2017 IEEE 56th Annual Conference on Decision and Control (CDC)
December 12-15, 2017, Melbourne, Australia

Geographically Coordinated Frequency Control

Joshua Comden[1], Tan N. Le[1,2], Yue Zhao[1], Bong Jun Choi[1,2], and Zhenhua Liu[1]
{joshua.comden, tan.le, yue.zhao.2, bongjun.choi, zhenhua.liu}@stonybrook.edu
[1]Stony Brook University, [2]SUNY Korea

Abstract— **Primary Frequency Control (PFC) is a fast acting mechanism used to ensure high-quality power for the grid that is becoming an increasingly attractive option for load participation. Because of the speed requirement, PFC requires distributed control laws to be used instead of a more centralized design. Current PFC designs assume that costs at each geographic location are *independent*. Unfortunately for many networked systems such as cloud computing, the decisions made among locations are *interdependent* and therefore require geographic coordination. In this paper, distributed control laws are designed for geo-distributed loads such as data centers in PFC. The controlled frequencies are provably stable, and the final equilibrium point is proven to strike an optimal balance between load participation and the frequency's deviation from its nominal set point. We evaluate the proposed control laws with realistic numerical simulations. Results highlight significant cost savings over existing approaches under a variety of settings.**

I. INTRODUCTION

In the electrical power grid, keeping a stable frequency at a set nominal value is important for supplying reliable high-quality power and maintaining safe grid infrastructure operation. The frequency can drift away from its set point if there is a power imbalance anywhere in the grid. To stabilize and return the frequency back to its nominal value, Frequency Control (FC) [1] is used to correct this power imbalance as fast as possible in a smooth manner. FC as a whole involves different mechanisms working at different timescales depending on the speed of the frequency drift. For frequency drifting slowly in the timescale of several seconds, Automatic Generation Control (AGC) [2] is used to *centrally* decide and change the set power injections of the generators. However, if the frequency drifts much faster due to a sudden power imbalance, e.g., generator failure, power injection modifications must be made faster than can be centrally decided and disseminated to all of the control points by AGC. In this case, Primary Frequency Control (PFC) [3] is used to stop the drift and stabilize the frequency which is done by controlling each generator's power injection independently according to a function of its *locally* measured deviated frequency. Usually this stabilized equilibrium frequency is deviated from its nominal value in which AGC is then employed to return it back.

Traditionally, FC is done on the generation side, while it is becoming an attractive opportunity for load participation. Similar to generation-side, demand-side PFC works by setting devices to independently adjust their individual power consumptions according to some function of the locally measured frequency. However when a device deviates its power consumption, there is an associated loss of utility to the owner of that device. Finding an optimal balance between the cost of load participation and the stabilized equilibrium's deviated frequency is a major challenge of demand-side PFC. It has been shown that with well designed control laws, this optimal balance can be made in a fully distributed fashion and is provably stable as long as these costs are assumed to be *independent* of each other [4].

However, the assumption of independent costs is somehow restrictive. For instance, in some networked systems such as in cloud computing, not all costs can be considered independent. In general, networks of data centers make up the infrastructure that supplies the computing resources necessary for making the cloud run. These data centers are located around the world and groups of them may be connected to the same electrical power network. User workloads requesting computing resources are distributed to different data centers in a way that depends on data availability, server utilization, network delay, etc. Since servers essentially convert electrical power into computational power, the distribution of IT workloads among the data centers has a direct impact on the distribution of their power consumptions. Through Geographic Load Balancing (GLB), networks of data centers can dynamically redistribute workloads depending on data center and power network conditions [5]. In other words, some workloads that cannot be processed in one data center can be served by other data centers, which significantly increases system reliability and flexibility in workload distribution. However, if some fraction of the workload is not processed by *any* of the data centers, the whole system is penalized from the resulting loss of revenue, which can be much larger than the costs of redistribution [6]. This fraction of the workload not processed by any of the data centers is measured by the cloud provider and communicated back to all of the data centers.

Since the cloud has been steadily increasing its share of the total US electricity consumption to about 1.8% in 2014 and has precise energy management of its systems [7], it has great potential to be a large contributor to PFC. For this reason, the paper answers the following question: **How to coordinate primary frequency control that have geographic interdependencies?** While motivated by cloud computing, this question and associated solutions can be applied to general cases with interdependent costs.

For large-scale control with many decisions such as PFC, distributed control design is important. As networks grow, the increased communication overhead required for a centralized controller becomes infeasible. Additionally, privacy requirements may not allow a central entity to know the objectives and constraints of all the users. This is especially true in the case of deregulated power markets and load frequency control [8]. Distributed control is a major challenge for systems with interdependencies because each component needs information about the collective system-wide decisions.

In order to tackle these existing problems, we make the following contributions in this paper:

1) We formulate a primary frequency control problem that balances the extent of load control participation for a cloud computing network operating at different geographic locations with both independent and *interdependent* costs (Sections III and IV).
2) We study the frequency control problem's optimal solution characteristics, and design a set of distributed feedback control laws that a) has an optimal equilibrium point, and b) is asymptotically stable (Sections V and VI).
3) We evaluate our control scheme and show that it works on a realistic emulator system using the Power System Toolbox (PST) [9]. Furthermore, we show that our distributed control gives significant cost savings as interdependent costs become more prevalent (Section VII).

The next section gives a motivating example to show the impact of interdependent costs on a cloud computing network participating in PFC.

II. MOTIVATING EXAMPLE

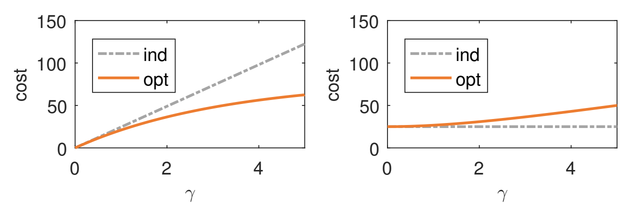
In order to demonstrate the importance of taking geographic interdependence into account for PFC design, we showcase a concrete example of a cloud computing workload running on a network of data centers which consume electrical power from the power grid.

Consider a network of two identical data centers where the only difference between them is their locations in the power grid and their efficiencies which is the ratio of computational power output to the electrical power input. The first data center is a high efficiency one with an efficiency of 0.9 while the second is an average one with an efficiency of 0.5. The first is 1.8× more efficient than the second which is consistent with those studied in [10], [7]. Therefore, the total computational power of the data centers is a linear combination of both power consumptions (d_1, d_2) weighted by their efficiencies. The workload that the data centers need to share requires 28 MW of computing power. Subtracting the workload size from the total computational power gives the excess computational power of the network $0.9d_1 + 0.5d_2 - 28$. A negative excess computational power means that some of the workload was not processed which results in a loss of revenue shared by all of the data centers. In this example it is represented by the quadratic cost function $\gamma\left(0.9d_1 + 0.5d_2 - 28\right)^2$ that features an increasing marginal cost where γ is the *interdependent* cost coefficient. Each data center also observes an *independent* quadratic cost of $(d_j - 20)^2$ for purchasing and processing d_j of electrical power into computational power. Therefore the total cost of both data centers is:

$$\gamma\left(0.9d_1 + 0.5d_2 - 28\right)^2 + (d_1 - 20)^2 + (d_2 - 20)^2.$$

The solution $d_1 = d_2 = 20$ MW minimizes the total cost with a value of 0 and gives an aggregate electrical power consumption of 40 MW. Let this be the power consumed when the data centers are not participating in FC.

Suppose that there is a sudden power disturbance in the grid and therefore it is required that the data centers reduce their aggregate power consumption by 10 MW for FC, i.e. $\Delta d_1 + \Delta d_2 = -10$. If only the independent costs (last two terms of the total cost) are considered for FC, then reducing both power consumptions by 5 MW minimizes the cost. However, if the interdependent cost (first term of the total cost) is also considered then reducing both power consumptions by 5 MW may be suboptimal as shown in Figure 1. Figure 1(a) shows that interdependent cost for FC that only minimizes the independent cost increases linearly with the interdependent cost coefficient γ while optimal FC increases sublinearly. This gap is not fully compensated by the difference in independent costs (cf. Figure 1(b)). Therefore it is important for FC to take into account interdependent costs for networked systems and the rest of this paper focuses on closing this gap. As a result, there is great opportunity to reduce the total cost by taking the interdependent costs into the optimization.



(a) Interdependent Costs (b) Independent Costs

Fig. 1. Interdependent (a) and Independent (b) costs vs. the cost coefficient γ of the interdependent cost for FC that only considers independent costs (dashed gray line) and of the optimal FC (solid orange line) .

III. MODEL AND NOTATION

A. Power network model

We consider a power network consisting of a set of buses $\mathcal{N}$ connected by a set of arbitrarily directed lossless lines $\mathcal{E}$, and only consider the real power injection at each bus and the real power flow across each line. We ignore reactive power since it is more closely related to voltage control as compared to frequency control [11]. For each bus $j \in \mathcal{N}$ we denote P_j as the real power injection, θ_j as the voltage phase angle from a standard reference point rotating at the set nominal frequency ω_0, ω_j as the frequency's deviation from the nominal set point ω_0 or also the time rate of change for the voltage phase angle

$$\omega_j := \frac{d\theta_j}{dt} \quad \forall j \in \mathcal{N}, \tag{1}$$

and we assume that the voltage magnitude $|V_j|$ remains constant during the time frame of PFC. For each directed line $(j,k) \in \mathcal{E}$ we denote x_{jk} as the reactance.

The power injection at each bus is split into three terms

$$P_j := p_j - D_j\omega_j - d_j \quad \forall j \in \mathcal{N} \tag{2}$$

where p_j is the frequency-insensitive part of the non-controllable power injection, the second term is the frequency-sensitive part of the non-controllable power injection, and d_j is the controllable load. We assume that p_j remains constant during the time frame of PFC. The second term approximates the frequency-sensitivity as a first-order dependence on the frequency's deviation where $D_j > 0$ is the linear coefficient. This is reasonable for small deviations [12].

The voltage phase angle difference across each line connected to bus $j \in \mathcal{N}$ determines the real power flow out from that bus into the rest of the power network:

$$\begin{aligned} F_j(\boldsymbol{\theta}) := &\sum_{k:(j,k)\in\mathcal{E}} Y_{jk}\sin(\theta_j - \theta_k) \\ &- \sum_{i:(i,j)\in\mathcal{E}} Y_{ij}\sin(\theta_i - \theta_j) \end{aligned} \tag{3}$$

where $Y_{jk} := \frac{|V_j||V_k|}{x_{jk}}$ is the maximum power flow across line $(j,k) \in \mathcal{E}$ as determined by the constant bus voltage magnitudes and line reactance.

B. Cloud computing model

We consider a set of data centers $\mathcal{D} \subseteq \mathcal{N}$ connected by a high speed communication network that provides cloud computing services for a workload incoming rate of size W. We assume that the incoming workload rate remains constant for a primary frequency control event time duration. Each data center $j \in \mathcal{D}$ processes some of the incoming workload at a rate of r_j. Subtracting the workload incoming rate from the sum of the data center processing rates we get the excess computational power of the network:

$$s := \sum_{j\in\mathcal{D}} r_j - W. \tag{4}$$

A negative s represents insufficient computational power to process all of the incoming workload which means that $-s$ of the workload may suffer a delay or remain unprocessed. This causes a loss of revenue captured by the cost function $g(s)$ which means that a specific cost level *depends on all* of the data center processing rates. We assume that each data center knows the interdependent cost function $g(\cdot)$ and can receive information about the excess computational power s via the communication network.

We model each data center j as a machine that converts electrical power d_j into computational power r_j with a linear usage profile:

$$d_j := \underline{d}_j + \frac{1}{a_j} r_j \tag{5}$$

where $\underline{d}_j$ is the constant overhead electrical power usage, and a_j is the conversion coefficient that can be considered the computational efficiency as described in Chapter 5 of [13]. Each data center has an upper bound on its electrical power consumption $\overline{d}_j$, and observes a cost of $c_j(d_j)$ associated with obtaining and processing the electrical power d_j. We assume that the value of d_j and the function $c_j(\cdot)$ are only known by data center j.

The excess computational power of the network (4) can now be expressed in terms of the individual electrical power consumptions in (5):

$$s = \sum_{j\in\mathcal{D}} a_j d_j - b \tag{6}$$

where $b := W + \sum_{j\in\mathcal{D}} a_j \underline{d}_j$.

C. System dynamics

The power network frequency dynamics at each bus are determined by the swing equation

$$M_j \frac{d\omega_j}{dt} = P_j - F_j(\boldsymbol{\theta}) \quad \forall j \in \mathcal{N} \tag{7}$$

where M_j is the physical interia of the rotating equipment.

Since stability is an essential feature of FC, we give the following definition for an equilibrium point of the previously described system.

Definition 1: A closed-loop equilibrium of the system (1) (2) (3) (6) (7), is any solution $(\boldsymbol{\theta}^*, \boldsymbol{\omega}^*, \mathbf{P}^*, \mathbf{d}^*, s^*)$ that further satisfies:

$$\frac{d\omega_j^*}{dt} = 0 \qquad \forall j \in \mathcal{N} \tag{8a}$$

$$\frac{dP_j^*}{dt} = 0 \qquad \forall j \in \mathcal{N} \tag{8b}$$

$$\omega_j^* = \omega^* \qquad \forall j \in \mathcal{N}. \tag{8c}$$

Note that (8a) makes the LHS of (7) equal to zero and (8c) synchronizes all deviations of frequency to a single value. (8b) implies that $d(d_j^*)/dt = 0 : \forall j \in \mathcal{N}$ and thus $ds^*/dt = 0$.

IV. GEOGRAPHIC FREQUENCY CONTROL PROBLEM

Before designing control laws for data center participation in PFC, we must first decide what an optimal balance is between the controllable data center loads and equilibrium deviations of frequency. Essentially, we form an optimization problem that minimizes the global cost of the system which is the total cost of the network of data centers plus the summed cost of the equilibrium deviations of frequency at each bus.

The total cost of the network of data centers is the interdependent cost of the excess computational power summed with the independent electrical power costs:

$$g(s) + \sum_{j\in\mathcal{D}} c_j(d_j).$$

For the cost of the deviations of frequency, we adopt the cost function developed by [14] which is a sum of the squared deviations at each bus weighted by its associated frequency-sensitive linear coefficient:

$$\sum_{j\in\mathcal{N}} \frac{D_j}{2}\omega_j^2$$

Given the above total cost of the network of data centers and equilibrium deviations of frequency, we state the Geographic Frequency Control (GFC) problem:

$$\min_{s,\mathbf{d},\boldsymbol{\omega}} \quad g(s) + \sum_{j\in\mathcal{N}} \left(c_j(d_j) + \frac{D_j}{2}\omega_j^2 \right) \tag{9a}$$

$$\text{s.t.} \quad \sum_{j\in\mathcal{N}} a_j d_j - b = s \tag{9b}$$

$$\sum_{j\in\mathcal{N}} (p_j - D_j\omega_j - d_j) = 0 \tag{9c}$$

$$\underline{d}_j \le d_j \le \overline{d}_j \quad \forall j \in \mathcal{N} \tag{9d}$$

where (9b) is the excess computational power (6), (9c) is the balance of electrical power on the grid, and (9d) are the box constraints on the data center electrical power consumptions. Note that for any bus $j \notin \mathcal{D}$, we set $a_j = \underline{d}_j = \overline{d}_j = 0$.

In order to take advantage of the structure of GFC (9) we use the following mild assumptions.

Assumption 1: $g(s)$ is strictly convex and twice continuously differentiable. For all $j \in \mathcal{N}$: $c_j(d_j)$ is strictly convex and twice continuously differentiable for all $d_j \in [\underline{d}_j, \overline{d}_j]$.

Assumption 2: GFC (9) has a feasible solution, and for any optimal solution there exists a feasible $\boldsymbol{\theta}$ such that:

$$F_j(\boldsymbol{\theta}) = p_j - D_j\omega_j - d_j \quad \forall j \in \mathcal{N}. \tag{10}$$

Convex cost functions are found in geographic load balancing optimization problems [5] and are consistent with concave disutility functions used in demand response programs [15], [16], [17]. Assumption 2 ensures that for any optimal solution of power injections, there exists a set of voltage phase angles that satisfy the solution.

With the above assumptions we now prove that GFC is a convex optimization problem.

Lemma 1: Given Assumption 1, then GFC (9) is a convex optimization problem and has a unique solution.

Proof: From Assumption 1, the objective function is strictly convex which means it has a unique minimizer. Additionally, the equality constraints (9b) (9c) are linear and the inequality constraints (9d) are convex which gives the result. ■

V. CHARACTERIZING THE OPTIMA

We now provide characterizations of the optimal solution of GFC. This will then motivate the design of decentralized algorithm later on. From Assumption 2 and Lemma 1, the Karush-Kuhn-Tucker (KKT) conditions for optimality for GFC (9) are applicable and can be determined with the dual variables $(\mu, \lambda, \underline{\boldsymbol{\kappa}}, \overline{\boldsymbol{\kappa}})$ for each constraint respectively:

$$g'(s) - \mu = 0 \tag{11a}$$

$$c_j'(d_j) + \mu a_j - \lambda - \underline{\kappa}_j + \overline{\kappa}_j = 0 \quad \forall j \in \mathcal{N} \tag{11b}$$

$$D_j\omega_j - \lambda D_j = 0 \quad \forall j \in \mathcal{N} \tag{11c}$$

$$\underline{\kappa}_j(\underline{d}_j - d_j) = 0 \quad \forall j \in \mathcal{N} \tag{11d}$$

$$\overline{\kappa}_j(d_j - \overline{d}_j) = 0 \quad \forall j \in \mathcal{N} \tag{11e}$$

$$\underline{\kappa}_j \ge 0 \quad \forall j \in \mathcal{N} \tag{11f}$$

$$\overline{\kappa}_j \ge 0 \quad \forall j \in \mathcal{N} \tag{11g}$$

$$\sum_{j\in\mathcal{N}} a_j d_j - b - s = 0 \tag{11h}$$

$$\sum_{j\in\mathcal{N}} (p_j - D_j\omega_j - d_j) = 0 \tag{11i}$$

$$\underline{d}_j - d_j \le 0 \quad \forall j \in \mathcal{N} \tag{11j}$$

$$d_j - \overline{d}_j \le 0 \quad \forall j \in \mathcal{N} \tag{11k}$$

where (11a) (11b) (11c) are the first-order stationary conditions, (11d) (11e) are the complementary slackness conditions, (11f) (11g) are the dual feasibility conditions, and (11h) (11i) (11j) (11j) are the primal feasibility conditions. Note that the operator $(\cdot)'$ denotes the derivative.

From the above KKT conditions we can infer the following properties of the optimal solution:

1) From (11c) we have that

$$\omega_j = \lambda \quad \forall j \in \mathcal{N} \tag{12}$$

which means that the deviation of frequency at each bus is equal to a single value, the same as the equilibrium condition (8c).

2) From (11a) and Assumption 1, we have that at equilibrium the excess computational power s is equal to

$$s = (g')^{-1}(\mu) \tag{13}$$

which is an increasing function of μ due to the strict convexity property. Note that the operator $(\cdot)^{-1}$ denotes the inverse.

3) If $\underline{d}_j < d_j < \overline{d}_j$ then from (11d) (11e) we have that $\underline{\kappa}_j = \overline{\kappa}_j = 0$ and from (11b) we have that:

$$d_j = (c_j')^{-1}(\omega_j - a_j\mu) \tag{14}$$

which is also an increasing function of $(\omega_j - a_j\mu)$ due to the strict convexity property.

VI. DISTRIBUTED FREQUENCY CONTROL

In this section, we first state the distributed control laws and then prove their optimality and stability in solving GFC.

A. Control Laws

At time t, let $(\omega_j(t), s(t))$ respectively be the local deviation of frequency measured by data center j, and the excess computational power of the network that is measured and broadcasted by the cloud provider through the communication network. We propose the following control law of demand response $d_j(t)$, with an auxiliary variable $\mu(t)$:

$$\mu(t) = \mu(0) + \beta \int_0^t \left(s(\tau) - (g')^{-1}(\mu(\tau))\right) d\tau \qquad (15a)$$

$$d_j(t) = \left[(c_j')^{-1}(\omega_j(t) - a_j\mu(t))\right]_{\underline{d}_j}^{\overline{d}_j} \quad \forall j \in \mathcal{N} \qquad (15b)$$

where the constant $\beta > 0$ controls the speed at which the auxiliary variable $\mu(t)$ reacts to the difference $\left(s(t) - (g')^{-1}(\mu(t))\right)$. Note that $(g')^{-1}(\cdot)$ and $(c_j')^{-1}(\cdot)$ are well defined because of the strict convexity assumption stated in Assumption 1.

Essentially the control laws work by first implementing the optimal solution's property (14) and relaxing property (13). It then gradually changes the auxiliary variable μ in (15a) based on how far (13) is from being satisfied. Changing the value of μ simultaneously changes the value of each electrical power consumption d_j from (15b) and therefore changes the excess computational power s from (6) in the direction of satisfying (13).

In order to better qualify an equilibrium point, we further define it to include the auxiliary variable μ when the optimal property (13) is satisfied and therefore is not changing with time. The following definition will be useful when proving stability in Section VI-C.

Definition 2: A closed-loop equilibrium of the system (1) (2) (3) (6) (7) (15a), is any solution $(\boldsymbol{\theta}^*, \boldsymbol{\omega}^*, \mathbf{P}^*, \mathbf{d}^*, s^*, \mu^*)$ that satisfies Definition 1 and

$$\frac{d\mu^*}{dt} = 0. \qquad (16)$$

B. Optimality

The following theorem states that an equilibrium point of the above distributed control laws is optimal to the GFC optimization problem (9).

Theorem 2: Given Assumptions 1 and 2, an equilibrium point from Definition 2 of the system (1) (2) (3) (6) (7) with control laws (15) is an optimal solution of GFC (9).

Please see Appendix A for the proof of the above theorem.

This is important because it shows that when the system reaches steady state, there is a guaranteed optimal balance between data center PFC load participation and the system wide deviated frequency that is equal to the deviated frequencies at every bus. Also at steady state, the marginal independent cost for each data center is the equilibrium deviation of frequency discounted by its marginal contribution to the interdependent cost (See (13), (14)). This means that a data center with a large marginal contribution to the interdependent cost results in a low marginal independent cost as compared to the other data centers.

C. Stability

To prove that the system is asymptotically stable with the distributed control laws, we give the following assumption which is found to be true under normal operating conditions [12].

Assumption 3: The equilibrium phase angle deviations between connected buses are bounded: $|\theta_i^* - \theta_j^*| < \frac{\pi}{2}$ for all $(i,j) \in \mathcal{E}$.

We use the Lypanunov method in the following theorem to prove that an equilibrium point, within the neighborhood described in the above assumption, is asymptotically stable. Additionally, we show that the system asymptotically converges to an equilibrium point that does not consider the specific phase angles. This is important because it guarantees that the system trajectory is always moving towards an equilibrium point.

Theorem 3: Given Assumptions 1 and 2, an equilibrium point (Definition 2) of the system (1) (2) (3) (6) (7) with feedback control (15) that satisfies Assumption 3 is asymptotically stable. In particular under the same assumptions, the trajectory of $(\boldsymbol{\omega}, \mathbf{P}, \mathbf{d}, s, \mu)$ such that $|\theta_i - \theta_j| < \frac{\pi}{2}$: $\forall (i,j) \in \mathcal{E}$ will asymptotically converge to an equilibrium point $(\boldsymbol{\omega}^*, \mathbf{P}^*, \mathbf{d}^*, s^*, \mu^*)$.

Due to space limitations we have provided the proof of the above theorem in an extended version of this paper [18].

VII. PERFORMANCE EVALUATION

We simulate a 50 MW generation loss at the 5 second time stamp and show that the proposed feedback control stabilizes the system to an equilibrium point with significant cost savings. We compare the proposed control with a decentralized load control OLC [4] and the system under no control mechanism denoted as "none".

A. Setup

In order to demonstrate the performance of the proposed control on a more realistic system model than described in Section III, we use the Power System Toolbox (PST) [9] to simulate it.

Power network: The IEEE 39-bus (New England) system was chosen as the test case for the evaluation which has 19 buses available to place controllable loads. The total power demand is set at 14 GW [19]. We use the following ten buses to place controllable data center loads: 3, 4, 7, 8, 15, 16, 18, 20, 21, and 23. To simulate a power disturbance, at the 5 second time stamp power drops 50 MW from a generator at bus 39.

Network of data centers: The ten data centers contain a total of 500k 300W servers in which each server's average power consumption relative to peak is 75% [13]. The data centers each have a nominal demand of 25 MW which in total is 1.8% of the total power demand [7]. Each data center has minimum and maximum demands of 15 MW and 30 MW respectively. Each data center's efficiency a_j is estimated based on the fact that typical data centers have a Power Usage Effectiveness (PUE) range between 1.1 and 2.1 with an average of 1.8 [10], [7], therefore we randomly select $1/a_j \in [1.1, 2.1]$ with $\mathbb{E}[1/a_j] = 1.8$.

Costs: The objective function used in analyzing the cost for control is defined as follows:

$$\gamma \frac{1}{2} \left(\sum_{j \in \mathcal{D}} a_j \delta_j \right)^2 + \sum_{j \in \mathcal{D}} \frac{\eta_j}{2} \delta_j^2 + \alpha \sum_{j \in \mathcal{N}} \frac{D_j}{2} \omega_j^2. \qquad (17)$$

where $\delta_j := d_j - 25$ which means that each data center has a nominal demand of 25 MW. The first term is the interdependent cost, the second term is the independent costs, and the third term is the cost associated with the deviations of frequency. The interdependent cost is approximated by the fact that the Amazon Web Service's revenue of 1.3 million servers [20] is \$2.5 billion for the first quarter in 2016 [21]. This gives a cost of \$2.47 per second for each 10k servers. For our system, this translates into maximum of estimated interdependent cost for each second is \$123.50. Hence assuming an average PUE of 1.8 and only half of a 100 MW power equivalent workload is processed, we set $\gamma = \$0.16/\text{MW}^2$ throughout the evaluation. The dependent cost for each data center includes its wasted cost for under/over-utilizing the pre-purchased electricity, operation, and maintenance. Based on the Total Cost of Ownership in [13], under utilizing a data center by 50% can cost \$0.28 per second for each 10k servers. For the average data center in our system with 50k servers this is \$1.40 which we equate to a 5 MW decrease of the total available 10 MW decrease. Hence, we randomly choose η_j such that $\mathbb{E}[\eta_j] = \$0.11/\text{MW}^2$. The cost for deviated frequency has been valued at \$15/MW for a 0.2 Hz deviation [22], [23]. Since $\sum_{j\in\mathcal{N}} \frac{D_j}{2}\omega_j$ is the aggregate frequency-sensitive load, we set α to \$75/MW-Hz.

Baselines: To show the benefit of incorporating interdependent load costs, we compare the proposed control to Optimal Load Control (OLC) described in [4] which is a decentralized control that does not take into account interdependent costs. In the stability analysis, we also compare it to the case called "none" where there are no primary frequency control system in place; it relies on frequency control at a larger timescale. In the equilibrium cost analysis, we also compare it to the case called "optimal" which is an estimated offline optimal solution to GFC (9) with $D_j = 0 : \forall j \in \mathcal{N}$.

B. Stability analysis

The proposed control stabilizes the frequency and each load decision within 30 seconds of the generation loss. In Figure 2(a), both the proposed control and OLC converge to an equilibrium frequency within 30 seconds. On the other hand, for the system trial with no control, the frequency continues to drop even well after the 50 second time stamp and does not converge to a final equilibrium frequency. Note that for each control scheme in Figure 2(a), the frequencies for the separate buses are so close that they cannot be distinguished.

Additionally in Figure 2(b), all of the loads reach their equilibrium points at different times which are all within 30 seconds of the generation loss. It is worth noting that the most efficient data center DC 1 actually increases its load instead of decreasing it. This is because the value of $(\omega_1 - a_1\mu)$ remains positive throughout the simulation due to the product of a negative auxiliary variable μ and a relatively large computational efficiency a_1 counteracting and surpassing the negative deviation of frequency ω_1. That positive value is used in the proposed control (15b) which is an increasing function.

C. Equilibrium cost analysis

The proposed control shows a 24% decrease in total cost from OLC (cf. Figure 3(a)). While on the other hand, it has a larger deviation of frequency than OLC (cf. Figure 2(a)) because OLC underestimates the actual total cost of decreasing the load consumptions. This is caused by OLC minimizing only for independent and deviated frequency costs (i.e. last two terms of (17)) as compared to the proposed control which minimizes the sum of all three cost types. This results in OLC having a 38% lower independent cost and a 83% smaller equilibrium deviation of frequency, but a 46% higher interdependent cost.

Also, the proposed control will utilize smaller power deviations from higher efficient data centers than OLC. Higher values of computational efficiency a_j cause power deviations to have larger marginal increases to the interdependent cost (i.e. the first term of (17)) than for lower values of a_j. This causes the OLC power deviations to be more negative than the proposed control for data centers with higher than average efficiencies (DC 1-3), and less negative for all other data centers (cf. Figure 3(c)).

Also shown in Figure 3(a) is that the proposed control is close to but not at the estimated optimal cost. This is rooted in the fact that the frequency sensitive loads in PST exhibit nonlinear behavior as opposed to the linear model used in GFC (9). By setting $D_j = 0 : \forall j \in \mathcal{N}$ in GFC (9) for the "optimal" case comparison, we get an optimal solution where all the data centers must have an aggregate load consumption decrease *equal* to the power drop. However, when using the proposed control in PST, we get an aggregate load consumption decrease that is *greater* than the power drop. This means that the estimated optimal has a lower cost since it does not need to decrease as much load consumption as was done by the proposed control in PST.

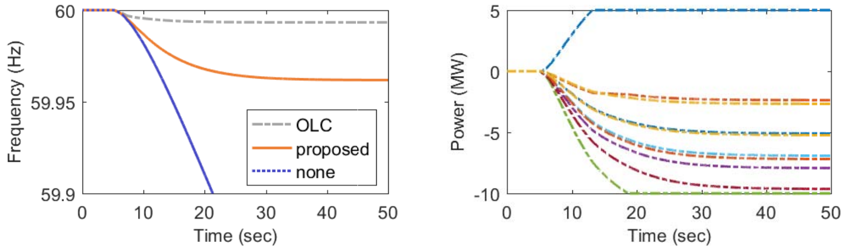
D. Sensitivity analysis

Impact of interdependence: In order to measure the effect that the interdependent cost has on the cost savings of the proposed control, we vary $\gamma \in [0.03, 0.3]$ (cf. Figure 4(a)). While low interdependent costs ($\gamma \sim 0.03$) make the cost savings between the proposed control and OLC insignificant, higher interdependent costs result in larger cost savings for the proposed control.

Demand flexiblity (X%): As data centers may not be able to reduce all of their demand at this fast timescale, next we evaluate the impacts of data centers' demand flexibility, which is the fraction of the nominal load that can be changed, i.e. $d_j \in [25 * (1 - X/100), 30]$ (cf. Figure 4(b)). Observe that the cost savings remain significant for most of the range except just above 20% demand flexibility. This is where both proposed and OLC total costs converge because the box constraints (9d) become active and every data center reduces its demand to its lowest allowed value.

VIII. CONCLUSION

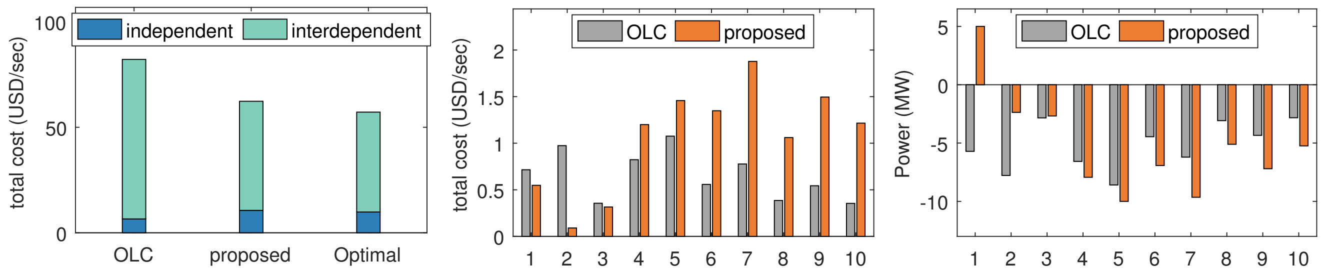
Frequency control is an important class of mechanisms to ensure high-quality power for the grid and has been an increasingly attractive option for load participation. Primary



(a) Frequency

(b) Power demand change, δ_j

Fig. 2. Trajectories of the system state variables: (a) Bus frequencies for each of the ten buses containing a data center under the three different control schemes; (b) changes in load for each of the ten data centers under the proposed control.

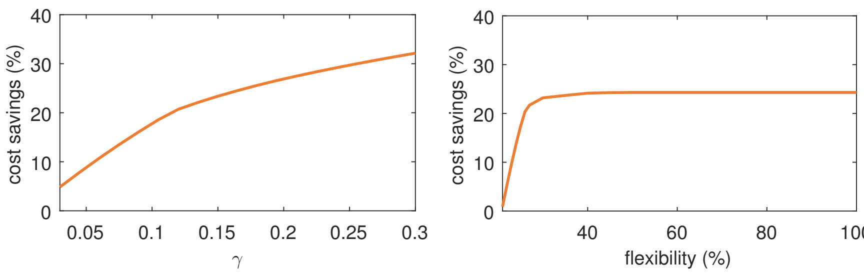


(a) Total cost

(b) Independent costs

(c) Load deviations

Fig. 3. (a) Total equilibrium cost separated by independent and interdependent costs of OLC, the proposed control, and the optimal offline solution of GFC (9). For each data center (b) and (c) give the independent costs and equilibrium load deviations respectively under OLC and the proposed control. Note: The data centers are numbered in decreasing order of efficiency a_j.



(a) Impact of γ (interdependent cost)

(b) Impact of demand flexibility

Fig. 4. Sensitivity analysis in terms of cost savings by using the proposed control instead of OLC for: (a) Interdependent cost coefficient; (b) Demand flexibility.

Frequency Control, being the fastest of these, requires that the decisions be made in a distributed fashion because of the speed requirement. Prior work on distributed load control assumed that the cost for changing load demand at each geographic location is independent from the rest. However, in some networked systems such as a network of data centers that support cloud computing, the decisions made at each data center affect the group because of interdependent costs and so require geographic coordination. In this paper, we designed a set of distributed control laws that can handle interdependent costs in such a way that is provably stable. Also, we proved that the final equilibrium point optimally balances the cost of load participation with the frequency level deviated from its nominal set point. We tested our control laws for a network of data centers on a realistic emulator, Power System Toolbox, and found that there is significant cost savings with the proposed control law over existing benchmarks that do not account for interdependent costs.

The results presented in this paper open up three distinct future research directions. The first is to explore how other interdependent systems (e.g. electric mass transit, thermal grids) can be used to help increase the reliability of the grid. The second is to investigate how a network of data centers that are located in multiple disjoint power grids can utilize their interconnectedness to enhance the reliability in those grids. The third is to apply the distributed control laws to a system with a higher-order transient stability model. For our future work, we plan to extend the proposed control laws to take into account further network effects such as power flow constraints across lines and network losses.

IX. ACKNOWLEDGMENTS

We would like to thank Dr. Changhong Zhao for his valuable and insightful comments on this paper and the reviewers for their suggestions. This research is supported by NSF grants CNS-1464388, CNS-1617698, CNS-1730128, CNS-1717558, and was partially funded by MSIP, Korea, grants IITP-2015-R0346-15-1007 and NRF-2015R1C1A1A01053788.

References

[1] H. Shayeghi, H. Shayanfar, and A. Jalili, "Load frequency control strategies: A state-of-the-art survey for the researcher," *Energy Conversion and management*, vol. 50, no. 2, pp. 344–353, 2009.

[2] N. Jaleeli, L. S. VanSlyck, D. N. Ewart, L. H. Fink, and A. G. Hoffmann, "Understanding automatic generation control," *IEEE transactions on power systems*, vol. 7, no. 3, pp. 1106–1122, 1992.

[3] A. Molina-Garcia, F. Bouffard, and D. S. Kirschen, "Decentralized demand-side contribution to primary frequency control," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 411–419, 2011.

[4] C. Zhao and S. Low, "Optimal decentralized primary frequency control in power networks," in *53rd IEEE Conference on Decision and Control*, pp. 2467–2473, IEEE, 2014.

[5] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. Andrew, "Greening geographical load balancing," in *Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems*, pp. 233–244, ACM, 2011.

[6] A.-F. Antonescu, P. Robinson, and T. Braun, "Dynamic sla management with forecasting using multi-objective optimization," in *Integrated Network Management (IM 2013), 2013 IFIP/IEEE International Symposium on*, pp. 457–463, IEEE, 2013.

[7] A. Shehabi, S. Smith, N. Horner, I. Azevedo, R. Brown, J. Koomey, E. Masanet, D. Sartor, M. Herrlin, and W. Lintner, "United states data center energy usage report," *Lawrence Berkeley National Laboratory, Berkeley, California. LBNL-1005775 Page*, vol. 4, 2016.

[8] E. Camponogara, D. Jia, B. H. Krogh, and S. Talukdar, "Distributed model predictive control," *IEEE Control Systems*, vol. 22, no. 1, pp. 44–52, 2002.

[9] J. Chow, G. Rogers, and K. Cheung, "Power system toolbox," 2009.

[10] I. H. Cheung, S. Greenberg, R. Mahdavi, R. Brown, and W. Tschudi, "Energy efficiency in small server rooms: Field surveys and findings," 2014.

[11] Y. G. Rebours, D. S. Kirschen, M. Trotignon, and S. Rossignol, "A survey of frequency and voltage control ancillary servicespart i: Technical features," *IEEE Transactions on power systems*, vol. 22, no. 1, pp. 350–357, 2007.

[12] A. R. Bergen and D. J. Hill, "A structure preserving model for power system stability analysis," *IEEE Transactions on Power Apparatus and Systems*, no. 1, pp. 25–35, 1981.

[13] L. A. Barroso, J. Clidaras, and U. Hölzle, "The datacenter as a computer: An introduction to the design of warehouse-scale machines," *Synthesis lectures on computer architecture*, vol. 8, no. 3, pp. 1–154, 2013.

[14] C. Zhao, U. Topcu, N. Li, and S. Low, "Design and stability of load-side primary frequency control in power systems," *IEEE Transactions on Automatic Control*, vol. 59, no. 5, pp. 1177–1189, 2014.

[15] E. Bitar, K. Poolla, P. Khargonekar, R. Rajagopal, P. Varaiya, and F. Wu, "Selling random wind," in *System Sciences, Hawaii International Conference on*, pp. 1931–1937, IEEE, 2012.

[16] L. Jiang and S. Low, "Multi-period optimal energy procurement and demand response in smart grid with uncertain supply," in *Decision and Control and European Control Conference*, pp. 4348–4353, IEEE, 2011.

[17] N. Li, L. Chen, and S. H. Low, "Optimal demand response based on utility maximization in power networks," in *2011 IEEE power and energy society general meeting*, pp. 1–8, IEEE, 2011.

[18] J. Comden, T. N. Le, Y. Zhao, B. J. Choi, and Z. Liu, "Geographically Coordinated Frequency Control technical report," *arXiv preprint arXiv:1709.06137*, 2017.

[19] "2016 Monthly Data by Load Zone." Online, March 2017. `https://www.iso-ne.com/isoexpress/web/reports/load-and-demand/-/tree/zone-info`.

[20] L. Dignan, "Aws cloud computing ops, data centers, 1.3 million servers creating efficiency flywheel," 2016.

[21] J. NOVET, "Amazon web services grabs $2.5 billion in revenue in q1 2016, up 63.8 % over last year," 2016.

[22] "Enhanced frequency response seminar." Presenation Slides, June 2016.

[23] R. Lin, "Energy storage for fast, flexible enhanced frequency response," Jan 2016.

Appendix

A. Proof of Theorem 2

Proof: Since GFC is a convex optimization problem (Lemma 1) and is feasible (Assumption 2), we need to show that the proposition satisfies the KKT conditions for optimality (11).

From the strict convexity property in Assumption 1, the function $(c_j')^{-1}(\cdot)$ is well defined and increasing. Therefore, control law (15b) can be separated into three cases: (i) $d_j^* \in (\underline{d}_j, \overline{d}_j)$, thus $d_j^* = (c_j')^{-1}(\omega_j^* - a_j\mu^*)$; (ii) $d_j^* = \underline{d}_j$, thus $(c_j')^{-1}(\omega_j^* - a_j\mu^*) \le \underline{d}_j$; (iii) $d_j^* = \overline{d}_j$, thus $(c_j')^{-1}(\omega_j^* - a_j\mu^*) \ge \overline{d}_j$. Case (i) results in:

$$\underline{d}_j < (c_j')^{-1}(\omega_j^* - a_j\mu^*) < \overline{d}_j$$
$$c_j'(\underline{d}_j) < \omega_j^* - a_j\mu^* < c_j'(\overline{d}_j).$$

since $c_j'(\cdot)$ and $(c_j')^{-1}(\cdot)$ are increasing functions. Likewise, Case (ii) results in:

$$(c_j')^{-1}(\omega_j^* - a_j\mu^*) \le \underline{d}_j$$
$$\omega_j^* - a_j\mu^* \le c_j'(\underline{d}_j)$$

and Case (iii) results in:

$$\overline{d}_j \le (c_j')^{-1}(\omega_j^* - a_j\mu^*)$$
$$c_j'(\overline{d}_j) \le \omega_j^* - a_j\mu^*.$$

Let us define the following two variables:

$$\underline{\kappa}_j := [c_j'(\underline{d}_j) - (\omega_j^* - a_j\mu^*)]^+$$
$$\overline{\kappa}_j := [(\omega_j^* - a_j\mu^*) - c_j'(\overline{d}_j)]^+$$

which satisfy (11f) (11g). Note that in: Case (i), $\underline{d}_j - d_j^* \neq 0$, $d_j^* - \overline{d}_j \neq 0$, $\underline{\kappa}_j = 0$, $\overline{\kappa}_j = 0$; Case (ii), $\underline{d}_j - d_j^* = 0$, $d_j^* - \overline{d}_j \neq 0$, $\underline{\kappa}_j \ge 0$, $\overline{\kappa}_j = 0$; Case (iii) $\underline{d}_j - d_j^* \neq 0$, $d_j^* - \overline{d}_j = 0$, $\underline{\kappa}_j = 0$, $\overline{\kappa}_j \ge 0$. Therefore each case satisfies (11d) (11e).

Additionally, each case from their definitions satisfies:

$$c_j'(d_j^*) - \underline{\kappa}_j + \overline{\kappa}_j = \omega_j^* - a_j\mu^* \tag{18}$$

and we can use the equilibrium condition (8c) to define the following variable since each frequency deviation must be equal to a single value:

$$\lambda := \omega^* = \omega_j^* \quad \forall j \in \mathcal{N}.$$

which is equivalent to (11c) when multiplied by D_j. Therefore, substituting λ for ω_j^* in (18) becomes (11b).

The time derivative of (15a) at equilibrium gives:

$$\frac{d\mu^*}{dt} = \beta\left(s^* - (g')^{-1}(\mu^*)\right)$$

From the strict convexity property in Assumption 1, the function $(g')^{-1}(\cdot)$ is well defined. Since $d\mu^*/dt = 0$ from equilibrium condition (16), then we have that $(g')^{-1}(\mu^*) = s$ which is equivalent to (11a).

System equation (6) is equivalent to (11h).

To get (11i), start with (7) and apply (2), and equilibrium condition (8a):

$$p_j - d_j^* - D_j\omega_j^* - F_j(\boldsymbol{\theta}^*) = 0.$$

Summing the above equation for all $j \in \mathcal{N}$ gets (11i) because $\sum_{j\in\mathcal{N}} F_j(\boldsymbol{\theta}^*) = 0$ from each term canceling when summing (3).

Because the range of (15b) is $[\underline{\delta}_j, \overline{\delta}_j]$, (11j) (11k) are satisfied.

Since all of the KKT conditions have been satisfied, we get the resultant. ■