Grid frequency regulation with deferrable loads – an $\mathcal{H}_2$-optimal control approach.

Federico Bliman, Fernando Paganini and Andres Ferragut
Universidad ORT Uruguay

Abstract— **Power grid operators require fast balancing actions to regulate grid frequency; in a Smart Grid scenario this service may be provided by flexible loads. We consider a load aggregator that manages a large quantity of loads, and must align their overall consumption with an operator reference. Starting from a stochastic queueing model for the set of loads, we derive a macroscopic differential equation model, excited by noise, that represents the aggregate, and includes a scalar deferral action variable. An $\mathcal{H}_2$-optimal control method is then used to optimize the tracking error in power consumption while keeping the control variable within its hard limits. The controller is tested with real-world frequency regulation signals, and achieves high performance using the relevant industry performance metric. An implementation is given which achieves these benefits with mild communication requirements.**

I. Introduction

The development of the Smart Grid [10], [18] is opening new possibilities for Demand Response [11], in which consumer loads may provide services to the grid, with the aid of information and communication technologies.

We focus here on the *frequency regulation* service, which involves power balancing actions at a fast time-scale. The terminology stems from the fact that imbalance leads to deviations from nominal frequency of the grid. The system operator (SO), measuring these deviations, generates a power reference signal to be tracked by regulation providers. From the perspective of these providers, the problem becomes one of tracking such power directives; grid frequency itself will thus not explicitly appear in what follows.

Traditionally, regulation is provided by fast-responding generators; here we are interested in the possibility that a load aggregator entity (e.g. [5], [8]) may track the power reference by controlling consumption of a set of loads under its authority. Candidate loads for such task are those whose service can be partially *deferred*.

We briefly survey some recent research in this area. A possible class of loads for this task are electric vehicles [10], [21]. In particular [21] estimates the frequency regulation capacity of a fleet of EVs while being charged. Another suitable category are thermostatically controlled loads (TCLs) [9], [12], [13], [20], which exploit thermal inertia to modify instantaneous power without affecting user comfort. [9] characterizes a collection of TCLs as an equivalent battery, showing how this approach can provide frequency regulation in a real scenario. A recent reference on decentralized control of generic TCLs is [20], while [13] focuses on commercial building HVAC systems to provide frequency regulation.

References which consider load deferability in general are [7], [14], [19]. In [14] the aim is to characterize the aggregate flexibility of a cluster of loads in terms of electricity storage. [19] compares classical scheduling algorithms with a model predictive control proposal. [7] uses decentralized control of generic deferrable loads to make the aggregate total power consumption on the grid as smooth as possible.

In our previous work [1] we proposed the use of a scalar control signal (fraction of service power) to manage an aggregate of loads in a decentralized way, which we deem more practical than individual load scheduling by the aggregator. We obtained excellent regulation tracking with a simple control based on a first order differential equation model; its main limitation is that, without load micro-management, deadlines are only met in a mean sense. By incorporating a separate state for loads which can no longer be deferred, in [2] we explored the performance of a classical controller that tracks the reference while strictly enforcing deadlines.

In this paper we begin in Section II by providing more substantiation for the two-state model in [2], as a macroscopic approximation of a natural stochastic queueing model. Then we move to a systematic control design with tools of $\mathcal{H}_2$-optimal control. First in Section III we design a regulator to reduce power variance, and in Section IV we tackle the main problem of tracking an exogenous power signal. In Section V we discuss how such controller can be implemented in a decentralized way. Conclusions are given in Section VI.

II. Macroscopic dynamic model

The object of study is an aggregator entity that manages a set of customer loads; it has procured in advance a mean consumption power, assumed known. The effective load is, however, uncertain in advance: it materializes as a sequence of requests characterized by an arrival time, a service time, a deadline before which each must be served, and a nominal power. The purpose of the controlled service deferral is to align the consumption of the aggregate to a desired trajectory.

In this section we develop a *macroscopic* dynamic model of the loads under control, in the sense that it only attempts to track aggregate quantities; this will make it suitable for control purposes. To develop it we begin with a microscopic, discrete counterpart, and then indicate how to approximate this behavior with a differential equation model.

E-mail:{bliman,paganini,ferragut}@ort.edu.uy.
Research supported by ANII Uruguay under grant FSE_1_2014_1_102426 and POS_NAC_2013_1_11675.

A. Markov chain model

Consider first the following stochastic queueing model from [6]. Load requests arrive as a Poisson process of rate λ; for simplicity assume they all have the same nominal power p_0, but may differ in the required energy. Let τ_k denote the nominal service time of load k, i.e. the time (energy/p_0) it would take to serve it at full nominal power. The possibility of deferring a certain load is specified through its spare time or *laxity* L_k; in other words, the load's hard deadline for completing service is $\tau_k + L_k$. We assume here that τ_k, L_k are independent, exponential random variables of respective means τ, L, also independent of the arrival process.

The deferral action is defined by the *service level*, a scalar variable $u \in (0, 1]$ specifying the fraction of nominal power at which loads are served. [1] After an interval of time dt, a load served at power up_0 will have reduced its required service time by udt, but also will have consumed $dt - udt = (1 - u)dt$ of its spare time. Figure 1 shows the trajectory in (service-time/laxity) space when u is constant.

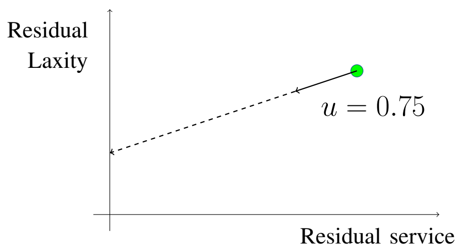

Fig. 1. Service-laxity trajectory under service level u

A trajectory reaching the vertical axis completes service and leaves the system. If, instead, the horizontal axis is reached first, laxity expires and to keep its deadline the load can no longer be deferred. We will denote by $n(t)$ the population of loads that at time t still have remaining laxity, and thus are served at level u; the remainder of loads $m(t)$ with expired laxity will be served at full power.

With these notations, under the independence and exponential distribution assumptions the behavior of the population variables over time is described by the continuous-time Markov chain with transition rates depicted in Figure 2. This represents basically two $M/M/\infty$ queues with a partial transition between the two, as is now explained.

- $(n, m) \mapsto (n + 1, m)$ is a new Poisson(λ) arrival.
- $(n, m) \mapsto (n, m - 1)$ represents a departure from the m queue. Service here is at full power, so service times are $exp(\frac{1}{\tau})$, invoking the memoryless property of exponentials. The minimum of m such exponentials is $exp(\frac{m}{\tau})$, justifying the transition rate $\frac{m}{\tau}$.
- $(n, m) \mapsto (n-1, m)$ represents a load from the n queue completing service. Since these are served at fractional power u, their individual service time is $exp(\frac{u}{\tau})$, which yields the transition rate $\frac{nu}{\tau}$.

[1] Alternatively, one could serve a fraction of loads at nominal power.

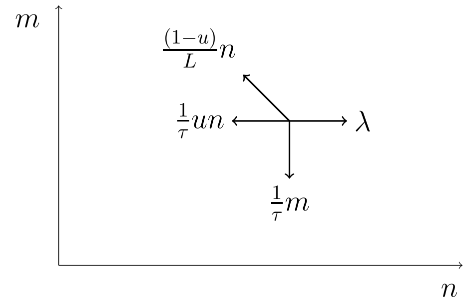

Fig. 2. Markov state diagram with transition rates

- $(n, m) \mapsto (n - 1, m + 1)$ represents the transition between the n and m queues due to expiration of laxity. Since laxity is consumed at rate $(1 - u)$, the time for this occurrence in one load is distributed as $exp(\frac{1-u}{L})$, which yields the transition rate $\frac{n(1-u)}{L}$.

In [6] some analysis was provided for the above Markov chain in the case of a fixed u; in particular it admits a stationary distribution of product form. However in this paper we are interested in *controlling* u to achieve a desired regulation objective; for this purpose a more tractable model involves replacing the Markov chain by a differential equation.

B. Fluid flow model

The fluid-flow counterpart to the Markov chain in Figure 2 is obtained by interpreting n and m as continuous variables, and replacing the transition rates with different contributions to their drift, as follows:

$$\dot{n} = \lambda - \frac{nu}{\tau} - \frac{n(1-u)}{L}, \tag{1a}$$

$$\dot{m} = \frac{n(1-u)}{L} - \frac{m}{\tau}. \tag{1b}$$

The formal relationship between the two models is beyond the scope of this paper. We state briefly that the solution to the differential equation can be seen as the limit of the scaled stochastic processes $(\frac{1}{k}n^k(t), \frac{1}{k}m^k(t))$ as $k \to \infty$, where the (n^k, m^k) correspond to the Markov chain under scaled arrival parameter $k\lambda$, and suitably scaled initial condition. For details on such procedure we refer to [17].

In the above system the deferral action u can be seen as a *control input*; varying this quantity (at a slower time-scale than the microscopic load dynamics) influences the macroscopic population states, and through them the main output of interest: the aggregate power consumed:

$$p = p_0(nu + m). \tag{2}$$

A first step in the analysis is to find the equilibrium of this model for the case of a fixed control $u(t) \equiv u^*$. Introducing

$$\nu := \frac{u^*}{\tau} + \frac{(1 - u^*)}{L}, \tag{3}$$

we have the equilibrium point

$$n^* = \frac{\lambda}{\nu}, \qquad m^* = \frac{\lambda\tau(1 - u^*)}{\nu L}. \tag{4}$$

The equilibrium power of the cluster is then found to be

$$p^* = p_0(n^*u^* + m^*) = \lambda p_0 \tau. \tag{5}$$

Note that the latter expression does not depend on u^*, and matches the mean exogenous demand for power (load arrival rate times mean energy). All arriving loads must be served sooner or later, the only influence our control can have is over the variations around such equilibrium power.

In what follows we will analyze the system locally around this equilibrium, through the corresponding linearization around the operating point (n^*, m^*). Writing $\delta(nu) \approx u^*\delta n + n^*\delta u$, and $\delta(n(1-u)) \approx (1-u^*)\delta n - n^*\delta u$ in the incremental quantities, we obtain the linearized dynamics:

$$\dot{\delta n} = -\underbrace{\left[\frac{u^*}{\tau} + \frac{1-u^*}{L}\right]}_{\nu}\delta n + \left[\frac{n^*}{L} - \frac{n^*}{\tau}\right]\delta u; \tag{6a}$$

$$\dot{\delta m} = \frac{1-u^*}{L}\delta n - \frac{1}{\tau}\delta m - \frac{n^*}{L}\delta u; \tag{6b}$$

$$\delta p = p_0(u^*\delta n + \delta m) + p_0 n^* \delta u. \tag{6c}$$

C. Introducing noise

The preceding models are purely deterministic, having removed all randomness from the original Markov chain. For a more accurate description around the operating point we will introduce random noise, that results from a diffusion approximation of the Markov chain dynamics. Formally, (see [17]) this noise process is the limit in distribution of

$$\sqrt{k}\left(\frac{n^k(t)}{k} - n(t), \frac{m^k(t)}{k} - m(t)\right)$$

where (n^k, m^k) is the scaled process mentioned before, and (n, m) the fluid limit. Again this is outside our scope, but we can motivate our noise model by reviewing the case of a Poisson process $a(t)$, such as the arrivals to our system. Its diffusion approximation satisfies the stochastic differential equation $da = \lambda dt + \sqrt{\lambda} dW$, where $W(t)$ is Brownian motion. More informally we can write equation

$$\dot{a} = \lambda + \sqrt{\lambda}\, w(t),$$

where $w(t)$ is unit white noise. In classical terms, the fluid model $\dot{a} = \lambda$ for Poisson arrivals is modified by additive white noise of power spectral density equal to the arrival rate itself. What we are looking for is the analogous modification to the model (6) to track the fluctuations of the process (n, m), locally around its equilibrium (n^*, m^*).

First, the Poisson arrivals will introduce a noise term $v_1(t) = \sqrt{\lambda}\, w_1(t)$ in (6a). The two departure terms in (1a) will also introduce noise terms, with power spectral density equal to the transition rate, evaluated at equilibrium:

$$v_2 = \sqrt{\frac{n^*u^*}{\tau}}\, w_2, \quad v_3 = \sqrt{\frac{n^*(1-u^*)}{L}}\, w_3;$$

here w_2, w_3 are independent unit white noises. With some algebra we can also rewrite the above as

$$v_2 = \sqrt{\alpha\lambda}\, w_2, \quad v_3 = \sqrt{(1-\alpha)\lambda}\, w_3, \tag{7}$$

$$\text{where} \quad \alpha := \frac{u^*}{\tau\nu} = P\left[\frac{\tau_k}{u^*} \le \frac{L_k}{1-u^*}\right]$$

is the probability that a load finishes service before expiring its laxity. The interpretation of (7) is that departures of the n queue are equivalent to two Poisson processes, where the rate is "thinned" by the probability of, respectively, leaving the system and joining the m queue. The term v_3 will also appear as noise in arrivals to the dynamics (6b) for $m(t)$, and

$$v_4 = \sqrt{\frac{m^*}{\tau}}\, w_4 = \sqrt{(1-\alpha)\lambda}\, w_4$$

will represent noise in departures from this second queue. The resulting dynamics with noise is thus

$$\dot{\delta n} = -\nu\delta n + \left[\frac{n^*}{L} - \frac{n^*}{\tau}\right]\delta u + v_1 - v_2 - v_3, \tag{8a}$$

$$\dot{\delta m} = \frac{1-u^*}{L}\delta n - \frac{1}{\tau}\delta m - \frac{n^*}{L}\delta u + v_3 - v_4, \tag{8b}$$

$$\delta p = p_0(u^*\delta n + \delta m) + p_0 n^*\delta u. \tag{8c}$$

D. State space model and variance calculation

The standard state-space form of the above model is:

$$\overbrace{\begin{bmatrix}\dot{\delta n}\\ \dot{\delta m}\end{bmatrix}}^{\dot{x}} = \overbrace{\begin{bmatrix}-\nu & 0\\ \frac{1-u^*}{L} & -\frac{1}{\tau}\end{bmatrix}}^{A}\overbrace{\begin{bmatrix}\delta n\\ \delta m\end{bmatrix}}^{x} + B_1 \overbrace{\begin{bmatrix}w_1\\ w_2\\ w_3\\ w_4\end{bmatrix}}^{w} + \overbrace{\begin{bmatrix}-\frac{n^*}{\tau} + \frac{n^*}{L}\\ -\frac{n^*}{L}\end{bmatrix}}^{B_2}\delta u; \tag{9a}$$

$$\delta p = \overbrace{\begin{bmatrix}p_0 u^* & p_0\end{bmatrix}}^{C}\begin{bmatrix}\delta n\\ \delta m\end{bmatrix} + \overbrace{p_0 n^*}^{D}\delta u, \tag{9b}$$

where

$$B_1 = \begin{bmatrix}\sqrt{\lambda} & -\sqrt{\alpha\lambda} & -\sqrt{(1-\alpha)\lambda} & 0\\ 0 & 0 & \sqrt{(1-\alpha)\lambda} & -\sqrt{(1-\alpha)\lambda}\end{bmatrix}.$$

A first analysis question is to quantify the natural fluctuations of consumed power when there is no active control δu (i.e. we take a constant deferral action u^*). To compute this we find the steady-state covariance matrix of the state, given (see e.g. [4]) by the solution Q to the Lyapunov equation

$$AQ + QA^T + B_1B_1^T = 0;$$

the resulting variance of the output p is $E\left[(\delta p)^2\right] = CQC^T$. Carrying out the calculations for the given matrices gives the following result:

$$E\left[(\delta p)^2\right] = p^* p_0\left[1 - \frac{1}{\frac{1}{1-u^*} + \frac{\tau}{Lu^*}}\right]. \tag{10}$$

Note that the choice of u^* affects the variance. So even with no real-time control, there is a variance reduction obtained through fixed deferral, its optimum occurring at

$$u^*_{opt} = \frac{\sqrt{\tau}}{\sqrt{L} + \sqrt{\tau}}. \tag{11}$$

In the remainder of the paper we will add active control of the signal $\delta u(t)$ to reduce the variance even further, or more importantly for tracking an exogenous reference.

III. $\mathcal{H}_2$ CONTROL FOR REDUCING POWER VARIANCE

In the present section we assume that the regulation objective is to stay as close as possible to a constant power consumption, i.e. to reduce the variability caused by the randomness in loads. Later on we will extend the solution to tracking variable external references.

In particular, we design here a state-feedback controller that minimizes a compromise between output variance and control effort, expressed by the weighted objective

$$J := \mathbf{E}[(k_1\delta p)^2 + (\delta u)^2]. \tag{12}$$

Setting up the problem in the standard form for $\mathcal{H}_2$-control (see e.g. [22]), we have a generalized plant

$$G(s) = \left[\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & 0 & D_{12} \\ I & 0 & 0 \end{array}\right], \tag{13}$$

where A, B_1, B_2 are the same as in (9a), and we introduce

$$C_1 = \frac{1}{k_2}\begin{bmatrix} k_1 C \\ 0 \end{bmatrix} = \frac{1}{k_2}\begin{bmatrix} k_1 p_0 u^* & k_1 p_0 \\ 0 & 0 \end{bmatrix},$$
$$D_{12} = \frac{1}{k_2}\begin{bmatrix} k_1 D \\ 1 \end{bmatrix} = \frac{1}{k_2}\begin{bmatrix} k_1 p_0 n^* \\ 1 \end{bmatrix}.$$

The penalized output corresponds to the cost in (12), except for the constant $k_2 = (1 + (k_1 p_0 n^*)^2)^{\frac{1}{2}}$ which provides the simplifying normalization $D_{12}^* D_{12} = 1$.

We assumed that the state (n, m) is available for feedback, later on we discuss the implementation requirements. Under these conditions the $\mathcal{H}_2$-optimal control is (see [22]) the static state feedback

$$\delta u = -Fx = -(B_2^* X + D_{12}^* C_1)\begin{bmatrix} \delta n \\ \delta m \end{bmatrix} \tag{14}$$

where X is the stabilizing solution to the Algebraic Riccati Equation

$$\begin{aligned}(A^* - C_1^* D_{12} B_2^*)X + X(A^* - B_2 D_{12}^* C_1) \\ -X B_2 B_2^* X + C_1^*(I - D_{12} D_{12}^*) C_1 = 0.\end{aligned} \tag{15}$$

Obtaining a parametric solution of this equation would be cumbersome so we analyze it numerically. The parameters we will use are $\lambda = 0.2\, loads/s$, $\tau = 1800s$, $L = 3600s$, $p_0 = 2kW$. For u^* we choose the optimal value from (11) in the uncontrolled case, which yields $u^* = 0.41$. It follows that $n^* = 511$, $m^* = 151$ and $p^* = 720kW$.

Since our main objective is reducing variance, k_1 should be increased as much as possible in relation to the unit penalty on control effort. The main limitation is that $\delta u(t)$ becomes too large nonlinear effects come into play, in particular saturation of $u(t) \in (0, 1]$.

We carried out a linear search for k_1 over multiples of $\frac{1}{p_0 n^*}$ (this choice keeps units normalized), simulating the nonlinear dynamics. We found that when k_1 reaches $\frac{10}{p_0 n^*}$, the controller exhibits "actuator windup" with u remaining at zero and consequently deterioration of performance.

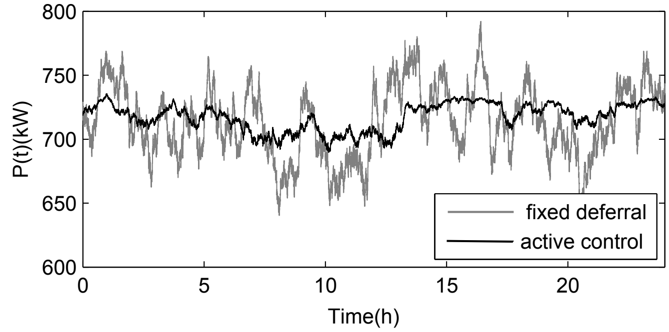

Fig. 3. One day simulation for the system with $k_1 = \frac{5}{u^* n^*}$.

From this evaluation, we settled on the value $k_1 = \frac{5}{p_0 n^*}$, which gives the optimal feedback

$$F = \left[\frac{0.735u^*}{n^*}, \frac{0.838}{n^*}\right].$$

In Fig. 3 we see the simulated trajectory of $p(t)$ under this feedback control, in comparison with the one obtained with the static deferral action $u = u^*$ from (11). We observe that the variance of the output power is significantly reduced. Also the trajectory for $u(t)$ (not shown here) operates in the interval $[0.33, 0.5]$, thus varying moderately around its nominal value $u^* \approx 0.41$.

IV. $\mathcal{H}_2$ CONTROL FOR REFERENCE TRACKING

We present now the central part of this work, where we replace the objective of keeping a constant power with a more ambitious one: offering a frequency regulation service by having the aggregate of deferrable loads follow a power reference signal provided by the system operator (SO).

Specifically, a provider of this ancillary service must commit to varying its power consumption up to a fraction θ of its nominal power p^*, in response to a real-time signal $\rho(t) \in [-1, 1]$ that it receives every few seconds from the SO. Upon receiving this signal the load should ideally become

$$p(t) = p^*(1 + \theta\rho(t)) = p^* + \underbrace{\theta p^* \rho(t)}_{r(t)}. \tag{16}$$

A. Maximum offered regulation

A frequency regulator provider is rewarded by the maximum deviation θp^* it is able to offer, thus it is to our convenience to make θ as large as possible. The maximum theoretical value is $\theta = 1$, which would imply varying the power in the range $[0, 2p^*]$.

In our system of deferrable loads this value is not achievable, because consumed power must lie within the bounds

$$p_0 m(t) \le p(t) \le p_0[n(t) + m(t)];$$

In particular the lower bound is always positive since we have chosen not to defer the loads $m(t)$ with expired laxity, and the upper bound is constrained by loads currently present. Both bounds are time-varying, but we can get an estimate of the achievable margin by applying the equilibrium values.

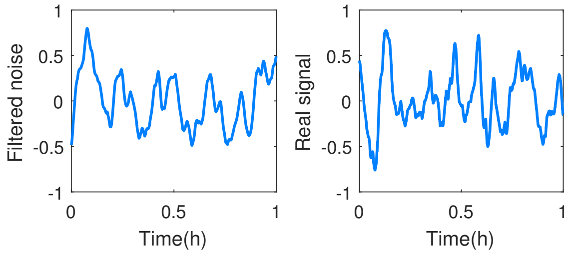

Fig. 4. Artificial regulation signal from filtered white noise, in comparison with a real regulation signal from PJM.

In particular, imposing that the committed minimal power $p^*(1-\theta)$ is above $p_0 m^*$ and recalling $p^* = p_0(n^*u^* + m^*)$ leads to the bound

$$\theta \leq \frac{n^*u^*}{n^*u^* + m^*} = \frac{Lu^*}{Lu^* + \tau(1-u^*)}.$$

Similarly, the upper bound $p^*(1+\theta) \leq p_0[n^* + m^*]$ gives

$$\theta \leq \frac{n^*(1-u^*)}{n^*u^* + m^*} = \frac{L(1-u^*)}{Lu^* + \tau(1-u^*)}.$$

The above upper bounds on θ are, respectively, increasing and decreasing in u^*, and they become equal in $u^* = \frac{1}{2}$; therefore this choice is the value that provides the maximum (symmetric) regulation capability, namely $\theta_{\max} = \frac{L}{L+\tau}$. We will use this choice of u^* in what follows; note that it need not coincide with the value from (11) providing minimal open-loop power variability.

B. Regulation signal characterization

Having decided on the amplitude of reference signals we are offering to track, the next key requirement for a good tracking controller is to characterize their spectral content.

For this purpose we turn to a particular family of real-life regulation signals $\rho(t)$ taken from PJM [15], a regional transmission operator in the US. We performed a spectral density estimation based on these PJM signals using MATLAB's signal identification toolbox. A first observation is that they have band limited energy, with cutoff frequency $\omega_r \approx 1.65 \times 10^{-2}\, rad/s$, after which they present a roll-off of $40\, db/dec$, indicating a second-order filtering. A closer inspection shows a resonance in the cutoff frequency with a damping factor of $\zeta \approx 0.4$. We therefore approximated the practical signals as generated by white noise through the frequency weighting filter

$$W_\rho(s) = \frac{\kappa_r \omega_r^2}{s^2 + s2\zeta\omega_r + \omega_r^2}, \tag{17}$$

where $\kappa_r \approx 3$ was chosen to match the mean signal power.

In Fig. 4 we can see a 1- hour simulation of filtered white noise along with a real regulation signal, with a qualitatively similar behavior.

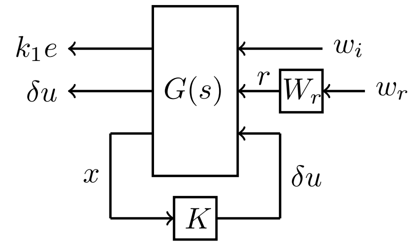

Fig. 5. Controller design for tracking the regulation signal

C. $\mathcal{H}_2$-optimal control

We now set-up our $\mathcal{H}_2$-optimal control problem through the generalized plant model in Fig. 5. Here, the input weight is $W_r := \theta p^* W_\rho$ with W_ρ in (17), consistent with $r(t)$ in (16), and is driven by a white noise signal $w_r(t)$, independent of the previously considered noise signals for the loads.

The tracking error signal is

$$e(t) = r(t) - \delta p(t) = p^*(1 + \theta\rho(t)) - p(t); \tag{18}$$

the penalized variables for $\mathcal{H}_2$ control correspond to the cost function $J_2 := E[(k_1 e)^2 + (\delta u)^2]$.

We now augment our state-space realization (13) to :

$$G^r(s) = \left[\begin{array}{c|cc} A^r & B_1^r & B_2^r \\ \hline C_1^r & 0 & D_{12}^r \\ I & 0 & 0 \end{array}\right]. \tag{19}$$

The state vector $x = [\delta n,\, \delta m,\, r, \dot{r}]^T$ now incorporates the frequency weight, and is assumed available, see below for implementation details. The augmented matrices are:

$$A^r = \left[\begin{array}{c|c} A & 0 \\ \hline 0 & A_{22}^r \end{array}\right],\; B_1^r = \left[\begin{array}{c|c} B_1 & 0 \\ \hline 0 & B_{12}^r \end{array}\right],\; B_2^r = \left[\begin{array}{c} B_2 \\ \hline 0 \end{array}\right],$$

$$\text{with} \quad A_{22}^r = \begin{bmatrix} 0 & 1 \\ -\omega_r^2 & -2\zeta\omega_r \end{bmatrix}, \quad B_{12}^r = \begin{bmatrix} 0 \\ 3\omega_r^2 \end{bmatrix};$$

$$C_1^r = \frac{k_1}{k_2}\begin{bmatrix} -p_0 u^* & -p_0 & \theta p^* & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, D_{12}^r = \frac{1}{k_2}\begin{bmatrix} -k_1 p_0 n^* \\ 1 \end{bmatrix}.$$

The resulting $\mathcal{H}_2$-optimal control law is

$$\delta u = -F^r x = -(B_2^{r*} X^r + D_{12}^{r}{}^* C_1^r)x \tag{20}$$

where X^r is the solution to the corresponding Algebraic Riccati Equation.

As before, the tradeoff parameter k_1 is set with the aid of simulations, seeking the least possible tracking error variance that keeps $u(t)$ from saturating. For the load parameters of the previous section, and an offered regulation of $\theta = \theta_{max} = 0.66$, $k_1 = \frac{6}{n^*u^*}$ provides an adequate tradeoff.

V. Distributed Implementation and Performance

A first implementation question is how does the aggregator obtain the state vector information required for this feedback law. In regard to the states n and m, they are relatively easy to track: the aggregator only needs to be notified when each load arrives, runs out of laxity or leaves the system, a minimal communication requirement.

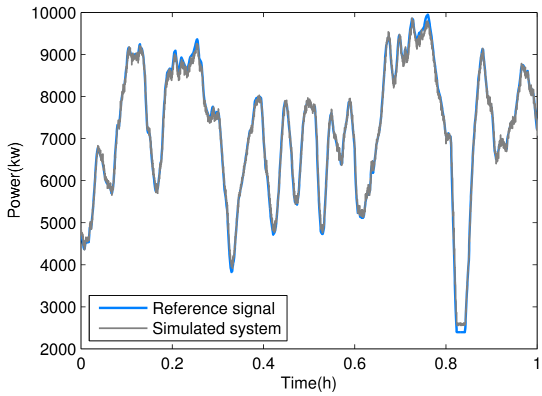

Fig. 6. Tracking a regulation signal with deferrable, interruptible loads.

The regulation signal $r(t) = \theta p^* \rho(t)$ is obtained directly from the SO. While $\dot{r}(t)$ is not directly available, we note that in practice, $\rho(t)$ is communicated very frequently, e.g. every $T_s = 4$ seconds, a sampling rate of $0.25Hz$. On the other hand the bandwidth of r is much lower, in the previously fit model we have $f_r \approx 2.6.10^{-3}Hz$. This implies that a simple estimate $\dot{r}(t) \approx \frac{r(t)-r(t-T_s)}{T_s}$ has high accuracy for control purposes, so it is not justified to employ a more complex dynamic control (including the corresponding Kalman filter) solely for tracking this state variable.

The remaining issue is implementing the actuation part, i.e. how to impose the deferral action u on the loads. The most faithful implementation is to have the aggregator broadcast the signal $u(t)$ every T_s when the value of r changes. Upon receiving it, loads with adjustable power can set their consumption to up_0; loads of the ON/OFF type can make such decision randomly with probability u. For a large enough number of loads, the aggregate service corresponds to nu active loads. In addition, any load with expired laxity turns itself on, and thus is guaranteed to meet its deadline.

In Fig. 6 we show one hour of a day-long simulation using this technique and our $\mathcal{H}_2$ controller. All parameters are as before except $\lambda = 2\, loads/s$. The RMS error was found to be $113kW$, a 2.4% of the total regulation offered.

For additional validation we computed for our controller a performance score used by PJM [16] to rank regulation resources. This score is calculated comparing the reference signal with the actual response form the system and is the average of three components: correlation, delay and precision; all of them measured in a scale from 0 to 1. A score of 0.75 is required for participation in the market, and values above 0.9 are considered excellent. The score corresponding to our control system was found to be 0.97, indicating a highly satisfactory performance.

One limitation of the preceding solution in regard to ON/OFF loads, is that they may be interrupted multiple times during their service. A different approach is required to handle non-interruptible loads, or those with a limit on the number of interruptions. Space limitations preclude us from addressing this here, we refer to [3] for more details.

VI. Conclusions

We have shown a methodology for an aggregator of deferrable loads to provide frequency regulation to the electric grid. Its main feature is the simplicity of the macroscopic model, in terms of an ODE with a scalar control signal. This enables the use of $\mathcal{H}_2$-optimal control to design a controller to track power directives from the SO, which achieves an excellent performance using the industry standard. It also allows for a streamlined interface between aggregator and loads for a decentralized implementation.

The main implementation challenge which remains is to handle constraints on service interruptions. Partial results in this direction are given in [3]; improving them is a topic of future research.

References

[1] F. Bliman, A. Ferragut, and F. Paganini, "Controlling aggregates of deferrable loads for power system regulation," in *Proc. of the American Control Conference*, 2015.

[2] ——, "Managing load deferability to provide power regulation," in *Proc. IEEE ISGT-LA*, 2015.

[3] F. Bliman, "Frequency regulation in electric power systems using deferrable loads," Master's thesis, Universidad ORT Uruguay, 2016.

[4] J. Burl, *Linear Optimal Control: $\mathcal{H}_2$ and $\mathcal{H}_\infty$ methods*. Menlo Park, CA: Addison Wesley, 1999.

[5] EnergyPool, http://www.energy-pool.eu/en/.

[6] A. Ferragut and F. Paganini, "Queueing analysis of service deferrals for load management in power systems," in *Proc. of Allerton Conference, Monticello, IL*, 2015.

[7] S. Goguri, J. Hall, R. Mudumbai, and S. Dasgupta, "A distributed, real-time and non-parametric approach to demand response in the smart grid." in *Proc. Conf. on Inf. Sciences and Systems (CISS)*, 2015.

[8] GoodEnergy, http://www.goodenergy.com.

[9] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent, "Frequency Regulation from Flexible Loads: Potential, Economics, and Implementation," in *Proc. of the American Control Conference*, 2014.

[10] A. Ipakchi and F. Albuyeh, "Grid of the Future," *IEEE power & energy magazine*, vol. 7, pp. 52–62, 2009.

[11] C. ISO, "Demand Response and Energy Efficency Roadmap: Maximizing preferred resources," http://www.caiso.com/documents/dr-eeroadmap.pdf.

[12] S. Koch, J. Mathieu, and D. Callaway, "Modeling and Control of Aggregated Heterogeneous Thermostatically Controlled Loads for Ancillary Services," in *Power Systems Computation Conference*, 2011.

[13] Y. Lin, P. Barooah, S. Meyn, and T. Middelkoop, "Demand Side Frequency Regulation from Commercial Building HVAC Systems: An Experimental Study," in *American Control Conference*, 2015.

[14] A. Nayyar, J. Taylor, A. Subramanian, K. Poolla, and P. Varaiya, "Aggregate Flexibility of a Collection of Loads," in *Proc. of the 52nd IEEE Conference on Decision and Control*, 2013.

[15] PJM, "Fast Response Regulation Signal," http://www.pjm.com/ markets-and-operations/ ancillary-services /mkt-based-regulation/ fast-response-regulation-signal.aspx.

[16] ——, "PJM Manual 12: Balancing Operations," http://www.pjm.com/ /media/documents/manuals/m12.ashx.

[17] P. Robert, *Stochastic networks and queues*. Berlin: Springer-Verlag, 2003.

[18] E. Santacana, G. Rackliffe, L. Tang, and X. Feng, "Getting Smart," *IEEE power and energy magazine*, vol. 8, pp. 41–48, 2010.

[19] A. Subramanian, M. Garcia, D. Callaway, K. Poolla, and P. Varaiya, "Real-Time Scheduling of Distributed Resources," *IEEE Transactions on Smart Grid*, vol. 4, pp. 2122–2130, 2013.

[20] S. Tindemans, V. Trovato, and G. Strbac, "Decentralized Control of Thermostatic Loads for Flexible Demand Response," *IEEE Transactions On Control Systems Technology*, vol. 23, pp. 1685–1700, 2015.

[21] H. Zarkoob, S. Keshav, and C. Rosenberg, "Optimal contracts for providing load-side frequency regulation service using fleets of electric vehicles," *Journal of Power Sources*, vol. 241, pp. 94–111, 2013.

[22] K. Zhou, J. Doyle, and K. Glover, *Robust and Optimal Control*. Prentice Hall, 1996.