

Optimal Load Control via Frequency Measurement and Neighborhood Area Communication

Changhong Zhao, Ufuk Topcu, *Member, IEEE*, and Steven H. Low, *Fellow, IEEE*

Abstract—We propose a decentralized optimal load control mechanism that provides contingency reserve in the presence of sudden demand-supply mismatch. The mechanism takes advantage of flexibility of frequency responsive loads and neighborhood area communication to solve an optimization problem that balances demand and supply while minimizing end-use disutility. Frequency measurement enables individual loads to estimate the total demand-supply mismatch. Neighborhood area communication helps mitigate effects of inconsistencies in the local estimates due to, e.g., frequency measurement noise. Case studies show that the proposed mechanism can balance demand with supply and restore the frequency on a timescale of seconds following a contingency, even when the loads use a highly simplified system model in their algorithms. We also investigate tradeoffs between the amount of communication and performance of the mechanism through simulation-based experiments.

Index Terms—Frequency responsive load, neighborhood area communication, distributed control, optimization

I. INTRODUCTION

In power systems, contingency reserves restore the generation and load balance after sudden unexpected loss of a generator or transmission line; otherwise the frequency may deviate from its nominal level. Such frequency deviations, if not resolved in a short period, may lead to instability or even cause damage to facilities [1]. Hence, deployment of contingency reserves is important in power systems.

Recovery mechanisms following contingencies have traditionally been provided by generators. Generators are equipped with speed governors, which adjust turbines instantaneously in response to a change in frequency caused by imbalance between generation and load [2]. Governors provide limited adjustment around the setpoint of generators. Hence, it is a resort to maintain machine stability rather than a contingency reserve. In case of a major generation disruption, spinning reserve is required where a set of power sources, which have already been connected and synchronized to the grid, increase their generation level. Spinning reserve is initiated within seconds in response to a major generator or transmission outage, and typically reaches full output within 10 minutes. However, the grid-connected and partially-loaded generators providing spinning reserve are of low efficiency and increase fuel costs and emissions [3]. Other contingency reserves, such as supplemental reserve and replacement reserve, are provided

by offline generators. They are much slower with a response time of 10 minutes to hours [5].

To complement these generation-side measures, loads provide an attractive alternative toward matching demand with supply. In 2002, Long Island Power Authority (LIPA) developed LIPAedge, which was the largest residential load control program in US. LIPAedge provided 24.9 MW of demand reduction during hot summer days and could deliver 75 MW of 10-minute spinning reserve [4]. Later studies demonstrated that loads can provide even faster service and save large amount of communication by using local frequency measurement. Through advanced metering, loads can sense the frequency as a proxy of demand-supply mismatch and respond accordingly in less than a second [6]. The Grid Friendly Appliance Controller [7] developed by Pacific Northwest National Laboratory suggests that appliances can provide fast reserve within seconds by responding to certain trends in the frequency. Trudnowski *et al.* designed a mechanism in which loads take proportional negative-feedback of frequency deviation [8][9]. Simulation-based studies show that the mechanism recovers generation and load balance within several seconds after a sudden generation drop.

In this paper, we propose a frequency responsive load control mechanism that provides fast contingency reserve with moderate amount of communication. We formulate an optimization problem that minimizes global end-use disutility caused by change in loads, and matches loads with generation. Then, we develop a decentralized algorithm to solve the optimization problem. In the algorithm, loads estimate total demand-supply mismatch from local frequency measurement, based on a power system model characterizing the dynamics of frequency deviation. Such local estimates are used as the increments in a distributed gradient method to solve the dual problem of the proposed optimization problem. To compensate for inconsistencies between local estimates caused by, e.g., frequency measurement noise, loads exchange estimates with their neighbors via a communication network. With moderate assumptions on the disutility function, dynamic model, measurement noise and communication graph, the almost sure convergence of the proposed algorithm is proved. Case studies show that the proposed mechanism recovers the balance between generation and load within seconds after a sudden generation loss. The mechanism is robust to modeling inaccuracies in the sense that it performs well when loads use a highly simplified and inaccurate dynamic model to compute the local estimates. Moreover, tradeoffs between the amount of communication and performance of the load control mechanism are discussed. Simulations show that moderate

This work is supported by NSF NetSE grants CNS 0911041, Southern California Edison (SCE), Okawa Foundation, Boeing Corporation, and Cisco. C. Zhao, U. Topcu and S. H. Low are with the Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA. e-mail: (czhao@caltech.edu, utopcu@cds.caltech.edu, slow@caltech.edu).

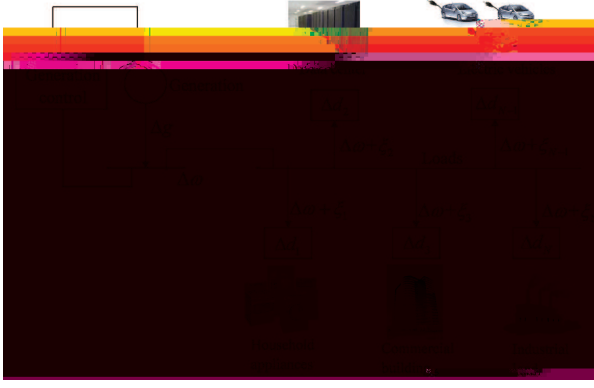


Fig. 1. Schematic diagram of the power system under consideration.

amount of communication ensures good performance in the presence of measurement noise.

The paper is organized as follows. Section II describes a power system model and the optimal load control problem. Section III introduces the approach of estimating total demand-supply mismatch from local frequency measurement. Section IV presents the decentralized load control algorithm and proves its convergence. Section V shows the results of case studies. Finally, Section VI provides concluding remarks and discusses interesting problems for future work.

II. PROBLEM SETUP

We now introduce a dynamic power system model in which frequency responsive loads are controlled. We then set up an optimal load control problem, which is solved by the load control mechanism that will be proposed later.

A. Setup

We consider a setup of power system model shown in Fig. 1. In the model, we assume all generators and loads are within a strongly connected area [10]. Hence, there are negligible differences between phase angles at different sites, and a single frequency appears throughout the system. Therefore, it is reasonable to assume that the model contains a major generating unit which has equivalent dynamics to all the generators combined. The major generating unit is equipped with generation control components, such as a speed governor. In the model, there are a number of controllable loads, each of which may also be considered as an aggregate of multiple, smaller responsive loads that can be controlled [3][6].

Let $V = \{1, 2, \dots, N\}$ denote the set of loads. In case of a contingency, a change in real power supply, denoted by Δg , occurs. To compensate for Δg , load i will be changed by Δd_i through load control. Let $\Delta \omega$ denote the frequency deviation from its nominal value (e.g., 60Hz). Load i locally measures this frequency deviation. Let $\Delta \bar{\omega}_i$ be this measured value, which may be different from $\Delta \omega$ by a measurement noise ξ_i . Besides, the model contains a stochastic disturbance to generation, denoted by ζ , which may come from changes in system parameters caused by environment influence [11].

The load control mechanism we are to propose is a sampled control system, i.e., loads sample $\Delta \omega$ and make decisions once

every Δt time. Hence, we use $t = 0, 1, 2, \dots$ to denote the time instances $t = 0, \Delta t, 2\Delta t, \dots$. Let $\Delta d_i(t)$ stand for the amount of decrease in load i at time t . Then, the total mismatch between demand and supply at time t is

$$u(t) := - \sum_{i \in V} \Delta d_i(t) - \Delta g(t). \quad (1)$$

The dynamic model of the relation between u , $\Delta \omega$ and $\Delta \bar{\omega}_i$ is described by

$$\begin{aligned} \mathbf{x}(t+1) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}u(t) + \zeta(t) \\ \Delta \omega(t) &= \mathbf{C}\mathbf{x}(t) \\ \Delta \bar{\omega}_i(t) &= \Delta \omega(t) + \xi_i(t), \end{aligned} \quad (2)$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ is the internal states of the system at time t .

We assume, for all $t, s \geq 0$ and all $i, j \in V$, that the process disturbance ζ and the measurement noise ξ_i have zero mean, and their covariances satisfy

$$\mathbb{E}[\zeta(t)\zeta(s)^T] = \mathbf{Q}\delta_{ts}, \quad \mathbb{E}[\xi_i(t)\xi_j(s)] = \mathbf{W}\delta_{ts}\delta_{ij}, \quad (3)$$

where $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is positive semi-definite, $\mathbf{W} \geq 0$, and δ_{ts} and δ_{ij} denote the Kronecker delta function.

B. Optimal load control problem

We now present the optimal load control problem. Suppose a sudden change in generation, $\Delta g < 0$ without loss of generality, occurs at time 0. In response to such a change, the load control mechanism sheds load i by $\Delta d_i \in [0, \bar{d}_i]$, which involves a disutility or service cost $D_i(\Delta d_i)$, as a result of interrupting the normal usage or compromising end-use function of appliances [3][4]. Here \bar{d}_i is the maximum decrease in load i allowed by appliance design or user preference. The optimal load control problem, denoted by **PP**, is

$$\mathbf{PP} \left\{ \begin{array}{ll} \min_{\Delta d_i \in [0, \bar{d}_i]} & \sum_{i=1}^N D_i(\Delta d_i) \\ \text{subject to} & \sum_{i=1}^N \Delta d_i + \Delta g = 0. \end{array} \right.$$

The objective of the problem is to minimize the global disutility of shedding the loads in a way that eliminates the load-generation mismatch $u = - \sum_{i=1}^N \Delta d_i - \Delta g$.

For feasibility of **PP**, we assume $\sum_{i \in V} \bar{d}_i + \Delta g > 0$. This assumption holds if a large enough group of loads participate in load control. Additionally, we make the following two assumptions on the disutility functions D_i , so that **PP** is a convex problem which can be solved using the decentralized algorithm proposed later.

Assumption 1: For $i \in V$, D_i is increasing, strictly convex and twice continuously differentiable over $[0, \bar{d}_i]$.

Assumption 2: For $i \in V$, there exists $\alpha_i > 0$ so that $D'_i(\Delta d_i) \geq 1/\alpha_i$ for $\Delta d_i \in [0, \bar{d}_i]$.

Power system operators usually solve a problem like **PP** using a centralized method, which requires a central coordinator and two-way communication between loads and the coordinator. Load-to-coordinator communication is used to collect time-varying information about availability, flexibility or the disutility function of loads. The coordinator then formulates

an optimization problem like **PP** and computes the optimal dispatch of loads. Then, the dispatch signals are issued to loads via coordinator-to-load communication. To avoid a large number of simultaneous connections overwhelming the communication system, the coordinator staggers the time intervals during which it communicates with different loads. Hence, it may take a long time for a large population of loads to respond. For example, it typically requires 90 minutes for 20,000 thermostats to respond in LIPAnedge [4].

To speed up load response, we design a decentralized algorithm where a central coordinator is no longer essential. To this end, we consider solving the dual problem of **PP**. Taking \mathbf{p} as the dual variable, the dual problem of **PP** is

$$\max_{\mathbf{p} \in \mathbb{R}} \Psi(\mathbf{p}) := \sum_{i=1}^N \Psi_i(\mathbf{p}) - \mathbf{p} \Delta \mathbf{g}, \quad (4)$$

where

$$\Psi_i(\mathbf{p}) := \min_{\Delta \mathbf{d}_i \in [0, \bar{\mathbf{d}}_i]} \mathbf{D}_i(\Delta \mathbf{d}_i) - \mathbf{p} \Delta \mathbf{d}_i.$$

Under Assumption 1, given $\mathbf{p} \in \mathbb{R}$, the problem

$$\min_{\Delta \mathbf{d}_i \in [0, \bar{\mathbf{d}}_i]} \mathbf{D}_i(\Delta \mathbf{d}_i) - \mathbf{p} \Delta \mathbf{d}_i \quad (5)$$

has a unique minimizer given by

$$\Delta \mathbf{d}_i(\mathbf{p}) = \min\{\max\{(\mathbf{D}_i')^{-1}(\mathbf{p}), 0\}, \bar{\mathbf{d}}_i\}. \quad (6)$$

Note the inverse of \mathbf{D}_i' exists over $[\mathbf{D}_i'(0), \mathbf{D}_i'(\bar{\mathbf{d}}_i)]$ since \mathbf{D}_i is continuous and strictly increasing by Assumption 1. Since \mathbf{D}_i is convex for all $i \in V$ and **PP** has affine constraints, Slater's condition implies that there is zero duality gap between **PP** and its dual problem in (4), and the optimal solution of (4), denoted by \mathbf{p}^* , is attained [14, Sec. 5.5.3]. It follows that $\Delta \mathbf{d}(\mathbf{p}^*) := [\Delta \mathbf{d}_1(\mathbf{p}^*), \dots, \Delta \mathbf{d}_N(\mathbf{p}^*)]^T$ is primal feasible and optimal [14, Sec. 5.5.2]. Moreover, it is easy to show that, for any given $\underline{\mathbf{p}}$ and $\bar{\mathbf{p}}$ such that $\underline{\mathbf{p}} \leq \min_i \mathbf{D}_i'(0)$ and $\bar{\mathbf{p}} \geq \max_i \mathbf{D}_i'(\bar{\mathbf{d}}_i)$, the problem in (4) has at least one optimal point $\mathbf{p}^* \in [\underline{\mathbf{p}}, \bar{\mathbf{p}}]$. Hence, we can constrain \mathbf{p} to $[\underline{\mathbf{p}}, \bar{\mathbf{p}}]$. Therefore, we focus on solving the following problem

$$\mathbf{DP} \left\{ \begin{array}{l} \max_{\mathbf{p} \in [\underline{\mathbf{p}}, \bar{\mathbf{p}}]} \Psi(\mathbf{p}) = \sum_{i=1}^N \Psi_i(\mathbf{p}) - \mathbf{p} \Delta \mathbf{g}. \end{array} \right. \quad (7)$$

The decentralized algorithm can be given informally as follows (see Section IV-A for a formal treatment). Each load $i \in V$ updates its value of dual variable \mathbf{p} at time t as

$$\mathbf{p}_i(t) = \max\{\min\{\mathbf{p}_i(t-1) + \gamma(t)\mathbf{u}(t-1), \bar{\mathbf{p}}\}, \underline{\mathbf{p}}\}, \quad (8)$$

where $\gamma(t) > 0$ is some stepsize. Then, load i calculates its amount of load shedding at time t as $\Delta \mathbf{d}_i(t) = \Delta \mathbf{d}_i(\mathbf{p}_i(t))$, where $\Delta \mathbf{d}_i(\cdot)$ is defined in (6). As shown in [13, Section III-A], the total demand-supply mismatch at time $t-1$, denoted by $\mathbf{u}(t-1)$, is the gradient of the dual objective function Ψ . Therefore, this decentralized algorithm is essentially a gradient projection method [14] applied on the dual problem **DP**. The decentralized algorithm does not rely on communication between loads and a central coordinator, thus can improve load response speed according to the discussion before. To implement this algorithm with frequency responsive loads,

loads should be able to estimate \mathbf{u} from local frequency measurement. A candidate estimation mechanism is introduced in Section III.

III. ESTIMATING DEMAND-SUPPLY MISMATCH

We now introduce a mechanism through which loads can estimate total demand-supply mismatch \mathbf{u} from measured frequency deviation. Since \mathbf{u} is the input in the model given by (2), we call this mechanism an *input estimator*.

The input estimator of load i uses noisy frequency measurement $\Delta \bar{\omega}_i(1), \dots, \Delta \bar{\omega}_i(t)$ to estimate $\mathbf{u}(0), \dots, \mathbf{u}(t)$ using the power system model in (2). Use $\hat{\mathbf{x}}_i(t|s)$ to denote the estimate of $\mathbf{x}(t)$ and $\hat{\mathbf{u}}_i(t|s)$ to denote the estimate of $\mathbf{u}(t)$ with measurement up to time s . Starting from $s=0$, the input estimator of load i is given recursively by

$$\begin{aligned} \hat{\mathbf{u}}_i(t-1|t) &= \mathbf{M}(\Delta \bar{\omega}_i(t) - \mathbf{C}\hat{\mathbf{x}}_i(t|t-1)) \\ \hat{\mathbf{x}}_i(t|t) &= \hat{\mathbf{x}}_i(t|t-1) + \mathbf{B}\hat{\mathbf{u}}_i(t-1|t) \\ \hat{\mathbf{x}}_i(t+1|t) &= \mathbf{A}\hat{\mathbf{x}}_i(t|t), \end{aligned} \quad (9)$$

where $\mathbf{M} = (\mathbf{CB})^{-1}$. We assume $\mathbf{CB} \neq 0$. As a particular case, this assumption holds for the power system model we have used, in particular for the model in [15].

The input estimator is essentially the same as proposed by Kitanidis [15], which gives unbiased minimum variance estimate of the state and the input. The covariance of $\mathbf{x}_i(t|t)$, denoted by $\Sigma_{i|t}^i \in \mathbb{R}^{n \times n}$, is given recursively by

$$\begin{aligned} \Sigma_{i+1|t+1}^i &= (\mathbf{I}_n - \mathbf{BMC})(\mathbf{A}\Sigma_{i|t}^i \mathbf{A}^T + \mathbf{Q}) + \mathbf{BWM}^T \mathbf{B}^T, \end{aligned} \quad (10)$$

where \mathbf{Q} and \mathbf{W} are defined as in (3). The estimate error by $\mathbf{e}_i(t) := \mathbf{x}_i(t) - \hat{\mathbf{x}}_i(t|t)$

has a unique, positive definite solution Σ^* . Additionally, $\lim_{t \rightarrow \infty} \Sigma_t^i$ exists and is Σ^* . By (12), we have

$$\lim_{t \rightarrow \infty} \mathbb{E} \left[(e_i(t))^2 | \mathbf{F}_{t-1} \right] = \frac{CA\Sigma^*A^TC^T + W}{(CB)^2},$$

where the right hand side is independent of i , and can be determined from A, B, C, Q and W . ■

For any power system model in the form of (2), we can check whether the condition in Proposition 1 is satisfied a priori. As a partial justification, it holds for the model we use in the case studies. Understanding implications of this condition needs future study.

To prove the convergence of Algorithm 1 below, we need to bound the variance of the input estimate error. The following corollary, which is a straightforward consequence of Proposition 1, gives such a bound.

Corollary 1: If the condition for Proposition 1 holds, then $\mathbb{E} \left[(e_i(t))^2 | \mathbf{F}_{t-1} \right] \leq \bar{\sigma}^2$ for all $i \in V$ and all $t \geq 0$, where $\bar{\sigma}^2$ is some constant determined by A, B, C, Q, W and the initial covariance Σ_{00}^i of all $i \in V$.

With the input estimator introduced above, we now give the decentralized load control algorithm formally and prove its convergence in Section IV.

IV. LOAD CONTROL ALGORITHM

In this section, we first give a formal introduction of the decentralized algorithm that solves the optimal load control problem. Then, we discuss the communication architecture that supports the information exchange in this algorithm. Finally, we present the main result regarding the convergence of the proposed algorithm.

A. Decentralized load control algorithm

The decentralized algorithm has already been informally discussed in Section II-B. The dual variable update in (8) calls for estimating u locally. As shown in Section III, there may be inconsistencies between local estimates due to frequency measurement noise. We use neighborhood communication between loads to compensate for such inconsistencies. Define the set of loads that communicate directly with load i at time t to be $\mathbf{N}(i, t)$. Load i assigns a weight $r_{ij}(t)$ for all $j \in \mathbf{N}(i, t)$, and a weight $r_{ii}(t)$ for itself. Conditions on the weights will be discussed later in Section IV-B. Through neighborhood communication, load i receives the value of dual variable $p_j(t)$ from all $j \in \mathbf{N}(i, t)$, and calculates an average value of dual variable, denoted by $q_i(t)$, as

$$q_i(t) = \sum_{j=i, j \in \mathbf{N}(i, t)} r_{ij}(t) p_j(t). \quad (15)$$

This averaging procedure is typically used in consensus algorithms [17]. The load control algorithm with neighborhood communication is presented formally as follows.

Algorithm 1: Decentralized optimal load control algorithm

At time $t = 0$, the following information is known to all loads $i \in V$: the power system model which contains matrices A, B and C , the lower bound \underline{p} and upper bound \bar{p} defined



Fig. 2. (From [16]) The home area network (HAN) involves the communication between appliances and smart meters. The neighborhood area network (NAN), which is used in the load control mechanism, aids the communication between utilities and smart meters within distribution networks. The wide area network (WAN) aids the long range communication between substations.

in Section II-B, and a sequence of positive stepsize $\{\gamma(t), t = 1, 2, \dots\}$ which is the same for all loads.

Each load i starts from arbitrary initial state estimate $\hat{x}_i(1|0)$ and initial value of dual variable $q_i(0)$.

At times $t = 1, 2, \dots$, load i :

- 1) Measures the frequency deviation $\Delta\bar{\omega}_i(t)$, and calculates $\hat{u}_i(t-1|t)$ using the input estimator (9).
- 2) Updates the dual variable according to

$$p_i(t) = \max\{\min\{q_i(t-1) + \gamma(t)\hat{u}_i(t-1|t), \bar{p}\}, \underline{p}\}, \quad (16)$$

and transmits $p_i(t)$ to all loads $j \in \mathbf{N}(i, t)$.

- 3) Receives the value of $p_j(t)$ from all $j \in \mathbf{N}(i, t)$, and calculates $q_i(t)$ as (15).
- 4) Computes the change in its load as $\Delta d_i(t) = \Delta d_i(q_i(t))$.

Before presenting the main result regarding Algorithm 1, we first introduce the neighborhood area communication architecture supporting the information exchange in (15), as well as some conditions on the weights $r_{ij}(t)$ for the convergence proof of Algorithm 1.

B. Neighborhood area communication

As an example, we take the smart grid communication architecture proposed by Trilliant Inc. [16] shown in Fig. 2. Algorithm 1 does not rely on communication between loads and a central coordinator, and instead takes advantage of communication between a load and its neighbors. This neighborhood communication uses mainly the neighborhood area network (NAN). In NAN, reliable, scalable, fast responding and cost-effective communication technologies such as 802.2.15.4/ZigBee are widely used [16], which greatly facilitate the implementation of decentralized load control.

For the convergence proof of Algorithm 1, we make the following assumptions on the weights $r_{ij}(t)$ in (15).

Assumption 3: There exists a scalar $0 < \eta < 1$ such that for all $i = 1, \dots, N$ and all $t \geq 0$, $r_{ij}(t) \geq \eta$ if $j = i$ or $j \in \mathbf{N}(i, t)$, and $r_{ij}(t) = 0$ otherwise.

Assumption 4: For all $i = 1, \dots, N$ and all $t \geq 0$, $\sum_{j=1}^N r_{ij}(t) = 1$, and $\sum_{j=1}^N r_{ji}(t) = 1$.

Note that with Assumption 3, equation (15) simplifies to

$$q_i(t) = \sum_{j=1}^N r_{ij}(t) p_j(t). \quad (17)$$

Moreover, in order to make the information at load j impact load i infinitely often, we assume that within any finite period of time, the set of links which have ever appears form a strongly connected graph. Define $E_t := \{(i, j) | r_{ij}(t) > 0\}$ to be the set of links at time t . The connectivity requirement above is formally stated in the following assumption.

Assumption 5: There exists a integer $Q \geq 1$ such that the graph $(V, \bigcup_{\tau=1, \dots, Q} E_{t+\tau-1})$ is strongly connected for all t .

Define $R(t)$ to be the matrix with (i, j) -th entry $r_{ij}(t)$, and define $\Phi(t, s) := R(t)R(t-1)\dots R(s+1)$. The following result given by [17, Lemma 3.2] will be used in the convergence proof of Algorithm 1:

$$\left| [\Phi(t, s)]_{ij} - \frac{1}{N} \right| \leq \theta \beta^{t-s}, \quad (18)$$

where

$$\theta = \left(1 - \frac{\eta}{4N^2}\right)^{-2}, \quad \beta = \left(1 - \frac{\eta}{4N^2}\right)^{\frac{1}{Q}}. \quad (19)$$

C. Convergence of Algorithm 1

Now we present some results regarding the convergence of Algorithm 1. We first consider the case where the stepsize sequence $\{\gamma(t), t = 1, 2, \dots\}$ converges to some nonnegative constant. Theorem 1 gives a bound between the maximum expected value of the dual objective function Ψ and the optimal value of **DP**, denoted by Ψ^* .

Theorem 1: Suppose Assumptions 1-5 hold. If $\lim_{t \rightarrow \infty} \gamma(t) = \gamma \geq 0$ and $\sum_{t=1}^{\infty} \gamma(t) = \infty$, then, for all $i \in V$,

$$\limsup_{t \rightarrow \infty} \mathbb{E}[\Psi(p_i(t))] \geq \Psi^* - \frac{\gamma(G^2 + \bar{\sigma}^2)}{2} - \gamma G(\bar{\alpha}NL + G) \left(2 + \frac{N\theta\beta}{1-\beta}\right), \quad (20)$$

where $G := \max\{\sum_{i=1}^N \bar{d}_i + \Delta g, |\Delta g|\}$, and $L := \bar{p} - \underline{p}$. Recall that $\bar{\alpha} = \max_{i \in V} \alpha_i$, with α_i defined in Assumption 2, and $\bar{\sigma}^2$ is defined in Corollary 1.

Due to space limitation, we skip the proof of Theorem 1 and have it in our technical report [19, Appendix A].

Taking $\gamma = 0$ in (20), it implies that when the stepsize $\gamma(t)$ diminishes with increasing t , the maximum expected value of Ψ achieved by Algorithm 1 is exactly the optimal value of the dual problem **DP**.

Corollary 2: Suppose Assumptions 1-5 hold. If $\lim_{t \rightarrow \infty} \gamma(t) = 0$ and $\sum_{t=1}^{\infty} \gamma(t) = \infty$, then, for all $i \in V$,

$$\limsup_{t \rightarrow \infty} \mathbb{E}[\Psi(p_i(t))] = \Psi^*.$$

With further restrictions on the stepsize $\gamma(t)$, the sequence generated by each load using Algorithm 1 almost surely converges to the same optimal point of the dual problem **DP**. In addition, the change in load $\Delta d_i(t)$ almost surely converges to the solution of the optimal load control problem **PP**. This result is formally stated in Theorem 2.

Theorem 2: Suppose Assumptions 1-5 hold. When $\sum_t \gamma(t) = \infty$ and $\sum_{t=1}^{\infty} \gamma(t)^2 < \infty$, the sequence $\{q_i(t), t = 1, 2, \dots\}$ of each load $i \in V$ converges to the same optimal point of the dual problem **DP** with probability 1 and in mean square. Moreover, define $\Delta d(t) = [\Delta d_1(t), \dots, \Delta d_N(t)]^T$. Then the sequence $\{\Delta d(t), t = 1, 2, \dots\}$ converges to the optimal point of the primal problem **PP** with probability 1.

Proof: With Assumptions 1-5, Equation (11) and Corollary 1, the conditions for [17, Theorem 6.2] are satisfied. We follow the same technique and get that for all $i \in V$, the sequence $\{q_i(t), t = 1, 2, \dots\}$ converges to the same optimal point of the dual problem **DP** with probability 1 and in mean square. Define $\Delta d(q(t)) = [\Delta d_1(q_1(t)), \dots, \Delta d_N(q_N(t))]^T$. By Section II-B, the sequence $\{\Delta d(q(t)), t = 1, 2, \dots\}$ converges to the optimal point of the primal problem **PP** with probability 1. By $\Delta d(t) = \Delta d(q(t))$, the theorem follows. ■

In Algorithm 1, communication is used to mitigate the effect of measurement noise. Now we show a special case where the disturbance ζ and the measurement noise ξ_i for all $i \in V$ are small enough to be omitted. In such a case, Algorithm 1 still converges while no communication between loads is required. This result is formally stated as follows.

Theorem 3: Suppose $\zeta(t) = 0$ and $\xi_i(t) = 0$ for all $i \in V$ and $t \geq 0$. Suppose Assumptions 1 and 2 hold. In Algorithm 1, load $i \in V$ has $r_{ii}(t) = 1$ and $N(i, t) = \emptyset$ for all $i \in V$ and $t \geq 0$. If constant stepsize $\gamma(t) = \gamma$ is used, where γ satisfies $0 < \gamma < \frac{2}{\bar{\alpha}N}$, then the sequence $\{q_i(t), t = 1, 2, \dots\}$ for all $i \in V$ converges to the same optimal point of the dual problem **DP**. Moreover, the sequence $\{\Delta d(t), t = 1, 2, \dots\}$ converges to the optimal point of the primal problem **PP**.

Due to space limitation, we skip the proof of Theorem 3 and have it in our technical report [13, Appendix A].

V. CASE STUDIES

In this section, we take a relatively detailed power system model for simulation-based experiments. We use Algorithm 1 to control the loads when a sudden generation drop occurs, and observe frequency, change in load and global end-use disutility to evaluate its performance. Additionally, we test the robustness of Algorithm 1 to modeling inaccuracies by making the loads use a simplified power system model. We also discuss tradeoffs between the amount of communication and performance of Algorithm 1.

A. System settings

We now present the power system model under consideration as shown in Fig. 3. The model contains a speed governor with the transfer function

$$G_{gov}(s) = -\frac{1}{R(1+sT_G)}$$

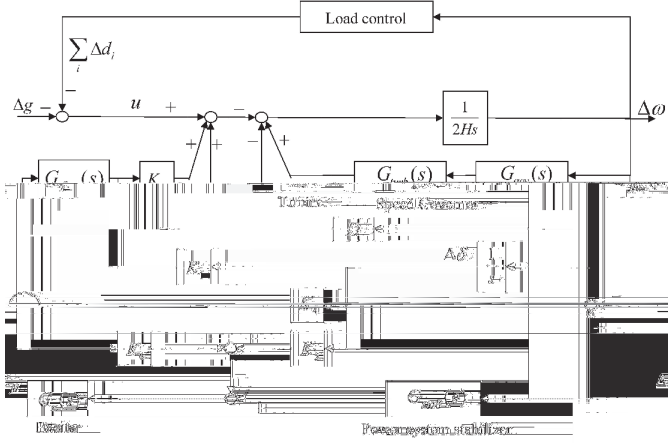


Fig. 3. Power system model used in the case studies.

TABLE I
PARAMETERS USED IN THE CASE STUDIES

Param.	Value	Param.	Value (s)	Param.	Value (s)
K_A	200	H	5	T_1	0.2
K_1	1.0755	T_A	0.04	T_2	0.02
K_2	1.2578	T_B	12	T_3	0.4
K_3	0.3072	T_C	1	T_4	0.04
K_4	1.7124	τ'_{d0}	5.9	T_w	10
Param.	Value	Param.	Value (s)	Param.	Value (pu)
K_5	-0.0409	T_G	0.2	D	1
K_6	0.4971	T_{CH}	0.3	R	0.05
K_w	20	T_{RH}	7	F_{HP}	0.3

and a turbine with the transfer function

$$G_{turb}(s) = \frac{(1 + sF_{HP}T_{RH})}{(1 + sT_{CH})(1 + sT_{RH})}.$$

The generator is equipped with a power system stabilizer (PSS) with the transfer function

$$G_{stab}(s) = \frac{sK_w(1 + sT_1)(1 + sT_3)}{(1 + sT_w)(1 + sT_2)(1 + sT_4)}.$$

The output voltage of the generator is regulated via an IEEE AC4A exciter [2], which has the transfer function

$$G_x(s) = \frac{K_A(1 + sT_C)}{(1 + sT_A)(1 + sT_B)}.$$

Moreover, the flux decay transfer function of the generator is

$$G_{flux}(s) = \frac{K_3}{1 + K_3\tau'_{d0}s}.$$

Table I gives parameters used in the transfer functions above. We take a sampling time $\Delta t = 0.5$ s. With the given parameters, we calculate the matrices A , B and C in (2), using the technique in [13, Section II-A].

There are $N = 100$ loads. Each load $i \in V = \{1, 2, \dots, N\}$ communicates directly to loads $\max\{i - K, 1\}, \dots, \min\{i + K, N\}$, as shown in Fig. 4. Load i has a disutility function $D_i(\Delta d_i) = (\Delta d_i)^2 / (2\bar{\alpha}_i)$, where $\bar{\alpha}_i$ is a random positive number, e.g., uniformly distributed on $[1, 3]$ in our experiments. The baseline power is $P_{base} = 200$ MVA. For $i \in V$, $\Delta d_i \in$

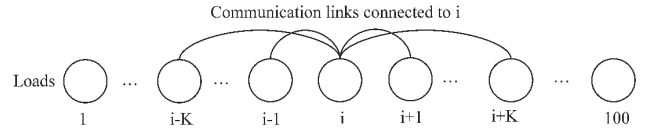


Fig. 4. Communication graph of loads in the case studies.

$[0, \bar{d}_i]$. In the case studies, we let \bar{d}_i to be a positive random number and choose $\sum_{i \in V} \bar{d}_i = 0.30$ per unit (pu). The change in generation $\Delta g(t)$ contains two step drops resembling sudden generation loss events:

$$\Delta g(t) = \begin{cases} 0 & 0 \leq t < 20 \text{ s} \\ -0.05 \text{ pu} & 20 \text{ s} \leq t < 50 \text{ s} \\ -0.15 \text{ pu} & t \geq 50 \text{ s}. \end{cases}$$

Process disturbance ζ has covariance $Q = B(0.002 \text{ pu})^2 B^T$ for B obtained above. Measurement noise ξ_i for all $i \in V$ has variance $W = (0.001 \text{ pu})^2$. In Algorithm 1, loads use a diminishing stepsize $\gamma(t) = \gamma(0)/(t^{0.8})$ for some arbitrarily selected $\gamma(0) > 0$, so $\sum_{t=1}^{\infty} \gamma(t) = \infty$ and $\sum_{t=1}^{\infty} \gamma(t)^2 < \infty$. Therefore, all the conditions in Theorems 1 and 2 are satisfied.

B. Performance and robustness to modeling inaccuracies

We compare the performance of Algorithm 1 between the settings of “accurate modeling” and “simplified modeling”. Under the setting of accurate modeling, loads use the accurate, high-order power system model given by matrices A , B and C in Algorithm 1. Under the setting of simplified modeling, loads use a highly simplified, second-order model instead. By comparison, we test the robustness of Algorithm 1 to modeling inaccuracies due to practical consideration that the utility company may not reveal the exact system information to users for privacy reasons. There are multiple ways to simplify the system model. For example, using the technique in [13, Section IV-B], we derive a simplified transfer function

$$\tilde{G}(s) = -\frac{0.1555s + 0.0222}{s^2 + 0.9918s + 0.4666}$$

for the model shown in Fig. 3.

Fig. 5-7 respectively show frequency, total change in load and global end-use disutility with loads using the accurate model or the simplified model, all with $K = 5$ in the communication graph. All the results under simplified modeling are quite close to those under accurate modeling, which suggests the proposed mechanism is robust to modeling inaccuracies. The results also show that load control drives frequency around 60 Hz, makes load following generation and minimizes global end-use disutility within seconds following generation drops, much faster than the case without load control.

The main reason for the high recovery speed is that loads are capable of sensing frequency locally and adjusting their power consumption more rapidly than synchronous machines with large inertia can ramp up. The decentralized computation and short-distance communication required in our frequency-based load control ensures that its implementation will be practical.

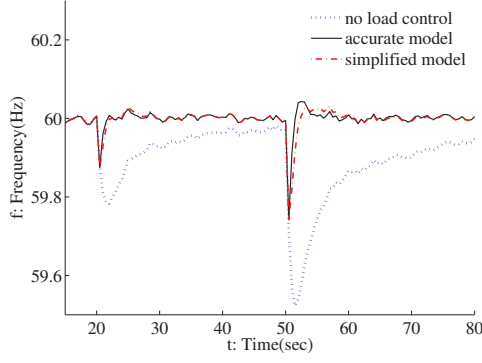


Fig. 5. Frequency curve. The dotted line is the curve without load control. The solid and dash-dot lines are the curves with loads using different models.

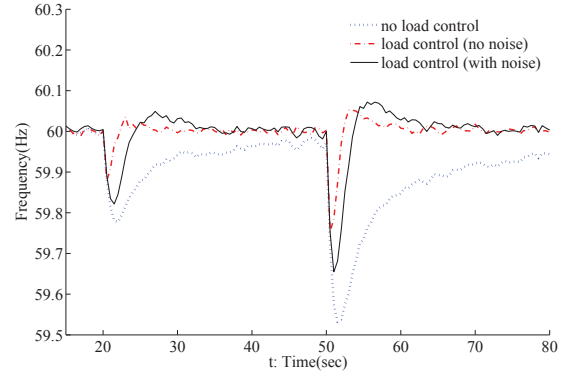


Fig. 8. Frequency curve without communication. The dotted line is the curve without load control. The solid line and the dash-dot line respectively correspond to the case with and without frequency measurement noise.

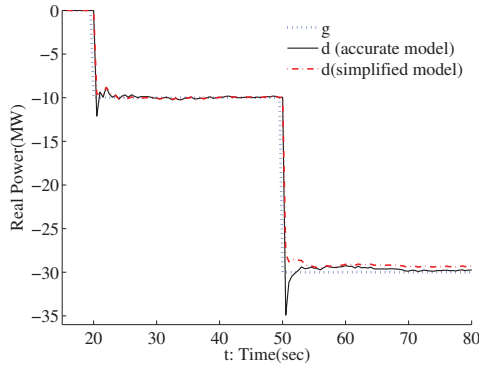


Fig. 6. Change in load. The dotted line is the change in generation. The solid and dash-dot lines are the change in load with loads using different models.

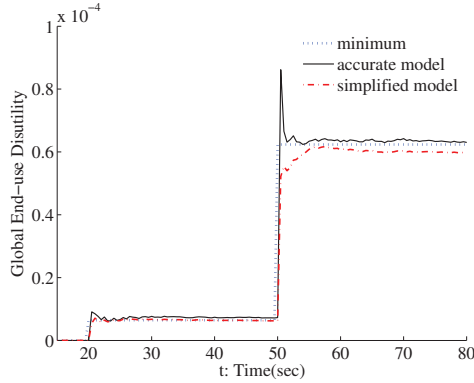


Fig. 7. Global end-use disutility. The dotted line is the minimal disutility. The solid and dash-dot lines are disutility achieved by the load control with loads using different models.

C. Tradeoff between communication and performance

Theorem 3 states the convergence of Algorithm 1 without communication between loads in the ideal case where there is no disturbance and measurement noise. However, communication is required to guarantee satisfactory performance of the proposed mechanism under practical settings where measurement noise cannot be omitted. Fig. 8-10 respectively show frequency, change in load and global end-use disutility when loads perform Algorithm 1 with $\mathbf{K} = 0$, i.e.,

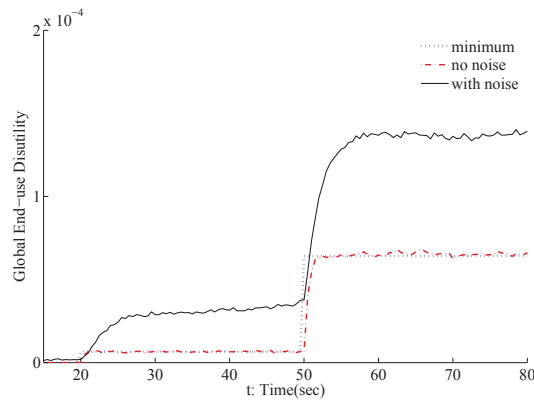


Fig. 10. Global end-use disutility without communication. The dotted line is the minimal disutility. The solid and dash-dot lines are respectively disutility achieved by the load control with and without frequency measurement noise.

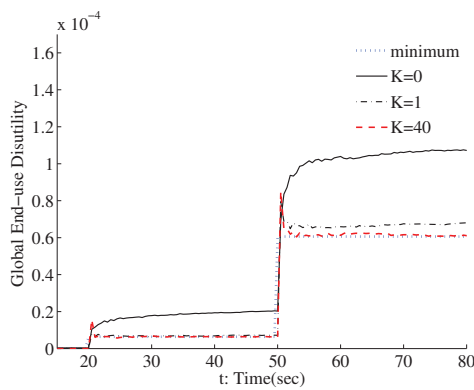


Fig. 11. Global end-use disutility with different connectivity in the communication graph. The dotted line is the minimal disutility. The solid line, the dash-dot line and the dashed line respectively correspond to the case $K=0$, $K=1$ and $K=40$.

the frequency measurement noise effectively using a moderate amount of neighborhood communication.

VI. CONCLUSIONS AND FUTURE WORK

We have proposed an optimal load control mechanism for contingency reserve. In case of a sudden drop in generation, loads are shedded to match demand with supply and minimize global end-use disutility by a decentralized algorithm. In the algorithm, loads estimate global demand-supply mismatch from local frequency measurement and accordingly shed their power. To mitigate the estimation inaccuracy caused by frequency measurement noise, inter-load communication is used. The convergence of the proposed algorithm has been proved. Simulation-based experiments have shown that the proposed mechanism could match change in load with change in generation and restore frequency to nominal value on a timescale of seconds following a contingency event, faster than spinning reserve via generators. Moreover, the proposed mechanism perform well even when loads use a simplified power system model instead of the accurate model. We have discussed the necessity of communication under measurement noise, and showed tradeoffs between the amount of communication and performance of the proposed mechanism. Case studies have

indicated that a moderate amount of communication could improve performance significantly.

Currently the load control is in a strongly connected area that contains dynamics of a single frequency. In the future, we will work on multiple interconnected areas with different frequencies at different sites. In that case, the external input of each area, denoted by $\Delta g(t)$ in this paper, becomes the power interchange between control areas. Distributed algorithm and communication requirement under such settings remain to be explored. Besides, physical network constraints such as line capacity constraints, voltage level constraints, KCL and KVL will add extra correlations between loads. We will study how the load control problem will change and what will be an effective solution when more practical network settings are taken into consideration.

REFERENCES

- [1] A. Molina-Garcia, F. Bouffard and D. S. Kirschen, "Decentralized demand-side contribution to primary frequency control," *IEEE Trans. on Power Systems*, vol. 26, no. 1, pp. 411-419, 2001.
- [2] P. Kundur, *Power System Stability and Control*. New York: McGraw Hill, Inc., 1994.
- [3] D. S. Callaway and I. A. Hiskens, "Achieving controllability of electric loads," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 184-199, 2011.
- [4] B. Kirby, "Spinning Reserve From Responsive Loads," *Report of Oak Ridge National Laboratory*, ORNL/TM-2003/19, March 2003.
- [5] G. Heffner, C. Goldman, B. Kirby and M. Kintner-Meyer, "Loads providing ancillary services: Review of international experience," *Report of Lawrence Berkeley National Laboratory*, LBNL-62701, May 2007.
- [6] A. Brooks, E. Liu, D. Reicher, C. Spirakis and B. Wehl, "Demand dispatch: Using real-time control of demand to help balance generation and load," *IEEE Power&Energy Magazine*, vol. 8, no. 3, pp. 21-30, 2010.
- [7] PNNL, *Grid Friendly™ Controller Helps Balance Energy Supply and Demand*. [Online]. Available: <http://www.gridwise.pnl.gov/docs/pnnlsa36565.pdf>.
- [8] D. Trudnowski, M. Donnelly and E. Lightner, "Power-system frequency and stability control using decentralized intelligent loads," in *Proc. of the 2005 IEEE T&D Conf. Expo.*, Dallas, TX, May 2006.
- [9] D. Trudnowski, M. Donnelly and E. Lightner, "Power-system frequency and stability control using decentralized intelligent loads: Benefits and pitfalls," in *Proc. of the 2010 IEEE Power and Energy Society General Meeting*, Minneapolis, MN, Jul. 2010.
- [10] M. Ilic and Q. Liu, "Toward sensing, communications and control architectures for frequency regulation in systems with highly variable resources," *Control and Optimization Theory for Electric Smart Grids*. New York: Springer, 2011.
- [11] A. Rubtsov, "Approach to stochastic modeling of power systems," *Scientific Journal of Riga Technical University, Power and Electrical Engineering*, Vol. 26, 2010.
- [12] A. R. Bergen and V. Vittal, *Power Systems Analysis*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2000.
- [13] C. Zhao, U. Topcu and S. H. Low, "Frequency-based load control in power systems," *Technical Report*, California Institute of Technology, 2011. [CaltechCDSTR:2011.007]. Available: <http://resolver.caltech.edu/CaltechCDSTR:2011.007>.
- [14] S. Boyd and L. Vandenberghe, *Convex Optimization*. U.K.: Cambridge University Press, 2004.
- [15] P. K. Kitanidis, "Unbiased minimum-variance linear state estimation," *Automatica*, vol. 23, no. 6, pp. 775-778, 1987.
- [16] Trilliant Inc., "The Multi-Tier Smart Grid Architecture," [Online]. Available: <http://www.trilliantinc.com/solutions/multi-tier-architecture/>.
- [17] S. S. Ram, A. Nedic and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of Optimization Theory and Applications*, vol. 147, no. 3, pp. 516-545, 2010.
- [18] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation*. Upper Saddle River, NJ: Prentice Hall, 1989.
- [19] C. Zhao, U. Topcu and S. H. Low, "Fast load control with stochastic frequency measurement," *Technical Report*, California Institute of Technology, 2011. [Online]. Available: <http://www.cds.caltech.edu/~utopcu/ZTLPEs12.html>.