Privacy-Aware Load Ensemble Control: A Linearly-Solvable MDP Approach

Ali Hassan, Deepjyoti Deka and Yury Dvorkin

Abstract—Demand response (DR) programs engage distributed demand-side resources, e.g., controllable residential and commercial loads, in providing ancillary services for electric power systems. Ensembles of these resources can help reducing system load peaks and meeting operational limits by adjusting their electric power consumption. To equip utilities or load aggregators with adequate decision-support tools for ensemble dispatch, we develop a Markov Decision Process (MDP) approach to optimally control load ensembles in a privacy-preserving manner. To this end, the concept of differential privacy is internalized into the MDP routine to protect transition probabilities and, thus, privacy of DR participants. The proposed approach also provides a trade-off between solution optimality and privacy guarantees, and is analyzed using real-world data from DR events in the New York University microgrid in New York, NY.

NOMENCLATURE

A. Sets and Indices

\(\mathcal{A}\) Set of discretized space states
\(\alpha, \beta, \nu\) States in set \(\mathcal{A}\)
\(\mathcal{T}\) Set of operating time intervals, indexed by \(t\)
\(\Delta_n\) Output space of the Dirichlet mechanism
\(\Omega_1, \Omega_2\) Two disjoint sets that constitute output space \(\Delta_n\)

B. Parameters and Variables

\(\mathcal{P}\) Non-private default transition probability matrix
\(\mathcal{P}^{\alpha\beta}\) Elements of non-private default transition probability matrix \(\mathcal{P}\)
\(\zeta^{\beta}\) Row vector of non-private default transition probability matrix \(\mathcal{P}\)
\(\eta^{\beta}\) Vector adjacent to row vector \(\zeta^{\beta}\)
\(\mathcal{P}^{\alpha\beta}\) Elements of private default transition probability matrix \(\mathcal{P}\)
\(\gamma\) Penalty parameter for discomfort cost
\(h\) Maximum scalar difference between two vectors
\(k\) Scaling parameter of the Dirichlet mechanism
\(\psi\) Parameter defining sets \(\Omega_1, \Omega_2\) in output space \(\Delta_n\)
\(\omega, \psi\) Auxiliary parameters of the Dirichlet mechanism
\(\zeta^{\beta}\) Output vector of the Dirichlet mechanism
\(\mathcal{P}_t\) Non-private controlled transition probability matrix
\(\mathcal{P}_t^{\alpha\beta}\) Elements of non-private controlled transition probability matrix \(\mathcal{P}_t\)
\(\hat{\mathcal{P}}_t^{\alpha\beta}\) Private controlled transition probability matrix
\(\mathcal{P}_t^{\alpha\beta}\) Elements of private controlled transition probability matrix \(\mathcal{P}_t\)
\(\rho_t\) Non-private vector containing probabilities of the load ensemble being in a certain state such as \(\alpha, \beta \in \mathcal{A}\) at time \(t\)
\(\rho_t^{\alpha\beta}\) Elements of non-private vector \(\rho_t\)
\(\hat{\rho}_t^{\alpha\beta}\) Elements of private vector \(\hat{\rho}_t\)
\(U_t^{\beta}\) Utility of the aggregator in state \(\beta\) at time \(t\)
\(\Delta C\) Cost of privacy

C. Functions and Operators

\(\varphi_t^{\beta}\) Non-private value function of state \(\beta\) at time \(t\)
\(z_t^{\beta}\) Non-private desirability function of state \(\beta\) at time \(t\)
\(~\varphi_t^{\beta}\) Private value function of state \(\beta\) at time \(t\)
\(~z_t^{\beta}\) Private desirability function of state \(\beta\) at time \(t\)
\(B(.)\) Multivariate beta function
\(\Gamma(.)\) Gamma function
\(\psi(.)\) Digamma function
\(KL[.\] Kullback–Leibler divergence
\(P(.)\) Probability of a random variable
\(E(.)\) Expectation of a random variable
\(\text{Var(.)}\) Variance of a random variable

I. INTRODUCTION

Electric power utilities deploy Demand Response (DR) programs to reduce dependency on conventional generation assets for meeting peak electricity demand or avoiding network congestion. DR programs provide an alternative to using these resources by encouraging electricity customers to change their power consumption in exchange for an incentive \([1]\). This incentive can come in a form of direct monetary compensations (e.g. payment or electricity bill rebate/credit) or else (e.g. voucher, gift card, etc) \([2], [3]\). In total, the U.S. Energy Information Administration (EIA) reports that DR programs yielded energy savings of 1,462,735 MWh in 2019, an increase of 9.4% from 2016 \([4]\).

To model residential DR programs, which must deal with a large number of small-scale appliances with various nameplate parameters and characteristics, the Markov Decision Process (MDP) framework has been adopted and then field-validated as an effective and scalable decision-making framework for controlling large ensembles of such appliances and buildings \([5]–[16]\). The advantage of this framework is in connecting (i) modeling convenience of the MDP, which makes it easier to obtain high-fidelity and data-driven representations of load and building ensembles, and (ii) solution simplicity of dynamic programming. However, while the MDP is useful to operators of DR programs \([5]–[16]\, it has relatively large data
requirements to properly construct a Markov Process (MP) underling the MDP. This, in turn, raises privacy and security concerns that the data used for constructing a MP of the load ensemble can be breached and sensitive features of individual DR participants can be revealed [17], [18].

Privacy issues are not unique to DR programs. For example, in 2007, Netflix released a data set of their user ratings as part of a competition to see if anyone can outperform their collaborative filtering algorithm. Although all personal identifying information was removed for privacy protection, two researchers from the University of Austin were still able to breach privacy and recovered 99% of the removed personal information using reference data from IMDb [19]. This exposed the ability of compromising user privacy by reconstructing useful information from other, potentially unprotected, data sets. Similarly, large volumes of data generated and gathered for DR programs can be used to compromise the privacy of DR participants and reveal their sensitive information. For example, detailed electricity consumption data can reveal lifestyle profiles and choices, including but not limited to a number of people living in a household, absence or presence of household members, working hours, meal times, types of appliances and even religious affiliations in some cases [20]. National Institute of Standards and Technology (NIST) considers four different dimensions or aspects of privacy as – personal information, personal privacy, behavioral privacy, and personal communication – that can be infringed upon using the electricity consumption and DR program data [18].

Accordingly, motivated by the data breaches as in [17], [18], [20], DR programs should be immunized against risks of exposing their customers. Previously, privacy concerns in DR programs have been studied in different contexts. The notable techniques employed to protect the privacy involve data aggregation [21]–[24], data anonymization [25]–[27] and data perturbation [28]–[30]. Data aggregation and anonymization schemes can provide a false sense of privacy. For example, even though they can ensure the security of data during communication, privacy could still be compromised by specific filters such as Collaborative Filtering [31] or reconstruction theorems or attacks [32], and the aggregated data can be manipulated to release sensitive information about individuals (see the Netflix data breach example mentioned above [19]). In [33], Jawurek et al. show that even after pseudonymization, which is a data management procedure promoted by the European Union’s General Data Protection Regulation (GDPR), data items can still be traced to their respective individuals sources. Data perturbation means adding random noise to the data to protect the original data, but this procedure per se does not always guarantee privacy as original data can be recovered from the perturbed data [34] (e.g. by means of de-noising [35]). This motivates us to explore such techniques that can guarantee privacy, while ensuring the usability of the aggregated data. Differential privacy (DP) is a mathematically rigorous framework that allows for sharing aggregated data (e.g. parameters of a given load ensemble), while preserving features of individuals (e.g. loads in the ensemble). In other words, DP makes it possible to quantify to what extent privacy of an individual in the aggregated data is preserved, while sharing useful aggregated information about the data set [36]. This property, as explained in [37], ensures that DP is capable of “learning nothing about an individual, while learning useful information about a population”. This property makes DP suitable for internalizing privacy restrictions into the MDP framework applied to DR programs, which deals with the aggregated behavior of DR participants. Additionally, DP performance hardly changes when a single individual joins or leaves a given ensemble, i.e. its output will be roughly the same, regardless if an individual is part of the data set or not. Notably, despite being a fairly recent invention, DP has already been widely used in privacy-critical applications, e.g. by US Census Bureau [38], Apple [39], LinkedIn [40], Facebook [41], Microsoft [42] and Google [43].

A DP analysis of a method or an algorithm is usually called a mechanism. Any DP mechanism is randomized to ensure that anyone observing its output will essentially make the same inference about any individual’s data whether or not that individual’s data is included in its input. As such, a given mechanism is usually randomized by adding some noise to the input or output of the algorithm depending upon the given framework or application [44]. Some common DP mechanisms include Laplace and Gaussian mechanisms, where added noises are drawn from Laplace and Gaussian distributions. While these mechanisms are widely employed in practice, their application in the MDP is challenging as adding a random noise may not comply with the integrality requirement on the MDP, i.e. the cumulative probability of moving from any given state to all future possible states is equal to one. Recently, this challenge has been overcome by means of the Dirichlet mechanism [45], which was subsequently introduced into the MDP [46]. Since the Dirichlet distribution is simplex-constrained, it allows for maintaining the integrality requirement for the MDP. However, the result in [46] deals with conventional MDPs, i.e. it searches over all actions for each future state, which is computationally demanding due to the exponential growth of future states. Furthermore, MDPs often do not allow for analytical solutions and require using numerical methods. To overcome this challenge, a conventional MDP can be reduced to a linearly-solvable MDP (LS-MDP), under suitable action costs. Unlike in a conventional MDP, the optimal policy derived from the LS-MDP is not a mapping of states to action variables, but rather a mapping of a current state into a next-state distribution. Additionally, LS-MDPs are reduced to linear eigenvalue problems and can be solved using analytical optimal policy, thus reducing computational requirements, while providing accurate approximations to conventional MDPs [47].

The current literature on DP implementation for DR programs adds noise to the input or aggregated data [48]–[50]. Such methods don’t consider the optimization problem, limiting their application since their solutions are not certified feasible. In contrast, we internalize the noise into our optimization problem and utilize stochastic programming [51] for privacy-aware DR solutions. This paper introduces DP into a LS-MDP framework, where DP is employed to privatize default transition probabilities, which are obtained from the MP derived from electricity consumption data, by means of adding a noise from a specified Dirichlet distribution to protect
the privacy of DR participants. The main contributions are as follows:

1) We leverage stochastic programming to derive privatized optimal policies. These optimal policies are the generalized expressions of the stochastic policies derived in \[13\] to internalize uncertainty, using first and second statistical moments, but extended to privatized versions with privacy guarantees on default transition probabilities.

2) While the use of stochastic programming enables the aggregators to protect the privacy of DR participants in their optimization problems, they may need to send their data to external entities (e.g., web/cloud computing or report data to authorities for audits [52]). This motivates us for exploring the average value method for DP, our second contribution. This method solves the LS-MDP problem for a set of randomly generated samples of default transition probabilities, and provides randomized optimal policies for these samples. This also allows the aggregator to outsource the optimization step to a third entity without compromising the privacy of DR participants. Unlike in the stochastic methods described above, the average value method does not require specifying first and second statistical moments as exactly known for the randomized policies. The average of all the optimal randomized policies is computed to compare the outcome of the average value method to the stochastic approach.

3) The cost of privacy for the DP LS-MDP is also derived for both the stochastic and average value methods, and the loss of optimality, also known as the cost of privacy, that occurs as the result of providing privacy to DR participants, is analyzed.

The usefulness of the proposed framework is demonstrated on the real-life data based on New York University (NYU) campus building, which participates in Consolidated Edison operated DR events.

The rest of this paper is organized as follows. Section II presents the fundamentals of LS-MDP and derives a non-private control policy. Section III introduces the basic concepts of differential privacy, describes the Dirichlet mechanism and related properties. Sections IV and V describe the proposed stochastic and average value approaches respectively and derive privacy-aware control policies and their cost of privacy expressions. Section VI presents the case study to show the effectiveness of the proposed models. Section VII concludes the paper.

II. A LS-MDP FORMULATION

The load ensemble can be represented by a discrete-time, discrete-space MDP similar to our prior work in [9]–[12], [15], [16]. We consider a LS-MDP with state space \( \mathcal{A} \), time space \( T \), default transition probability matrix \( \mathcal{P} \in \mathbb{R}^{n \times n}, n = |\mathcal{A}| \), controlled transition probability matrix \( \mathcal{P}_t \in \mathbb{R}^{n \times n}, n = |\mathcal{A}| \) and utility of the aggregator \( U_t^\beta \). State space \( \mathcal{A} = \{\alpha, \beta, \cdots\} \) is obtained by discretizing controllable parameters of the loads in the ensemble (e.g., power consumption or temperature levels) and time space \( T = \{t, t+1, \cdots\} \) represents discrete time intervals (e.g., hours or minutes). Fig. 1 illustrates a MP with eight states with possible transitions from state 1 to other states. Default transition probability matrix \( \mathcal{P} \in \mathbb{R}^{n \times n} \) represents the evolution of the ensemble in the state space without controlled decisions of the aggregator, i.e. it represents a MP that corresponds to the endogenous dynamics of the load ensemble. On the other hand, controlled transition probability matrix \( \mathcal{P}_t \in \mathbb{R}^{n \times n} \) represents the evolution of the ensemble under the control actions of the aggregator. Notably, matrix \( \mathcal{P}_t \) is time-variable to enable dynamic control of the ensemble. More specifically, \( \mathcal{P}^\alpha_\beta \) and \( \mathcal{P}_t^\alpha_\beta \) are elements of matrix \( \mathcal{P} \) and \( \mathcal{P}_t \) associated with a transition from state \( \beta \) to state \( \alpha \neq \beta \in \mathcal{A} \).

While matrix elements \( \mathcal{P}^\alpha_\beta \) can be retrieved from processing historical observations (measurements) of the load ensemble [5], matrix elements \( \mathcal{P}_t^\alpha_\beta \) are the outcomes of the decision-making process of the aggregator, which aims to maximize its utility function given by \( U_t^\beta \). This decision-making process can be cast as:

\[
\begin{align*}
\min_{\rho_t, \mathcal{P}_t} & \quad O^A := \mathbb{E}_\rho \sum_{t \in T} \sum_{\alpha \in \mathcal{A}} \left( -U_t^\alpha + \sum_{\beta \in \mathcal{A}} \gamma \log \frac{\mathcal{P}_t^\alpha_\beta}{\mathcal{P}^\alpha_\beta} \right) \quad (1a) \\
\text{s.t.} & \quad \rho_t^\alpha = \sum_{\beta \in \mathcal{A}} \mathcal{P}_t^\alpha_\beta \rho_t^\beta, \quad \forall \alpha \in \mathcal{A}, t \in T \\
& \quad \sum_{\alpha \in \mathcal{A}} \mathcal{P}_t^\alpha_\beta = 1, \quad \forall \beta \in \mathcal{A}, t \in T - 1, \quad (1c)
\end{align*}
\]

where \( \rho_t \in \mathbb{R}^n, n = |\mathcal{A}| \), is a vector with entries \( \rho_t^\alpha \geq 0 \) and \( \rho_t^\beta \geq 0 \) representing the probabilities that the load ensemble is operated in states \( \alpha, \beta \in \mathcal{A} \) at times \( t+1 \) and \( t \) respectively, and are related via controlled transition probabilities \( \mathcal{P}_t^\alpha_\beta \). Eq. (1a), where operator \( \mathbb{E}_\rho \) denotes the expectation over \( \rho \), maximizes the utility of the DR aggregator and minimizes the discomfort cost of DR participants. The discomfort cost is modeled as the Kullback-Leibler (KL) divergence [53], i.e. the logarithmic difference between the default transition probabilities and controlled transition probabilities weighted by cost penalty \( \gamma \), which can be interpreted as a lost utility due to the need to deviate from the default behavior of electric loads in the ensemble. For example, this interpretation can be viewed as a change in temperature settings of air-conditioning units from user-defined comfort levels to the levels optimized by the DR aggregator. Furthermore, if \( \mathcal{P}_t^\alpha_\beta = 0 \), i.e. a transition from state \( \beta \) to \( \alpha \) and ratio has not been observed in the past, the model in (1) restricts \( \mathcal{P}_t^\alpha_\beta = 0 \) and excludes such transitions from
the optimization \(^1\) Eq. (15) describes the temporal evolution of the load ensemble (e.g., changes in temperature settings of air-conditioning units) from time \(t\) to \(t+1\) over time horizon \(T\). Eq. (13) imposes the integrality constraint on the transition decisions optimized by the DR aggregator such that their total probability is equal to one. As shown in [13], Eq. (1) can be optimized as follows:

**Theorem 1.** Let (1) model a load ensemble as a LS-MDP. Then, the optimal control policy \(\forall \beta, \alpha \in \mathcal{A}, \forall t \in T\), is computed as:

\[
\mathcal{P}^\alpha_\beta = \frac{\mathcal{P}^\alpha_\beta z^\alpha_{\beta+1}}{\sum_{\alpha \in \mathcal{A}} \mathcal{P}^\alpha_\beta z^\alpha_{\beta+1}},
\]

where \(\forall \beta \in \mathcal{A}, \forall t \in T\), and

\[
z^\beta_t = \exp\left(\frac{U^\beta_t}{\gamma}\right) \sum_{\alpha \in \mathcal{A}} \mathcal{P}^\alpha_\beta z^\alpha_{t+1},
\]

**Proof.** See proof in Appendix A

Here, \(z^\beta_0\) is a desirability function representing value function \(\varphi^\beta_t\) of state \(\beta\) at time \(t\), i.e., \(z^\beta_t = \exp(-\frac{U^\beta_t}{\gamma})\). Similarly, \(z^\alpha_{t+1}\) is a desirability function for state \(\alpha\) at time \(t+1\).

### III. Differentially Private LS-MDP

The LS-MDP developed in Section II to model an ensemble of controllable loads is prone to privacy breaches and can reveal sensitive data of DR participants. In this section, we incorporate DP into the LS-MDP formulation in (1) to protect privacy of DR participants and derive differentially private optimal policy as summarized in Fig. 2. We explain the use of DP for LS-MDPs by starting with the following definitions:

**Definition III.1** (Adjacent vectors). Let \(\vec{\zeta}^\beta \in \mathbb{R}^n, n = |\mathcal{A}|\), be a row vector of matrix \(\overline{\mathcal{P}}\), i.e., \(\vec{\zeta}^\beta = \{\mathcal{P}^\beta_\alpha, \forall \alpha \in \mathcal{A}\}\), and let vector \(\vec{\eta}^\beta \in \mathbb{R}^n, n = |\mathcal{A}|\), have the same elements as vector \(\vec{\zeta}^\beta\) except for two randomly chosen elements. Then vectors \(\vec{\zeta}^\beta\) and \(\vec{\eta}^\beta\) are adjacent if:

\[
||\vec{\zeta}^\beta - \vec{\eta}^\beta||_1 \leq h,
\]

where \(||\cdot||_1\) denotes the first norm of a vector and \(0 < h < 1\) represents the maximum scalar difference between the two vectors. Using \(h = 0\) will result in two identical vectors i.e., \(\vec{\zeta}^\beta = \vec{\eta}^\beta\), and \(h = 1\) will violate the integrality constraint on the row vector \(\vec{\eta}^\beta\). We denote two adjacent vectors using symbol \(\simeq\), i.e. \(\vec{\zeta}^\beta \simeq \vec{\eta}^\beta\).

Consistently with [45], [46]. **Definition III.1** ensures that vectors \(\vec{\zeta}^\beta\) and \(\vec{\eta}^\beta\) are different by no more than value \(h\), which is user defined. This definition departs from the traditional notion of differential privacy [37], where two data sets are different in only one entry. This modification is needed to ensure that if one element in vector \(\vec{\eta}^\beta\) is modified to ensure DP, the integrality constraint in (1c) holds, i.e. the probability of the ensemble being in a given state is equal to one. In other words, we modify two elements in vector \(\vec{\eta}^\beta\) because in the LS-MDP context, the vector elements are interpreted as probabilities and, thus, must add up to one.

**Definition III.2** (Differential Privacy). Random mechanism \(\mathcal{M}\) satisfies \((\epsilon, \delta)\)-probabilistic differential privacy, if for all adjacent input vectors \(\vec{\zeta}^\beta \simeq \vec{\eta}^\beta\) and output vectors \(\vec{\chi}^\beta\), the following holds:

\[
P[\mathcal{M}(\vec{\zeta}^\beta) = \vec{\chi}^\beta] \leq e^{\epsilon} P[\mathcal{M}(\vec{\eta}^\beta) = \vec{\chi}^\beta],
\]

where \(P\) denotes the probability of an event and \(\epsilon \geq 0\) represents the privacy loss quantified by DP, where low values of \(\epsilon\) provide more privacy at the expense of accuracy.

As applied to the default transition probabilities of the LS-MDP optimization in (1). Definition III.2 treats this process as mechanism \(\mathcal{M}\) and its input \(\overline{\mathcal{P}}^\alpha_\beta\) and output \(\mathcal{P}^\alpha_\beta\) as components of vectors \(\vec{\zeta}^\beta\) and \(\vec{\chi}^\beta\), respectively. Thus, this DP definition guarantees that a randomized mechanism behaves similarly on adjacent inputs, thus protecting the individual’s privacy in the database by not providing enough information on the output of the randomized mechanism that can relate it (or trace back) to a particular input database.

**Definition III.3** (Probabilistic Differential Privacy). Random mechanism \(\tilde{\mathcal{M}}\) satisfies \((\epsilon, \delta)\)-probabilistic differential privacy, if for all vectors \(\vec{\chi}^\beta\) output space \(\Delta_n\) of \(\mathcal{M}\) can be divided into sets \(\Omega_1, \Omega_2\) such that:

\[
P[\mathcal{M}(\vec{\zeta}^\beta) \in \Omega_2] \leq \delta,
\]

and for all \(\vec{\zeta}^\beta \simeq \vec{\eta}^\beta\) and for all \(\vec{\chi}^\beta \in \Omega_1\):

\[
\log \frac{P[\mathcal{M}(\vec{\zeta}^\beta) = \vec{\chi}^\beta]}{P[\mathcal{M}(\vec{\eta}^\beta) = \vec{\chi}^\beta]} \leq \epsilon.
\]

where \(\epsilon \geq 0\), and \(0 \leq \delta \leq 1\) represents the maximum probability of breaching \(\epsilon\)-differential privacy.

This definition of \((\epsilon, \delta)\)-probabilistic differential privacy guarantees that \(\epsilon\)-differential privacy is achieved with at least \((1 - \delta)\) probability [54]. Both parameters \(\epsilon\) and \(\delta\) are user-defined and can be selected to satisfy privacy preferences of a particular decision-maker. A smaller value of \(\epsilon\) reflects stronger privacy for a relatively less accurate response, and \(\delta\) is usually chosen very small (\(0.05\)) to ensure high satisfaction of \(\epsilon\)-differential privacy.

**Definition III.4** (Dirichlet Mechanism). Let \(\tilde{\mathcal{M}}\) be a Dirichlet mechanism with input vector \(\vec{\zeta}^\beta\) and output vector \(\vec{\chi}^\beta\) according to the Dirichlet distribution, i.e.:

\[
P[\tilde{\mathcal{M}}(k\vec{\zeta}^\beta) = \vec{\chi}^\beta] =
\]

\[
\frac{1}{B(k)} \prod_{i=1}^{n-1} (\vec{\chi}^\beta_i)^{k\vec{\zeta}^\beta_i - 1}(1 - \sum_i^{n-1} \vec{\chi}^\beta_i)^{k\vec{\zeta}^\beta_{n-1}},
\]

where \(k \geq 0\) is a user-defined parameter that can be adjusted to provide a trade-off between privacy and the accuracy level of the Dirichlet mechanism, and \(B(k)\) is the normalizing constant, which is a multivariate beta function, expressed as:

\[
B(k) = \prod_{i=1}^{n} \Gamma(k\vec{\zeta}^\beta_i) / \Gamma(k \sum_{i=1}^{n} \vec{\zeta}^\beta_i)
\]

where \(\Gamma(.)\) is the gamma function [55]. The Dirichlet mech-
anism in Definition III.4 is a randomized mechanism that takes a vector scaled by parameter $k$ as an input and provides an output vector according to the Dirichlet distribution. The output vector is more concentrated around the input vector for greater values of $k$, that is the value of $k$ controls the amount of noise added by the Dirichlet mechanism. When $k$ is small, the output vector tends to depart from the input vector, indicating better privacy, but this comes at the expense of the solution accuracy. So, physically the effect of differential privacy on the load ensemble represented by a LS-MDP is its perturbation via Dirichlet noise without compromising its properties. To ensure privacy guarantee for this mechanism, we state:

**Theorem 2 (Privacy Guarantee [45]).** The Dirichlet mechanism in Definition III.4 satisfies $(\epsilon, \delta)$-probabilistic differential privacy with:

$$
\epsilon = \log \left( \frac{B(\omega, k(1 - \omega - \omega))}{B(\omega + \frac{h}{2}, k(1 - \omega - \omega - \frac{h}{2}))} \right) + \frac{k}{2}h \log \left( 1 - \left( \frac{|W| - 1}{\psi} \right) \right)
$$

and

$$
\delta = 1 - \min \left\{ \mathcal{F}[\tilde{M}(\alpha)] \in \Omega_1 \right\}
$$

where $\omega, \omega, \psi \in (0, 1)$, $h \in (0, 1]$ and $W \subseteq [n - 1]$. Note that output space $\Delta_n$ of Dirichlet mechanism $M$ is split in two disjoint sets $\Omega_1 := \{\tilde{\alpha} \in \Delta_n | \tilde{\alpha} \leq \psi, \forall \beta \in A\}$ and $\Omega_2 := \{\Delta_n \setminus \Omega_1\}$.\hfill \Box

**Remark.** Alternately, $\delta$ in Theorem 2 can be selected as the maximum probability of privacy failure to divide the output space $\Delta_n$ of $M$ in disjoint sets $\Omega_1$ and $\Omega_2$ based on the value obtained for $\psi$, and $\epsilon$ is calculated accordingly. Interested readers are referred to [45] for more details.

Owing to Theorem 2, the Dirichlet mechanism in Definition III.4 makes it possible to obtain private default transition probabilities $(\tilde{P})$ to be used in (1), which satisfy $(\epsilon, \delta)$-probabilistic differential privacy and lead to a private optimal policy using the property of post processing. This is possible since differential privacy is immune to post-processing [37].

**Definition III.5 (Post-Processing).** Let $\tilde{M}$ be a random, $(\epsilon, \delta)$-differential private mechanism. Let $f$ be an arbitrary mapping. Then $f \circ \tilde{M}$ is $(\epsilon, \delta)$-differential private.

As a result of Definition III.5, if the LS-MDP optimization in (1) is fed with default transition probabilities $P^{\alpha\beta}$ and applies the Dirichlet mechanism as in Definition III.4 to the default transition probabilities, then Theorem 2 and Definition III.5 ensure that the LS-MDP optimization in (1) outputs private controlled transition probabilities, which we represent by $\tilde{P}^{\alpha\beta}$ as shown in Fig. 2. Sections IV and V derive differentially private optimal policies using the application of the Dirichlet mechanism.

**IV. Stochastic Approach**

We leverage stochastic optimization to solve the LS-MDP when DP is employed to privatize default transition proba-

![Figure 2. LS-MDP with differential privacy](image)

bilities. Accordingly, the objective function in Eq. (1a) is re-written for the DP LS-MDP as:

$$
\min_{\tilde{P}, \tilde{\rho}} \mathbb{E}_{\tilde{P}, \tilde{\rho}} \sum_{t \in T} \sum_{\alpha \beta \in A} \left( -U_{t+1}^{\alpha} + \gamma \sum_{\beta \in A} \tilde{P}^{\alpha\beta} - \mathbb{E}_{\tilde{P}, \tilde{\rho}} [\log \tilde{P}^{\alpha\beta}] \right)
$$

Given (12), the optimization in (1) is recast as:

$$
\min_{\tilde{P}, \tilde{\rho}} \mathbb{E}_{\tilde{P}, \tilde{\rho}} \sum_{t \in T} \sum_{\alpha \beta \in A} \left( -U_{t+1}^{\alpha} + \gamma \sum_{\beta \in A} \tilde{P}^{\alpha\beta} - \mathbb{E}_{\tilde{P}, \tilde{\rho}} [\log \tilde{P}^{\alpha\beta}] \right)
$$

s.t. Eq. (1b) – (13b).

where $\tilde{P}^{\alpha\beta}$ represents privatized default transition probabilities and the decision-maker employing (13) can select a certain policy on how to randomize the original default transition probabilities. For the reasons stated in Section III we use the Dirichlet mechanism for this randomization as given in Definition III.4. To solve the DP LS-MDP formulation in (13), we prove:

**Theorem 3.** Let (13) model a load ensemble as a differential private LS-MDP. Then, the differentially private optimal control policy $\forall \beta, \alpha \in A, \forall t \in T$, is computed as:

$$
\tilde{P}^{\alpha\beta}_t = \frac{\exp(\mathbb{E}_{\tilde{P}, \tilde{\rho}} [\log \tilde{P}^{\alpha\beta}]) \tilde{z}_t^{\alpha\beta}}{\sum_{\alpha \in A} \exp(\mathbb{E}_{\tilde{P}, \tilde{\rho}} [\log \tilde{P}^{\alpha\beta}]) \tilde{z}_t^{\alpha\beta}},
$$

where $\forall \beta, \alpha \in A, \forall t \in T$:

$$
\tilde{z}_t^{\alpha\beta} = \exp \left( \frac{U_t^{\alpha\beta}}{\gamma} \right) \sum_{\alpha} \exp(\mathbb{E}_{\tilde{P}, \tilde{\rho}} [\log \tilde{P}^{\alpha\beta}]) \tilde{z}_t^{\alpha\beta},
$$

**Proof.** See proof in Appendix A.\hfill \Box

We note that the optimal control policy in (14) achieves differential privacy because the stochastic LS-MDP internalizes the Dirichlet noise, thus privatizing the optimization itself and making its output private [37]. Notably, the outcome of Theorem 3 depends on the evaluation of $\mathbb{E}_{\tilde{P}, \tilde{\rho}} [\log \tilde{P}^{\alpha\beta}]$. We, therefore, propose two methods for computing $\mathbb{E}_{\tilde{P}, \tilde{\rho}} [\log \tilde{P}^{\alpha\beta}]$ in Theorem 3 under a given Dirichlet distribution and, thus,

2The term $\mathbb{E}_{\tilde{P}, \tilde{\rho}} [\log \tilde{P}^{\alpha\beta}]$ is generic and other distributions such as Normal and Laplace distributions can be used to further evaluate it. However, the output of non-Dirichlet distributions (where output does not belong to the unit simplex) needs to be adjusted or normalized to ensure that every row in the transition probability matrix remains equal to one. In other words, the probability of moving from present state $\beta$ to all possible next states $\alpha$ is equal to one.
for enabling its computational tractable solution. As described below, the first method is based on the Taylor approximation, while the second one is based on Digamma functions and yields an exact equivalent.

A. Taylor Approximation

We approximate \( \mathbb{E}_{\mathcal{P}^{\alpha \beta}}[\log \mathcal{P}^{\alpha \beta}] \) by the second-order Taylor expansion to represent this term as a function of the first and second statistical moments (mean and variance):

\[
\mathbb{E}_{\mathcal{P}^{\alpha \beta}}[\log \mathcal{P}^{\alpha \beta}] \approx \log \mathbb{E}_{\mathcal{P}^{\alpha \beta}}[\mathcal{P}^{\alpha \beta}] - \frac{\text{Var}(\mathcal{P}^{\alpha \beta})}{2} \mathbb{E}_{\mathcal{P}^{\alpha \beta}}[\mathcal{P}^{\alpha \beta}]^2, \tag{16}
\]

where \( \text{Var}(\cdot) \) denotes variance. Next, using the notion of the Dirichlet distribution, we invoke for the mean \( \mathbb{E}_{\mathcal{P}^{\alpha \beta}}[\mathcal{P}^{\alpha \beta}] \) in Eq. (16):

\[
\mathbb{E}_{\mathcal{P}^{\alpha \beta}}[\mathcal{P}^{\alpha \beta}] = \sum_{\alpha \in A} k\mathcal{P}^{\alpha \beta} = k(1) = \mathcal{P}^{\alpha \beta}. \tag{17}
\]

Similarly, using the notion of variance for the Dirichlet distribution, we obtain for the variance term in Eq. (16):

\[
\text{Var}(\mathcal{P}^{\alpha \beta}) = \frac{k\mathcal{P}^{\alpha \beta}(k \sum_{\alpha \in A} \mathcal{P}^{\alpha \beta} - k \mathcal{P}^{\alpha \beta})}{(k \sum_{\alpha \in A} \mathcal{P}^{\alpha \beta})^2 (k \sum_{\alpha \in A} \mathcal{P}^{\alpha \beta} + 1)} = \frac{\mathcal{P}^{\alpha \beta}(1 - \mathcal{P}^{\alpha \beta})}{k + 1}. \tag{18}
\]

Given (17) and (18), the resulting value of Eq. (16) can be approximated as:

\[
\mathbb{E}_{\mathcal{P}^{\alpha \beta}}[\log \mathcal{P}^{\alpha \beta}] \approx \log \mathcal{P}^{\alpha \beta} - \frac{1 - \mathcal{P}^{\alpha \beta}}{2(\mathcal{P}^{\alpha \beta})(k + 1)} \tag{19}
\]

Given the DP LS-MDP formulation in (13) and expression in (19), we propose:

**Proposition 1.** Let (13) model a load ensemble as a differentially private LS-MDP and let \( \mathbb{E}_{\mathcal{P}^{\alpha \beta}}[\log \mathcal{P}^{\alpha \beta}] \) be approximated by the second-order Taylor approximation as in (19). Then, the differentially private optimal control policy \( \forall \beta, \alpha \in A, \forall t \in \mathcal{T} \), is computed as:

\[
\hat{P}^{\alpha \beta}_t = \frac{\mathcal{P}^{\alpha \beta} \exp \left[ -\frac{1 - \mathcal{P}^{\alpha \beta}}{2(\mathcal{P}^{\alpha \beta})(k + 1) - \mathcal{P}^{\alpha \beta}} \right]}{\sum_{\alpha \in A} \mathcal{P}^{\alpha \beta} \exp \left[ -\frac{1 - \mathcal{P}^{\alpha \beta}}{2(\mathcal{P}^{\alpha \beta})(k + 1) - \mathcal{P}^{\alpha \beta}} \right]}, \tag{20}
\]

where \( \forall \beta \in A, \forall t \in \mathcal{T} \) and:

\[
\hat{z}^{\beta}_t = \exp \left( \frac{U^{\beta}_t}{\gamma} \right) \sum_{\alpha \in A} \mathcal{P}^{\alpha \beta} \exp \left[ -\frac{1 - \mathcal{P}^{\alpha \beta}}{2(\mathcal{P}^{\alpha \beta})(k + 1) - \mathcal{P}^{\alpha \beta}} \right]. \tag{21}
\]

**Proof.** Our proof follows immediately from Theorem 3. By using Eq. (19) in Eqs. (14) and (15), we end up with the results in Eqs. (20) and (21) and conclude the proof. \(\blacksquare\)

Since (13) handles the noise from Dirichlet distribution introduced for DP, its outcome is more costly as compared to the non-private optimal policy from Theorem 1. The difference in cost between the private and non-private policies, therefore, constitutes the cost of privacy that the decision-maker will pay for privatizing its data. Notably, the cost of privacy incurred by (13) can be evaluated a priori as follows:

**Corollary 1.** Given the differentially private optimal control policy in (20), the cost of privacy for the stochastic approach using the second-order Taylor approximation is calculated as:

\[
\Delta C = \sum_{\alpha \in A} \frac{\gamma \log \sum_{\alpha \in A} \mathcal{P}^{\alpha \beta} \exp \left[ -\frac{1 - \mathcal{P}^{\alpha \beta}}{2(\mathcal{P}^{\alpha \beta})(k + 1) - \mathcal{P}^{\alpha \beta}} \right]}{\mathcal{P}^{\alpha \beta}(k + 1)} \tag{22}
\]

**Proof.** See proof in Appendix B. \(\blacksquare\)

While the result in (22) is unwieldy, it is valuable for evaluating the effect of the Dirichlet distribution on the operating cost prior to implementing the noise for ensuring differentially private solutions. Notably, this method requires only a characterization of the first and second statistical moments, which are often available in practice [56], and tuning of parameter \(k\).

B. Digamma Function Equivalent

While using the Taylor approximation in Section A provides a reasonably accurate approximation for \( \mathbb{E}_{\mathcal{P}^{\alpha \beta}}[\log \mathcal{P}^{\alpha \beta}] \), the decision-maker can alternatively derive analytical results for the differentially private control policy by utilizing the notion of Digamma functions [55] to represent the value of \( \mathbb{E}_{\mathcal{P}^{\alpha \beta}}[\log \mathcal{P}^{\alpha \beta}] \). However, the exactness of this policy comes at the expense of more complex representations that somewhat obstruct intuitive interpretations of the results. Accordingly, the use of Digamma functions leads to:

\[
\mathbb{E}_{\mathcal{P}^{\alpha \beta}}[\log \mathcal{P}^{\alpha \beta}] = \psi(k\mathcal{P}^{\alpha \beta}) - \psi(k) \tag{23}
\]

where \(\psi(.)\) represents the Digamma function, defined as the logarithmic derivative of the following form:

\[
\psi(x) = \frac{d}{dx} \log[\Gamma(x)] = \frac{\Gamma'(x)}{\Gamma(x)} \tag{24}
\]

where \(\Gamma(.)\) is the gamma function. Given the expression in (23), we prove for the DP LS-MDP formulation in (13).

**Proposition 2.** Let (13) model a load ensemble as a differentially private LS-MDP and let \( \mathbb{E}_{\mathcal{P}^{\alpha \beta}}[\log \mathcal{P}^{\alpha \beta}] \) be evaluated using the Digamma function as in (23). Then, the differentially private optimal control policy \( \forall \beta, \alpha \in A, \forall t \in \mathcal{T} \), is computed as:

\[
\hat{P}^{\alpha \beta}_t = \frac{\exp \left[ \psi(k\mathcal{P}^{\alpha \beta}) - \psi(k) \right]}{\sum_{\alpha \in A} \exp \left[ \psi(k\mathcal{P}^{\alpha \beta}) - \psi(k) \right]} \tag{25}
\]

where \( \forall \beta \in A, \forall t \in \mathcal{T} \) and:

\[
\hat{z}^{\beta}_t = \exp \left( \frac{U^{\beta}_t}{\gamma} \right) \sum_{\alpha \in A} \exp \left[ \psi(k\mathcal{P}^{\alpha \beta}) - \psi(k) \right]. \tag{26}
\]
Proof. Similarly to Proposition 1, this proof follows immediately from Theorem 3. By using Eq. (23) in Eqs. (14) and (15), we end up with the results in Eqs. (25) and (26) and conclude the proof. 

Similarly to Proposition 1, this optimal control policy in Proposition 2 internalizes the Dirichlet noise, but in a more accurate fashion since Digamma function provides an equivalent expression for \( \mathbb{E}_{P_2^{\alpha \beta}}[\log \tilde{P}_2^{\alpha \beta}] \). This enhancement in accuracy of the optimal control policy, relative to the Taylor approximation above, results in added privacy, which, in turn, increases the cost of privacy as our later case study shows. Furthermore, the result of Proposition 2 only requires tuning of parameter \( k \) and does not necessitate any knowledge on the first- and second statistical moments. Using the result of Proposition 2, the cost of privacy can then be computed as:

**Corollary 2.** Given the differential private optimal control policy in (25), the cost of privacy for the stochastic approach using the Digamma function equivalent is calculated as:

\[
\Delta C = \gamma \sum_{\alpha,\bar{\alpha}} \frac{\exp \left( \psi(k\tilde{P}_2^{\alpha \beta}) - \psi(k) \right) \tilde{z}_{i+1}^{\alpha \beta}}{\sum_{\alpha,\bar{\alpha}} \exp \left( \psi(k\tilde{P}_2^{\alpha \beta}) - \psi(k) \right) \tilde{z}_{i+1}^{\alpha \beta} - \gamma \log \sum_{\alpha,\bar{\alpha}} \exp \left( \psi(k\tilde{P}_2^{\alpha \beta}) - \psi(k) \right) \tilde{z}_{i+1}^{\alpha \beta}} \tag{27}
\]

Proof. See proof in Appendix B.

Similarly to Corollary 1, the cost of privacy in (27) depends on the default transition probabilities \( P_2^{\alpha \beta} \), parameter \( k \) of the Dirichlet mechanism, penalty parameter \( \gamma \), and the value function represented as the desirability function \( \tilde{z}_{i+1}^{\alpha \beta} \). Furthermore, we note that the cost of privacy for the Digamma function equivalent, represented by (27), is expected to be more than the one calculated with the Taylor approximation in (22). Note that this because the Digamma function equivalent provides an exact equivalent for \( \mathbb{E}_{P_2^{\alpha \beta}}[\log \tilde{P}_2^{\alpha \beta}] \) resulting in more privacy, while the Taylor approximation upper bounds the actual value of this term.

V. AVERAGE VALUE APPROACH

While the stochastic formulations described in Section IV provides useful analytical insights in the effects of differential privacy on the optimal control policy, these formulations are limited by the need to specify either statistical moments or invoke the notion of the digamma function, which requires differentiability. In practice, however, these assumptions can be restrictive and require the knowledge of ground-truth default transition probabilities. Therefore, we use the average value approach (57), (58) that does not require knowing actual default transition probabilities and can establish privacy at the aggregation level. Thus, in the average value approach, we generate samples to obtain a set of \( N \) differentially private default transition probabilities \( \{\tilde{P}_j^{\alpha \beta}\}_{j \in N} \), where \( j \) is a sample index and each sample \( \tilde{P}_j^{\alpha \beta} \) is a realization of differentially private default transition probabilities \( P_2^{\alpha \beta} \). These samples are generated via the Dirichlet mechanism with default transition probabilities \( \tilde{P}_j^{\alpha \beta} \) as input to the mechanism and samples as the output. Using the generated samples, we then solve the optimization problem in (1) and obtain the expected differential private optimal control policy by averaging the optimal solutions obtained for each sample. That is, as given by Theorem 1 for each sample we obtain the following optimal solution from (1):

\[
\tilde{P}_t^{\alpha \beta} = \frac{\sum_{\alpha,\bar{\alpha}} \tilde{P}_j^{\alpha \beta} \tilde{z}_{i+1}^{\alpha \beta}}{\sum_{\alpha,\bar{\alpha}} \tilde{P}_j^{\alpha \beta}}, \tag{28}
\]

The optimal control policy in (28) is differentially private since we feed the LS-MDP with privatized default transition probabilities \( \tilde{P}_j^{\alpha \beta} \) and the property that DP is immune to post-processing (57), see Definition III.5. The expected differential private optimal control policy is calculated as:

\[
\mathbb{E}_{P_2^{\alpha \beta}}[\tilde{P}_t^{\alpha \beta}] \approx \frac{1}{N} \sum_{j=1}^{N} \tilde{P}_j^{\alpha \beta} \tag{29}
\]

The expected policy obtained in (29) requires very large number of samples \( N \) to obtain accurate results which could be time consuming for MDP problems. Therefore, we derive an expression for the expected differential private optimal control policy and propose:

**Proposition 3.** Given a differential private optimal control policy in (28) for each sample, we derive its expected value as:

\[
\mathbb{E}_{P_2^{\alpha \beta}}[\tilde{P}_t^{\alpha \beta}] = \frac{\sum_{\alpha,\bar{\alpha}} \tilde{P}_j^{\alpha \beta} \tilde{z}_{i+1}^{\alpha \beta}}{\sum_{\alpha,\bar{\alpha}} \tilde{P}_j^{\alpha \beta} \tilde{z}_{i+1}^{\alpha \beta}} \approx \frac{\tilde{P}_j^{\alpha \beta} \tilde{z}_{i+1}^{\alpha \beta}}{\sum_{\alpha,\bar{\alpha}} \tilde{P}_j^{\alpha \beta} \tilde{z}_{i+1}^{\alpha \beta}} \tag{30}
\]

Proof. See proof in Appendix C.

Accordingly, we demonstrate the expected cost of privacy for the average value approach as follows:

**Corollary 3.** Given a differential private optimal control policy in (28), the expected cost of privacy for the average
value approach is calculated as:
\[
\Delta C \approx \gamma \sum_{\alpha \in A} \left[ (z_{t+1}^{\alpha})^2 \mathcal{P}^{\alpha \beta} (1 - \mathcal{P}^{\alpha \beta}) (k + 1) \sum_{\alpha \in A} (\mathcal{P}^{\alpha \beta} z_{t+1}^{\alpha})^2 \right]
\]
\[
2\mathcal{P}^{\alpha \beta} z_{t+1}^{\alpha} \left\{ (z_{t+1}^{\alpha})^2 \mathcal{P}^{\alpha \beta} (1 - \mathcal{P}^{\alpha \beta}) - \sum_{\nu \neq \alpha \in A} (z_{t+1}^{\nu})^2 \mathcal{P}^{\alpha \beta} \mathcal{P}^{\nu \beta} \right\}
\]
\[
- \left( k + 1 \right) \sum_{\alpha \in A} (\mathcal{P}^{\alpha \beta} z_{t+1}^{\alpha})^3 \left\{ \sum_{\alpha \in A} (z_{t+1}^{\alpha})^2 \mathcal{P}^{\alpha \beta} (1 - \mathcal{P}^{\alpha \beta}) - \sum_{\alpha \in A, \nu \neq \alpha \in A} (z_{t+1}^{\nu})^2 \mathcal{P}^{\alpha \beta} \mathcal{P}^{\nu \beta} \right\}
\]
\[
= \frac{(k + 1)}{4} \sum_{\alpha \in A} \left( \mathcal{P}^{\alpha \beta} z_{t+1}^{\alpha} \right)^4
\]
\[
- \mathbb{E}_{\mathcal{P}^{\alpha \beta}} \left[ \tilde{P}_t^{\alpha \beta} \right] \left( \log \mathcal{P}^{\alpha \beta} - z_{t+1}^{\alpha} \right) + \gamma \log \sum_{\alpha \in A} \mathcal{P}^{\alpha \beta} z_{t+1}^{\alpha}
\]
where \( \mathbb{E}_{\mathcal{P}^{\alpha \beta}} \left[ \tilde{P}_t^{\alpha \beta} \right] \) is given by Eq. (30).

\[\Box\]

VI. CASE STUDY

The case study uses electricity power consumption data and available control means (e.g. HVAC, lightning, elevators) from one campus building of New York University (NYU) located in Manhattan, NY [59]. This data spans across 2018 and includes the participation in several DR events of the commercial system relief program (CSRP) operated by Consolidated Edison, the local electric power utility. Using this data, we generate an ensemble of 100 synthetic buildings, where energy consumption data is randomized by adding a random Gaussian noise to the original data such that the resulting synthetic consumption remains within \( \pm 10\% \) of the original values. The aggregated power consumption data for this ensemble of 100 buildings is displayed in Fig. 3 (a) and is used to construct the MP with 20 Markovian states for the summer season (from mid-June to late-September) when all DR events in this region occurred in 2018. The resulting default transition probabilities \((\mathcal{P}^{\alpha \beta})\) are shown in Fig. 3 (b).

For simulating a realistic DR event, we isolate the period from 11:00 AM to 3:00 PM on July 2nd, 2018 when such an event happened in Manhattan, NY [60] and NYU buildings were called upon to provide DR. We also price electricity consumption during the DR event using real-world tariff from Consolidated Edison, [60], and set the discomfort penalty paid to individual DR participants \(\gamma = \$15\).

1) Effect of privatization: We first illustrate the effect of privatization with the Dirichlet distribution on the the default transition probabilities \((\mathcal{P}^{\alpha \beta})\) given in Fig. 3 (b). To this end, we take row vector 18 from the default transition probability matrix in Fig. 3 (b), which corresponds to the state with power consumption of 32.1 MW, as vector \(\tilde{e}^\beta\) and obtain its adjacent vector \(\tilde{e}^{\beta}_i\) by using Definition III.1 and setting \(h = 0.03\). This means we subtract 0.015 from one of the entries (entry 18 in our case) in vector \(\tilde{e}^\beta\) and add 0.015 to one of the remaining entries (entry 19 in our case). We set \(k = 50\) and \(k = 200\) to perform 1000 independent executions of the Dirichlet mechanism for non-zero elements of both \(\tilde{e}^\beta\) and \(\tilde{e}^{\beta}_i\) and plot the output private default transition probabilities in Fig. 4. We note that, if \(k = 50\) and the maximum probability of privacy failure is \(\delta = 0.05\), the mechanism is (0.88, 0.05)-differentially private as per Theorem 2 and it is impossible to distinguished between the outcomes, especially with a high probability, of \(\tilde{e}^\beta\) or \(\tilde{e}^{\beta}_i\). However, if \(\delta = 0.05\) for \(k = 200\), we found the mechanism is (5.45, 0.05)-differentially private, i.e. less private then in the case with \(k = 50\). In this case, the outcomes of the mechanism for \(\tilde{e}^\beta\) and \(\tilde{e}^{\beta}_i\) are relatively distinguishable (see Fig. 4). Thus, we can achieve stronger privacy guarantees for lower values of \(k\).

2) Effect on the optimal control policy: Next, we implement the stochastic approach to obtain analytical differentially private optimal control policies. Fig. 5 compares the resulting policies for the two adjacent input vectors \(\tilde{e}^\beta\) and \(\tilde{e}^{\beta}_i\) with \(h = 0.03\). We plot the optimal control policy for the state \(\beta = 18\) at time 12:45 PM and show optimal transitions from \(\beta\) to all other non-zero states \(\alpha\). We note that the two adjacent inputs provide similar control policies (as outputs) for both the Taylor approximation and Digamma equivalent methods. In fact, the \(L_1\)-norm distance between the two output vectors in Fig. 5 is 0.0493 when \(k = 50\) and 0.0482 when \(k = 200\) for the Taylor approximation, and 0.0494 when \(k = 50\) and 0.0482 when \(k = 200\) for the Digamma equivalent. Similarly, based on private default transition probabilities \((\tilde{\mathcal{P}}^{\alpha \beta})\) in Fig. 4 we implement the average value approach to obtain private optimal control policies. The results of 1000 independent runs of the LS-MDP method for 1000 independent samples of DP...
DP Optimal Control Policy for Stochastic Approach ($h=0.03$ for the average value method. The plots are the optimal transition probabilities from state $\beta=18$ to states 17, 18 and 19 at 12:45 pm.

![Figure 5. Comparison of the DP control policies for adjacent vectors ($h=0.03$) for (a) Taylor and (b) Digamma methods from state $\beta=18$ to states 17, 18 and 19 at 12:45 pm.](image)

Figure 6. Comparison of the DP optimal control policies for adjacent vectors ($h=0.03$) for the average value method. The plots are the optimal transition probabilities from state $\beta=18$ to states 17, 18 and 19 at 12:45 pm. Each red and blue dot in (a) and (b) represents an independent run for inputs $\eta$ and $\zeta$ for $k=50$ and $k=200$, respectively, and the shaded triangle defines the output space of the Dirichlet mechanism exhibiting integrality of a unit simplex. The red and blue dots in (c) and (d) represent the expected optimal policy for both inputs $\eta$ and $\zeta$ for $k=50$ and $k=200$, respectively.

![Figure 6. Comparison of the DP optimal control policies for adjacent vectors ($h=0.03$) for the average value method. The plots are the optimal transition probabilities from state $\beta=18$ to states 17, 18 and 19 at 12:45 pm. Each red and blue dot in (a) and (b) represents an independent run for inputs $\eta$ and $\zeta$ for $k=50$ and $k=200$, respectively, and the shaded triangle defines the output space of the Dirichlet mechanism exhibiting integrality of a unit simplex. The red and blue dots in (c) and (d) represent the expected optimal policy for both inputs $\eta$ and $\zeta$ for $k=50$ and $k=200$, respectively.](image)

default transition probabilities are shown in Fig. 6 (a) for $K=50$ and Fig. 6 (b) for $k=200$, where red dots correspond to input vector $\eta$ and blue dots to $\zeta$. We computed the expected DP optimal policy for the average value approach and illustrated it in Fig. 6 (c) for $k=50$ and Fig. 6 (d) for $k=200$. The $L_1$-norm distance between the two outputs for $k=50$ is 0.0479 and for $k=200$ is 0.0476.

3) Effect of privacy on a DR event: Using both the stochastic and average value approaches, we obtain the optimal control policies for one DR event (11:00 AM to 3:00 PM on July 2nd, 2018) in which the ensemble of 100 NYU buildings participated, and present our results in Fig. 7. Note that in each case we implement the DR protocol 30 minutes before the DR event to mirror the real-life conditions and without perfect foresight of the future. First, observe that all MDP-based policies extract more capacity from the ensemble than the ad-hoc current practice. Second, the DR curtailments are slightly decreased for privacy-aware solutions as expected due to the introduction of the Dirichlet noise. We also observe that both the Digamma and Taylor approaches lead to similar DR capacity extracted during the DR event with the least deviation from the non-privatized response obtained with the standard MDP optimization. On the other hand, the average value approach leads to the extraction of $\sim 15\%$ less capacity at the peak of the DR event than the analytical policies obtained with the Digamma and Taylor approaches.

This performance difference between the stochastic and average value approaches is also observed in their privacy guarantees and costs, where generally more private policies yield less DR capacity extracted from the ensemble and greater privacy costs. We present the summary of privacy guarantees and privacy costs in Fig. 8 for different values of parameter $k$. Thus, among the analytical policies, the Digamma method is slightly more expensive than the Taylor approximation, but provides more accurate solutions. On the other hand, the average value method is more expensive as compared to both the Taylor and Digamma methods. In general, privacy decreases for all methods with a decrease in the value of $k$, indicating a trade-off between solution privacy and cost as a function of the added Dirichlet noise.

VII. CONCLUSION

This paper extended existing MDP formulations for dispatching ensembles of electrical loads to incorporate the notion of differential privacy that safeguards energy consumption data of DR participants. This is achieved by leveraging the Dirichlet mechanism, which preserves unit simplex (i.e., the sum of transition probabilities for a given state is equal to one), and its coupling with the two variances of the stochastic MDP formulations. The first variant – stochastic approach – privatizes the optimization routine of the aggregator, but involves more data requirements. The second variant – average value approach – reduces data constraints, but ensures differential privacy using its post-processing property. The case study is based on real-world demand response data and demonstrates that both the stochastic and average value approaches provide
privacy-cognizant solutions at the expense of a slight increase in operating costs, reflected in the cost of privacy, but provide robust privacy guarantees and allow for trading-off between solution quality and privacy risks.

REFERENCES

[1] M. H. Albadi and E. F. El-Saadany, “Demand response in electricity markets: An overview,” in 2007 IEEE Pow. En. Soc. Gen. Me., pp. 1–5.

[2] Energize Connecticut. [Online]. Available: https://www.energizect.com/your-home/solutions-list/demand-response

[3] Clean Power Alliance. [Online]. Available: https://cleanpoweralliance.org/wp-content/uploads/2020/02/Residential-Small-Business-Customers-FAQs.pdf

[4] Energy Information Administration. “Electric Power Annual 2019.”

[5] S. Koch, J. L. Mathieu, and D. S. Callaway, “Modeling and control of aggregated heterogeneous thermostatically controlled loads for ancillary services,” in Proc. Power Systems Computation Conf, 2011.

[6] A. Bušić and S. Meyn, “Ordinary differential equation methods for markov decision processes and application to Kullback–Leibler control cost,” SIAM Jour. on Cont. and Opt., vol. 56, no. 1, pp. 343–366, 2018.

[7] S. P. Meyn et al., “Ancillary service to the grid using intelligent deferrable loads,” IEEE Tran. on Aut. Cont., vol. 60, no. 11, pp. 2847–2862, 2015.

[8] S. Meyn et al., “Ancillary service from deferrable loads: The case for intelligent pool pumps in Florida,” in IEEE Conf. on Dec. and Cont., 2013, pp. 6946–6956.

[9] M. Chertkov, V. Y. Chernyak, and D. Deka, “Ensemble control of cycling energy loads: Markov decision approach,” in Energy Markets and Responsive Grids: Modeling, Cont., and Opt., 2018, pp. 363–382.

[10] M. Chertkov, D. Deka, and Y. Dvorkin, “Optimal ensemble control of loads in distribution grids with network constraints,” 2018 Power Systems Computation Conference (PSCC), pp. 1–7, June 2018.

[11] A. Hassan et al., “Optimal load ensemble control in chance-constrained optimal power flow,” IEEE Tran. on Smart Grid, vol. 10, no. 5, pp. 5186–5195, 2019.

[12] R. Pop et al., “A markov process approach to ensemble control of smart buildings,” in 2019 IEEE Milan PowerTech, 2019, pp. 1–6.

[13] E. Benenati et al., “A tractable formulation for multi-period linearized optimal power flow in presence of thermostatically controlled loads,” in IEEE Conf. on Dec. and Cont. (CDC), 2019, pp. 4189–4194.

[14] K. Turitsyn et al., “Smart finite state devices: A modeling framework for demand response technologies,” in IEEE Conf. on Dec. and Cont., 2011, pp. 7–14.

[15] A. Hassan, R. Mieth, D. Deka, and Y. Dvorkin, “Stochastic and distributionally robust load ensemble control,” IEEE Transactions on Power Systems, vol. 35, no. 6, pp. 6478–6488, 2020.

[16] A. Hassan et al., “Data-driven learning and load ensemble control,” Electric Power Systems Research, vol. 189, p. 106780, 2020.

[17] A. Molina-Markham, P. Shenoy, K. Fu, E. Cechet, and D. Irwin, “Private memoirs of a smart meter,” ser. BuildSys ’10. New York, NY, USA: Association for Computing Machinery, 2010, p. 61–66.

[18] A. Gopstein et al., “NIST framework and roadmap for smart grid interoperability standards, release 4.0,” 07 2020.

[19] A. Narayanan and V. Shmatikov, “Robust de-anonymization of large sparse datasets,” in 2008 IEEE Sym. on Sec. and Pri., pp. 111–125.

[20] F. D. Garcia and B. Jacobs, “Privacy-friendly energy-metering via homomorphic encryption,” in Security and Trust Management. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 226–238.

[21] C. I. Fan, S.-Y. Huang, and Y.-L. Lai, “Privacy-protected data aggregation scheme against internal attackers in smart grid,” IEEE Transactions on Industrial Informatics, vol. 10, no. 1, pp. 666–675, 2014.

[22] F. Li, B. Luo, and P. Liu, “Secure information aggregation for smart grids using homomorphic encryption,” in 2010 First IEEE International Conference on Smart Grid Communications, 2010, pp. 327–332.

[23] Y. Ding, B. Wang, Y. Wang, K. Zhang, and H. Wang, “Secure metering data aggregation with batch verification in industrial smart grid,” IEEE Transactions on Industrial Informatics, vol. 16, no. 10, pp. 6607–6616, 2020.

[24] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, “Eppa: An efficient and privacy-preserving aggregation scheme for secure smart grid communications,” IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 11, pp. 1621–1631, 2012.

[25] C. Efthymiou and G. Kalogridis, “Smart grid privacy via anonymization of smart metering data,” in 2010 First IEEE International Conference on Smart Grid Communications, 2010, pp. 238–243.

[26] S. Afrin and S. Mishra, “An anonymized authentication framework for smart metering data privacy,” in 2016 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2016, pp. 1–5.

[27] K. Okada and H. Nishi, “Big data anonymization method for demand response technologies,” in 2019 IEEE Milan PowerTech, 2019, pp. 881–886.

[28] J. A. Calandrino, A. Kilzer, A. Narayanan, E. W. Felten, and V. Shmatikov, “You might also like: privacy risks of collaborative filtering,” in 2011 IEEE Symposium on Security and Privacy, 2011, pp. 231–246.

[29] I. Dinur and K. Nissim, “Revealing information while preserving privacy,” 01 2003, pp. 202–210.

[30] M. Jawurek, M. John, and K. Rieck, “Smart metering de-pseudonymization,” in Proceedings of the 27th Annual Computer Security Applications Conference, ser. ACSAC ’11. New York, NY, USA: Association for Computing Machinery, 2011, p. 227–236.

[31] K. Liu, C. Giannella, and H. Kargupta, “A survey of attack techniques on privacy-preserving data perturbation methods,” Boston, MA: Springer US, 2008, pp. 359–381. [Online]. Available: https://doi.org/10.1007/978-0-387-70992-5_15

[32] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, “Random-data perturbation techniques and privacy-preserving data mining,” Knowledge and Information Systems, vol. 7, pp. 387–414, 2004.

[33] C. Dwork, “Differential privacy: A survey of results,” in Theory and Applications of Models of Computation,” TAMC, ser. Lecture Notes in Computer Science, vol. 4978. Springer Verlag, April 2008, pp. 1–19.

[34] C. Dwork and A. Roth, “The algorithmic foundations of differential privacy,” Foundations and Trend® in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211–407, 2014.

[35] U.S. Census Bureau, “Statistical Safeguards.” [Online]. Available: https://www.census.gov/about/policies/privacy/statistical_safeguards.html

[36] Apple. [Online]. Available: https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf

[37] K. Kenthapadi, T. Tran, M. Dietz, and I. Koepp., “Privacy-preserving analytics and reporting at LinkedIn,” 2019.

[38] C. Nayak, “New privacy-protected Facebook data for independent research on social media’s impact on democracy,” 2020.

[39] Microsoft. [Online]. Available: https://www.microsoft.com/en-us/ai/ai-lab-differential-privacy

[40] M. Guevara et al., “Expanding our Differential Privacy Library,” 2020. [Online]. Available: https://opensource.googleblog.com/2020/06/expanding-our-differential-privacy.html

[41] K. Chaudhuri et al., “Differentially private empirical risk minimization,” Jour. of Mach. Lear. Res., vol. 12, pp. 1069–1109, 2011.

[42] P. Gohari, B. Wu, C. Hawkins, M. Hale, and U. Topcu, “Differential privacy on the unit simplex via the dirichlet mechanism,” IEEE Tran. on Information Forensics and Security, vol. 16, pp. 2326–2340, 2021.

[43] P. Gohari, M. Hale, and U. Topcu, “Privacy-preserving policy synthesis in markov decision processes,” in 2020 59th IEEE Conf. on Dec. and Cont. (CDC), 2020, pp. 6266–6271.

[44] E. Todorov, “Linearly-solvable markov decision problems,” in Advances in Neural Information Processing Systems 19, B. Schölkopf, J. C. Platt, and T. Hoffman, Eds. MIT Press, 2007, pp. 1369–1376.

[45] J. Li, K. Zhang, K. Alharbi, X. Lin, N. Zhang, and X. S. Shen, “Differentially private smart metering with fault tolerance and range-based filtering,” IEEE Transactions on Smart Grid, vol. 8, no. 5, pp. 2483–2493, 2017.
Exponentiating Eq. (36) leads to the following form:

\[
\alpha^\beta_t = \exp\left(\frac{U^\beta_t}{\gamma} \right) \sum_{z} P^\alpha z^\alpha_{t+1}.
\]  

(37)

Since the value of \(Z^{\alpha \beta}\) varies for Theorems 1 and 3, we derive theorem-specific results for each case below:

For Theorem 1 We have \(Z^{\alpha \beta} = \tilde{P}^{\alpha \beta}\), and using this value of \(Z^{\alpha \beta}\) in (35) concludes the proof for Theorem 1.

For Theorem 3 We have \(Z^{\alpha \beta} = \mathbb{E}_{\tilde{P}^{\alpha \beta}}[\log \tilde{P}^{\alpha \beta}]\), and using this value of \(Z^{\alpha \beta}\) in (35) concludes the proof for Theorem 3.

APPENDIX A

PROOFS OF THEOREMS 1 AND 3

We follow the same procedure to prove Theorems 1 and 3 and denote theorem-specific terms as \(Z^{\alpha \beta}\), where \(Z^{\alpha \beta}\) for each theorem is derived at the end of this appendix. We write the following Bellman equation \(\forall t \in \mathbb{N}\) and \(\forall \alpha \in A\):

\[
\frac{\alpha^\beta_t}{\gamma} = \min_{P \in \mathcal{P}} \left( -U^\beta_t + \mathbb{E}_{P^{\alpha \beta}}[\gamma \log P^\alpha - \gamma \log Z^{\alpha \beta} + \varphi^\alpha_{t+1}] \right),
\]  

(32)

where \(\varphi^\alpha_{t+1}\) is the value function of the load ensemble at present state \(\beta\) at time \(t\) and \(\varphi^\alpha_{t+1}\) is the value function at next state \(\alpha\) at time \(t+1\). By introducing desirability function \(z^\beta_t = \exp(\frac{-\alpha^\beta_t}{\gamma})\) in (32) we obtain:

\[
-\log z^\beta_t = \min_{P \in \mathcal{P}} \left( -U^\beta_t + \mathbb{E}_{P^{\alpha \beta}}[\log P^\alpha - \gamma \log z^\alpha_{t+1}] \right).
\]  

(33)

After normalizing as \(G^\beta_t(z) = \sum_{\alpha \in A} z^\alpha z^\alpha_{t+1}\), (33) is recast as:

\[
-\log(z^\beta_t) = -\frac{U^\beta_t}{\gamma} + \min_{P \in \mathcal{P}} KL[P^\alpha_t | z^\alpha_t] - \log G^\beta_t(z).
\]  

(34)

By equating the two distributions in the KL divergence, i.e. setting \(KL[\cdot | \cdot] = 0\), the optimal policy follows as:

\[
P^\alpha_t = \frac{z^\alpha_{t+1}}{G^\beta_t(z)} = \frac{z^\alpha_{t+1}}{\sum_{\alpha \in A} z^\alpha_{t+1}}.
\]  

(35)

Then, this optimal policy converts (34) to:

\[
- \log(z^\beta_t) = -\left\{ \frac{U^\beta_t}{\gamma} + \log \left( \sum_{\alpha} z^\alpha z^\alpha_{t+1} \right) \right\}.
\]  

(36)

APPENDIX B

PROOF OF COROLLARIES 1-2

Let \(\varphi^\beta_t\) represents the value function for the non-private LS-MDP and \(\varphi^\beta_t\) for the DP LS-MDP. We start with the non-private LS-MDP and \(\forall \beta \in A, t \in T\):

\[
\varphi^\beta_t = -U^\beta_t + \sum_{\alpha \in A} P^\alpha_t (\gamma \log P^\alpha_t - \gamma \log \tilde{P}^\alpha_t + \varphi^\alpha_{t+1}).
\]  

(38)

By using (2) from Theorem 1 in (38), we obtain:

\[
\varphi^\beta_t = -U^\beta_t - \gamma \log \sum_{\alpha \in A} \tilde{P}^\alpha z^\alpha_{t+1}.
\]  

(39)

Similarly, for the DP LS-MDP and \(\forall \beta \in A, t \in T\):

\[
\varphi^\beta_t = -U^\beta_t - \gamma \log \sum_{\alpha \in A} P^\alpha_t (\gamma \log P^\alpha_t - \gamma \log \tilde{P}^\alpha_t + \varphi^\alpha_{t+1}).
\]  

(40)

By using (13) from Theorem 3 in (40), we obtain:

\[
\varphi^\beta_t = -U^\beta_t - \gamma \log \sum_{\alpha \in A} P^\alpha_t (\gamma \log \tilde{P}^\alpha_t) - \gamma \log \sum_{\alpha \in A} \tilde{P}^\alpha_t \log \tilde{P}^\alpha_t - \gamma \log \sum_{\alpha \in A} \exp(\mathbb{E}_{\tilde{P}^\alpha_t} [\log \tilde{P}^\alpha_t]) \tilde{z}^\alpha_{t+1}.
\]  

(41)

We define the cost of privacy as the difference between the value functions of the non-private and DP LS-MDPs:

\[
\Delta C = \Delta \varphi^\beta_t = \varphi^\beta_t - \varphi^\beta_t = \gamma \sum_{\alpha \in A} \tilde{P}^\alpha_t \mathbb{E}_{\tilde{P}^\alpha_t} [\log \tilde{P}^\alpha_t] - \gamma \sum_{\alpha \in A} P^\alpha_t \log \tilde{P}^\alpha_t - \gamma \log \sum_{\alpha \in A} \exp(\mathbb{E}_{\tilde{P}^\alpha_t} [\log \tilde{P}^\alpha_t]) \tilde{z}^\alpha_{t+1} + \gamma \log \sum_{\alpha \in A} \tilde{P}^\alpha_t \tilde{z}^\alpha_{t+1}.
\]  

(42)

For Corollary 1 Using (20) in (42), we conclude the proof.

For Corollary 2 Using (25) in (42), we conclude the proof.

APPENDIX C

PROOF OF PROPOSITION 3

Recall the multi-variate Taylor expansion:

\[
\mathbb{E}[X|Y] \approx \mathbb{E}[X] - \frac{\text{Cov}[X,Y]}{\mathbb{E}[Y]^2} + \frac{\mathbb{E}[X\text{Var}[Y]]}{\mathbb{E}[Y]^3}
\]  

(43)
where Cov[X, Y] represents the covariance between X and Y. Using (43), we proceed as:

\[
\mathbb{E}_{P^\alpha_\beta} \left[ \tilde{P}_{t+1}^{\alpha_\beta} \right] \approx \mathbb{E}_{P^\alpha_\beta} \left[ \tilde{P}_{t+1}^{\alpha_\beta} \right] = \frac{\sum_{\alpha \in A} \mathbb{E}_{P^\alpha_\beta} [P^\alpha_\beta] \tilde{z}^\alpha_{t+1}}{\sum_{\alpha \in A} \mathbb{E}_{P^\alpha_\beta} [P^\alpha_\beta] \tilde{z}^\alpha_{t+1}} - \text{Cov}[\tilde{P}^{\alpha_\beta} \tilde{z}^\alpha_{t+1}, \mathbb{E}_{P^\alpha_\beta} [P^\alpha_\beta] \tilde{z}^\alpha_{t+1}] + \frac{\sum_{\alpha \in A} \mathbb{E}_{P^\alpha_\beta} [P^\alpha_\beta] \tilde{z}^\alpha_{t+1} \mathbb{E}_{P^\alpha_\beta} [P^\alpha_\beta] \tilde{z}^\alpha_{t+1}}{\sum_{\alpha \in A} \mathbb{E}_{P^\alpha_\beta} [P^\alpha_\beta] \tilde{z}^\alpha_{t+1}}
\]

where \( \mathbb{E}_{P^\alpha_\beta} [\tilde{P}^\alpha_\beta] \) is already computed and given by (30), and we derive \( \mathbb{E}_{P^\alpha_\beta} [\tilde{P}^\alpha_\beta] \log \tilde{P}_t^{\alpha_\beta} \) as follows:

\[
\mathbb{E}_{P^\alpha_\beta} [\tilde{P}^\alpha_\beta] \log \tilde{P}_t^{\alpha_\beta} \approx \mathbb{E}_{P^\alpha_\beta} [(\tilde{P}^\alpha_\beta)^2 - \tilde{P}_t^{\alpha_\beta}] = \mathbb{E}_{P^\alpha_\beta} [(\tilde{P}^\alpha_\beta)^2] - \mathbb{E}_{P^\alpha_\beta} [\tilde{P}_t^{\alpha_\beta}]
\]

Now, we evaluate \( \text{Var}(\tilde{P}_t^{\alpha_\beta}) \) in (50) as follows:

\[
\text{Var}(\tilde{P}_t^{\alpha_\beta}) = \text{Var}\left( \frac{\sum_{\alpha \in A} \tilde{P}^\alpha_\beta \tilde{z}^\alpha_{t+1}}{\sum_{\alpha \in A} \tilde{P}^\alpha_\beta \tilde{z}^\alpha_{t+1}} \right) = \frac{\sum_{\alpha \in A} \mathbb{E}_{P^\alpha_\beta} [\tilde{P}^\alpha_\beta] \tilde{z}^\alpha_{t+1} \mathbb{E}_{P^\alpha_\beta} [\tilde{P}^\alpha_\beta] \tilde{z}^\alpha_{t+1}}{\sum_{\alpha \in A} \mathbb{E}_{P^\alpha_\beta} [\tilde{P}^\alpha_\beta] \tilde{z}^\alpha_{t+1}} \sum_{\alpha \in A} \mathbb{E}_{P^\alpha_\beta} [\tilde{P}^\alpha_\beta] \tilde{z}^\alpha_{t+1}
\]

(44) We have already computed terms \( \mathbb{E}_{P^\alpha_\beta} [\cdot], \text{Var}[\cdot], \) and \( \text{Cov}[\cdot] \) used in (51) as part of Appendix C. By using respective terms in (51), inserting (51) in (50), and then using (50) in (49), we conclude the proof.

### APPENDIX D

**PROOF OF COROLLARY 3**

We start with the non-private LS-MDP and \( \forall \beta \in A, t \in T \):

\[
\varphi_t^\beta = -U_t^\beta - \gamma \log \sum_{\alpha \in A} P^\alpha_\beta \tilde{z}^\alpha_{t+1}
\]

(47)

Similarly, for the DP LS-MDP and \( \forall \beta \in A, t \in T \):

\[
\tilde{\varphi}_t^\beta = -U_t^\beta + \gamma \sum_{\alpha \in A} \tilde{P}_t^{\alpha_\beta} (\log \tilde{P}_t^{\alpha_\beta} - \log P^\alpha_\beta) + \tilde{\varphi}^\alpha_{t+1} / \gamma
\]

(48)

Then, the expected cost of privacy is derived as:

\[
\Delta C = \mathbb{E}_{P^\alpha_\beta} [\Delta \varphi_t^\beta] = \mathbb{E}_{P^\alpha_\beta} [\tilde{\varphi}_t^\beta] - \varphi_t^\beta = \gamma \sum_{\alpha \in A} \mathbb{E}_{P^\alpha_\beta} [\tilde{P}_t^{\alpha_\beta} \log \tilde{P}_t^{\alpha_\beta}] - \gamma \sum_{\alpha \in A} \mathbb{E}_{P^\alpha_\beta} [\tilde{P}_t^{\alpha_\beta}] \log \tilde{P}_t^{\alpha_\beta} - \gamma \sum_{\alpha \in A} \mathbb{E}_{P^\alpha_\beta} [\tilde{P}_t^{\alpha_\beta}] \tilde{z}^\alpha_{t+1} + \gamma \log \sum_{\alpha \in A} \tilde{P}_t^{\alpha_\beta} \tilde{z}^\alpha_{t+1},
\]

(49)