Estimation and Control of Quality of Service in Demand Dispatch

Yue Chen, Ana Bušić, Sean Meyn

Estimation and Control of Quality of Service in Demand Dispatch

Yue Chen, Ana Bušić, and Sean Meyn

Abstract—**It is now well known that flexibility of energy consumption can be harnessed for the purposes of grid-level ancillary services. In particular, through distributed control of a collection of loads, a balancing authority regulation signal can be tracked accurately, while ensuring that the quality of service (QoS) for each load is acceptable *on average*. Subject to distributed control approaches advocated in recent research, the histogram of QoS is approximately Gaussian, and consequently each load will eventually receive poor service. Statistical techniques are developed to estimate the mean and variance of QoS as a function of the power spectral density of the regulation signal.**

It is also shown that additional local control can eliminate risk: The histogram of QoS is *truncated* through this local control, so that strict bounds on service quality are guaranteed. While there is a tradeoff between the grid-level tracking performance (capacity and accuracy) and the bounds imposed on QoS, it is found that the loss of capacity is minor in typical cases.

Index Terms—**Demand dispatch, demand response, ancillary services, mean field control.**

I. INTRODUCTION

The power grid requires regulation to ensure that supply matches demand. Regulation is required by each *balancing authority* (BA) on multiple time-scales, corresponding to the time-scales of volatility of both supply and demand for power. Resources that supply these regulation services are collectively known as *ancillary services*. FERC orders 755 and 745 are examples of recent federal policy intended to provide incentives for the provision of these services.

A number of papers have explored the potential for extracting ancillary service through the inherent flexibility of loads. Examples of loads with sufficient flexibility to provide service to the grid are aluminum manufacturing, plug-in electric vehicles, heating and ventilation (HVAC), and water pumping for irrigation [1], [2], [3], [4], [5], [19]. Even with direct load control, there may be delay and dynamics, so harnessing ancillary services from flexible loads amounts to a control problem: The BA wishes to design some signal to be broadcast to loads, so that deviation in power consumption tracks a reference signal. It has been argued that a randomized control architecture at each load can simplify this control problem [4], [6], [7]. Randomized algorithms also appear in recent grid-solutions from industry, such as the mechanism for load shedding for voltage control in one of Schneider's patents [18].

This research is supported by the NSF grants CPS-1259040 and EPCN-1609131, the French National Research Agency grant ANR-12-MONU-0019, and US-Israel BSF Grant 2011506.

Y.C. and S.M. are with the Department of Electrical and Computer Engg. at the University of Florida, Gainesville. A.B. is with Inria and the Computer Science Dept. of École Normale Supérieure, Paris, France.

Fig. 1 shows a schematic of the *Demand Dispatch* control architecture adopted in [4], in which each load operates according to a randomized policy based on its internal state, and a common control signal ζ. Theoretical results and examples in this prior work demonstrate that local randomized policies can be designed to simplify control of the aggregate to provide reliable ancillary services.

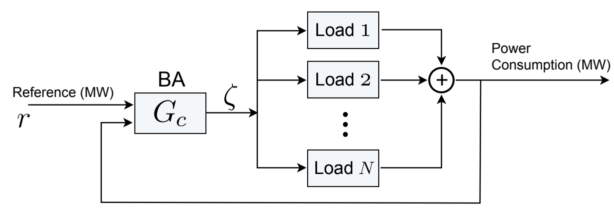


Fig. 1. Control architecture for Demand Dispatch

Absent in prior work is any detailed analysis of risk for an individual load (with the exception of the preliminary work [8] on which the present paper is built). In the setting of [4] it can be argued that the quality of service (QoS) for each load is acceptable only *on average*. Strict bounds on QoS are addressed in [9] for a deterministic model. The *service curves* considered there are of similar flavor to the QoS metrics used in the present work.

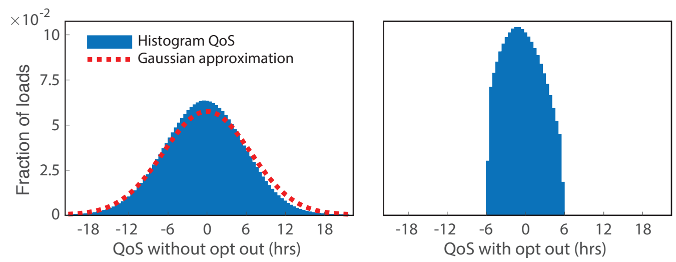


Fig. 2. Discounted QoS (10) with and without local opt-out control.

The main contributions of this paper are summarized as follows:

(i) QoS metrics are proposed in Section II. Techniques to approximate their second order statistics are developed in Section III.

(ii) An approach is proposed to restrict QoS to pre-specified bounds: A load will opt-out of service to the grid temporarily, whenever its QoS is about to exit a violate these bounds. In addition, techniques are introduced to condition the grid-level reference signal so that opt-out is rare.

(iii) These conclusions are illustrated in numerical results surveyed in Section IV.

Fig. 2 shows histograms of QoS based on simulation experiments described in Section IV. The plot on the left hand side shows that a Gaussian approximation is a good fit with empirical results when there is no local opt-out control. The figure on the right shows how the histogram is truncated when opt-out control is in place. It is found that opt-out control has little impact on grid-level performance, provided QoS bounds are not overly restrictive.

The methodology used in this work employs a linear representation of the controlled Markov chain that represents the state of an individual load [10]. Similar representations are used in [11], [12] to obtain nonlinear filters for state estimation.

The numerical examples in this paper all concern a homogeneous collection of a particular load: residential pool pumps. The homogeneity assumption simplifies the analysis, but is by no means essential. This is imposed only to obtain an approximation for the statistics of the signal broadcast to the loads. The extension to certain other types of loads, such as refrigeration or water heating is straightforward [12], [16]. For other loads, estimation of QoS remains a frontier. For example, air-conditioning appears to be more challenging because load is highly time-varying, so that a stationary setting for analysis may not be appropriate.

We begin with a brief survey of a portion of results from [4], and a precise definition of QoS for a load.

II. Randomized control and mean-field models

A. Randomized control

The system architecture considered in this paper is illustrated in Fig. 1, based on the following components:

(i) There are N homogeneous loads that receive a common scalar command signal from the balancing authority (BA), denoted $\boldsymbol{\zeta} = \{\zeta_t\}$ in the figure.

(ii) Each load evolves as a controlled Markov chain on the finite state space $\mathsf{X} = \{x^1, \dots, x^d\}$. Its transition probability is determined by its own state, and the BA signal $\boldsymbol{\zeta}$. The common dynamics are defined by a controlled transition matrix $\{P_\zeta : \zeta \in \mathbb{R}\}$. For the ith load, there is a state process $\boldsymbol{X}^i$ whose transition probability is given by,

$$\mathsf{P}\{X^i_{t+1} = x' \mid X^i_r, \zeta_r : r \le t\} = P_\zeta(x, x'), \qquad (1)$$

for each $x' \in \mathsf{X}$, $X^i_t = x \in \mathsf{X}$ and $\zeta_t = \zeta \in \mathbb{R}$.

(iii) The BA has measurements of the other two scalar signals shown in the figure: The aggregate power consumption $\boldsymbol{y}$ and desired deviation power consumption $\boldsymbol{r}$.

Part (i) is assumed to simplify exposition. Control of a heterogeneous population is treated in [16]. Statistics of QoS for an individual can then be estimated and/or controlled using the methods introduced in the present paper.

An approach to construction of $\{P_\zeta : \zeta \in \mathbb{R}\}$ was proposed in [4] based on information-theoretic arguments. The *nominal behavior* is defined as the dynamics with $\boldsymbol{\zeta} \equiv 0$. A specific construction of P_ζ is not required here, but the following assumptions are imposed in our main results.

A1: The transition matrix P_ζ is twice continuously differentiable (C^2) in a neighborhood of $\zeta = 0$, and the second derivative is Lipschitz continuous. In addition, the nominal transition matrix P_0 is irreducible and aperiodic.

The first and second order derivatives of the transition matrix at $\zeta = 0$ are denoted,

$$\mathcal{E} = \frac{d}{d\zeta} P_\zeta \Big|_{\zeta=0}, \qquad \mathcal{E}^{(2)} = \frac{d^2}{d\zeta^2} P_\zeta \Big|_{\zeta=0}. \qquad (2)$$

A2: $\zeta_t = \varepsilon \zeta^1_t$, where $0 \le \varepsilon < 1$ and $\boldsymbol{\zeta}^1 = \{\zeta^1_t : t \in \mathbb{Z}\}$ is a real-valued stationary stochastic process with zero mean. The following additional assumptions are imposed:

(i) It is bounded, $|\zeta^1_t| \le 1$ for all t with probability one. Hence $\sigma^2_\zeta = \mathsf{E}[\zeta^2_t] \le \varepsilon^2$.

(ii) Its auto-covariance satisfies, for each t,

$$|\Sigma_\zeta(t)| \le \varepsilon^2 b \rho^{|t|}, \text{ with } b < \infty, \text{ and } |\rho| < 1.$$

Assumption **A1** is ensured by design in all prior work. Assumption **A2** is a modeling assumption, imposed to study how QoS is impacted by the command signal $\boldsymbol{\zeta}$.

It is assumed that the power consumption at time t from load i is equal to some function of the state, denoted $\mathcal{U}(X^i_t)$. The normalized power consumption is denoted,

$$y^N_t = \frac{1}{N} \sum_{i=1}^N \mathcal{U}(X^i_t). \qquad (3)$$

Under Assumption **A1**, P_0 has a unique pmf (probability mass function) π_0. The value $\overline{y}^0 := \sum_x \pi_0(x)\mathcal{U}(x)$ is interpreted as the average nominal power usage. On combining the ergodic theorem for Markov chains with the Law of Large Numbers for i.i.d. sequences we can conclude that $y^N_t \approx \overline{y}^0$ when both N and t are large, and $\boldsymbol{\zeta} \equiv 0$.

It is assumed that the signal $\boldsymbol{r}$ is also normalized so that tracking amounts to choosing the signal $\boldsymbol{\zeta}$ so that $\tilde{y}^N_t \approx r_t$ for all t, where $\tilde{y}^N_t = y^N_t - \overline{y}^0$ is the deviation from nominal behavior. For example, we might use error feedback of the form,

$$\zeta_t = G_c e_t, \qquad e_t = r_t - \tilde{y}^N_t\,, \qquad (4)$$

where G_c is the control transfer function [13], [14], [4]. Adopting terminology from control engineering, $\boldsymbol{r}$ will be called the *reference signal*.

A state space model approximating the dynamics of the aggregate is obtained as the *mean-field model* defined next.

B. Mean-field model

The mean-field model is based on the empirical pmfs:

$$\mu^N_t(x) := \frac{1}{N} \sum_{i=1}^N \mathbb{I}\{X^i_t = x\}, \quad x \in \mathsf{X}. \qquad (5)$$

Each entry of the vector μ^N_t represents the fraction of loads in a particular state. Under very general conditions on the input sequence $\boldsymbol{\zeta}$, it can be shown that as $N \to \infty$, the empirical distributions converge to a solution to a nonlinear state space model with state denoted $\{\mu_t\}$, and input $\{\zeta_t\}$. It is convenient to express μ_t as a d-dimensional row vector, since then the evolution equations become

$$\mu_{t+1} = \mu_t P_{\zeta_t}, \qquad t \ge 0. \qquad (6)$$

That is, for each state $x \in \mathsf{X}$ and $t \geq 0$, the value $\mu_{t+1}(x)$ is an approximation of the empirical pmf (or histogram) (5), and is related to the previous empirical pmf via

$$\mu_{t+1}(x) = \sum_{x' \in \mathsf{X}} \mu_t(x') P_{\zeta_t}(x', x).$$

The normalized power consumption (3) converges to the value $y_t = \sum_x \mu_t(x)\mathcal{U}(x)$.

The unique equilibrium with $\boldsymbol{\zeta} \equiv 0$ is $\mu_t \equiv \pi_0$ and $y_t \equiv \overline{y}^0$. The linearization about this equilibrium is described by the linear state space model,

$$\begin{aligned} \Phi_{t+1} &= A\Phi_t + B\zeta_t \\ \gamma_t &= C\Phi_t \end{aligned} \tag{7}$$

in which $A = P_0^{\tau}$, C is a row vector of dimension $d = |\mathsf{X}|$ with $C_i = \mathcal{U}(x^i)$ for each i, and B is a d-dimensional column vector with entries $B_j = \sum_x \pi_0(x)\mathcal{E}(x, x^j)$, where the matrix $\mathcal{E}$ is defined in (2). In the state equations (7), the state Φ_t is d-dimensional, and $\Phi_t(i)$ is intended to approximate $\mu_t(x^i) - \pi_0(x^i)$ for $1 \leq i \leq d$. The output γ_t is an approximation of $\tilde{y}_t = y_t - \overline{y}^0$.

Recall that a primary goal is to design the signal $\boldsymbol{\zeta}$ so that $\tilde{y}_t^N \approx r_t$ for all t. The linear model is invaluable for design of the feedback compensator G_c appearing in (4). It is convenient to base this design on the transfer function from $\boldsymbol{\zeta}$ to $\boldsymbol{\gamma}$:

$$G_p(z) := C[Iz - A]^{-1}B \tag{8}$$

C. QoS for an individual and the population

The QoS metrics considered in this paper are specified by a function $\ell : \mathsf{X} \to \mathbb{R}$, and a stable transfer function denoted $H_{\mathcal{L}}$. For example, the function ℓ may represent temperature, cycling, or power consumption as a function of $x \in \mathsf{X}$. The QoS of the ith load at time t is denoted $\mathcal{L}_t^i$, and defined by passing $\boldsymbol{L}^i := \{L_t^i = \ell(X_t^i) : t \in \mathbb{Z}\}$ through the transfer function $H_{\mathcal{L}}$.

Two classes of transfer functions $H_{\mathcal{L}}$ are considered in numerical experiments:

(i) Summation over a finite time horizon T_f:

$$\mathcal{L}_t^i = \sum_{k=0}^{T_f} \ell(X_{t-k}^i). \tag{9}$$

(ii) Discounted sum:

$$\mathcal{L}_t^i = \sum_{k=0}^{\infty} \beta^k \ell(X_{t-k}^i)\,, \tag{10}$$

where the discount factor satisfies $\beta \in [0, 1)$.

In particular, setting $T_f = 0$ or $\beta = 0$ gives $\mathcal{L}_t^i = \ell(X_t^i)$.

Unless elsewhere specified, the function ℓ is chosen to reflect the power consumption of a load,

$$\ell(X_t^i) = \mathcal{U}(X_t^i), \tag{11}$$

and its normalized form is also considered,

$$\ell(X_t^i) = \mathcal{U}(X_t^i) - \overline{y}^0 \tag{12}$$

where $\overline{y}^0$ is defined after (3).

The average QoS at time t is denoted,

$$\bar{\mathcal{L}}_t = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}_t^i\,,$$

and the filtered signal is denoted $\boldsymbol{R} = H_{\mathcal{L}}\boldsymbol{r}$. The following result follows from the definitions:

Proposition 2.1: Suppose there is perfect tracking: $\tilde{y}_t^N = r_t$ for all t. Then, under the definition of $\ell(X_t^i)$ in (12),

$$\bar{\mathcal{L}}_t = R_t. \tag{13}$$

□

In practice we can only expect the approximation $\tilde{y}_t^N \approx r_t$, which will imply the corresponding approximation $\bar{\mathcal{L}}_t \approx R_t$. This leads to a useful heuristic notion of capacity of the aggregate of loads as a function of time:

Battery analogy: With ℓ defined in (12), the signal $\bar{\mathcal{L}}_t$ is similar to the SOC (state of charge) of a battery. In particular, a large value of $\bar{\mathcal{L}}_t$ suggests loads have a large capacity for "discharge". Since R_t approximates $\bar{\mathcal{L}}_t$ by design, the former can be used as an indicator of SOC of loads.

D. Example: Intelligent pools

The paper [4] on which the present work is based considered an application of this system architecture in which the loads are a collection of pools. The motivation for considering pools is the inherent flexibility of pool cleaning, and because the total load in a region can be very large. The maximum load is approximately 1GW in the state of Florida.

In this prior work, the state space was the finite set,

$$\mathsf{X} = \{(\kappa, j) : \kappa \in \{\oplus, \ominus\},\ j \in \{1, 2, \ldots, \mathcal{I}\}\}. \tag{14}$$

The load samples the grid signal periodically (the sampling increments are assumed to be deterministic, or i.i.d. and distributed according to a geometric distribution). At the time of the kth sample, if $X_k = (\ominus, j)$ then the load has remained off for the past j sampling times, and was turned off at sampling time $k - j$. The interpretation of $X_k = (\oplus, j)$ is symmetrical, with "$\oplus$" indicating that the load is currently consuming power.

A technique to construct the transition matrix P_ζ was introduced in [4] using an optimal-control approach, and improved in [12], [15], [16] to reduce the dimension of the state space, and extend application beyond this class of loads. It has been demonstrated through extensive simulation that tracking at the grid level is nearly perfect, provided that the reference signal is scaled and filtered [8], [4], [12]. What was left out in prior work is any detailed consideration of the quality of service to individual loads. The pool pump example is ideal for illustrating the possibility of risk for an individual load, and illustrating how this risk can be reduced or even eliminated.

In a numerical experiment conducted with 10^4 pools, each pool pump was expected to clean the pool 12 hour/day and assumed to consume 1 kW when operating. Hence the function $\mathcal{U}$ that defines (11) and (12) is the indicator function $\mathcal{U}(x) = \sum_j \mathbb{I}\{x = (\oplus, j)\}$. Fig. 3 shows the histogram of the QoS metric $\mathcal{L}_t^i$ based on (9) and (11), with T_f corresponding to 10 days.

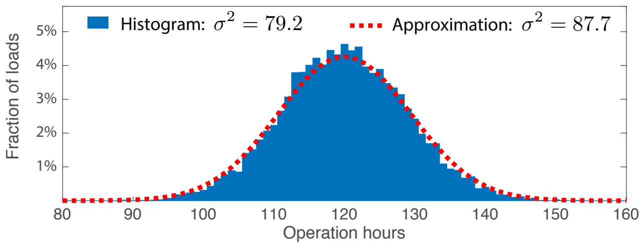


Fig. 3. Histogram of the moving-window QoS metric (9).

The histogram appears Gaussian, with a mean of approximately 120 hours — this is consistent with the time horizon used in this experiment, given the nominal 12 hour/day cleaning period. It is evident that a fraction of pools are over-cleaned or under cleaned by 24 hours or more.

An analysis of QoS is presented in the next section based on a model of an individual load in the mean field limit.

III. QoS ANALYSIS AND OPT-OUT CONTROL

In the mean-field limit, the aggregate dynamics are deterministic, following the discrete-time nonlinear control model (6). The behavior of each load remains probabilistic.

A. Mean field model for an individual load

The mean field model for *one load* is defined by replacing $\boldsymbol{\zeta}^N$ with the sequence $\boldsymbol{\zeta}$ that arises in the mean-field limit. The justification is that we have a very large number of loads, but our interest is in the statistics of an individual.

The super-script i is dropped in our analysis of a single load. Hence $\boldsymbol{X}$ denotes the controlled Markov chain whose transition probabilities are defined consistently with (1):

$$\mathsf{P}\{X_{t+1} = x' \mid X_r, \zeta_r : r \le t\} = P_{\zeta_t}(x, x'), \quad x = X_t \tag{15}$$

The construction of the mean field model (6) is based on lifting the state space from the d-element set $\mathsf{X} = \{x^1, \cdots, x^d\}$, to the d-dimensional simplex S. For the i^{th} load at time t, the element $\Gamma_t \in \mathsf{S}$ is the degenerate distribution whose mass is concentrated at x if $X_t = x$; that is, $\Gamma_t = \delta_x$.

This lifting from X to $\mathbb{R}^d$ will be applied in several ways. First, we observe that nonlinear functions of the state become linear:

$$L_t = \ell(X_t) = \langle \Gamma_t, \ell \rangle := \sum_i \Gamma_t(x^i)\ell(x^i).$$

The dynamics of the load remain random, but evolve as a linear system, similar to (6):

$$\Gamma_{t+1} = \Gamma_t G_{t+1} \tag{16}$$

in which Γ_t is interpreted as a d-dimensional row vector. The $d \times d$ matrix G_t has entries 0 or 1 only, with $\sum_{x' \in \mathsf{X}} G_t(x, x') = 1$ for all $x \in \mathsf{X}$. It is conditionally independent of $\{\Gamma_0, \cdots, \Gamma_t\}$, given ζ_t, with

$$\mathsf{E}[G_{t+1} | \Gamma_0, \cdots, \Gamma_t, \zeta_t] = P_{\zeta_t}. \tag{17}$$

The random linear system (16) can be described as a linear system driven by "white noise":

$$\Gamma_{t+1} = \Gamma_t P_{\zeta_t} + \Delta_{t+1} \tag{18}$$

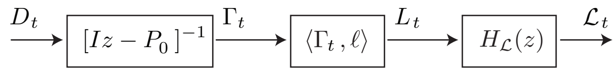


Fig. 4. Approximate linear model of QoS evolution.

where, $\{\Delta_{t+1} = \Gamma_t(G_{t+1} - P_{\zeta_t}) : t \ge 0\}$ is a martingale difference sequence.

A Taylor-series approximation of P_{ζ_t} leads to a useful approximation of (18). Recall that the first and second derivatives $\mathcal{E}$ and $\mathcal{E}^{(2)}$ were introduced in (2). The proof follows from the definitions.

Proposition 3.1: The nonlinear system (18) admits the LTI approximation,

$$\Gamma_{t+1} = \Gamma_t P_0 + D_{t+1} + O(\varepsilon^3) \tag{19}$$

in which,

$$D_{t+1} := B_t^{\intercal} \zeta_t + V_t^{\intercal} \zeta_t^2 + \Delta_{t+1}, \tag{20}$$

with $B_t^{\intercal} = \Gamma_t \mathcal{E}$ and $V_t^{\intercal} = \frac{1}{2}\Gamma_t \mathcal{E}^{(2)}$. □

The proposition implies a linear systems approximation for QoS that is illustrated in Fig. 4, and applied in the next set of results.

B. Steady-state QoS statistics

The goal of this section is to estimate the second order statistics of $\{\mathcal{L}_t\}$ for an arbitrary function ℓ and stable filter $H_{\mathcal{L}}$. These approximations are obtained for a stationary realization.

Theorem 3.2: Under assumptions **A1** and **A2**, there exists a realization $\{\zeta_t, X_t, D_t, \mathcal{L}_t : -\infty < t < \infty\}$ that is jointly stationary. □

The proof of the existence of the stationary realization $\{\zeta_t, X_t, D_t : -\infty < t < \infty\}$ follows from Proposition 2.3 of [10]. Stationarity of the joint process $\{\zeta_t, X_t, D_t, \mathcal{L}_t\}$ follows from the assumption that $\{\mathcal{L}_t\}$ is obtained from $\{\ell(X_t)\}$ through a stable filter.

The PSD for the stationary realization of the stochastic process $\boldsymbol{D}$ can be approximated by a second order Taylor series expansion: Applying Theorem 2.4 of [10], a function $\mathrm{S}_D^{(2)} \colon [-\pi, \pi] \to \mathbb{C}^{d \times d}$ can be constructed such that

$$\mathrm{S}_D(\theta) = \mathrm{S}_D^{\bullet}(\theta) + \varepsilon^2 \mathrm{S}_D^{(2)}(\theta) + o(\varepsilon^2), \quad \theta \in [-\pi, \pi], \tag{21}$$

in which $\mathrm{S}_D^{\bullet}$ is the PSD obtained with $\boldsymbol{\zeta} \equiv 0$. This is independent of θ since the sequence $\boldsymbol{D}$ is white in this case.

Approximations for second order statistics of QoS follow:

Proposition 3.3:

(i) The mean QoS admits the approximation,

$$\mathsf{E}[\mathcal{L}_t] = H_{\mathcal{L}}(1)\mathsf{E}[L_t], \quad \mathsf{E}[L_t] = \sum_x \ell(x)\pi_0(x) + O(\varepsilon^2)$$

where $H_{\mathcal{L}}(1)$ is the DC gain of the transfer function $H_{\mathcal{L}}$.

(ii) The PSD for QoS is given by,

$$\mathrm{S}_{\mathcal{L}}(\theta) = |H_{\mathcal{L}}(e^{j\theta})|^2 \mathrm{S}_L(\theta), \quad \theta \in \mathbb{R},$$

in which $\mathrm{S}_L(\theta) = K(e^{j\theta})^{\intercal} \mathrm{S}_D(\theta) K(e^{-j\theta}) + o(\varepsilon^2)$, and K is the $d \times 1$ transfer function with entries

$$K_i(z) = \sum_{j=1}^{d} [Iz - P_0]^{-1}_{i,j} \ell(x^j)$$

Proof: Part (i) follows from Proposition 2.5 of [10].

Part (ii) follows from (i), Prop. 3.1, and Fig. 4: In this block diagram, K is the transfer function from $\boldsymbol{D}$ to $\boldsymbol{L}$. □

Based on this PSD approximation we obtain approximations for the variance:

Theorem 3.4: The variance of QoS admits the following representation and approximations:

(i) The variance of QoS is the average of its PSD:

$$\mathcal{V}_{\mathcal{L}} = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{S}_{\mathcal{L}}(\theta)\, d\theta. \tag{22}$$

(ii) In the special case of the moving time-horizon (9), it admits the approximation

$$\frac{\mathcal{V}_{\mathcal{L}}}{T_f} \approx \mathrm{S}_L(0) \quad \text{if } T_f \approx \infty, \tag{23}$$

where S_L is the PSD of $\boldsymbol{L}$.

(iii) A Taylor series of the right hand side of (22) implies the approximation,

$$\mathcal{V}_{\mathcal{L}} = \mathcal{V}_{\mathcal{L}}^{\bullet} + \varepsilon^2 \mathcal{V}_{\mathcal{L}}^{(2)} + o(\varepsilon^2), \tag{24}$$

where the terms $\mathcal{V}_{\mathcal{L}}^{\bullet}$ and $\mathcal{V}_{\mathcal{L}}^{(2)}$ are obtained using the QoS variance formula in (22), and the approximation of the PSD $\mathrm{S}_{\mathcal{L}}$ is obtained using Prop. 3.3 (ii). □

Recall that the QoS histogram shown in Fig. 3 is based on the moving time-horizon (9), with T_f corresponding to 10 days. The approximation of the QoS variance was obtained using (23) and the second-order Taylor series approximation of $\mathrm{S}_L(0)$.

C. Opt-out control

An extra layer of control is required to truncate the two tails of the QoS histograms observed in experiments.

The local control considered in this paper is a simple "opt-out" mechanism, based on pre-defined upper and lower limits b_+ and b_-. A load ignores a command from the grid operator at time t if it may result in $\mathcal{L}_{t+1} \notin [b_-, b_+]$, and takes an alternative action so that $\mathcal{L}_{t+1} \in [b_-, b_+]$ with probability one. This ensures that the QoS metric of each load remains within the predefined interval for all time.

In many cases, the poor QoS revealed by the two tails of the histogram represents only a small portion of loads. Therefore, the impact from local opt-out control is insignificant at the grid level if the QoS interval $[b_-, b_+]$ is carefully chosen.

IV. Numerical Experiments

Numerical experiments were conducted on the pool pump model to illustrate the main technical conclusions. Simulation results are summarized below:

(i) Applications of Theorem 3.4 show that the approximations of QoS closely match observed QoS.

(ii) Opt-out control ensures that QoS lies within strict bounds, and tracking remains nearly perfect in most cases. In some extreme cases, capacity is reduced with the introduction of opt-out control. However, it is found through simulations that it is far less than might be predicted by the approximation in (i).

(iii) The opt-out control can be applied to multiple QoS metrics. The capacity is further reduced with the introduction of an additional QoS metric, but the reduction is found to be minor.

(iv) Approaches are proposed and tested in Section IV-E to condition the reference signal to reduce potential capacity reduction caused by opt-out local control.

A. Simulation setup

The simulation used $N = 10^4$ homogeneous Markov models: Each pool pump is operated under a 12 hours/day cleaning cycle, and consumes 1 kW during operation. The sampling time is 5 minutes in all of the experiments here. The justification is that these loads are used to track a lower frequency component $\boldsymbol{r}$ of the overall signal $\boldsymbol{r}^{\bullet}$ so that they will not endure excessive cycling.

Two QoS metrics are considered, differentiated by the function ℓ appearing in the definition $L_t^i = \ell(X_t^i)$: In the first QoS function, the normalized power consumption (12) is considered so that if $\mathcal{L}_t^i > 0$ ($\mathcal{L}_t^i < 0$) then the pool has been over-cleaned (under-cleaned).

The second QoS function is introduced to capture the on/off cycling of loads:

$$\ell^c(X_t^i, X_{t+1}^i) = \Big|\sum_j \Big(\mathbb{I}\{X_{t+1}^i = (\oplus, j)\} - \mathbb{I}\{X_t^i = (\oplus, j)\}\Big)\Big| \tag{25}$$

These two QoS metrics can be applied to many other loads, such as air conditioners, refrigerators, and water heaters [15], [16]. Consideration of the QoS metric (25) is postponed to Section IV-F.

The discounted sum (10) was used to define $\mathcal{L}_t^i$ in these experiments. The following interpretation is used to obtain insight on the choice of the discount factor in this QoS metric.

Let ξ denote a random variable that is independent of $\boldsymbol{X}^i$. Its distribution is taken to be geometric on $\mathbb{Z}_+$, with parameter β. Its mean is thus

$$\mathsf{E}[\xi] = \sum_{k=0}^{\infty} \mathsf{P}(\xi \ge k) = \sum_{k=0}^{\infty} \beta^k = \frac{1}{1-\beta}. \tag{26}$$

By independence we also have,

$$\begin{aligned}\mathsf{E}\Big[\sum_{k=0}^{\xi} \ell(X_{t-k}^i)\Big] &= \sum_{k=0}^{\infty} \mathsf{E}[\ell(X_{t-k}^i)\mathbb{I}(\xi \ge k)] \\ &= \sum_{k=0}^{\infty} \mathsf{E}[\ell(X_{t-k}^i)]\mathsf{P}(\xi \ge k) \\ &= \sum_{k=0}^{\infty} \mathsf{E}[\ell(X_{t-k}^i)]\beta^k = \mathsf{E}[\mathcal{L}_t^i]\end{aligned}$$

Hence the discounted sum (10) is similar to the moving window QoS metric (9). In the following experiments we took $\mathsf{E}[\xi] = 10$ days, which corresponds to 2880 time samples (given the 5 minute sampling period). Solving (26) gives $\beta = 1 - 1/2880$.

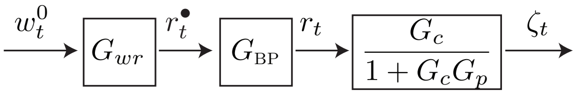


Fig. 5. The input $\boldsymbol{\zeta}$ modeled as a stationary stochastic process

B. Model for the stationary input

A linear model for the stationary input $\boldsymbol{\zeta}$ was constructed based on the block diagram shown in Fig. 5. As in the prior work [4], [8], it is assumed that the signal $\boldsymbol{r}$ is obtained by filtering the regulation signal $\boldsymbol{r}^\bullet$; the latter is modeled as filtered white noise with transfer function G_{wr}.

The BPA (Bonneville Power Authority [17]) balancing reserves, deployed in January 2015, were taken for the regulation signal $\boldsymbol{r}^\bullet$.

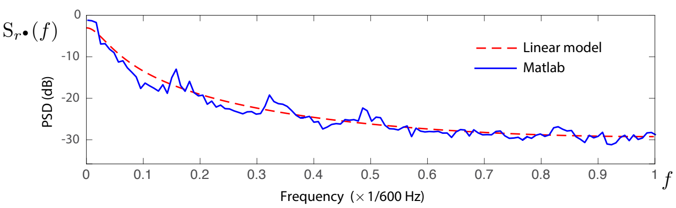


Fig. 6. BPA signal: Power spectral density and its approximation.

It is assumed that $\boldsymbol{r}^\bullet$ evolves as the ARMA (autoregressive moving average) model,

$$r_t^\bullet + a_1 r_{t-1}^\bullet + a_2 r_{t-2}^\bullet = w_t + b_1 w_{t-1} \qquad (27)$$

in which $\boldsymbol{w}$ is white noise with variance σ_w^2. The extended least squares (ELS) algorithm was used to estimate the coefficients a_1, a_2, b_1, and the variance σ_w^2 based on the BPA balancing reserves. The algorithm terminated at $[a_1, a_2, b_1]^T = [-1.16, 0.2301, -0.2489]^{\intercal}$, and $\sigma_w^2 = 4.36 \times 10^{-3}$. In the z-domain, its transfer function is expressed

$$G_{wr}(z) = \frac{1 - 0.2489z^{-1}}{1 - 1.6z^{-1} + 0.2301z^{-2}}. \qquad (28)$$

The dashed and solid lines in Fig. 6 represent the estimates of the spectrum given by $|S_{r^\bullet}(e^{j\theta})|^2 = \sigma_w^2 |G_{wr}(e^{j\theta})|^2$ and Matlab's `psd` command, respectively.

The transfer function $G_{\rm BP}$ in Fig. 5 is a filter designed to smooth the balancing reserve signal. A low pass filter was adopted with crossover frequency near the nominal period of a single load: $1/(24 \text{ hours})$ in this example. In most of the experiments reported here, $G_{\rm BP}$ is the first-order Butterworth low-pass filter with cut-off frequency $f_c = 1/(1000 \text{ minutes})$:

$$G_{\rm BP}(z) = 0.0155 \frac{1 + z^{-1}}{1 - 0.9691z^{-1}}. \qquad (29)$$

The reference signal must also be scaled so that the desired goal $\tilde{y}_t^N \approx r_t$ is feasible for all t. Denote by $\boldsymbol{r}^1$ the signal obtained using the largest scaling, while also ensuring that this signal can be tracked by the collection of pools. This was obtained through trial and error.

In the tracking plots, such as Fig. 11, the signals $\tilde{y}_t^N$ and r_t are re-scaled to units in MWs.

The construction of a stationary model for $\boldsymbol{\zeta}$ was based on the linearized mean-field model, and the scaled reference signal defined by $r_t = \varepsilon r_t^1$, $t \in \mathbb{Z}$. The linear state space model (7) leads to the approximation $\boldsymbol{\gamma} \approx G_p \boldsymbol{\zeta}$ (see (8)). (recall that $\boldsymbol{\gamma}$ is intended to approximate $\tilde{\boldsymbol{y}}$). For the linear control model (4) we obtain $\boldsymbol{\zeta} = G_c \boldsymbol{e} \approx G_c \boldsymbol{r} - G_c \boldsymbol{\gamma}$. On eliminating $\boldsymbol{\gamma}$ we obtain the linear equation $G_c \boldsymbol{r} \approx (1 + G_c G_p)\boldsymbol{\zeta}$. This is the justification for the transfer function $G_c/(1+G_cG_p)$ appearing in Fig. 5.

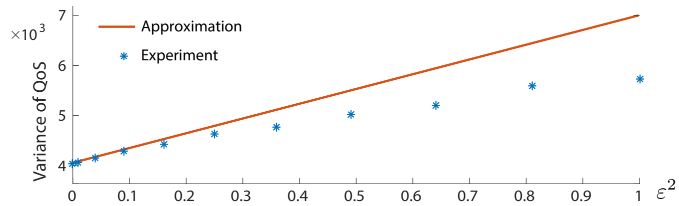


Fig. 7. Linear approximation for the variance of QoS.

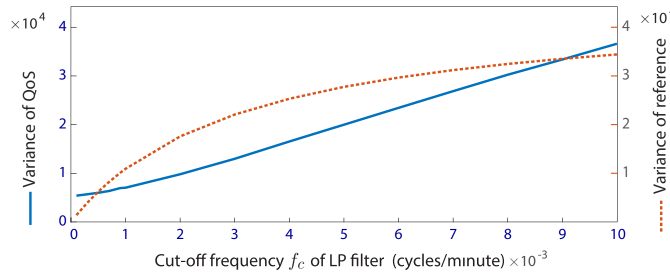


Fig. 8. QoS variance increases with bandwidth of the reference signal

In the experiments that follow, G_c is the PI controller with proportional gain 50 and integral gain 1.5.

C. Individual QoS

The QoS of pools without opt-out control is illustrated using the histogram on the left-hand-side of Fig. 2 using the reference signal $\boldsymbol{r}^1$. The histogram is based on samples of the QoS function (10) from each load over one month. The dashed lines represent Gaussian densities with zero mean, and variances $\sigma_{\mathcal{L}}^2$ obtained using the approximation given in (22).

The conclusions of Theorem 3.4 are illustrated in Fig. 7, where the QoS variances from simulation and Gaussian approximations are presented for several values of ε. It is seen that the variance is approximately linear in ε^2 for small $\varepsilon > 0$, with slope as predicted in Theorem 3.4.

The role of bandpass filter: A range of cut-off frequencies were considered in order to investigate the impact of the bandpass filter that is used to obtain $\boldsymbol{r}$. Fig. 8 shows a comparison of the variance of the reference and the variance of QoS as a function of the cut-off frequency f_c over a range of frequencies: The linear growth in QoS variance compared to the slow growth of the variance of the reference signal justifies a filter with $f_c < 10^{-2}$ cycles/minute.

The remaining numerical results that follow are based on the low pass filter (29) based on $f_c = 10^{-3}$ cycles/minute.

D. Grid level performance and opt-out control

We present in this section experimental results with opt-out control to ensure that QoS is subject to strict constraints.

Four QoS intervals were considered corresponding to constraints of, respectively, 5%, 10%, 15%, and 20%. A window of width 5% corresponds to ± 3 hours — a very tight constraint over a ten-day time horizon.

Recall that ten days corresponds to 2880 time samples. Following the notation in Section III-C, these percentages are converted to intervals for local opt-out control as $[b_-, b_+] = [-36, +36]$, $[-72, +72]$, $[-108, +108]$, and $[-144, +144]$, respectively.

Fig. 9 shows four histograms of QoS with reference signal $\boldsymbol{r}^1$. These histograms are truncated to the predefined QoS intervals as desired.

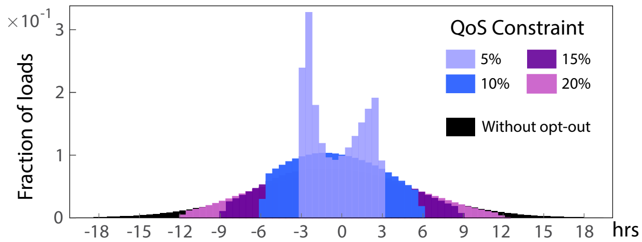


Fig. 9. QoS histograms with opt-out control using different QoS bounds.

The root-mean-square (RMS) of a signal $\boldsymbol{f}$ over a time horizon T is defined as

$$\mathrm{RMS}(\boldsymbol{f}) = \sqrt{\frac{1}{T}\sum_{k=1}^{T} f_k^2}$$

The value of T is consistent with the data from January 2015 at BPA. Given the 5 minute sampling interval, this gives $T = 31 * 24 * 12 = 8928$. A normalized root mean square error (NRMSE) was adopted as the metric of grid level tracking performance:

$$\mathrm{NRMSE} = \frac{1}{\varepsilon}\frac{\mathrm{RMS}(\boldsymbol{e}) - \mathrm{RMS}(\boldsymbol{e}^0)}{\mathrm{RMS}(\boldsymbol{r}^1)}, \tag{30}$$

The division by ε is used because $r_t = \varepsilon r_t^1$ by construction. The process $\boldsymbol{e}^0$ is the tracking error sequence obtained without opt-out control, and $\varepsilon = 0$. That is, $\boldsymbol{e} = \boldsymbol{e}^0$ when $\varepsilon = 0$. The approximation $\mathrm{RMS}(\boldsymbol{e}^0) \approx 8.35$ kW was obtained through simulation. This is a very small fraction of the nominal mean power consumption 5 MW (corresponding to 10^4 loads).

The grid level tracking performance with and without opt-out control is illustrated in Fig. 10. The tracking performance with 10%, 15%, or 20% QoS interval remains nearly perfect. This is surprising, given the improvement in QoS shown in Fig. 9. The explanation is that very few loads opt out: For example, in the simulation with 10% QoS interval and reference signal $\boldsymbol{r}^1$, no more than 1% of loads opt out at any time.

However, there are limitations on the capacity of ancillary service from a collection of loads, and experiments reveal that opt-out control can reduce capacity. As seen from Fig. 10, the additional opt-out control with 5% QoS interval degrades grid level tracking performance, especially when the reference signal is large.

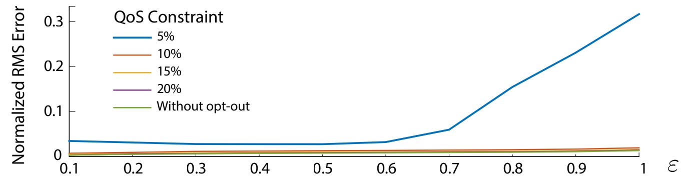


Fig. 10. Tracking NRMSEs (30), with and without local opt-out control.

We next apply Prop. 2.1 to better understand the grid-level impact of opt-out control, and to design an algorithm to reshape the reference signal so that loads are less likely to opt out of service.

E. Re-shaping the reference input

Recall the SOC heuristic introduced following Prop. 2.1. The proposition implies that $\bar{\mathcal{L}}_t \approx R_t$ for all t when there is near perfect tracking. Conversely, this result suggests that many loads will opt-out, leading to poor tracking, if R_t is near the boundary of the QoS interval $[b_-, b_+]$.

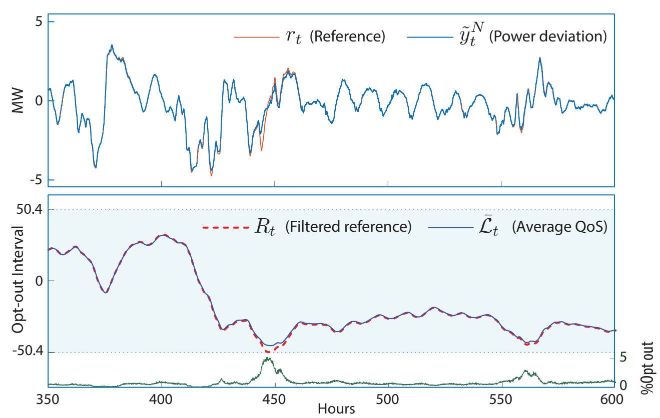


Fig. 11. Poor tracking results when R_t reaches its boundaries.

To illustrate the application of these concepts, consider the case $[b_-, b_+] = [-50.4, 50.4]$, corresponding to a 7% QoS interval, and reference signal $\boldsymbol{r}^1$. Results are provided in Fig. 11. Most of the time, the tracking results are nearly perfect and hence the average QoS approximates the filtered reference signal, $\bar{\mathcal{L}}_t \approx R_t$. However, at time ≈ 450 hr., R_t falls close to the lower bound $b_- = -50.4$. During this time period, many loads opted out, which resulted in degraded tracking.

In conclusion, to ensure good grid level tracking, the reference signal must respect any QoS constraints. The grid operator should re-shape the reference signal to ensure that $\boldsymbol{R}$ does not approach its limits. A smooth transformation is required because the reference signal for this example should not have significant high frequency content.

Here is one approach, based on two non-negative parameters: a threshold $\tau < 1$ and a gain parameter $\delta > 0$. The following recursive algorithm is designed to increase (decrease) the reference signal $\boldsymbol{r}$ when $\boldsymbol{R}$ reaches its lower (upper) threshold. The re-shaped reference signal is defined as follows:

$$\bar{r}_t = \begin{cases} [r_t - \delta(R_t - \tau b_+)]_+ & R_t > \tau b_+ \text{ and } r_t > 0 \\ [r_t + \delta(R_t - \tau b_-)]_- & R_t < \tau b_- \text{ and } r_t < 0 \\ r_t & \text{otherwise} \end{cases} \tag{31}$$

where z_+ denote the positive part of a real number z, and z_- the negative part.

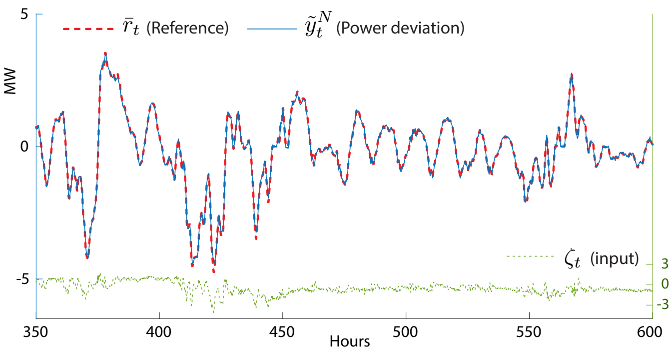


Fig. 12. Tracking performance with the re-shaped reference signal.

The plots in Fig. 12 illustrate the tracking performance with 7% QoS interval, and the re-shaped reference signal $\bar{\boldsymbol{r}}^1$ (using $\delta = 0.006$ and $\tau = 0.65$). Compared to $\boldsymbol{r}^1$ in Fig. 11, $\bar{\boldsymbol{r}}^1$ increases the reference signal $\boldsymbol{r}^1$ at time ≈ 450 hr. This small change avoids a large number of opt outs observed in Fig. 11, and the tracking performance is nearly perfect.

F. Multiple QoS metrics

We next investigate the impact of opt-out control based on two QoS metrics: the cleaning QoS function (12) and the cycling QoS function (25). The lower bound of cycling QoS was set to $-\infty$. The opt-out priority is assigned to the cleaning QoS metric when both QoS metrics reach their bounds at the same time. Thus, the cycling QoS metric occasionally exceeds its predefined upper bound.

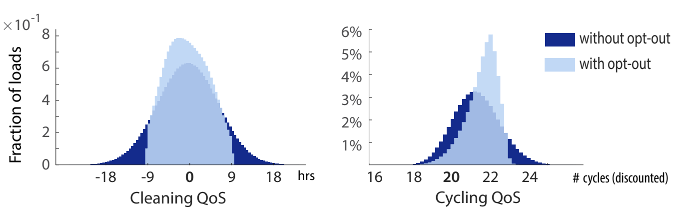


Fig. 13. QoS bounds are maintained using local opt-out control.

Four error tolerances $\{5\%, 10\%, 15\%, 20\%\}$ are considered for determining the opt-out interval $[b_-, b_+]$ for each of the two QoS metrics. Fig. 13 illustrates an example of QoS improvement based on a 15% error tolerance on both QoS metrics, using the reference signal $\boldsymbol{r}^1$. Tracking in this case was nearly perfect.

Tracking performance for a range of opt-out parameters is summarized in Fig. 14 using 16 colored bars, distinguished by each pair of QoS constraints. Each bar represents tracking

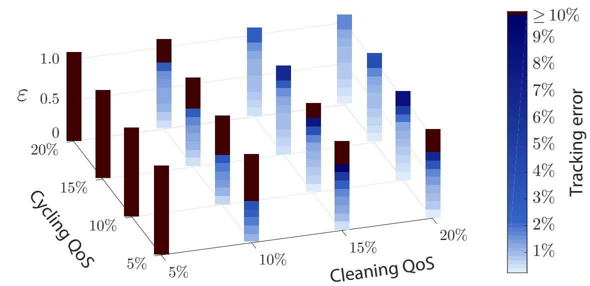


Fig. 14. Tracking performance with two QoS constraints.

errors for different reference signal scaling factors, $0.1 \leq \varepsilon \leq 1$. The darkest color represents NRMSE (30) of 10% or greater, and lighter colors represent smaller values of NRMSE (indicated on the color bar label). Results in this figure show that opt-out control based on these two QoS metrics have little impact on tracking error over a large range of opt-out intervals. For those cases that local opt-out largely degrades grid-level tracking, we can either reduce the reference signal or relax QoS constraints to maintain good tracking.

V. Conclusions

The main technical contribution of this paper is the approximation of QoS for an individual load. It is remarkable that it is possible to obtain accurate estimates of first and second order statistics for an individual load, taking into account second order statistics of exogenous inputs, along with correlation introduced by the Markovian model. It is also remarkable that strict bounds on QoS can be guaranteed while retaining nearly perfect grid-level tracking.

Under certain conditions, the overall QoS of a collection of loads is predictable to the grid operator. With this information, the grid operator can estimate the flexibility/capacity of loads and modify the reference signal if necessary, to maintain high quality of both grid-level tracking and QoS of loads.

Open problems remain in the area of estimation and control. It is hoped that estimates of the statistics of QoS can be obtained in the presence of opt-out control. It is also likely that control performance can be improved further with a more sophisticated approach to opt-out control.

References

[1] R. Malhame and C.-Y. Chong, “Electric load model synthesis by diffusion approximation of a high-order hybrid-state stochastic system,” *IEEE Trans. Automat. Control*, vol. 30, no. 9, pp. 854 – 860, Sep 1985.

[2] J. Mathieu, S. Koch, and D. Callaway, “State estimation and control of electric loads to manage real-time energy imbalance,” *IEEE Trans. Power Systems*, vol. 28, no. 1, pp. 430–440, 2013.

[3] A. Kizilkale and R. Malhame, “Mean field based control of power system dispersed energy storage devices for peak load relief,” in *52nd IEEE Conference on Decision and Control (CDC)*, 2013, pp. 4971–4976.

[4] S. Meyn, P. Barooah, A. Bušić, Y. Chen, and J. Ehren, “Ancillary service to the grid using intelligent deferrable loads,” *IEEE Trans. Automat. Control*, vol. 60, no. 11, pp. 2847–2862, Nov 2015.

[5] H. Hao, Y. Lin, A. Kowli, P. Barooah, and S. Meyn, “Ancillary service to the grid through control of fans in commercial building HVAC systems,” *IEEE Trans. on Smart Grid*, vol. 5, no. 4, pp. 2066–2074, July 2014.

[6] P. Barooah, A. Bušić, and S. Meyn, “Spectral decomposition of demand-side flexibility for reliable ancillary services in a smart grid,” in *Proc. 48th Annual Hawaii International Conference on System Sciences (HICSS)*, Kauai, Hawaii, 2015, pp. 2700–2709.

[7] Y. Chen, “Markovian demand dispatch design for virtual energy storage to support renewable energy integration,” Ph.D. dissertation, University of Florida, Gainesville, FL, USA.

[8] Y. Chen, A. Bušić, and S. Meyn, “Individual risk in mean field control with application to automated demand response,” in *53rd IEEE Conference on Decision and Control*, Dec 2014, pp. 6425–6432.

[9] J.-Y. Le Boudec and D.-C. Tomozei, “Demand response using service curves,” in *International Conference and Exhibition on Innovative Smart Grid Technologies (ISGT Europe)*, 2011, pp. 1–8.

[10] Y. Chen, A. Bušić, and S. Meyn, “Ergodic theory for controlled Markov chains with stationary inputs,” *ArXiv e-prints and Annals of Applied Prob.* (under revision), April 2016.

[11] N. V. Krylov, R. S. Lipster, and A. A. Novikov, “Kalman filter for Markov processes,” in *Statistics and Control of Stochastic Processes*. New York: Optimization Software, inc., 1984, pp. 197–213.

[12] Y. Chen, A. Bušić, and S. Meyn. State estimation for the individual and the population in mean field control with application to demand dispatch. *IEEE Transactions on Automatic Control*, 62(3):1138–1149, March 2017.

[13] R. Couillet, S. Perlaza, H. Tembine, and M. Debbah, “Electrical vehicles in the smart grid: A mean field game analysis,” *IEEE J. Selected Areas in Comm.*, vol. 30, no. 6, pp. 1086–1096, 2012.

[14] Z. Ma, D. Callaway, and I. Hiskens, “Decentralized charging control for large populations of plug-in electric vehicles: Application of the Nash certainty equivalence principle,” in *IEEE International Conf. on Control Applications (CCA)*, 2010, pp. 191–195.

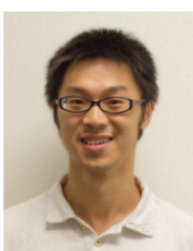
[15] A. Bušić and S. Meyn. Distributed randomized control for demand dispatch. In *IEEE Conference on Decision and Control*, pages 6964–6971, Dec 2016.

[16] J. Mathias, A. Bušić, and S. Meyn. Demand dispatch with heterogeneous intelligent loads. In *Proc. 50th Annual Hawaii International Conference on System Sciences* http://hdl.handle.net/10125/41537, 2017.

[17] “BPA balancing authority,” Online, http://tinyurl.com/BPAgenload, http://tinyurl.com/BPAbalancing.

[18] J. Sharp. Electrical load disconnect device with electronic control, Dec. 2012. US Patent 8,328,110.

[19] C. Ziras, E. Vrettos, and G. Andersson. Primary frequency control with refrigerators under startup dynamics and lockout constraints. In *IEEE Power & Energy Society General Meeting*, pages 1–5, 2015.



Yue Chen (S’ 14) received the B.E. degree in electronic and information engineering from Harbin Engineering University, Harbin, China, in 2008, the M.E. degree in electrical engineering from University of Detroit Mercy, Detroit, USA, in 2010, the M.S. degree in Electrical and Computer Engineering in 2012 from University of Florida, Gainesville, USA. During the Fall of 2015 he was an intern at Los Alamos National Laboratory conducting research on control techniques for demand dispatch.

He received the Ph.D. degree in electrical and computer engineering at the University of Florida in 2016.

Ana Bušić is a Research Scientist at Inria Paris – Rocquencourt, and a member of the Computer Science Department at Ecole Normale Supérieure, Paris, France. She received a M.S. degree in Mathematics in 2003, and a Ph.D. degree in Computer Science in 2007, both from the University of Versailles. She was a post-doctoral fellow at Inria Grenoble – Rhône-Alpes and at University Paris Diderot-Paris 7. She joined Inria in 2009. She is a member of the Laboratory of Information, Networking and Communication Sciences, a joint lab between Alcatel-Lucent, Inria, Institut Mines-Télécom, UPMC Sorbonne Universities, and SystemX. Her research interests include stochastic modeling, simulation, performance evaluation and optimization, with applications to communication networks and energy systems.

Sean P. Meyn (S’85, M’87, SM’95, F’02) received the B.A. degree in Mathematics Summa Cum Laude from UCLA in 1982, and the PhD degree in Electrical Engineering from McGill University in 1987 (with Prof. P. Caines). He held a two year postdoctoral fellowship at the Australian National University in Canberra, and was a professor at the University of Illinois from 1989-2011. Since January, 2012 he has been a professor and has held the Robert C. Pittman Eminent Scholar Chair in Electrical and Computer Engineering at the University of Florida. He is coauthor with Richard Tweedie of the monograph Markov Chains and Stochastic Stability, Springer-Verlag, London, and received jointly with Tweedie the 1994 ORSA/TIMS Best Publication In Applied Probability Award. His research interests include stochastic processes, optimization, complex networks, and information theory, with applications to smarter cities and smarter grids.

Prof. Meyn and Prof. Bušić share a 2015 Google Research Award: *to foster collaboration on demand dispatch for renewable energy integration*.

Appendix

Abbreviations

BA	balancing authority
BPA	Bonneville Power Administration [17]
LTI	linear time-invariant
pmf	probability mass function
PSD	power spectral density
QoS	quality of service
SOC	state of charge

Nomenclature

N	The number of loads
ε	Scaling factor $\in [0, 1]$
$\boldsymbol{r}^\bullet$	Overall regulation signal
$\boldsymbol{r}$	Reference signal
$\boldsymbol{\zeta}$	Control signal computed by BA
P_ζ	Transition matrix with input ζ
$\mathcal{E}$	The first order derivative of the transition matrix P_ζ evaluated at $\zeta = 0$
$\mathcal{E}^{(2)}$	The second order derivative of the transition matrix P_ζ evaluated at $\zeta = 0$
P_0	Nominal transition matrix
π_0	Invariant pmf for the nominal transition matrix P_0
X_t^i	State of the ith load at time t (in a d-element set)
μ_t^N	Empirical pmf of load states $X_t^i, i = 1, \dots, N$
μ_t	The limit of μ_t^N, as the load number $N \to \infty$
Φ_t	State vector of the LTI aggregate system
γ_t	Output of the LTI aggregate system
y_t^N	Normalized power consumption of N loads
$\overline{y}^0$	Steady-state power consumption for nominal model
$\tilde{y}_t^N$	Normalized power deviation, $\tilde{y}_t^N = y_t^N - \overline{y}^0$
ℓ	One-step QoS function
$\mathcal{L}_t^i$	QoS of the ith load
$\bar{\mathcal{L}}_t$	Average QoS of N loads
Γ_t	Individual load state in the d-dimensional simplex
D_t	Disturbance term in the LTI individual load model