution $P_X$. Thus, $\{Y_t\}_{t\geq 1}$ is statistically indistinguishable from $\{X_t\}_{t\geq 1}$.

See Remarks 3 and 5 in Section IV-B for proof.

*Property 3:* If the power demand has a symmetric PMF, i.e., for any $x \in \mathcal{X}$, $P_X(x) = P_X(m_x - x)$, then the optimal $\theta^*$ in Theorem 3 is also symmetric, i.e., for any $s \in \mathcal{S}$, $\theta^*(s) = \theta^*(m_s - s)$.

*Proof:* For $\theta \in \mathcal{P}_S$, define $\bar{\theta}(s) = \theta(m_s - s)$. Let $X \sim P_X$, $S \sim \theta$ and $\bar{S} \sim \bar{\theta}$. Then, by symmetry

$$I(S - X; X) = I(\bar{S} - X; X). \tag{18}$$

For any $\lambda \in (0, 1)$, let $\theta_\lambda(s) = \lambda\theta(s) + (1 - \lambda)\bar{\theta}(s)$ denote the convex combination of $\theta$ and $\bar{\theta}$. Let $S_\lambda \sim \theta_\lambda$. By Property 1, if $\theta \neq \bar{\theta}$, then

$$\begin{aligned} I(S_\lambda - X; X) &< \lambda I(S - X; X) + (1 - \lambda)I(\bar{S} - X; X) \\ &= I(S - X; X), \end{aligned}$$

where the last equation uses (18).

Thus, if $\theta \neq \bar{\theta}$, we can strictly decrease the mutual information by using $\theta_\lambda$. Hence, the optimal distribution must have the property that $\theta^*(s) = \theta^*(m_s - s)$. □

Given a distribution $\mu$ on some alphabet $\mathcal{M}$, we say that the distribution is ***almost*** *symmetric and unimodal* around $m^* \in \mathcal{M}$ if

$$\mu_{m^*} \geq \mu_{m^*+1} \geq \mu_{m^*-1} \geq \mu_{m^*+2} \geq \mu_{m^*-2} \geq \dots$$

where we use the interpretation that for $m \notin \mathcal{M}$, $\mu_m = 0$. Similarly, we say that the distribution is *symmetric and unimodal* around $m^* \in \mathcal{M}$ if

$$\mu_{m^*} \geq \mu_{m^*+1} = \mu_{m^*-1} \geq \mu_{m^*+2} = \mu_{m^*-2} \geq \dots$$

Note that a distribution can be symmetric and unimodal only if its support is odd.

*Property 4:* If the power demand is symmetric and unimodal around $\lfloor m_x/2 \rfloor$, then the optimal $\theta^*$ in Theorem 3 is *almost* symmetric and unimodal around $\lfloor m_s/2 \rfloor$. In particular, if $m_s$ is even, then

$$\theta^*_{m^*} \geq \theta^*_{m^*+1} = \theta^*_{m^*-1} \geq \theta^*_{m^*+2} = \theta^*_{m^*-2} \geq \dots$$

and if $m_s$ is odd, then

$$\theta^*_{m^*} = \theta^*_{m^*+1} \geq \theta^*_{m^*-1} = \theta^*_{m^*+2} \geq \theta^*_{m^*-2} = \dots$$

where $m^* = \lfloor m_s/2 \rfloor$.

*Proof:* Let $\bar{X} = -X$. Then, $I(S - X; S) = H(S - X) - H(S) = H(S + \bar{X}) - H(S)$. Note that $P_{\bar{X}}$ is also symmetric and unimodal around $\lfloor m_x/2 \rfloor$.

Let $S^\circ$ and $\bar{X}^\circ$ denote the random variables $S - \lfloor m_s/2 \rfloor$ and $\bar{X} - \lfloor m_x/2 \rfloor$. Then $\bar{X}^\circ$ is also symmetric and unimodal around the origin and

$$I(S - X; X) = H(S^\circ + \bar{X}^\circ) - H(S^\circ).$$

Now given any distribution $\theta^\circ$ of $S^\circ$, let $\theta^+$ be a permutation of $\theta^\circ$ that is almost symmetric and unimodal with a positive bias around origin. Then by [27, Corollary III.2], $H(P_X * \theta^\circ) \geq H(P_X * \theta^+)$. Thus, the optimal distribution must have the property that $\theta^\circ = \theta^+$ or, equivalently, $\theta$ is almost unimodal and symmetric around $\lfloor m_s/2 \rfloor$.

Combining this with the result of Property 3 gives the characterization of the distribution when $m_s$ is even or odd. □

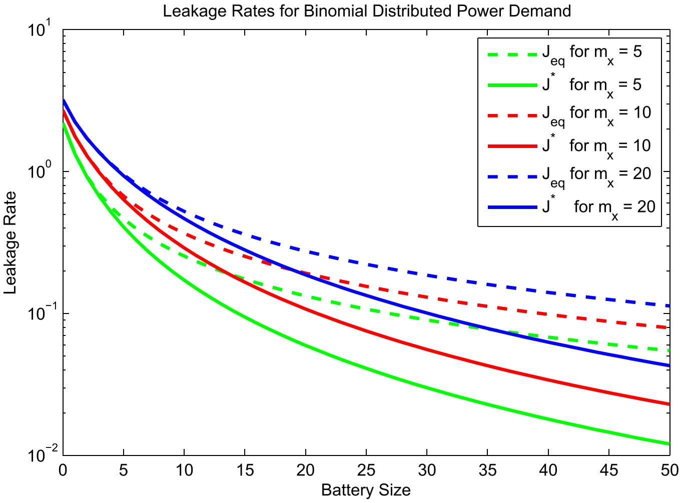

Fig. 3. A comparison of the performance of $q_{eq} \in \mathcal{Q}_B$ as defined in (19) with the optimal leakage rate for i.i.d. Binomial distributed demand Binomial $(m_x, 0.5)$ for $m_x = \{5, 10, 20\}$.

### *G. Numerical Example: i.i.d. Demand*

Suppose there are $n$ identical devices in the house and each is on with probability $p$. Thus, $X \sim \text{Binomial}(n, p)$. We derive the optimal policy and optimal leakage rate for this scenario under the assumption that $\mathcal{Y} = \mathcal{X}$. We consider two specific examples, where we numerically solve (14).

Suppose $n = 6$ and $p = 0.5$.

1) Consider $\mathcal{S} = [0{:}5]$. Then, by numerically solving (14), we get that the optimal leakage rate $J^*$ is is 0.4616 and the optimal battery charge distribution $\theta^*$ is

$$\{0.1032, 0.1747, 0.2221, 0.2221, 0.1747, 0.1032\}.$$

2) Consider $\mathcal{S} = [0{:}6]$. Then, by numerically solving (14), we get that the optimal leakage rate $J^*$ is is 0.3774 and the optimal battery charge distribution $\theta^*$ is

$$\{0.0773, 0.1364, 0.1847, 0.2031, 0.1847, 0.1364, 0.0773\}.$$

Note that both results are consistent with Properties 3 and 4.

We next compare the performance with the following time-homogeneous benchmark policy $\mathbf{q}_{eq} \in \mathcal{Q}_B$: for all $y \in \mathcal{Y}$, $w \in \mathcal{W}$,

$$q_t(y_t|w_t) = \frac{\mathbb{1}_{\mathcal{Y}_\circ(w_t)}\{y_t\}}{|\mathcal{Y}_\circ(w_t)|}. \tag{19}$$

This benchmark policy chooses all feasible values of $Y_t$ with equal probability. For that reason we call it *equi-probable policy* and denote its performance by $J_{eq}$.

In Fig. 3, we compare the performance of $\mathbf{q}^*$ and $\mathbf{q}_{eq}$ as a function of battery sizes for different demand alphabets.

Under $\mathbf{q}_{eq}$, the MDP converges to a belief state that is approximately uniform. Hence, in the low battery size regime, $J_{eq}$ is close to optimal but its performance gradually worsens with increasing battery size.

### *H. Numerical Example: Continuous Valued Alphabets*

Although our setup assumes discrete alphabets, it can shown that Theorem 3 can be extended to continuous alphabets under mild technical conditions assuming that the density function for $X$, say $f_X(x)$ exists; see [28] for details. As such we provide two achievability proofs for Theorem 3. The *weak achievability* in Section IV-B assumes that the distribution of the initial state $S_1$ can be selected by the user. The *strong achievability* in Section IV-C assumes that the initial state $S_1$ can be arbitrary. The proof of the weak achievability extends immediately to continuous valued alphabets. Furthermore the proof of the converse based on information theoretic arguments in Section IV-E also extends to continuous valued alphabets. We provide a numerical example involving a continuous valued input.

Let $\mathcal{X} = [0, 1]$ be continuous valued, and let $f_X(x) = 1$ for $x \in [0, 1]$. We assume that $\mathcal{S} = [0, B]$ where $B$ denotes the storage capacity. We will assume that $B \geq 2$ for convenience. Following Theorem 3, it suffices to take the output alphabet to be $\mathcal{Y} = [0, 1]$. Note that for $W = S - X$, we have that $\mathcal{W} = [-1, B]$ and that the support of the output for any given $w$ is given by:

$$\mathcal{Y}_\circ(w) = \begin{cases} [-w, 1], & -1 \leq w \leq 0, \\ [0, 1], & 0 \leq w \leq B-1, \\ [0, B-w], & B-1 \leq w \leq B. \end{cases} \tag{20}$$

Let $\theta^*(\cdot)$ be the density function that minimizes (14) and $S^*$ denote the random variable with this density, independent of $X$. Let $W = S^* - X$ and $\xi^*(w)$ be the associated density. Then it follows from (15) that the optimal policy is given by

$$b^*(y|w) = \frac{\theta^*(y+w)}{\xi^*(w)}, \quad y \in \mathcal{Y}_\circ(w). \tag{21}$$

Instead of computing $\theta^*(\cdot)$ numerically, which is cumbersome due to the density functions, we provide an analytical lower bound on the leakage rate. Note that the objective in (14) can be expressed as:

$$I(S-X; X) = h(S-X) - h(S) \tag{22}$$

where $h(\cdot)$ is the differential entropy. Using the entropy power inequality [29], since $S$ and $X$ are independent, we have for any density $f_S(\cdot)$ that:

$$2^{2h(S-X)} \geq 2^{2h(S)} + 2^{2h(X)}, \tag{23}$$

where we use the fact that $h(-X) = h(X)$. Substituting in (22) we have

$$\begin{aligned} I(S-X; X) &\geq \frac{1}{2}\log_2\left(2^{2h(S)} + 2^{2h(X)}\right) - h(S) \\ &= \frac{1}{2}\log_2\left(1 + 2^{2h(X)-2h(S)}\right) \\ &= \frac{1}{2}\log_2\left(1 + 2^{-2h(S)}\right) & (24) \\ &\geq \min_S \frac{1}{2}\log_2\left(1 + 2^{-2h(S)}\right) \\ &= \frac{1}{2}\log_2\left(1 + 2^{-2\{\max_S h(S)\}}\right) & (25) \\ &= \frac{1}{2}\log_2\left(1 + \frac{1}{B^2}\right) & (26) \end{aligned}$$

where we use the fact that when $X \sim \text{Unif}[0, 1]$, we have that $h(X) = 0$, see e.g., [30] in (24) and the fact that the expression in (25) is decreasing in $h(S)$. Finally in (26) we use the fact that a uniform distribution maximizes the differential entropy when $\mathcal{S} = [0, B]$ is fixed and the maximum value is $h(S) = \log_2 B$.

For the achievability, we evaluate the leakage rate by selecting $\theta^*(s)$ to be a uniform distribution over $[0, B]$ and using (21). The resulting leakage rate is given by

$$L^+ = h(W) - h(S) = h(\xi) - \log_2(B) \tag{27}$$

where the density $\xi(w)$ for $W = S - X$ is as follows:

$$\xi(w) = \begin{cases} \dfrac{1+w}{B}, & -1 \leq w \leq 0 \\ \dfrac{1}{B}, & 0 \leq w \leq B-1 \\ \dfrac{B-w}{B}, & B-1 \leq w \leq B \end{cases}$$

and we use that when $\theta(s)$ is a uniform density over $[0, B]$, it follows that $h(S) = \log_2(B)$. Through straightforward computations it can be shown that

$$h(\xi) = \frac{1}{2B\ln 2} + \log_2(B) \tag{28}$$

and thus it follows that

$$L^+ = \frac{1}{2B\ln 2}$$

is achievable with a uniform distribution on the state. We note that the analytical lower bound on the leakage rate in (26) decays as $1/B^2$ for large $B$ while the achievable rate decays as $1/B$. It will be interesting in future to determine the structure of the optimal input distribution and study the associated leakage rate.

## III. Proof of Theorem 1

One of the difficulties in obtaining a dynamic programming decomposition for Problem A is that the objective function is not of the form $\sum_{t=1}^{T} \text{cost}(\text{state}_t, \text{action}_t)$. We show that there is no loss of optimality to restrict attention to a class of policies $\mathcal{Q}_B$ and for any policy in $\mathcal{Q}_B$, the mutual information may be written in an additive form.

### A. Simplification of Optimal Charging Policies

Let $\mathcal{Q}_B \subset \mathcal{Q}_A$ denote the set of charging policies that choose consumption based only on the consumption history, current demand, and battery state. Thus, for $\mathbf{q} \in \mathcal{Q}_B$, at any time $t$, given history $(x^t, s^t, y^{t-1})$, the consumption $Y_t$ is $y$ with probability $q_t(y \mid x_t, s_t, y^{t-1})$. Then the joint distribution on $(X^T, S^T, Y^T)$ induced by $\mathbf{q} \in \mathcal{Q}_B$ is given by

$$\begin{aligned} &\mathbb{P}^{\mathbf{q}}(S^T = s^T, X^T = x^T, Y^T = y^T) \\ &= P_{S_1}(s_1)P_{X_1}(x_1)q_1(y_1 \mid x_1, s_1)\prod_{t=2}^{T}\Big[\mathbb{1}_{s_t}\{s_{t-1}-x_{t-1}+y_{t-1}\} \\ &\qquad \times Q(x_t|x_{t-1})q_t(y_t \mid x_t, s_t, y^{t-1})\Big]. \end{aligned}$$

*Proposition 1: In Problem A, there is no loss of optimality in restricting attention to charging policies in $\mathcal{Q}_B$. Moreover,* *for any* $\mathbf{q} \in \mathcal{Q}_B$*, the objective function takes an additive form:*

$$L_T(\mathbf{q}) = \frac{1}{T}\sum_{t=1}^{T} I^{\mathbf{q}}(X_t, S_t; Y_t \mid Y^{t-1})$$

*where*

$$I^{\mathbf{q}}(X_t, S_t; Y_t \mid Y^{t-1}) = \sum_{\substack{x_t\in\mathcal{X}, s_t\in\mathcal{S}\\ y^t\in\mathcal{Y}^t}} \mathbb{P}^{\mathbf{q}}(X_t = x_t, S_t = s_t, Y^t = y^t) \times \log \frac{q_t(y_t \mid x_t, s_t, y^{t-1})}{\mathbb{P}^{\mathbf{q}}(Y_t = y_t \mid Y^{t-1} = y^{t-1})}.$$

See Appendix II for proof. The intuition behind why policies in $\mathcal{Q}_B$ are no worse that others in $\mathcal{Q}_A$ is as follows. For a policy in $\mathcal{Q}_A$, observing the realization $y^t$ of $Y^t$ gives partial information about the history $(x^t, s^t)$ while for a policy in $\mathcal{Q}_B$, $y_t$ gives partial information only about the current state $(x_t, s_t)$. The dependence on $(x_t, s_t)$ cannot be removed because of the conservation constraint (1).

Proposition 1 shows that the total cost may be written in an additive form. Next we use an approach inspired by [16]–[18] and formulate an equivalent sequential optimization problem.

### B. An equivalent sequential optimization problem

Consider a system with state process $\{X_t, S_t\}_{t\geq 1}$ where $\{X_t\}_{t\geq 1}$ is an exogenous Markov process as before and $\{S_t\}_{t\geq 1}$ is a controlled Markov process as specified below. At time $t$, a decision maker observes $Y^{t-1}$ and chooses a distribution valued action $A_t \in \mathcal{A}$, where $\mathcal{A}$ is given by (7), as follows:

$$A_t = f_t(Y^{t-1}) \tag{29}$$

where $\mathbf{f} = (f_1, f_2, \dots)$ is called the decision policy.

Based on this action, an auxiliary variable $Y_t \in \mathcal{Y}$ is chosen according to the conditional probability $a_t(\cdot \mid x_t, s_t)$ and the state $S_{t+1}$ evolves according to (1).

At each stage, the system incurs a per-step cost given by

$$c_t(x_t, s_t, a_t, y^t; \mathbf{f}) := \log \frac{a_t(y_t \mid x_t, s_t)}{\mathbb{P}^{\mathbf{f}}(Y_t = y_t \mid Y^{t-1} = y^{t-1})}. \tag{30}$$

The objective is to choose a policy $\mathbf{f} = (f_1, \dots, f_T)$ to minimize the total finite horizon cost given by

$$\tilde{L}_T(\mathbf{f}) := \frac{1}{T}\mathbb{E}^{\mathbf{f}}\left[\sum_{t=1}^{T} c_t(X_t, S_t, A_t, Y^t; \mathbf{f})\right] \tag{31}$$

where the expectation is evaluated with respect to the probability distribution $\mathbb{P}^{\mathbf{f}}$ induced by the decision policy $\mathbf{f}$.

*Proposition 2: The sequential decision problem described above is equivalent to Problem A. In particular,*

1) *Given* $\mathbf{q} = (q_1, \dots, q_T) \in \mathcal{Q}_B$*, let* $\mathbf{f} = (f_1, \dots, f_T)$ *be*

$$f_t(y^{t-1}) = q_t(\cdot \mid \cdot, \cdot, y^{t-1}).$$

*Then* $\tilde{L}_T(\mathbf{f}) = L_T(\mathbf{q})$.

2) *Given* $\mathbf{f} = (f_1, \dots, f_T)$*, let* $\mathbf{q} = (q_1, \dots, q_T) \in \mathcal{Q}_B$ *be*

$$q_t(y_t \mid x_t, s_t, y^{t-1}) = a_t(y_t \mid x_t, s_t), \text{ where } a_t = f_t(y^{t-1}).$$

*Then* $L_T(\mathbf{q}) = \tilde{L}_T(\mathbf{f})$.

*Proof:* For any history $(x^t, s^t, y^{t-1})$, $a^t \in \mathcal{A}$, and $s_{t+1} \in \mathcal{S}$,

$$\begin{aligned}\mathbb{P}(S_{t+1} = s_{t+1} \mid X^t = x^t, S^t = s^t, Y^t = y^t, A^t = a^t)\\ = \sum_{y_t\in\mathcal{Y}} \mathbb{1}_{s_{t+1}}\{s_t + y_t - x_t\}\, a_t(y_t \mid x_t, s_t)\\ = \mathbb{P}(S_{t+1} = s_{t+1} \mid X_t = x_t, S_t = s_t, A_t = a_t).\end{aligned} \tag{32}$$

Thus, the probability distribution on $(X^T, S^T, Y^T)$ induced by a decision policy $\mathbf{f} = (f_1, \dots, f_T)$ is given by

$$\begin{aligned}\mathbb{P}^{\mathbf{f}}(S^T = s^T, X^T = x^T, Y^T = y^T)\\ = P_{S_1}(s_1)P_{X_1}(x_1)q_1(y_1 \mid x_1, s_1)\prod_{t=2}^{T}\Big[\mathbb{1}_{s_t}\{s_{t-1} - x_{t-1} + y_{t-1}\}\\ \times Q(x_t|x_{t-1})a_t(y_t|x_t, s_t)\Big].\end{aligned}$$

where $a_t = f_t(y^{t-1})$. Under the transformations described in the Proposition, $\mathbb{P}^{\mathbf{f}}$ and $\mathbb{P}^{\mathbf{q}}$ are identical probability distributions. Consequently, $\mathbb{E}^{\mathbf{f}}[c_t(X_t, S_t, A_t, Y^t; \mathbf{f})] = I^{\mathbf{q}}(S_t, X_t; Y_t \mid Y^{t-1})$. Hence, $L_T(\mathbf{q})$ and $\tilde{L}_T(\mathbf{f})$ are equivalent. □

Eq. (32) implies that $\{X_t, S_t\}_{t\geq 1}$ is a controlled Markov process with control action $\{A_t\}_{t\geq 1}$. In the next section, we obtain a dynamic programming decomposition for this problem. For the purpose of writing the dynamic program, it is more convenient to write the policy (29) as

$$A_t = f_t(Y^{t-1}, A^{t-1}). \tag{33}$$

Note that these two representations are equivalent. Any policy of the form (29) is also a policy of the form (33) (that simply ignores $A^{t-1}$); any policy of the form (33) can be written as a policy of the form (29) by recursively substituting $A_t$ in terms of $Y^{t-1}$. Since the two forms are equivalent, in the next section we assume that the policy is of the form (33).

### C. A Dynamic Programming Decomposition

The model described in Section III-B above is similar to a POMDP (partially observable Markov decision process): the system state $(X_t, S_t)$ is partially observed by a decision maker who chooses action $A_t$. However, in contrast to the standard cost model used in POMDPs, the per-step cost depends on the observation history and *past policy*. Nonetheless, if we consider the belief state as the information state, the problem can be formulated as a standard MDP.

For that matter, for any realization $y^{t-1}$ of past observations and any choice $a^{t-1}$ of past actions, define the belief state $\pi_t \in \mathcal{P}_{X,S}$ as follows: For $s \in \mathcal{S}$ and $x \in \mathcal{X}$,

$$\pi_t(x, s) = \mathbb{P}^{\mathbf{f}}(X_t = x, S_t = s | Y^{t-1} = y^{t-1}, A^{t-1} = a^{t-1}).$$

If $Y^{t-1}$ and $A^{t-1}$ are random variables, then the belief state is a $\mathcal{P}_{X,S}$-valued random variable.

The belief state evolves in a state-like manner as follows.

*Lemma 1: For any realization* $y_t$ *of* $Y_t$ *and* $a_t$ *of* $A_t$*,* $\pi_{t+1}$ *is given by*

$$\pi_{t+1} = \varphi(\pi_t, y_t, a_t) \tag{34}$$

*where $\varphi$ is given by*

$$\varphi(\pi, y, a)(x', s') = \frac{\sum_{x \in \mathcal{X}} Q(x'|x) a(y|x, s' - x + y) \pi(x, s' - x + y)}{\sum_{(x,s) \in \mathcal{X} \times \mathcal{S}} a(y|x, s - x + y) \pi(x, s - x + y)}.$$

*Proof:* For ease of notation, we use $\mathbb{P}(x_t, s_t | y^{t-1}, a^{t-1})$ to denote $\mathbb{P}(X_t = x_t, S_t = s_t | Y^{t-1} = y^{t-1}, A^{t-1} = a^{t-1})$. Similar interpretations hold for other expressions as well. Consider

$$\pi_{t+1}(x_{t+1}, s_{t+1}) = \mathbb{P}(x_{t+1}, s_{t+1} | y^t, a^t) = \frac{\mathbb{P}(x_{t+1}, s_{t+1}, y_t, a_t | y^{t-1}, a^{t-1})}{\mathbb{P}(y_t, a_t | y^{t-1}, a^{t-1})} \quad (35)$$

Now, consider the numerator of the right hand side.

$$\begin{aligned}
&\mathbb{P}(x_{t+1}, s_{t+1}, y_t, a_t | y^{t-1}, a^{t-1}) \\
&\quad = \mathbb{P}(x_{t+1}, s_{t+1}, y_t, a_t | y^{t-1}, a^{t-1}, \pi_t) \\
&\quad = \sum_{(x_t, s_t) \in \mathcal{X} \times \mathcal{S}} \mathbb{P}(x_{t+1} | x_t) \mathbb{1}_{s_{t+1}}(s_t + x_t - y_t) \\
&\qquad \times a_t(y_t | x_t, s_t) \mathbb{1}_{a_t}(f_t(y^{t-1}, a^{t-1})) \pi_t(x_t, s_t) \quad (36)
\end{aligned}$$

Substituting (36) in (35) (and observing that the denominator of the right hand side of (35) is the marginal of the numerator over $(x_{t+1}, s_{t+1})$), we get that $\pi_{t+1}$ can be written in terms of $\pi_t$, $y_t$ and $a_t$. Note that if the term $\mathbb{1}_{a_t}(f_t(y^{t-1}, a^{t-1}))$ is 1, it cancels from both the numerator and the denominator; if it is 0, we are conditioning on a null event in (35), so we can assign any valid distribution to the conditional probability. □

Note that an immediate implication of the above result is that $\pi_t$ depends only on $(y^{t-1}, a^{t-1})$ and not on the policy $\mathbf{f}$. This is the main reason that we are working with a policy of the form (33) rather than (29).

*Lemma 2: The cost $\tilde{L}_T(\mathbf{f})$ in* (31) *can be written as*

$$\tilde{L}_T(\mathbf{f}) = \frac{1}{T} \mathbb{E}\left[ \sum_{t=1}^{T} I(A_t; \Pi_t) \right]$$

*where $I(a_t; \pi_t)$ does not depend on the policy $\mathbf{f}$ and is computed according to the standard formula*

$$I(a_t; \pi_t) = \sum_{\substack{x \in \mathcal{X}, s \in \mathcal{S}, \\ y \in \mathcal{Y}}} \pi_t(x, s) a_t(y \mid x, s) \times \log \frac{a_t(y|x, s)}{\sum_{(\tilde{x}, \tilde{s}) \in \mathcal{X} \times \mathcal{S}} \pi_t(\tilde{x}, \tilde{s}) a_t(y \mid \tilde{x}, \tilde{s})}.$$

*Proof:* By the law of iterated expectations, we have

$$\tilde{L}_T(\mathbf{f}) = \frac{1}{T} \mathbb{E}\left[ \sum_{t=1}^{T} \mathbb{E}[c_t(X_t, S_t, A_t, Y^t; \mathbf{f}) | Y^{t-1}, A^{t-1}] \right] \quad (37)$$

Now, from (30), each summand may be written as

$$\begin{aligned}
&\mathbb{E}^{\mathbf{f}}[c_t(X_t, S_t, A_t, Y^t; \mathbf{f}) \mid Y^{t-1} = y^{t-1}, A^t = a^t] \\
&\quad = \sum_{\substack{x \in \mathcal{X}, s \in \mathcal{S}, \\ y \in \mathcal{Y}}} \pi_t(x, s) a_t(y \mid x, s) \\
&\qquad \times \log \frac{a_t(y|x, s)}{\sum_{(\tilde{x}, \tilde{s}) \in \mathcal{X} \times \mathcal{S}} \pi_t(\tilde{x}, \tilde{s}) a_t(y \mid \tilde{x}, \tilde{s})} \\
&\quad = I(a_t; \pi_t).
\end{aligned}$$

Thus, $\mathbb{E}^{\mathbf{f}}[c_t(X_t, S_t, Y^t; \mathbf{f} \mid Y^{t-1}, A^t] = I(A_t, \Pi_t)$. Substituting this back in (37), we get the result of the Lemma. □

*Proof of Theorem 1:* Lemma 1 implies that $\{\Pi_t\}_{t \geq 1}$ is a controlled Markov process with control action $A_t$. In addition, Lemma 2 implies that the objective function can be expressed in terms of the *state* $\Pi_t$ and the action $A_t$. Thus, one can use Markov decision theory [21] to identify the optimal policy. Since both the state space and the action space are continuous valued, we need to verify the standard technical conditions.

Define the stochastic kernel $K \colon \mathcal{P}_{X,S} \times \mathcal{A} \to \mathcal{P}_{X,S}$ as follows. For any Borel subset $B$ of $\mathcal{P}_{X,S}$ and any $\pi \in \mathcal{P}_{X,S}$ and $a \in \mathcal{A}$,

$$K(B \mid \pi, a) = \sum_{\substack{x \in \mathcal{X}, s \in \mathcal{S}, \\ y \in \mathcal{Y}}} \pi(x, s) a(y \mid x, s) \mathbb{1}_B\{\varphi(\pi, y, a))\}.$$

The sequential model of Sec. III-B has the following properties.

1) Based on (32), we can write the controlled dynamics of the state $(X_t, S_t)$ as follows:

$$\mathbb{P}(X_{t+1} = x_+, S_{t+1} = s_+ \mid X_t = x, S_t = s, A_t = a) = \sum_{y \in \mathcal{Y}} Q(x_+|x) a(y) \mathbb{1}_{s_+}\{s + y - x\},$$

which is continuous in $a$.

2) The observations are given by

$$\mathbb{P}(Y_t = y | X_t = x, S_t = s, A_t = a) = a(y)$$

which is continuous in $a$.

3) The observations $Y_t$ are discrete.

Therefore, from [22, Sec 4.4 and Lemma 4.1], we get the following:

4) The stochastic kernel $K(d\pi_+ \mid \pi, a)$ is weakly continuous.

In addition, the model has the following properties:

5) The action set $\mathcal{A}$ is compact.

6) The per-step cost $I(a, \pi)$ is continuous and bounded below. (In fact, the per-step cost is also bounded above).

Properties 4)–6) imply [21, Condition 3.3.3], which by [21, Th. 3.3.5], implies the *measurable selection condition* [21, Assumption 3.3.1]. Under the measurable selection condition, the “inf” in the dynamic program can be replaced by a “min”.

The continuity of the value function in $\pi$ follows from the continuity of the per-step cost $I(a, \pi)$ and the controlled stochastic kernel $K(d\pi_+ \mid \pi, a)$. The concavity of the value functions is proved in Appendix III.

From standard results in Markov decision theorem (e.g., see [21, Th. 3.2.1]), it follows that the policy given in part 2) of the Theorem is optimal for the sequential model of Sec III-B. Proposition 2 implies that this policy is also optimal for Problem A. □

### D. Remarks About Numerical Solution

The dynamic programs of Theorems 1 and 2, both state and action spaces are distribution valued (and, therefore, subsets of Euclidean space). Although, an exact solution of the dynamic program is not possible, there are two approaches to obtain an approximate solution. The first is to treat it as a dynamic program of an MDP with continuous state and action spaces and use approximate dynamic programming [24], [31]. The second is to treat it as a dynamic program for a POMDP and use point-based methods [32]. The point-based methods rely on concavity of the value function, which was established in Theorem 1.

## IV. Proof of Theorem 3

We follow the standard approach and show that the proposed leakage rate is optimal by showing achievability and a converse. As a preliminary step, we first show that under Assumption 1, the objective can be rewritten in a simpler but equivalent form. To show achievability, we show that the proposed optimal policy belongs to a class of policies that satisfies a certain invariance property. Using this property the multi-letter mutual information expression can be reduced into a single-letter expression. For the converse we provide two proofs: the first uses dynamic programming and the second uses purely probabilistic and information theoretic arguments.

### A. Simplification of the Dynamic Program

Define

$$\theta_t(s) = \mathbb{P}^{\mathbf{f}}(S_t = s \mid Y^{t-1} = y^{t-1}, A^{t-1} = a^{t-1}).$$

Then, under Assumption 1, we can simplify the belief state $\pi_t$ as follows:

$$\begin{aligned}\pi_t(x,s) &= \mathbb{P}(X_t = x, S_t = s | Y^{t-1} = y^{t-1}, A^{t-1} = a^{t-1}) \\ &\overset{(a)}{=} \mathbb{P}(X_t = x | S_t = s, Y^{t-1} = y^{t-1}, A^{t-1} = a^{t-1}) \\ &\quad \times \mathbb{P}(S_t = s \mid Y^{t-1} = y^{t-1}, A^{t-1} = a^{t-1}) \\ &\overset{(b)}{=} P_X(x)\theta_t(s)\end{aligned}$$

where $(a)$ follows from the product rule of probability and $(b)$ uses Assumption 1 and the definition of $\theta_t$.

Since $\pi_t(x,s) = P_X(x)\theta_t(s)$, in principle, we can simplify the dynamic program of Theorem 2 by using $\theta_t$ as an information state. However, for reasons that will become apparent, we provide an alternative simplification that uses an information state $\xi_t \in \mathcal{P}_W$.

Recall that $W_t = S_t - X_t$ which takes values in $\mathcal{W} = \{s - x : s \in \mathcal{S}, x \in \mathcal{X}\}$. For any realization $(y^{t-1}, a^{t-1})$ of past observations and actions, define $\xi_t \in \mathcal{P}_W$ as follows: for any $w \in \mathcal{W}$,

$$\xi_t(w) = \mathbb{P}^{\mathbf{f}}(W_t = w \mid Y^{t-1} = y^{t-1}, A^{t-1} = a^{t-1}).$$

If $Y^{t-1}$ and $A^{t-1}$ are random variables, then $\xi_t$ is a $\mathcal{P}_W$-valued random variable. As was the case for $\pi_t$, it can be shown that $\xi_t$ does not depend on the choice of the policy $\mathbf{f}$.

*Lemma 3: Under Assumption 1, $\theta_t$ and $\xi_t$ are related as follows:*

$$\xi_t(w) = \sum_{(x,s)\in\mathcal{D}(w)} P_X(x)\theta_t(s). \tag{38}$$

*Proof:*

$$\begin{aligned}\xi_t(w) &= \mathbb{P}^{\mathbf{f}}(W_t = w \mid Y^{t-1} = y^{t-1}, A^{t-1} = a^{t-1}) \\ &= \mathbb{P}^{\mathbf{f}}(S_t - X_t = w \mid Y^{t-1} = y^{t-1}, A^{t-1} = a^{t-1}) \\ &= \sum_{(x,s)\in\mathcal{D}(w)} P_X(x)\theta_t(s).\end{aligned}$$

□

Since $\pi_t(x,s) = P_X(x)\theta_t(s)$, Lemma 3 shows that $\xi_t$ is a function of $\pi_t$. We will show that we can simplify the dynamic program of Theorem 2 by using $\xi_t$ as the information state instead of $\pi_t$. For such a simplification to work, we would have to use charging policies of the form $q_t(y_t|w_t, y^{t-1})$. We establish that restricting attention to such policies is without loss of optimality. For that matter, define $\mathcal{B}$ as follows:

$$\mathcal{B} = \left\{b \in \mathcal{P}_{Y|W} : b(\mathcal{Y}_\circ(w) \mid w) = 1,\ \forall w \in \mathcal{W}\right\}. \tag{39}$$

*Lemma 4: Given $a \in \mathcal{A}$ and $\pi \in \mathcal{P}_{X,S}$, define the following:*

- *$\xi \in \mathcal{P}_W$ as $\xi(w) = \sum_{(x,s)\in\mathcal{D}(w)} \pi(x,s)$*
- *$b \in \mathcal{B}$ as follows: for all $y \in \mathcal{Y}, w \in \mathcal{W}$*

$$b(y \mid w) = \frac{\sum_{(x,s)\in\mathcal{D}(w)} a(y \mid x,s)\pi(x,s)}{\xi(w)};$$

- *$\tilde{a} \in \mathcal{A}$ as follows: for all $y \in \mathcal{Y}, x \in \mathcal{X}, s \in \mathcal{S}$*

$$\tilde{a}(y|x,s) = b(y|s-x).$$

*Then under Assumption 1, we have*

1) *Invariant Transitions: for any $y \in \mathcal{Y}$, $\varphi(\pi,y,a) = \varphi(\pi,y,\tilde{a})$.*
2) *Lower Cost: $I(a;\pi) \geq I(\tilde{a};\pi) = I(b;\xi)$.*

*Therefore, in the sequential problem of Sec. III-B, there is no loss of optimality in restricting attention to actions $b \in \mathcal{B}$.*

*Proof:*

1) Suppose $(X,S) \sim \pi$ and $W = S - X$, $S_+ = W + Y$, $X_+ \sim P_X$. We will compare $\mathbb{P}(S_+|Y)$ when $Y \sim a(\cdot|X,S)$ with when $Y \sim \tilde{a}(\cdot|X,S)$. Given $w \in \mathcal{W}$ and $y \in \mathcal{Y}$,

$$\begin{aligned}\mathbb{P}^a(W = w, Y = y) &= \sum_{(x,s)\in\mathcal{D}(w)} a(y|x,s)\pi(x,s) \\ &= \sum_{(x,s)\in\mathcal{D}(w)} b(y|w)\pi(x,s) \\ &\overset{\text{(a)}}{=} \sum_{(x,s)\in\mathcal{D}(w)} \tilde{a}(y|x,s)\pi(x,s) \\ &= \mathbb{P}^{\tilde{a}}(W = w, Y = y)\end{aligned} \tag{40}$$

where (a) uses that for all $(x,s) \in \mathcal{D}(w)$, $s - x = w$. Marginalizing (40) over $W$, we get that $\mathbb{P}^a(Y = y) = \mathbb{P}^{\tilde{a}}(Y = y)$. Since $S_+ = W + Y$, Eq. (40) also implies

$\mathbb{P}^a(S_+ = s, Y = y) = \mathbb{P}^{\tilde{a}}(S_+ = s, Y = y)$. Therefore, $\mathbb{P}^a(S_+ = s | Y = y) = \mathbb{P}^{\tilde{a}}(S_+ = s | Y = y)$.

2) Let $(X, S) \sim \pi$ and $W = S - X$. Then $W \sim \xi$. Therefore, we have

$$I(a; \pi) = I^a(X, S; Y) \geq I^a(W; Y).$$

where the last inequality is the data-processing inequality. Under $\tilde{a}$, $(X, S) - W - Y$, therefore,

$$I(\tilde{a}; \pi) = I^{\tilde{a}}(X, S; Y) = I^{\tilde{a}}(W; Y).$$

Now, by construction, the joint distribution of $(W, Y)$ is the same under $a$, $\tilde{a}$, and $b$. Thus,

$$I^a(W; Y) = I^{\tilde{a}}(W; Y) = I^b(W; Y).$$

Note that $I^b(W; Y)$ can also be written as $I(b; \xi)$. The result follows by combining all the above relations. □

Once attention is restricted to actions $b \in \mathcal{B}$, the update of $\xi_t$ may be expressed in terms of $b \in \mathcal{B}$ as follows:

*Lemma 5: For any realization $y_t$ of $Y_t$ and $b_t$ of $B_t$, $\xi_{t+1}$ is given by*

$$\xi_{t+1} = \tilde{\varphi}(\xi_t, y_t, b_t) \tag{41}$$

*where $\tilde{\varphi}$ is given by*

$$\tilde{\varphi}(\xi, y, b)(w_+) = \frac{\sum_{x \in \mathcal{X}, w \in \mathcal{W}} P_X(x) \mathbb{1}_{w_+}\{y + w - x\} b(y \mid w) \xi(w)}{\sum_{w \in \mathcal{W}} b(y \mid w) \xi(w)}.$$

*Proof:* The proof is similar to the proof of Lemma 1. □

For any $b \in \mathcal{B}$ and $\xi \in \mathcal{P}_W$, let us define the Bellman operator $\tilde{\mathscr{B}}_b : [\mathcal{P}_W \to \mathbb{R}] \to [\mathcal{P}_W \to \mathbb{R}]$ as follows: for any $\tilde{V} : \mathcal{P}_W \to \mathbb{R}$ and any $\xi \in \mathcal{P}_W$,

$$[\tilde{\mathscr{B}}_b \tilde{V}](\xi) = I(b; \xi) + \sum_{y \in \mathcal{Y}, w \in \mathcal{W}} \xi(w) b(y \mid w) \tilde{V}\big(\tilde{\varphi}(\xi, y, b)\big).$$

*Theorem 4: Under Assumption 1, there is no loss of optimality in restricting attention to optimal policies of the form $q_t(y_t | w_t, \xi_t)$ in Problem A.*

1) *For the finite horizon case, we can identify the optimal policy $\mathbf{q}^* = (q_1^*, \dots, q_T^*)$ by iteratively defining* value functions $\tilde{V}_t : \mathcal{P}_W \to \mathbb{R}$. *For any $\xi \in \mathcal{P}_W$, $\tilde{V}_{T+1}(\xi) = 0$, and for $t = T, T-1, \dots, 1$,*

$$\tilde{V}_t(\xi) = \min_{b \in \mathcal{B}} [\tilde{\mathscr{B}}_b \tilde{V}_{t+1}](\xi). \tag{42}$$

*Let $f_t^\circ(\xi)$ denote the arg min of the right hand side of* (42)*. Then, the optimal policy $\mathbf{q}^* = (q_1^*, \dots, q_T^*)$ is given by*

$$q_t^*(y_t | w_t, \xi_t) = b_t(y_t | w_t), \text{ where } b_t = f_t^\circ(\xi_t).$$

*Moreover, the optimal (finite horizon) leakage rate is given by $\tilde{V}_1(\xi_1)/T$, where $\xi_1(w) = \sum_{(x,s) \in \mathcal{D}(w)} P_X(x) P_{S_1}(s)$.*

2) *For the infinite horizon, suppose that there exists a constant $\tilde{J} \in \mathbb{R}$ and a function $\tilde{v} : \mathcal{P}_S \to \mathbb{R}$ which satisfies the following fixed point equation:*

$$\tilde{J} + \tilde{v}(\xi) = \min_{b \in \mathcal{B}} [\tilde{\mathscr{B}}_b \tilde{v}](\xi), \ \forall \xi \in \mathcal{P}_W. \tag{43}$$

*Let $\mathbf{f}^\circ(\xi)$ denote the arg min of the right hand side of* (43)*. Then, the time-homogeneous policy $\mathbf{q}^* = (q^*, q^*, \dots)$ given by*

$$q^*(y_t | w_t, \xi_t) = b_t(y_t | w_t), \text{ where } b_t = f^\circ(\xi_t)$$

*is optimal. Moreover, the optimal (infinite horizon) leakage rate is given by $\tilde{J}$.* □

*Proof:* Lemma 5 implies that $\{\xi_t\}_{t \geq 1}$ is a controlled Markov process with control action $b_t$. Lemma 4, part 2), implies that the per-step cost can be written as

$$\frac{1}{T} \mathbb{E}\bigg[\sum_{t=1}^{T} I(b_t; \xi_t)\bigg].$$

Thus, by standard results in Markov decision theory [21], the optimal solution is given by the dynamic program described above. □

### *B. Weak Achievability*

To simplify the analysis, we assume that we are free to choose the initial distribution of the state of the battery, which could be done by, for example, initially charging the battery to a random value according to that distribution. In principle, such an assumption could lead to a lower achievable leakage rate. For this reason, we call it *weak* achievability. In the next section, we will show achievability starting from an arbitrary initial distribution, which we call *strong* achievability.

*Definition 1: [Constant-distribution policy] A time-homogeneous policy $\mathbf{f}^\circ = (f^\circ, f^\circ, \dots)$ is called a constant-distribution policy if for all $\xi \in \mathcal{P}_W$, $f^\circ(\xi)$ is a constant. If $f^\circ(\xi) = b^\circ$, then with a slight abuse of notation, we refer to $\mathbf{b}^\circ = (b^\circ, b^\circ, \dots)$ as a constant-distribution policy.*

Recall that under a constant-distribution policy $b \in \mathcal{B}$, for any realization $y^t$ of $Y^t$, $\theta_t$ and $\xi_t$ are given as follows:

$$\begin{aligned} \theta_t(s) &= \mathbb{P}(S_t = s \mid Y^{t-1} = y^{t-1}, B^{t-1} = b^{t-1}) \\ \xi_t(w) &= \mathbb{P}(W_t = w \mid Y^{t-1} = y^{t-1}, B^{t-1} = b^{t-1}). \end{aligned}$$

*1) Invariant Policies:* We next impose an invariance property on the class of policies. Under this restriction the leakage rate expression will simplify substantially. Subsequently we will show that the optimal policy belongs to this restricted class.

*Definition 2: [Invariance Property] For a given distribution $\theta_1$ of the initial battery state, a constant-distribution policy $b \in \mathcal{B}$ is called an* invariant *policy if for all $t$, $\theta_t = \theta_1$ and $\xi_t = \xi_1$, where $\xi_1$ and $\theta_1$ satisfy* (38).

*Remark 3:* An immediate implication of the above definition is that under any invariant policy $\mathbf{b}$, the conditional distribution $\mathbb{P}^{\mathbf{b}}(X_t, S_t, Y_t | Y^{t-1})$ is the same as the joint distribution $\mathbb{P}^{\mathbf{b}}(X_1, S_1, Y_1)$. Marginalizing over $(X, S)$ we get that $\{Y_t\}_{t \geq 1}$ is an i.i.d. sequence.

*Lemma 6: If the system starts with an initial distribution $\theta$ of the battery state, and $\xi$ and $\theta$ satisfy* (38)*, then an invariant policy $\mathbf{b} = (b, b, \dots)$ corresponding to $(\theta, \xi)$ achieves a leakage rate*

$$L_T(\mathbf{b}) = I(W_1; Y_1) = I(b; \xi)$$

*for any horizon $T$.*

*Proof:* The proof is a simple corollary of the invariance property in Definition 2. Recall from the dynamic program of Theorem 4, that the performance of any policy $\mathbf{b} = (b_1, b_2, \dots)$ such that $b_t \in \mathcal{B}$, is given by

$$L_T(\mathbf{b}) = \frac{1}{T}\mathbb{E}\bigg[\sum_{t=1}^{T} I(b_t; \xi_t)\bigg].$$

Now, we start with an initial distribution $\xi_1 = \xi$ and follow the constant-distribution structured policy $\mathbf{b} = (b, b, \dots)$. Therefore, by Definition 2, $\xi_t = \xi$ for all $t$. Hence, the leakage rate under policy $\mathbf{b}$ is

$$L_T(\mathbf{b}) = I(b; \xi). \qquad \square$$

*Remark 4:* Note that Lemma 6 can be easily derived independently of Theorem 4. From Proposition 1, we have that:

$$L_T(\mathbf{b}) = \frac{1}{T}\sum_{t=1}^{T} I^{\mathbf{b}}(S_t, X_t; Y_t | Y^{t-1}). \tag{44}$$

From Remark 3, we have that $I^{\mathbf{b}}(S_t, X_t; Y_t|Y^{t-1}) = I^b(S_1, X_1; Y_1) = I^b(W_1; Y_1)$, which immediately results in Lemma 6.

For invariant policies we can express the leakage rate in the following fashion, which is useful in the proof of optimality.

*Lemma 7: For any invariant policy b,*

$$I^b(W_1; Y_1) = I^b(W_1; X_1).$$

*Proof:* Consider the following sequence of simplifications:

$$\begin{aligned} I^b(W_1; Y_1) &= H^b(W_1) - H^b(W_1|Y_1) \\ &= H^b(W_1) - H^b(W_1 + Y_1|Y_1) \\ &\overset{\text{(a)}}{=} H^b(W_1) - H^b(S_2|Y_1) \\ &\overset{\text{(b)}}{=} H^b(W_1) - H^b(S_1) \\ &\overset{\text{(c)}}{=} H^b(W_1) - H^b(S_1|X_1) \\ &\overset{\text{(d)}}{=} H^b(W_1) - H^b(W_1|X_1) \\ &= I^b(W_1; X_1). \end{aligned}$$

where (a) is due to the battery update equation (1); (b) is because $b$ is an invariant; (c) is because $S_1$ and $X_1$ are independent; and (d) is because $S_1 = W_1 + X_1$. □

*2) Structured Policy:* We now introduce a class of policies that satisfy the invariance property in Def. 2. This will be then used in the proof of Theorem 3.

*Definition 3: [Structured Policy] Given $\theta \in \mathcal{P}_S$ and $\xi \in \mathcal{P}_W$, a constant-distribution policy $\mathbf{b} = (b, b, \dots)$ is called a* structured policy *with respect to $(\theta, \xi)$ if:*

$$b(y|w) = \begin{cases} P_X(y)\frac{\theta(y+w)}{\xi(w)}, & y \in \mathcal{X} \cap \mathcal{Y}_\circ(w) \\ 0, & \textit{otherwise}. \end{cases}$$

Note that it is easy to verify that the distribution $b$ defined above is a valid conditional probability distribution.

*Lemma 8: For any $\theta \in \mathcal{P}_S$ and $\xi \in \mathcal{P}_W$, the structured policy $\mathbf{b} = (b, b, \dots)$ given in Def. 3 is an invariant policy.*

*Proof:* Since $(\theta_t, \xi_t)$ are related according to Lemma 3, in order to check whether a policy is invariant it is sufficient to check that $\theta_t = \theta_1$ for all $t$. Furthermore, to check if a *time-homogeneous* policy is an invariant policy, it is sufficient to check that either $\theta_2 = \theta_1$.

Let the initial distributions $(\theta_1, \xi_1) = (\theta, \xi)$ and the system variables be defined as usual. Now consider a realization $s_2$ of $S_2$ and $y_1$ of $Y_1$. This means that $w_1 = s_2 - y_1$. Since $Y_1$ is chosen according to $b(\cdot|w_1)$, it must be that $y_1 \in \mathcal{X} \cap \mathcal{Y}_\circ(w_1)$. Therefore,

$$\begin{aligned} \mathbb{P}^{\mathbf{b}}(S_2 = s_2, Y_1 = y_1) &= \mathbb{P}^{\mathbf{b}}(S_2 = s_2, Y_1 = y_1, W_1 = s_2 - y_1) \\ &= \xi_1(s_2 - y_1) b(y_1 | s_2 - y_1) \\ &= P_X(y_1)\theta_1(s_2), \end{aligned} \tag{45}$$

where in the last equality we use the fact that $y_1 \in \mathcal{X} \cap \mathcal{Y}_\circ(s_2 - y_1)$. Note that if $y_1 \notin \mathcal{X} \cap \mathcal{Y}_\circ(s_2 - y_1)$, then $\mathbb{P}^{\mathbf{b}}(S_2 = s_2, Y_1 = y_1) = 0$. Marginalizing over $s_2$, we get $\mathbb{P}^{\mathbf{b}}(Y_1 = y_1) = P_X(y_1)$.

Consequently, $\theta_2(s_2) = \mathbb{P}^{\mathbf{b}}(S_2 = s_2 | Y_1 = y_1) = \theta_1(s_2)$. Hence, $\mathbf{b}$ is invariant as required. □

*Remark 5:* As argued in Remark 3, under any invariant policy, $\{Y_t\}_{t\geq 1}$ is an i.i.d. sequence. As argued in the proof of Lemma 8, for a structured policy the marginal distribution of $Y_t$ is $P_X$. Thus, an eavesdropper cannot statistically distinguish between $\{X_t\}_{t\geq 1}$ and $\{Y_t\}_{t\geq 1}$.

*Proposition 3: Let $\theta^*$, $\xi^*$, and $b^*$ be as defined in Theorem 3. Then,*

1) *$(\theta^*, \xi^*)$ satisfy (38);*
2) *$b^*$ is a structured policy with respect to $(\theta^*, \xi^*)$.*
3) *If the system starts in the initial battery state $\theta^*$ and follows the constant-distribution policy $\mathbf{b}^* = (b^*, b^*, \dots)$, the leakage rate is given by $J^*$.*

*Thus, the performance $J^*$ is achievable.*

*Proof:* The proofs of parts 1) and 2) follows from the definitions. The proof of part 3) follows from Lemmas 6 and 7. □

This completes the proof of the achievability of Theorem 3.

### *C. Strong Achievability*

*Lemma 9: Assume that for any $x \in \mathcal{X}$, $P_X(x) > 0$. Let $(\theta^\circ, \xi^\circ)$ be a pair satisfying* (38) *and $\mathbf{b}^\circ = (b^\circ, b^\circ, \dots)$ be the corresponding structured policy.*

*Assume that $\theta^\circ \in int(\mathcal{P}_S)$ or equivalently, for any $w \in \mathcal{W}$ and $y \in \mathcal{X} \cap \mathcal{Y}_\circ(w)$, $b^\circ(y|w) > 0$. Suppose the system starts in the initial state $(\theta_1, \xi_1)$ and follows policy $\mathbf{b}^\circ$. Then:*

1) *the process $\{\theta_t\}_{\geq 1}$ converges weakly to $\theta^\circ$;*
2) *the process $\{\xi_t\}_{\geq 1}$ converges weakly to $\xi^\circ$;*
3) *for any continuous function $c \colon \mathcal{P}_W \to \mathbb{R}$,*

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E[c(\xi_t)] = c(\xi^\circ). \tag{46}$$

4) *Consequently, the infinite horizon leakage rate under $\mathbf{b}^\circ$ is*

$$L_\infty(\mathbf{b}^\circ) = I(b^\circ, \xi^\circ).$$

*Proof:* The proof of parts 1) and 2) is presented in Appendix IV. From 2), $\lim_{t\to\infty} \mathbb{E}[c(\xi_t)] = c(\xi^\circ)$, which implies (46). Part 4) follows from part 3) by setting $c(\xi_t) = I(b^\circ, \xi_t)$. □

Proposition 1 implies that $\theta^*$ defined in Theorem 3 lies in $int(\mathcal{P}_S)$. Then, by Lemma 9, the constant-distribution policy $\mathbf{b}^* = (b^*, b^*, \dots)$ (where $b^*$ is given by Theorem 3), achieves the leakage rate $I(b^*, \xi^*)$. By Lemma 7, $I(b^*, \xi^*)$ is same as $J^*$ defined in Theorem 3. Thus, $J^*$ is achievable starting from any initial state $(\theta_1, \xi_1)$.

### D. Dynamic Programming Converse

We provide two converses. One is based on the dynamic program of Theorem 4, which is presented in this section; the other is based purely on information theoretic arguments, which is presented in the next section.

In the dynamic programming converse, we show that for $J^*$ given in Theorem 3, $v^*(\xi) = H(\xi)$, and any $b \in \mathcal{B}$,

$$J^* + v^*(\xi) \le [\tilde{\mathscr{B}}_b v^*](\xi), \quad \forall \xi \in \mathcal{P}_W, \tag{47}$$

Since $H(\xi)$ is bounded, from [21, Lemma 5.2.5(b)], we get that $J^*$ is a lower bound of the optimal leakage rate.

To prove (47), pick any $\xi \in \mathcal{P}_W$ and $b \in \mathcal{B}$. Suppose $W_1 \sim \xi$, $Y_1 \sim b(\cdot|W_1)$, $S_2 = Y_1 + W_1$, $X_2$ is independent of $W_1$ and $X_2 \sim P_X$ and $W_2 = S_2 - X_2$. Then,

$$\begin{aligned}[\tilde{\mathscr{B}}_b v^*](\xi) &= I(b;\xi) + \sum_{(w_1,y_1)\in\mathcal{W}\times\mathcal{Y}} \xi(w_1)b(y_1|w_1)v^*(\hat{\varphi}(\xi, y_1, b)) \\ &= I(W_1; Y_1) + H(W_2|Y_1) \end{aligned}\tag{48}$$

where the second equality is due to the definition of conditional entropy. Consequently,

$$\begin{aligned}[\tilde{\mathscr{B}}_b v^*](\xi) - v^*(\xi) &= H(W_2|Y_1) - H(W_1|Y_1) \\ &= H(W_2|Y_1) - H(W_1 + Y_1|Y_1) \\ &\overset{(a)}{=} H(S_2 - X_2|Y_1) - H(S_2|Y_1) \\ &\overset{(b)}{\ge} \min_{\theta_2 \in \mathcal{P}_S} \big[H(\tilde{S}_2 - X_2) - H(\tilde{S}_2)\big], \qquad \tilde{S}_2 \sim \theta_2 \\ &= J^* \end{aligned}\tag{49}$$

where (a) uses $S_2 = Y_1 + W_1$ and $W_2 = S_2 - X_2$; (b) uses the fact that $H(A_1|B) - H(A_1 - A_2|B) \ge \min_{P_{A_1}} \big[H(A_1) - H(A_1 - A_2)\big]$ for any joint distribution on $(A_1, A_2, B)$.

The equality in (49) occurs when $b$ is an invariant policy and $\theta_2$ is same as $\theta^*$ defined in Theorem 3. For $\xi$ that are not equivalent to $\theta^*$ via (38), the inequality in (49) is strict.

We have shown that Eq. (47) is true. Consequently, $J^*$ is a lower bound on the optimal leakage rate $\tilde{J}$.

### E. Information Theoretic Converse

Consider the following inequalities: for any admissible policy $\mathbf{q} \in \mathcal{Q}_B$, we have

$$\begin{aligned} I(S_1, X^T; Y^T) &= \sum_{t=1}^{T} I(S_t, X_t; Y_t|Y^{t-1}) \\ &\overset{(a)}{\ge} \sum_{t=1}^{T} I(W_t; Y_t|Y^{t-1}) \end{aligned}\tag{50}$$

where (a) follows from the fact that $W_t = X_t - S_t$ is a deterministic function of $(X_t, S_t)$ and that the mutual information is non-negative.

Now consider

$$\begin{aligned} I(W_t; Y_t|Y^{t-1}) &= H(W_t|Y^{t-1}) - H(W_t|Y^t) \\ &= H(W_t|Y^{t-1}) - H(W_t + Y_t|Y^t) \\ &\overset{(b)}{=} H(W_t|Y^{t-1}) - H(S_{t+1}|Y^t) \\ &\overset{(c)}{=} H(W_t|Y^{t-1}) - H(S_{t+1}|Y^t, X_{t+1}) \\ &= H(W_t|Y^{t-1}) - H(S_{t+1} - X_{t+1}|Y^t, X_{t+1}) \\ &\overset{(d)}{=} H(W_t|Y^{t-1}) - H(W_{t+1}|Y^t, X_{t+1}) \end{aligned}\tag{51}$$

where (b) follows from (1); (c) follows because of assumption (A); and (d) also follows from (1).

Substituting (51) in (50) (but expanding the last term as $H(W_T|Y^{T-1}) - H(W_T|Y^T)$, we get

$$I(S_1, X^T; Y^T) \ge \sum_{t=1}^{T} I(W_t; Y_t|Y^{t-1}) \tag{52}$$

$$= \sum_{t=1}^{T-1} \big[H(W_t|Y^{t-1}) - H(W_{t+1}|Y^t, X_{t+1})\big] + H(W_T|Y^{T-1}) - H(W_T|Y^T) \tag{53}$$

$$= H(W_1) + \sum_{t=2}^{T} \big[-H(W_t|Y^{t-1}, X_t) + H(W_t|Y^{t-1})\big] - H(W_T|Y^T) \tag{54}$$

$$= H(W_1) + \sum_{t=2}^{T} I(W_t; X_t|Y^{t-1}) - H(W_T|Y^T). \tag{55}$$

Now, we take the limit $T \to \infty$ to obtain a lower bound to the leakage rate:

$$\begin{aligned} L_\infty(\mathbf{q}) &= \limsup_{T\to\infty} \frac{1}{T} I(S_1, X^T; Y^T) \\ &\ge \limsup_{T\to\infty} \frac{1}{T}\Big[H(W_1) + \sum_{t=2}^{T} I(W_t; X_t|Y^{t-1}) - H(W_T|Y^T)\Big] \\ &\overset{(a)}{=} \limsup_{T\to\infty} \frac{1}{T}\Big[\sum_{t=2}^{T} I(W_t; X_t|Y^{t-1})\Big] \\ &\overset{(b)}{\ge} \min_{P_S \in \mathcal{P}_S} I(S - X; X) = J^* \end{aligned}$$

where (a) is because the entropy of any discrete random variable is bounded and (b) follows from the fact that each term in the summation satisfies:

$$I(W_t; X_t|Y^{t-1}) \ge \min_{P_S \in \mathcal{P}_S} I(S - X; X), \tag{56}$$

which can be justified as follows. First note that for any policy $\mathbf{q}$ we have that:

$$I(W_t; X_t|Y^{t-1}) = I(S_t - X_t; X_t|Y^{t-1})$$

depends on the joint distribution $P^{\mathbf{q}}_{S_t,X_t,Y^{t-1}}(s_t, x_t, y^{t-1})$ which factors as:

$$P^{\mathbf{q}}_{S_t,X_t,Y^{t-1}}(s_t, x_t, y^{t-1}) = P_X(x_t)P^{\mathbf{q}}_{S_t,Y^{t-1}}(s_t, y^{t-1}) \tag{57}$$

as $X_t$ is sampled i.i.d. from the distribution $P_X(\cdot)$ and from the state update equation (1), we have that $S_t$ is a function of

$(X^{t-1}, Y^{t-1})$, and thus independent of $X_t$. Now note that:

$$\begin{aligned}
&I(W_t; X_t|Y^{t-1})\\
&\quad = \sum_{y^{t-1}\in\mathcal{Y}^{t-1}} I(S_t - X_t; X_t|Y^{t-1} = y^{t-1})p(y^{t-1})\\
&\quad \geq \min_{y^{t-1}\in\mathcal{Y}^{t-1}} I(S_t - X_t; X_t|Y^{t-1} = y^{t-1})\\
&\quad \geq \min_{y^{t-1}\in\mathcal{Y}^{t-1}} \min_{p_{S_t,X_t|Y^{t-1}}(\cdot|y^{t-1})} I(S_t - X_t; X_t|Y^{t-1} = y^{t-1})\\
&\quad = \min_{y^{t-1}\in\mathcal{Y}^{t-1}, p_{S_t|Y^{t-1}}(\cdot|y^{t-1})=} I(S_t - X_t; X_t|Y^{t-1} = y^{t-1})
\end{aligned} \tag{58}$$

where (58) follows from (57) since $X_t$ is independent of $(S_t, Y^{t-1})$ and distributed according to $P_X(\cdot)$ Since (58) is simply equivalent to minimizing over the distribution $P_S(\cdot)$, the relation in (56) holds. This completes the proof. This shows that $J^*$ is a lower bound to the minimum (infinite horizon) leakage rate.

## V. Conclusion and Discussion

In this paper, we study a smart metering system that uses a rechargeable battery to partially obscure the user's power demand. Through a series of reductions, we show that the problem of finding the best battery charging policy can be recast as a Markov decision process. Consequently, the optimal charging policies and the minimum information leakage rate are given by the solution of an appropriate dynamic program.

For the case of i.i.d. demand, we provide an explicit characterization of the optimal battery policy and the leakage rate. In this special case it suffices to choose a memoryless policy where the distribution of $Y_t$ depends only on $W_t$. Our achievability results rely on restricting attention to a class of invariant policies. Under an invariant policy, the consumption $\{Y_t\}_{t\geq 1}$ is i.i.d. and the leakage rate is characterized by a single-letter mutual information expression. We then further restrict attention to what we call structured policies under which the marginal distribution of $\{Y_t\}_{t\geq 1}$ is $P_X$. Thus, under the structured policies, an eavesdropper cannot statistically distinguish between $\{X_t\}_{t\geq 1}$ and $\{Y_t\}_{t\geq 1}$. We provide two converses; one is based on the dynamic programming argument while the other is based on a purely information theoretic argument.

Extending of our MDP formulation to incorporate an additive cost, such as the price of consumption, is rather immediate. However, the approach presented in this work for explicitly characterizing the optimal leakage rate in the i.i.d. case may not immediately extend to such general cost functions. In another direction one can allow for a certain controlled wastage of energy drawn from the grid to increase privacy. It would be interesting to see how the leakage rate decreases with the wasted energy. The study of such problems, as well as finer implementation details of the proposed system, remains an interesting future direction.

## Appendix I
## Proof of Property 1

For any $\theta \in int(\mathcal{P}_S)$ and $\delta : \mathcal{S} \to \mathbb{R}$ such that $\sum_{s\in\mathcal{S}} \delta(s) = 0$. Let $\theta_\alpha(s) := \theta(s) + \alpha\delta(s)$. Then for small enough $\alpha$, $\theta_\alpha \in \mathcal{P}_S$. Given such a $\theta_\alpha$, let $\mathbb{P}_{W,X}(w, x) = \mathbb{P}_{W|X}(w|x)P_X(x) = \theta_\alpha(w + x)P_X(x)$. Then to show that $I(W; X)$ is strictly convex on $\mathcal{P}_S$ we require $\frac{d^2\ I(W;X)}{d\alpha^2} > 0$. Due to independence of $X$ and $S$, $I(W; X) = H(W) - H(S)$. Therefore,

$$\begin{aligned}
\frac{dI(W; X)}{d\alpha} &= \frac{d\,[-H(S) + H(W)]}{d\alpha}\\
&= \sum_{\tilde{s}} \delta(\tilde{s}) \ln \theta_\alpha(\tilde{s}) - \sum_{w\in\mathcal{W}, s\in\mathcal{S}} P_X(s - w)\delta(s) \ln P_W(w)
\end{aligned}$$

$$\frac{d^2\ I(W; X)}{d\alpha^2} = \sum_s \frac{\delta(s)^2}{\theta_\alpha(s)} - \sum_{w\in\mathcal{W}} \frac{\left(\sum_{\tilde{s}\in\mathcal{S}} P_X(\tilde{s} - w)\delta(\tilde{s})\right)^2}{P_W(w)}.$$

Let $a_w(s) = \delta(s)\sqrt{\frac{P_X(s-w)}{\theta_\alpha(s)}}$ and $b_w(s) = \sqrt{\theta_{\alpha(s)} P_X(s - w)}$. Using the Cauchy-Schwarz inequality, we can show that

$$\begin{aligned}
\frac{d^2\ I(W; X)}{d\alpha^2} &= \sum_s \frac{\delta(s)^2}{\theta_\alpha(s)} - \sum_{w\in\mathcal{W}} \frac{\left(\sum_{\tilde{s}\in\mathcal{S}} a_w(\tilde{s})b_w(\tilde{s})\right)^2}{P_W(w)}\\
&> \sum_s \frac{\delta(s)^2}{\theta_\alpha(s)} - \sum_{w\in\mathcal{W}} \frac{\left(\sum_{\tilde{s}\in\mathcal{S}} a_w(\tilde{s})^2\right)\left(\sum_{\hat{s}\in\mathcal{S}} b_w(\hat{s})^2\right)}{P_W(w)}\\
&= \sum_s \frac{\delta(s)^2}{\theta_\alpha(s)} - \sum_{w\in\mathcal{W}} \left(\sum_{\tilde{s}\in\mathcal{S}} a_w(\tilde{s})^2\right) = 0.
\end{aligned}$$

The strict inequality is because $a$ and $b$ cannot be linearly dependent. To see this, observe that $\frac{a(s)}{b(s)} = \frac{\delta(s)}{\theta(s)+\alpha\delta(s)}$ cannot be equal to a constant for all $s \in \mathcal{S}$ since $\delta$ must contain negative as well as positive elements.

## Appendix II
## Proof of Proposition 1

The proof of Proposition 1 relies on the following intermediate results (which are proved later):

*Lemma 2.1:* For any $\mathbf{q} \in Q_A$,

$$I^{\mathbf{q}}(S_1, X^T; Y^T) \geq \sum_{t=1}^{T} I^{\mathbf{q}}(X_t, S_t; Y_t|Y^{t-1})$$

with equality if and only if $q \in Q_B$.

*Lemma 2.2:* For any $\mathbf{q}_a \in Q_A$, there exists a $\mathbf{q}_b \in Q_B$, such that

$$\sum_{t=1}^{T} I^{\mathbf{q}_a}(X_t, S_t; Y_t|Y^{t-1}) = \sum_{t=1}^{T} I^{\mathbf{q}_b}(X_t, S_t; Y_t|Y^{t-1}).$$

Combining Lemmas 2.1 and 2.2, we get that for any $\mathbf{q}_a \in \mathcal{Q}_A$, there exists a $\mathbf{q}_b \in \mathcal{Q}_B$ such that

$$I^{\mathbf{q}_a}(S_1, X^T; Y^T) \geq I^{\mathbf{q}_b}(S_1, X^T; Y^T).$$

Therefore there is no loss of optimality in restricting attention to charging policies in $\mathcal{Q}_B$. Furthermore, Lemma 2.1 shows that for any $q \in \mathcal{Q}_B$, $L_T(q)$ takes the additive form as given in the statement of the proposition.

*Proof of Lemma 2.1:* For any $\mathbf{q} \in \mathcal{Q}_A$, we have

$$\begin{aligned} I^{\mathbf{q}}(S_1, X^T, Y^T) &\overset{(a)}{=} \sum_{t=1}^{T} I^{\mathbf{q}}(S_1, X^t; Y_t|Y^{t-1}) \\ &\overset{(b)}{=} \sum_{t=1}^{T} I^{\mathbf{q}}(X^t, S^t; Y_t|Y^{t-1}) \\ &\overset{(c)}{\geq} \sum_{t=1}^{T} I^{\mathbf{q}}(X_t, S_t; Y_t|Y^{t-1}) \end{aligned}$$

where (a) uses the chain rule of mutual information and the fact that $(S^{t-2}, Y^{t-1}) \to X_{t-1} \to X_t$[5]; (b) uses the fact that the battery process $S^t$ is a deterministic function of $S_1$, $X^{t-1}$, and $Y^{t-1}$ given by (1); and (c) uses the fact that removing terms does not reduce the mutual information. □

*Proof:* [Proof of Lemma 2.2] For any $\mathbf{q}_a = (q_1^a, q_2^a, \dots, q_T^a) \in \mathcal{Q}_A$, construct a $\mathbf{q}_b = (q_1^b, q_2^b, \dots, q_T^b) \in \mathcal{Q}_B$ as follows: for any $t$ and realization $(x^t, s^t, y^t)$ of $(X^t, S^t, Y^t)$ let

$$q_t^b(y_t|x_t, s_t, y^{t-1}) = \mathbb{P}^{\mathbf{q}_a}_{Y_t|X_t,S_t,Y^{t-1}}(y_t|x_t, s_t, y^{t-1}). \quad (59)$$

To prove the Lemma, we show that for any $t$,

$$\mathbb{P}^{\mathbf{q}_a}_{X_t,S_t,Y^t} = \mathbb{P}^{\mathbf{q}_b}_{X_t,S_t,Y^t}. \quad (60)$$

By definition of $\mathbf{q}_b$ given by (59), to prove (60), it is sufficient to show that

$$\mathbb{P}^{\mathbf{q}_a}_{X_t,S_t,Y^{t-1}} = \mathbb{P}^{\mathbf{q}_b}_{X_t,S_t,Y^{t-1}}. \quad (61)$$

We do so using induction.

For $t = 1$, $\mathbb{P}^{\mathbf{q}_a}_{X_1,S_1}(x, s) = P_{X_1}(x)P_{S_1}(s) = \mathbb{P}^{\mathbf{q}_b}_{X_1,S_1}(x, s)$. This forms the basis of induction. Now assume that (61) hold for $t$.

In the rest of the proof, for ease of notation, we denote $\mathbb{P}^{\mathbf{q}_a}_{X_{t+1},S_{t+1},Y^t}(x_{t+1}, s_{t+1}, y^t)$ simply by $\mathbb{P}^{\mathbf{q}_a}(x_{t+1}, s_{t+1}, y^t)$. For $t + 1$, we have

$$\begin{aligned} \mathbb{P}^{\mathbf{q}_a}(x_{t+1}, s_{t+1}, y^t) &= \sum_{(x_t,s_t)\in\mathcal{X}\times\mathcal{S}} \mathbb{P}^{\mathbf{q}_a}(x_{t+1}, x_t, s_{t+1}, s_t, y^t) \\ &= \sum_{(x_t,s_t)\in\mathcal{X}\times\mathcal{S}} Q(x_{t+1}|x_t)\mathbb{1}_{s_{t+1}}\{s_t - x_t + y_t\}q_a(y_t|x_t, s_t, y^{t-1}) \\ &\quad \times \mathbb{P}^{\mathbf{q}_a}(x_t, s_t, y^{t-1}) \\ &\overset{(a)}{=} \sum_{(x_t,s_t)\in\mathcal{X}\times\mathcal{S}} Q(x_{t+1}|x_t)\mathbb{1}_{s_{t+1}}\{s_t - x_t + y_t\}q_b(y_t|x_t, s_t, y^{t-1}) \\ &\quad \times \mathbb{P}^{\mathbf{q}_b}(x_t, s_t, y^{t-1}) \\ &= \mathbb{P}^{\mathbf{q}_b}(x_{t+1}, s_{t+1}, y^t) \end{aligned}$$

where (a) uses (59) and the induction hypothesis. Thus, (61) holds for $t+1$ and, by the principle of induction, holds for all $t$. Hence (60) holds and, therefore, $I^{\mathbf{q}_a}(X_t, S_t; Y_t|Y^{t-1}) = I^{\mathbf{q}_b}(X_t, S_t; Y_t|Y^{t-1})$. The statement in the Lemma follows by adding over $t$. □

## Appendix III
## Proof of Concavity of the Value Function

To prove the result, we show the following:

*Lemma 3.1:* For any action $a \in A$, if $V : \mathcal{P}_{X,S} \to \mathbb{R}$ is concave, then $\mathscr{B}_a V$ is concave on $\mathcal{P}_{X,S}$.

The concavity of the value functions follows from backward induction. $V_{T+1}$ is a constant and, therefore, also concave. Lemma 3.1 implies that $V_T, V_{T-1}, \dots, V_1$ are concave.

*Proof:* [Proof of Lemma 3.1] The first term $I(a; \pi)$ of $[\mathscr{B}_a V](\pi)$ is a concave function of $\pi$. We show the same for the second term.

Note that if a function $V$ is concave, then it's perspective $g(u, t) := tV(u/t)$ is concave in the domain $\{(u, t) : u/t \in \mathrm{Dom}(V), t > 0\}$. The second term in the definition of the Bellman operator (10)

$$\sum_{y\in Y}\left[\sum_{(x,s)\in\mathcal{X}\times\mathcal{S}} a(y|x, s)\pi(x, s)\right]V(\varphi(\pi, y, a))$$

has this form because the numerator of $\varphi(\pi, y, a)$ is linear in $\pi$ and the denominator is $\sum_{x,s} a(y|x, s)\pi(x, s)$ (and corresponds to $t$ in the definition of perspective). Thus, for each $y$, the summand is concave in $\pi$, and the sum of concave functions is concave. Hence, the second term of the Bellman operator is concave in $\pi$. Thus we conclude that concavity is preserved under $\mathscr{B}_a$. □

## Appendix IV
## Proof of Lemma 9

The proof of the convergence of $\{\xi_t\}_{t\geq 1}$ relies on a result on the convergence of partially observed Markov chains due to Kaijser [33] that we restate below.

*Definition 4: A square matrix $D$ is called* subrectangular *if for every pair of indices $(i_1, j_1)$ and $(i_2, j_2)$ such that $D_{i_1,j_1} \neq 0$ and $D_{i_2,j_2} \neq 0$, we have that $D_{i_2,j_1} \neq 0$ and $D_{i_1,j_2} \neq 0$.*

*Theorem 5 (Kaijser [33]): Let $\{U_t\}_{t\geq 1}$, $U_t \in \mathcal{U}$, be a finite state Markov chain with transition matrix $P^u$. The initial state $U_1$ is distributed according to probability mass function $P_{U_1}$. Given a finite set $\mathcal{Z}$ and an observation function $g : \mathcal{U} \to \mathcal{Z}$, define the following:*

- *The process $\{Z_t\}_{t\geq 1}$, $Z_t \in \mathcal{Z}$, given by*
$$Z_t = g(U_t).$$
- *The process $\{\psi_t\}_{t\geq 1}$, $\psi_t \in \mathcal{P}_{\mathcal{U}}$, given by*
$$\psi_t(u) = \mathbb{P}(U_t = u \mid Z^t).$$
- *A square matrix $M(z)$, $z \in \mathcal{Z}$, given by*
$$[M(z)]_{i,j} = \begin{cases} P^u_{ij} & \text{if } g(j) = z \\ 0 & \text{otherwise} \end{cases} \qquad i, j \in \mathcal{U}.$$

*If there exists a finite sequence $z_1^m$ such that $\prod_{t=1}^m M(z_t)$ is subrectangular, then $\{\psi_t\}_{t\geq 1}$ converges in distribution to a limit that is independent of the initial distribution $P_{U_1}$.*

We will use the above theorem to prove that under policy $\mathbf{b}^\circ$, $\{\xi_t\}_{t\geq 1}$ converges to a limit. For that matter, let $\mathcal{U} = \mathcal{S}\times\mathcal{Y}$, $\mathcal{Z} = \mathcal{Y}$, $U_t = (S_t, Y_{t-1})$ and $g(S_t, Y_{t-1}) = Y_{t-1}$.

[5] The notation $A \to B \to C$ is used to indicate that $A$ is conditionally independent of $C$ given $B$.

First, we show that $\{U_t\}_{t\geq 1}$ is a Markov chain. In particular, for any realization $(s^{t+1}, y^t)$ of $(S^{t+1}, Y^t)$, we have that

$$\begin{aligned} &\mathbb{P}^{\mathbf{b}^\circ}(U_{t+1} = (s_{t+1}, y_t) \mid U^t = (s^t, y^{t-1})) \\ &= \textstyle\sum_{\tilde{x}_t \in \mathcal{X}} P(U_{t+1} = (s_{t+1}, y_t), X_t = \tilde{x}_t \mid U^t = (s^t, y^{t-1})) \\ &= \textstyle\sum_{\tilde{x}_t \in \mathcal{X}} \mathbb{1}_{s_{t+1}}\{y_t + s_t - \tilde{x}_t\} b^*(y_t | s_t - \tilde{x}_t) P_X(\tilde{x}_t) \\ &= \mathbb{P}^{\mathbf{b}^\circ}(U_{t+1} = (s_{t+1}, y_t) \mid U_t = (s_t, y_{t-1})). \end{aligned}$$

Next, let $m = 2m_s$ and consider

$$z^m = \underbrace{111\cdots 1}_{m_s \text{ times}}\underbrace{000\cdots 0}_{m_s \text{ times}}.$$

We will show that this $z^m$ satisfies the subrectangularity condition of Theorem 5. The basic idea is the following. Consider any initial state $u_1 = (s, y)$ and any final state $u_m = (s', 0)$. We will show that

$$\mathbb{P}(S_{m_s} = m_s \mid U_1 = (s, y), Z^{m_s} = (111\ldots 1)) > 0, \quad (62)$$

and

$$\mathbb{P}(S_{2m_s} = s' \mid U_{m_s} = (s_m, 1), Z^{m_s}_{m_s+1} = (000\ldots 0)) > 0. \quad (63)$$

Eqs. (62) and (63) show that given the observation sequence $z^m$, for *any* initial state $(s, y)$ there is a positive probability of observing *any* final state $(s', 0)$.[6] Hence, the matrix $\prod_{t=1}^m M(z)$ is subrectangular. Consequently, by Theorem 5, the process $\{\psi_t\}_{t\geq 1}$ converges in distribution to a limit that is independent of the initial distribution $P_{U_1}$.

Now observe that $\theta_t(s) = \sum_{y\in\mathcal{Y}} \psi_t(s, y)$ and $(\theta_t, \xi_t)$ are related according to Lemma 3. Since $\{\psi_t\}_{t\geq 1}$ converges weakly independent of the initial condition, so do $\{\theta_t\}_{t\geq 1}$ and $\{\xi_t\}_{t\geq 1}$.

Let $\bar{\theta}$ and $\bar{\xi}$ denote the limit of $\{\theta_t\}_{t\geq 1}$ and $\{\xi_t\}_{t\geq 1}$. Suppose the initial condition is $(\theta^\circ, \xi^\circ)$. Since $b^\circ$ is an invariant policy, $(\theta_t, \xi_t) = (\theta^\circ, \xi^\circ)$ for all $t$. Therefore, the limits $(\bar{\theta}, \bar{\xi}) = (\theta^\circ, \xi^\circ)$.

*Proof of Eq.* (62)*:* Given the initial state $(s, y)$, define $\bar{s} = m_s - s$, and consider the sequence

$$x^{m_s} = \underbrace{000\cdots 0}_{\bar{s} \text{ times}}\underbrace{111\cdots 1}_{s \text{ times}}.$$

Under this sequence of demands, consider the sequence of consumption $y^{m_s-1} = (11\ldots 1)$, which is feasible because the state of the battery increases by 1 for the first $\bar{s}$ steps (at which time it reaches $m_s$) and then remains constant for the remaining $s$ steps. Therefore,

$$\begin{aligned} \mathbb{P}(S_{m_s} = m_s \mid U_1 = (s, y), \\ Y^{m_s-1} = (111\ldots 1), X^{m_s} = x^{m_s}) > 0. \end{aligned}$$

Since the sequnce of demands $x^m$ has a positive probability,

$$\begin{aligned} \mathbb{P}(S_{m_s} = m_s, X^{m_s} = x^{m_s} \mid U_1 = (s, y), \\ Y^{m_s-1} = (111\ldots 1)) > 0. \end{aligned}$$

Therefore,

$$\mathbb{P}(S_{m_s} = m_s \mid U_1 = (s, y), Y^{m_s-1} = (111\ldots 1)) > 0$$

which completes the proof. □

*Proof of Eq.* (63)*:* The proof is similar to the Proof of (62). Given the final state $(s', 0)$, define $\bar{s}' = m_s - s'$ and consider the sequence

$$x^{2m_s}_{m_s+1} = \underbrace{111\cdots 1}_{\bar{s}' \text{ times}}\underbrace{000\cdots 0}_{s' \text{ times}}.$$

Under this sequence of demands and the sequence of consumption given by $y^{2m_s-1}_{m_s} = (00\ldots 0)$, the state of the battery decreases by 1 for the first $\bar{s}'$ steps (at which time it reaches $s'$) and then remains constant for the remaining $s'$ steps. Since $x^{2m_s}_{m_s+1}$ has positive probability, we can complete the proof by following an argument similar to that in the proof of (62). □

[6] Note that given the observation sequence $z^m$, the final state must be of the form $(s', 0)$.

## 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] A. Prudenzi, “A neuron nets based procedure for identifying domestic appliances pattern-of-use from energy recordings at meter panel,” in *Proc. Power Eng. Soc. Winter Meeting*, vol. 2. Jan. 2002, pp. 941–946.

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

[4] H. Y. Lam, G. K. S. Fung, and W. K. Lee, “A novel method to construct taxonomy electrical appliances based on load signaturesof,” *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 653–660, May 2007.

[5] C. Efthymiou and G. Kalogridis, “Smart grid privacy via anonymization of smart metering data,” in *Proc. Smart Grid Commun.*, Oct. 2010, pp. 238–243.

[6] G. Kalogridis, C. Efthymiou, S. Z. Denic, T. A. Lewis, and R. Cepeda, “Privacy for smart meters: Towards undetectable appliance load signatures,” in *Proc. Smart Grid Commun.*, Oct. 2010, pp. 232–237.

[7] I. Csiszar and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[8] D. Varodayan and A. Khisti, “Smart meter privacy using a rechargeable battery: Minimizing the rate of information leakage,” in *Proc. Int. Conf. Acoust., Speech Signal Process.*, May 2011, pp. 1932–1935.

[9] O. Tan, D. Gündüz, and H. V. Poor, “Increasing smart meter privacy through energy harvesting and storage devices,” *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1331–1341, Jul. 2013.

[10] G. Giaconi, D. Gündüz, and H. V. Poor, “Smart meter privacy with an energy harvesting device and instantaneous power constraints,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2015, pp. 7216–7221.

[11] S. Han, U. Topcu, and G. J. Pappas, “Event-based information-theoretic privacy: A case study of smart meters,” in *Proc. Amer. Control Conf.*, Jul. 2016, pp. 2074–2079.

[12] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, “Optimal decoding of linear codes for minimizing symbol error rate,” *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 284–287, Mar. 1974.

[13] L. Sankar, S. R. Rajagopalan, and H. V. Poor, “Utility-privacy tradeoffs in databases: An information-theoretic approach,” *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 6, pp. 838–852, Jun. 2013.

[14] L. Sankar, S. R. Rajagopalan, S. Mohajer, and H. V. Poor, “Smart meter privacy: A theoretical framework,” *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 837–846, Jun. 2013.

[15] D. Gunduz and J. Gomez-Vilardebo, “Smart meter privacy in the presence of an alternative energy source,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013.

[16] S. Tatikonda and S. Mitter, “The capacity of channels with feedback,” *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 323–349, Jan. 2009.

[17] A. J. Goldsmith and P. P. Varaiya, “Capacity, mutual information, and coding for finite-state Markov channels,” *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 868–886, May 1996.

[18] H. H. Permuter, P. Cuff, B. Van Roy, and T. Weissman, “Capacity of the trapdoor channel with feedback,” *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3150–3165, Jul. 2008.

[19] J. Yao and P. Venkitasubramaniam, "The privacy analysis of battery control mechanisms in demand response: Revealing state approach and rate distortion bounds," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2417–2425, Sep. 2015.

[20] S. Li, A. Khisti, and A. Mahajan, "Structure of optimal privacy-preserving policies in smart-metered systems with a rechargeable battery," in *Proc. Int. Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2015, pp. 375–379.

[21] O. Hernández-Lerma and J. Lasserre, *Discrete-Time Markov Control Processes*. Berlin, Germany: Springer-Verlag, 1996.

[22] O. Hernández-Lerma, *Adaptive Markov Control Processes*. Berlin, Germany: Springer-Verlag, 1989.

[23] O. Hernández-Lerma, *Further Topics on Discrete-Time Markov Control Processes*. Berlin, Germany: Springer-Verlag, 1991.

[24] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 1, no. 2. Belmont, MA, USA: Athena Scientific, 1995.

[25] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 460–473, Jul. 1972.

[26] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 14–20, Jan. 1972.

[27] L. Wang and M. Madiman, "Beyond the entropy power inequality, via rearrangements," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5116–5137, Sep. 2014.

[28] S. Li, "Information theoretic privacy in smart metering systems using rechargeable batteries," M.S. thesis, Dept. Elect. Comput. Eng., Univ. Toronto, Toronto, ON, Canada, 2016.

[29] A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1501–1518, Nov. 1991.

[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.

[31] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, vol. 703. Hoboken, NJ, USA: Wiley, 2007.

[32] G. Shani, J. Pineau, and R. Kaplow, "A survey of point-based pomdp solvers," *Auto. Agents Multi-Agent Syst.*, vol. 27, no. 1, pp. 1–51, 2013.

[33] T. Kaijser, "A limit theorem for partially observed Markov chains," *Ann. Probab.*, vol. 3, no. 4, pp. 677–696, 1975.

**Simon Li** received his MASc degree from University of Toronto in 2015.

**Ashish Khisti** received his BASc Degree (2002) in Engineering Sciences (Electrical Option) from University of Toronto, and his S.M and Ph.D. Degrees in Electrical Engineering from the Massachusetts Institute of Technology. Between 2009-2015, he was an assistant professor in the Electrical and Computer Engineering department at the University of Toronto. He is presently an associate professor, and holds a Canada Research Chair in the same department. He is a recipient of an Ontario Early Researcher Award, the Hewlett-Packard Innovation Research Award and the Harold H. Hazen teaching assistant award from MIT. He presently serves as an associate editor for IEEE TRANSACTIONS ON INFORMATION THEORY and is also a guest editor for the Proceedings of the IEEE (Special Issue on Secure Communications via Physical-Layer and Information-Theoretic Techniques).

**Aditya Mahajan** (S'06–M'09–SM'14) is Associate Professor of Electrical and Computer Engineering at McGill University, Montreal, Canada. He received B.Tech degree in Electrical Engineering from the Indian Institute of Technology, Kanpur, India in 2003 and MS and PhD degrees in Electrical Engineering and Computer Science from the University of Michigan, Ann Arbor, USA in 2006 and 2008. From 2008 to 2010, he was postdoctoral researcher at Yale University, New Haven, CT, USA. From 2016 to 2017, he was a visiting scholar at the University of California, Berkeley, USA. He is the recipient of the 2015 George Axelby Outstanding Paper Award, 2014 CDC Best Student Paper Award (as supervisor), and the 2016 NecSys Best Student Paper Award (as supervisor). He presently serves as Associate Editor of Springer Mathematics of Control, Signal, and Systems. He was an Associate Editor of the IEEE Control Systems Society Conference Editorial Board from 2014 to 2017. His principal research interests include decentralized stochastic control, team theory, reinforcement learning, multi-armed bandits and information theory.