

Workload Transfer Strategy of Urban Neighboring Data Centers With Market Power in Local Electricity Market

Jun Sun, *Student Member, IEEE*, Minghua Chen[✉], *Senior Member, IEEE*, Haoyang Liu, Qinmin Yang, *Member, IEEE*, and Zaiyue Yang[✉], *Member, IEEE*

Abstract—This paper investigates how to minimize the operational cost of cloud service provider (CSP) that operates urban neighboring data centers (DCs) in the same electricity market and can conduct workload transfer among DCs. Due to the substantial electricity demand of DCs, their market power which can have impact on the locational marginal prices (LMPs) of the electricity market should be taken into consideration. We formulate a bilinear bilevel problem which regards the CSP as a price maker and explores cost-minimizing workload transfer strategies. The upper level is the operational cost minimization problem of CSP and the lower level corresponds to the economic dispatch problem of independent system operator (ISO) of electricity market which determines the electricity prices. It is challenging to directly solve the bilevel problem with bilinear term in the objective function. Hence, we first reformulate the original problem into a single level problem and then based on the property of the problem we develop a polytope cutting algorithm that attains the global optimal solution. The proposed algorithm solves linear optimizations iteratively by cutting the non-convex polytope feasible set into convex sets. In addition, considering the varying communication environment in practice, we analyze the impact of transfer price uncertainty on total cost of CSP, and show that the expected cost surprisingly decreases with the increasing uncertainty. Simulations based on the standard IEEE test cases show that the cost of CSP is significantly reduced and a win-win result for both the CSP and independent system operator (ISO) is possible.

Index Terms—Urban neighboring data centers, electricity market, locational marginal price, price maker.

NOMENCLATURE

Variables

\mathbb{N}	Bus node set
N_b	The number of buses
ε	Transmission line set
N_l	The number of transmission lines
\mathbf{Y}	Susceptance matrix
\mathbf{x}	Power injection by generators
Θ	Phase angles at buses
\mathbf{K}	Angle-to-power matrix
\mathbf{L}	Line capacity vector
\mathbf{X}	Generation capacity
\mathbf{c}_1	Generation price
N_d	The number of data centers
N_c	The number of communication lines
\mathbf{m}	Workload transfer
\mathbf{M}	Transmission bandwidth
\mathbf{z}^0	Workload originating at DCs
\mathbf{z}	Workload processed at DCs
\mathbf{B}	Incidence matrix
\mathbf{Z}	Workload processed capacity of DCs
\mathbf{c}_2	Workload transfer price
$\mu, \alpha, \bar{\alpha}, \underline{\gamma}, \bar{\gamma}, \underline{\gamma}, \bar{\gamma}$	Lagrange multipliers.

I. INTRODUCTION

THE RAPID growth of data intensive industry, such as Internet and finance has been driving the demand of data storing, transferring and processing continuously increasing in an exponential way. Cloud service providers (CSPs) including Amazon, Microsoft and Alibaba have built super-large scale data centers (DCs), which often locate in remote areas for cheap electricity and plenty of water that cools down IT facilities. Apart from super-large DCs, a large number of smaller DCs are built in urban areas by specialized CSPs, such as Verizon and Equinix, to provide more reliable services with low latency [1]. In this paper, we focus on this type of DCs, which receive little attention in previous studies.

To be specific, we consider the pragmatic situation where a CSP manages multiple DCs in an urban area. For example, more than 70 DCs are located around Washington D.C. as illustrated in Fig. 1, and Equinix alone holds 17 of them. As these DCs are distributed in a city, they often belong to

Manuscript received February 11, 2019; revised July 24, 2019 and October 23, 2019; accepted December 2, 2019. Date of publication January 20, 2020; date of current version June 19, 2020. This work was supported in part by the Natural Science Foundation of China under Grant 61873118, Grant U1609214, Grant 61673347, and Grant 61751205, in part by the Shenzhen Committee on Science and Innovations under Grant GJHZ20180411143603361, in part by the Department of Science and Technology of Guangdong Province under Grant 2018A050506003, in part by the Key Research and Development Program of Guangdong Province under Grant 2018B010107002, and in part by the Key Research and Development Program of Zhejiang Province under Grant 2019C01050. Paper no. TSG-00229-2019. (*Corresponding author: Zaiyue Yang.*)

Jun Sun and Qinmin Yang are with the College of Control Science and Engineering, State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China.

Minghua Chen is with the School of Data Science, City University of Hong Kong, Hong Kong.

Haoyang Liu is with the School of Sports Engineering, Beijing Sport University, Beijing 100084, China.

Zaiyue Yang is with the Department of Mechanical and Energy Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: yangzy3@sustech.edu.cn).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSG.2020.2967803

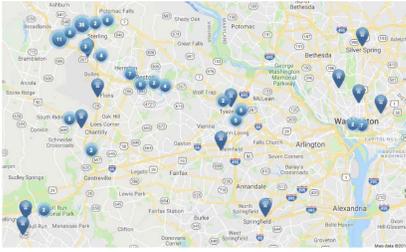


Fig. 1. The distribution of DCs around Ashburn [16].

the same electricity market and subject to locational marginal prices (LMP), which is widely adopted in America. LMP specifies different nodes different prices according to nodal demand and supply as well as physical constraints of power networks [2]. The CSP that owns multiple DCs in electricity market employing LMPs are known to possess market power, the ability to affect electricity prices [3]–[5]. Because, on the one hand, one single data center presents electricity consumption of about 10 MW, which is comparable to the household energy usage of 1000 US families [6]. On the other hand, the CSP can coordinate its DCs and redistribute the workload, which strengthens their ability of adjusting the electricity demands of some nodes. Therefore, the CSP is highly motivated by its considerable energy cost to leverage its two abilities: 1) transferring load among its DCs, and 2) affecting LMP, to cut down the total electricity bill that accounts for a major share of its operational cost. It is, however, technically challenging as the above two abilities are coupled together.

Related works on DC operation in electricity market can be categorized into two groups. Early studies often ignore the market power of DCs and regard them as price takers [7]–[14]. These works address optimization problems of DCs in temporal and spatial workload balancing [10], [11], [15], renewable energy accommodation [12], [13], and electricity procurement [14]. However, the price taker model only produces suboptimal solution in practice, because the prices are calculated after all bids of DCs have been collected. Sometimes, the resultant strategy is even worse than having no strategy, as illustrated by the example in Section II.

To overcome this drawback, recent years have witnessed the development of price maker model that accounts the market power of DCs. The competitions among individual participants in both cloud computing market and electricity market are simultaneously studied in [17] and then the data center capacity expansion problem is addressed with the electricity cost as the major consideration [18]. Studies [19], [20] consider how workload redistribution among DCs influences the load ratio of the grid, and adopt two-stage optimization to describe the interaction between DCs and grid. The market power of DCs is modeled using a supply function in [21]. Stackelberg game [22] and matching game [23] are employed to describe utility company choice by DCs and the real time price (RTP) set by different utility companies. DC is compared to large-scale storage in [24] and a prediction-based pricing is proposed to extract the potential of DC in DR. However, instead of admitting the pragmatic LMP model, all these studies are developed on simple price models which shelter the impact of workload transfer on LMP.

In summary, most of previous works on DC workload transfer in electricity market either treat DC as price taker, or assume simple price model rather than LMP to depict the market power of DC. To this end, the problem in this paper distinguishes from existing ones in threefold: 1) We consider a CSP possessing several urban neighboring DCs and transferring workload among them to reduce operational cost; 2) DCs are in the same electricity market, where the electricity pricing mechanism adopts LMP; 3) CSP is considered as a price maker that can anticipate how its strategy influences the LMP. To highlight the differences between this paper and previous works, we summarize the related work in Table I in Appendix D.

Our technical contributions are summarized below:

- We formulate a bilinear and bilevel problem to find the optimal workload transfer strategy of urban neighboring DCs owned by a CSP who can anticipate its influence on the electricity prices.
- To design effective solution, we convert the original problem into a single level one with a linear objective on a non-convex polytope feasible set. We propose an efficient polytope cutting algorithm to solve the converted problem.
- We analyze how the uncertainty of the workload transfer cost impacts the expected operational cost of CSP. It is shown that with the fluctuation of the transfer price increasing, the total cost of CSP does not increase as long as the mean of the transfer price keeps unchanged.
- The numerical results demonstrate that the total cost of CSP is reduced and the win-win situation for CSP and ISO is possible empirically.

The remainder of this paper is organized as follows. Section II describes the system model. In Section III, we formulate the price maker problem of CSP and transform it into a single level problem, for which a polytope cutting algorithm is proposed to obtain the global optimal solution. The influence of the workload transfer cost on total cost is analyzed in Section IV. In Section V, the simulation results validate the theoretical analysis.

II. SYSTEM MODEL

A. Power Network Model and Economic Dispatch

As shown in Fig. 2, we consider a power network which consists of a set of buses, $\mathbb{N} = \{1, 2, \dots, N_b\}$, and a set of transmission lines ε (the number of transmission line is N_l). Associated with the network is a susceptance matrix $\mathbf{Y} = \{\mathbf{Y}_{ij}\} \in \mathbb{R}^{N_b \times N_b}$, where $\mathbf{Y}_{ij} > 0$ is the susceptance of line (i, j) , if $(i, j) \in \varepsilon$; otherwise, $\mathbf{Y}_{ij} = 0$. In this paper we consider the direct current (dc) power flow model, which assumes that the voltages across the buses are constant, and the phase difference between any connected buses is small. The dc model is widely adopted by electricity markets such as, PJM and MISO, in their economic dispatch [2]. Without loss of generality, we assume at each bus there are one generator and one DC. Therefore, the market participants involved at bus i are a generator G_i , a DC, and other electricity consumers. Denoted by $\mathbf{x} \in \mathbb{R}^N$ and $\Theta \in \mathbb{R}^N$ are the power injected by generators and the phase

angle at buses, respectively. Let $\mathbf{z} \in \mathbb{R}^N$ and $\mathbf{d}^0 \in \mathbb{R}^N$ be the electricity consumption of DCs and other users, respectively.

The power flow is proportional to the circuit susceptance and the difference in voltage phase angles. When the network topology and parameters are specified, the power flow across circuits is determined by the difference in voltage phase angles between the terminating buses. The electricity is balanced at each bus, i.e.,

$$\mathbf{J}\Theta = \mathbf{x} - \mathbf{z} - \mathbf{d}^0 : \lambda, \quad (1)$$

where

$$\mathbf{J} = \text{Diag}(\mathbf{Y} \cdot \mathbf{1}) - \mathbf{Y}.$$

Here $\mathbf{1}$ is an all-one vector. In (1) and hereafter, the variable after the colon represents the Lagrange multiplier vector associated with the constraint. We assume that \mathbf{d}^0 is inelastic.

Each transmission line has a capacity, namely,

$$\mathbf{K}\Theta \leq \mathbf{L} : \mu, \quad (2)$$

where $\mathbf{L} \in \mathbb{R}^{2N_l}$ is the line capacity vector; matrix \mathbf{K} denotes the susceptance matrix with sign (slightly different from matrix \mathbf{Y}), and that the sign of $[\mathbf{K}]_{ij}$ is determined by the choice of power flow direction. When the network topology and parameters and direction are specified, the matrix \mathbf{K} is subsequently uniquely determined, that is, $|[\mathbf{K}]_{ij}| = \mathbf{Y}_{ij}$, if $(i, j) \in \varepsilon$, otherwise $[\mathbf{K}]_{ij} = 0$.

Every generator has a limited capacity denoted as \mathbf{X} and,

$$0 \leq \mathbf{x} \leq \mathbf{X} : (\underline{\alpha}, \bar{\alpha}). \quad (3)$$

Here $(\underline{\alpha}, \bar{\alpha})$ are the Lagrange multipliers associated with the left inequality and the right inequality, respectively.

We employ a linear function to model the generation cost, which is widely used in electricity market studies [10], [25], [26], as follows:

$$F(\mathbf{x}) = \mathbf{c}_1^T \mathbf{x}.$$

Conventionally, the economic dispatch (ED) problem of the ISO is to minimize the aggregated cost of the generation, given as follows:

$$\begin{aligned} \text{ED} : \min_{\mathbf{x}, \Theta} F(\mathbf{x}) \\ \text{s.t. } (1), (2), (3). \end{aligned}$$

Note that LMP at each bus corresponds to the Lagrange multiplier associated with the balance equations in (1). In this model, the price diversity of different nodes is caused by the congestion in the network.

B. Data Center Model

As shown in Fig. 2, suppose a CSP operates N_d DCs that are distributed across the power network and connected by N_c communication lines, through which the workload can be transferred. An undirected graph is used to describe the DC network, thus associated with each line are two nonnegative variables that indicate the workload transfer between two connected DCs in both directions, respectively. Let $\mathbf{m} \in \mathbb{R}^{2N_c}$

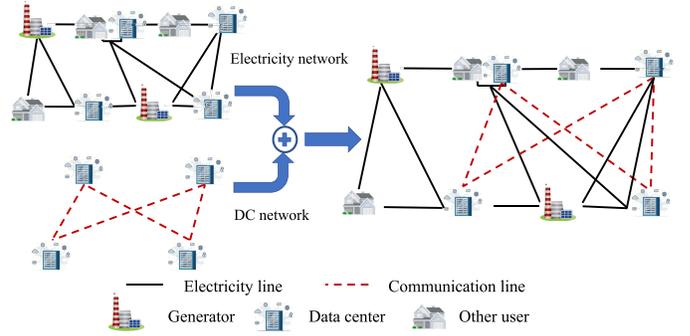


Fig. 2. The power network and data center network.

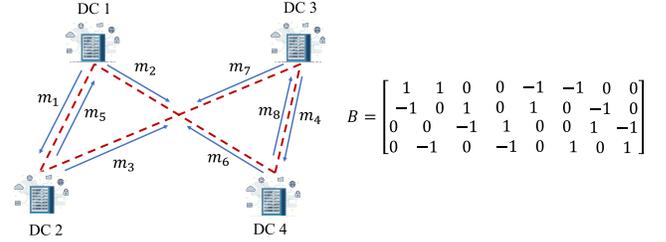


Fig. 3. Example of workload transfer and the incidence matrix.

denote workload transfer, which is limited by the transmission bandwidth,

$$0 \leq \mathbf{m} \leq \mathbf{M} : (\underline{\beta}, \bar{\beta}). \quad (4)$$

Our model is based on workload prediction, and assume the predicted workload originating at DCs is $\mathbf{z}^0, \mathbf{z}^0 \in \mathbb{R}^{N_b}$. Note that workload \mathbf{z}^0 is in terms of electricity demand. After workload transfer, the workload processed at DCs is \mathbf{z} within the capacity \mathbf{Z} :

$$\mathbf{z} = \mathbf{z}^0 - \mathbf{B}\mathbf{m},$$

and

$$0 \leq \mathbf{z} \leq \mathbf{Z} : (\underline{\gamma}, \bar{\gamma}), \quad (5)$$

where $\mathbf{B} \in \mathbb{R}^{N_b \times 2N_c}$ is the incidence matrix. As shown in Fig. 3, $\mathbf{B}_{ij} = 1$ (-1), if \mathbf{m}_j is the workload sent (received) by DC i to (from) one of its neighbors; otherwise, $\mathbf{B}_{ij} = 0$.

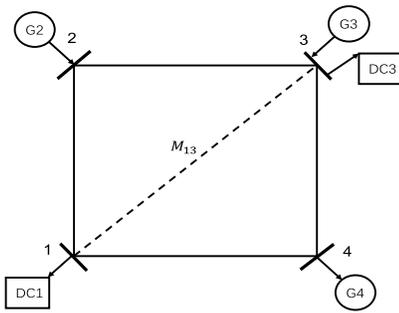
It is worth mentioning that the workload is pre-scheduled to be transferred among data centers when the CSP purchases electricity from day ahead market. When the workload arrive in real time, the CSP transfers the workload as planned. The mismatch of the predicted workload and the real workload is compensated by real time balancing or real time electricity procurement.

Assume workload transfer incurs a bandwidth and communication latency cost $R(\mathbf{m})$:

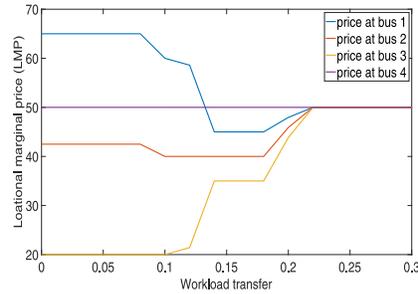
$$R(\mathbf{m}) = \mathbf{c}_2^T \mathbf{m}, \quad (6)$$

where \mathbf{c}_2 is a constant vector denoting the workload transfer price among DCs. Thus, the total cost of CSP consists of the workload transfer cost and the electricity cost:

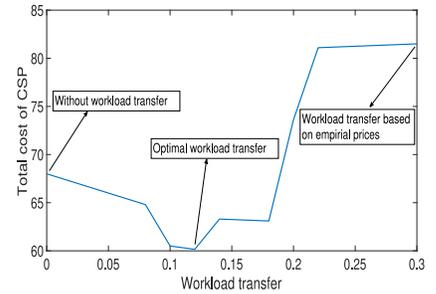
$$G(\mathbf{m}) = R(\mathbf{m}) + \lambda^T \mathbf{z} = \mathbf{c}_2^T \mathbf{m} - \lambda^T \mathbf{B}\mathbf{m} + \lambda^T \mathbf{z}^0. \quad (7)$$



(a) The configuration of the network.



(b) The LMP varying with the workload transfer.



(c) The total cost of the CSP.

Fig. 4. A four bus two DC example.

Observation: The communication lines can be regarded as special power transmission lines, as shifting workloads between two buses is equivalent to moving power between them in reverse direction. Meanwhile, such workload (equivalently power) transmission does not have to obey the Kirchhoff's Law. Thus, the integration of communication lines into the power system has its own advantages over directly building transmission lines.

There are two connections between DCs and economic dispatch: 1) energy, i.e., through the energy balance equation (1), and 2) economics, that is, solving economic dispatch problem yields electricity prices λ and the CSP transfer workload according electricity prices.

III. PRICE TAKER MODEL AND PRACTICAL ISSUE

A. A Noncooperative Game Between ISO and CSP

In this part, we assume that ISO and CSP do not share any information with each other and CSP is a price taker. CSP minimizes its cost by workload transfer according its estimation of LMP, which can be written as:

$$\begin{aligned} \text{CSP} : \min_{\mathbf{m}} G(\mathbf{m}) \\ \text{s.t. (4), (5).} \end{aligned}$$

In the noncooperative context of our model, both ISO and CSP are selfish and aim to maximize their own profit (or minimize their own cost) by their actions. Their best strategy depends on each other. Specifically, as rational players, provided the ISO has possessed how the CSP transfers its workload, it solves the ED to obtain best strategy. Meanwhile, given the CSP knows the strategy of ISO, the electricity price, it attains the best workload transfer strategy by minimizing $G(\mathbf{m})$. Therefore, there exists reciprocal dependency between the actions of ISO and CSP. This clarifies that the interaction of ISO and CSP can be characterized as a noncooperative game as characterized below.

- *Players:* ISO and CSP.
- *Strategy:* ISO: $(\mathbf{x}, \Theta, \lambda)$ satisfying constraints (1)-(3); CSP: \mathbf{m} satisfying constraints (4)-(5).
- *Payoff:* ISO: $-F(\mathbf{x})$; CSP: $-G(\mathbf{m})$.

The following theorem characterizes the property of the game.

Theorem 1: The game equilibrium is the optimal solution of the social cost minimization problem defined as

$$\begin{aligned} \text{SCM} : \min_{\mathbf{x}, \Theta, \mathbf{m}} F(\mathbf{x}) + R(\mathbf{m}) \\ \text{s.t. (1)-(5).} \end{aligned}$$

Please see the proof in Appendix A. The monetary cost or revenue is not included in social cost, because the monetary transfer between the ISO and CSP cancels out. (In other words, if the payment is taken into consideration, both ISO and CSP should include the monetary term in their total cost. The total social cost is: $F(\mathbf{x}) - \lambda^T \mathbf{z} + G(\mathbf{m}) = F(\mathbf{x}) - \lambda^T \mathbf{z} + R(\mathbf{m}) + \lambda^T \mathbf{z} = F(\mathbf{x}) + R(\mathbf{m})$). This result is appealing in that it means even in the strategic setting, there is no optimality loss in social cost as compared to the coordinating setting where the social cost is minimized. However, the social optimal solution may not be optimal for CSP and its cost can be worse than the situation where no workload is transferred, which is not ideal for the CSP. Moreover, to reach this solution usually requires a multi-round process where CSP and ISO exchange their price and demand information iteratively, which is not feasible in current electricity market.

B. Practical Issue

In today's electricity market, CSP has to report its nodal demands to ISO before LMP is computed and released, and the reported demands are not revisable. If the CSP implements optimal workload transfer strategy based on the LMP estimation from historical data, it may end up paying more than it does not transfer any workload. This situation is illustrated by the following example.

Example: (In this example the electricity and workload are in terms of MW and the cost is in dollar.) Consider a simple network with four buses and two DC as illustrated in Fig. 4(a), of which the electrical characteristics are drawn from the example in [27, p. 56]. Assume that the marginal cost of generators are constants, 40, 20, and 50, respectively. Their generation capacities are 1.5, 1.0, and 1.5. Suppose two DCs are owned by the same CSP, and the workload transfer capacity between them is 0.3 and the transfer price is 5. The workload originated at DC 1 and DC 2 are 1.0 and the load of other users at bus 1 is 1.1. Solving ED problem shows that if there is no workload transfer, the electricity price at bus 1 and bus 3 is 65 and 20, respectively. When CSP is a price taker,

it believes that its behavior has no influence on price. Thus it will transfer 0.3 workload to bus 3 and anticipate the total cost to be 68.0. However, the actual price at both bus 1 and 3 will change to 50, and the total cost of CSP is 81.5. Fig. 4(b) and Fig. 4(c) show how the workload transfer changes the nodal price and total cost of CSP.

IV. PRICE MAKER FORMULATION AND SOLUTION

A. Problem Formulation and Reformulation

In this section, we regard CSP as a price maker and further consider an ideal scenario, where CSP possesses an accurate model characterizing the relationship between LMP (λ) and its strategy (\mathbf{m}). In other words, CSP has learned all the parameters of problem **ED**, which is enabled by a large body of studies concerning how to accurately extract the power network information from limited observations [28]–[31]. Therefore, from the perspective of CSP the interaction between it and ISO can be regarded as a Stackelberg game, where CSP is the leader and ISO is the follower as formulated below. For brevity, we use $\Lambda = [\lambda, \mu, \underline{\alpha}, \bar{\alpha}]$ to collect all the Lagrange multipliers associated with equations (1)–(3).

$$\begin{aligned} \text{CSPa} : \min_{\mathbf{m}} G(\mathbf{m}) &= \mathbf{c}_2^T \mathbf{m} - \lambda^T \mathbf{Bm} + \lambda^T \mathbf{z}^0 \\ \text{s.t.} \quad (4), (5) \\ (\mathbf{x}, \Theta, \Lambda) &\text{ solves } \mathbf{ED}, \end{aligned} \quad (8)$$

where **ED** problem as the lower level problem is embedded in (8), and makes **CSPa** a two-level problem.

It is known that the two-level problem is NP hard even if the problems at both levels are LP [32]. Even more challenging, the upper level objective function $G(\mathbf{m})$ includes a nonconvex bilinear term, $\lambda^T \mathbf{Bm}$, which contains upper level decision variable, \mathbf{m} , and lower level dual variable, λ . The common way to deal with two-level problem is to replace the lower level problem by its KKT conditions and then convert it into a mixed integer problem by the big-M method [26]. However, directly employing this method results in a mixed integer bilinear nonconvex problem, for which it is difficult to solve.

In the following, we explore the structure of the problem and reformulate it into a special form that possesses linear objective and complementary constraints. First, this two-level problem can be converted into a single level problem by replacing the embedded **ED** problem by its KKT conditions:

$$\mathbf{c}_1^T - \lambda^T - \underline{\alpha}^T + \bar{\alpha}^T = 0; \quad (9a)$$

$$\lambda^T \mathbf{J} + \mu^T \mathbf{K} = 0; \quad (9b)$$

$$\mathbf{J}\Theta = \mathbf{x} - (\mathbf{z}^0 - \mathbf{Bm}) - \mathbf{d}^0; \quad (9c)$$

$$\mathbf{K}\Theta \leq \mathbf{L}; \quad (9d)$$

$$0 \leq \mathbf{x} \leq \mathbf{X}; \quad (9e)$$

$$\mu, \underline{\alpha}, \bar{\alpha} \geq 0; \quad (9f)$$

$$\mu^T (\mathbf{K}\Theta - \mathbf{L}) = 0; \quad (9g)$$

$$\underline{\alpha}^T \mathbf{x} = 0; \quad (9h)$$

$$\bar{\alpha}^T (\mathbf{x} - \mathbf{X}) = 0. \quad (9i)$$

Then, by multiplying (9a) by \mathbf{x} and together with (9h) and (9i) we have:

$$\begin{aligned} \lambda^T \mathbf{x} &= \mathbf{c}_1^T \mathbf{x} - \underline{\alpha}^T \mathbf{x} + \bar{\alpha}^T \mathbf{x} \\ &= \mathbf{c}_1^T \mathbf{x} + \bar{\alpha}^T \mathbf{x}. \end{aligned} \quad (10)$$

By multiplying (9b) by Θ and together with (9g) we have:

$$\lambda^T \mathbf{J}\Theta = -\mu^T \mathbf{K}\Theta = -\mu^T \mathbf{L}. \quad (11)$$

Multiplying (9c) by λ^T and combining (10) and (11) give:

$$\begin{aligned} \lambda^T (\mathbf{z}^0 - \mathbf{Bm}) &= -(\lambda^T \mathbf{J}\Theta - \lambda^T \mathbf{x} + \lambda^T \mathbf{d}^0) \\ &= \mu^T \mathbf{L} + \mathbf{c}_1^T \mathbf{x} + \bar{\alpha}^T \mathbf{x} - \lambda^T \mathbf{d}^0. \end{aligned} \quad (12)$$

Equation (12) indicates the bilinear term, $\lambda^T (\mathbf{z}^0 - \mathbf{Bm})$, can be replaced by a linear term. Therefore, the objective function of the upper level can be written as a linear function:

$$\begin{aligned} G(\mathbf{m}) &= \mathbf{c}_2^T \mathbf{m} + \lambda^T (\mathbf{z}^0 - \mathbf{Bm}) \\ &= \mathbf{c}_2^T \mathbf{m} + \mathbf{c}_1^T \mathbf{x} + \mathbf{L}^T \mu + \mathbf{X}^T \bar{\alpha} - \mathbf{d}^{0T} \lambda \\ &\triangleq g(\mathbf{m}, \mathbf{x}, \Theta, \Lambda) \end{aligned} \quad (13)$$

Thus, **CSPa** is equivalent to the following problem:

$$\begin{aligned} \text{CSPb} : \min_{\mathbf{m}, \mathbf{x}, \Theta, \Lambda} g(\mathbf{m}, \mathbf{x}, \Theta, \Lambda) \\ \text{s.t.} \quad (4), (5), (9) \end{aligned} \quad (14)$$

By above manipulations, **CSPa** has been equivalently converted into **CSPb**, a problem with linear objective and complementary constraints. The only barrier lies in the complementary constraints (9g)–(9i), for which we will develop a novel algorithm as detailed next.

B. Polytope Cutting Algorithm

It can be proved that the space shaped by constraints (4), (5) and (9) is a polytope [33]. It is based on this property that we develop our algorithm to solve **CSPb**. Below we will first give the definition of convex polytope and polytope [33].

Definition 1 (Convex Polytope): A convex polytope is defined as the intersection of a finite number of halfspaces, or a convex hull of any finite set of points.

Definition 2 (Polytope): A polytope is the union of finite number of convex polytope sets.

Let \mathbb{S}_1 and \mathbb{S}_2 denote the set described by (4), (5), (9a)–(9f) and (4), (5), (9), respectively. From above definitions, \mathbb{S}_1 is a convex polytope and \mathbb{S}_2 is a polytope, and the relationship between them is stated below.

Theorem 2: If the problem **CSPb** is feasible, \mathbb{S}_2 is at the face of \mathbb{S}_1 . (The face of a polytope is its surface, or boundary.)

The proof is in Appendix B. According to above theorem, \mathbb{S}_2 is a subset of \mathbb{S}_1 and can be regarded as a nonconvex polytope that consists of convex polytopes. Fig. 5(a) demonstrates this relationship in a two-dimension space. \mathbb{S}_1 is a convex polygon constructed by linear inequalities and \mathbb{S}_2 , the union of orange edges formed by linear inequalities and nonlinear equalities, is obvious non-convex. Nevertheless, each element of \mathbb{S}_2 , the edge, is convex.

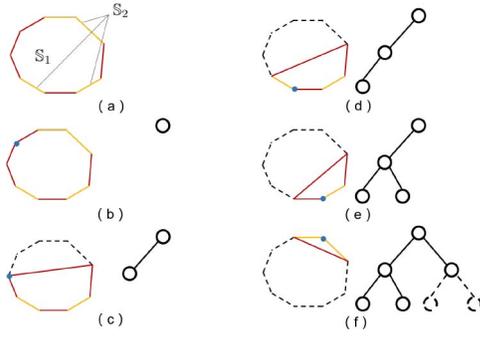


Fig. 5. An example of how a three-level tree develops.

The key idea of the algorithm is to find the solution by reducing the feasible set from \mathbb{S}_1 to the convex component of \mathbb{S}_2 . Step by step, \mathbb{S}_1 is cut into convex polytopes, upon which convex optimization problems are solved. For simplicity we define $\mathbf{t} = [\mathbf{L} - \mathbf{K}\Theta; \mathbf{x}; \mathbf{X} - \mathbf{x}] \in \mathbb{R}^{2(N_b + N_l)}$, then (9g)-(9i) can be written as

$$\Lambda^T \mathbf{t} = 0. \quad (15)$$

Let $u_i \in \{0, 1\}^i$, $i \in \{0, 1, 2, \dots, N\}$, where $N = 2(N_b + N_l)$. We use $\mathbb{S}(u_i)$ to denote the space shaped by fixing i components of \mathbf{t} to be 0:

$$\mathbb{S}(u_i) = \begin{cases} \Lambda^j = 0, \mathbf{t}^j \geq 0 & \text{if } u_i^j = 1; \\ \Lambda^j \geq 0, \mathbf{t}^j = 0 & \text{if } u_i^j = 0. \end{cases} \quad (16)$$

Here $[\cdot]^j$ denotes the j th component of a vector.

Then, it is straightforward that \mathbb{S}_2 can be written as:

$$\mathbb{S}_2 = \bigcup_{u_N \in \{0, 1\}^N} \mathbb{S}_1 \cap \mathbb{S}(u_N). \quad (17)$$

For each $u_N \in \{0, 1\}^N$, $\mathbb{S}_1 \cap \mathbb{S}(u_N)$ is a convex face of \mathbb{S}_1 . For each $u_i \in \{0, 1\}^i$, $i \in \{1, 2, \dots, N\}$, $\mathbb{S}_1 \cap \mathbb{S}(u_i)$ is a convex polytope containing at least one of the faces $\mathbb{S}_1 \cap \mathbb{S}(u_N)$, $u_N \in \{0, 1\}^N$.

Based on Theorem 2 and above analysis we design the following polytope cutting algorithm to obtain the solution of **CSPa**. We first define four sets: P_k containing the index of the complementary constraints that have been examined, $\Phi_k^+ = \{i | i \in P_k, \Lambda^i = 0\}$, $\Phi_k^- = \{i | i \in P_k, \mathbf{t}^i = 0\}$, and $\Phi_k^0 = \{i | i \notin P_k\}$. F_u is the candidate of optimal objective value.

- *Step 1 (Initialization)*: $k = 0$, set $\Phi_k^+ = \emptyset$, $\Phi_k^- = \emptyset$, $\Phi_k^0 = \{1, 2, \dots, N\}$, $P_k = \emptyset$, $F_u = \bar{F}$. (\bar{F} is a large enough constant.)
- *Step 2 (Iteration k)*: Set $\Lambda^i = 0$ for $i \in \Phi_k^+$ and $\mathbf{t}^i = 0$ for $i \in \Phi_k^-$. If problem **CSPLb** is infeasible, go to *Step 6*. Otherwise, solve **CSPLb** and obtain the solution $(\mathbf{m}_k^*, \mathbf{x}_k^*, \Lambda_k^*)$ and the corresponding objective value $g_k^* = g(\mathbf{m}_k^*, \mathbf{x}_k^*, \Lambda_k^*)$.
- *Step 3 (Fathoming)*: If $g_k^* \geq F_u$, go to *Step 6*.
- *Step 4 (Branching)*: If $\Lambda_k^i = 0$ hold for $i = 1, 2, \dots, N$, go to *Step 5*; Otherwise, set $i' = \arg \max_i \Lambda^i \mathbf{t}^i$. Update $\Phi_i^+ = \Phi_i^+ \cup \{i'\}$ and $\Phi_i^0 = \Phi_i^0 \setminus \{i'\}$, $P_k = [P_{k-1} \ i']$, and go to *Step 2*.
- *Step 5 (Updating)*: $F_u = g_k^*$.

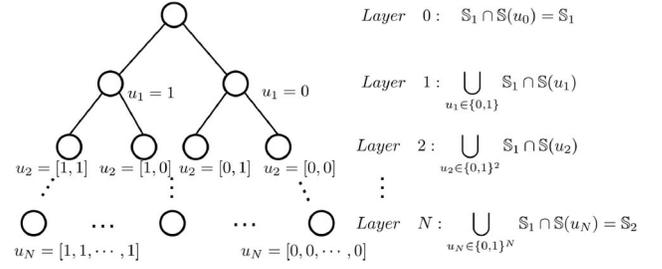


Fig. 6. The corresponding feasible sets of each layer in the branch-and-bound tree.

- *Step 6 (Backtracking)*: If there is no live node, i.e., $\Phi_k^0 = \emptyset$, go to *Step 7*. Otherwise trace back to the last node where new branch is possible, branch and update Φ_k^+ , Φ_k^- , Φ_k^0 , P_k accordingly, and go to 2.
- *Step 7 (Result)*: If $F_u = \bar{F}$, no feasible solution exists. Otherwise, the optimal solution is that with F_u .

In *Step 2*, problem **CSPLb** in iteration k is an LP problem and is defined as

$$\begin{aligned} \text{CSPLb:} \quad & \min_{\mathbf{m}, \mathbf{x}, \Theta, \Lambda} && g(\mathbf{m}, \mathbf{x}, \Theta, \Lambda) \\ & \text{s.t.} && (4), (5), (9a)–(9f), \\ & && \Lambda^i = 0, \forall i \in \Phi_k^+, \\ & && \mathbf{t}^i = 0, \forall i \in \Phi_k^-. \end{aligned} \quad (18)$$

The branch-and-bound tree is visualized in Fig. 6 (u_i represents one possible realization), where the proposed algorithm starts from the root node (\mathbb{S}_1) and goes down to lower layer (\mathbb{S}_2) to check if the optimal solution can be obtained at a node. The corresponding feasible sets of the layers from the top to the bottom shrink from \mathbb{S}_1 to \mathbb{S}_2 . The optimal solution may be found before reaching the bottom layer. In each iteration, at *Step 2*, the objective is minimized upon a convex polytope containing at least one of the components of \mathbb{S}_2 . *Step 4* checks if the solution lies in \mathbb{S}_2 ; if it does, the solution is a candidate for the optimum. *Step 6* checks if there exist nodes to be examined in next iteration. The proposed algorithm can find the global optimal solution, because in the worst case it searches all the nodes. Fortunately, in practice the algorithm usually finds the optimal solution in a small number of iterations, which is verified in our simulation in Section VI. An illustrative example is given below to show how the algorithm proceeds.

As illustrated in Fig. 5(a), the feasible set of the original problem, \mathbb{S}_2 , contains four edges of the convex polytope \mathbb{S}_1 . The root node corresponds to a convex optimization problem upon the feasible set \mathbb{S}_1 as shown in Fig. 5(b), where the blue point represents the solution. The solution is not feasible for the original problem, thus a constraint is added and the feasible set is cut into a new convex set shown in Fig. 5(c). Similarly, the feasible set is further cut, and a feasible solution is obtained in Fig. 5(d), and F_u is updated. Next, the tree branches as Fig. 5(e). Then it traces back to the root node and continues to branch as demonstrated in Fig. 5(f). Suppose here a smallest ever objective is obtained, then the nodes below (shown by the dashed ones) are eliminated, because they cannot provide a better solution. Therefore, the algorithm stops at this node and returns the solution found in Fig. 5(f).

The proposed algorithm is developed following the ideas of branch-and-bound and cutting plane methods, and it solves only a simple linear problem in each iteration. In terms of the difference between our algorithm and cutting plane method, usually cutting plane is commonly adopted to find integer solutions to mixed integer linear programming (MILP) problems, as well as to solve general, not necessarily differentiable convex optimization problems. In comparison, our algorithm, a combination of branch-and-bound and cutting plane method, solves our nonconvex problem which contains bilinear terms in the constraints. Note that Big-M method can linearize the bilinear term in constraints by introducing auxiliary integer variables. However, bilinear terms exist in both objective function and constraints in **CSPb**. Therefore, adopting conventional Big-M method [26], [34] or Benders decomposition [18], [35], [36] to solve the original problem requires us to solve a nonlinear problem in each iteration. Obviously, it is more expensive than our algorithm which solves linear problems instead. The proposed algorithm performs well for practical engineering problems, which is verified in our numerical studies. However, the limitation of the presented algorithm is that theoretically the computation grows exponentially with the problem size in worst case. To tackle this problem, in practical implementation, if the computation time to solve the problem is limited, the algorithm can stop before a global solution is found and return a local optimal solution. A local optimal solution is usually acceptable in practice considering the nonconvexity of the problem.

V. THE INFLUENCE OF TRANSFER PRICE UNCERTAINTY

In practice, the workload transfer price is not constant, because of the continuously vary network conditions. For example, congestion in communication lines induces a large delay of workload transfer, which corresponds to a high transfer price. Assume the transfer price \mathbf{c}_2 follows a distribution within a certain range, and the entries of \mathbf{c}_2 are independent. We use variance σ^2 to represent the fluctuation or uncertainty of the transfer price \mathbf{c}_2 . To investigate the impact of the fluctuation of transfer price on total cost of CSP, we assume that \mathbf{c}_2^a and \mathbf{c}_2^b are two transfer prices sampled from possibly different distributions, and they have the same mean but different variance σ_a^2 and σ_b^2 . Let $g^*(\mathbf{c}_2) = g(\mathbf{m}^*, \mathbf{x}^*, \Theta^*, \Lambda^*)$ denote the optimal objective value given \mathbf{c}_2 . The following theorem addresses the expected cost of CSP in presence of transfer price uncertainty.

Theorem 3: The larger transfer price fluctuation does not cause more total cost. Mathematically, assume that \mathbf{c}_2^a and \mathbf{c}_2^b are the transfer price with the same mean but different variance, i.e., $\mathbb{E}[\mathbf{c}_2^a] = \mathbb{E}[\mathbf{c}_2^b]$ and $\sigma_a^2 \geq \sigma_b^2$ (element wise larger than or equal to), then the expected cost under \mathbf{c}_2^a is not higher than that under \mathbf{c}_2^b , namely, $\mathbb{E}[g^*(\mathbf{c}_2^a)] \leq \mathbb{E}[g^*(\mathbf{c}_2^b)]$.

Please see Appendix C for the proof. It seems counterintuitive that the expected cost of the CSP will not increase with the increase of transfer price fluctuation, given the same mean of price. The intuition behind it is that when transfer price is low, the overall cost can be significantly reduced by workload transfer. With the increase of transfer price, the total cost will rise. But when the transfer price is too high, the total cost will

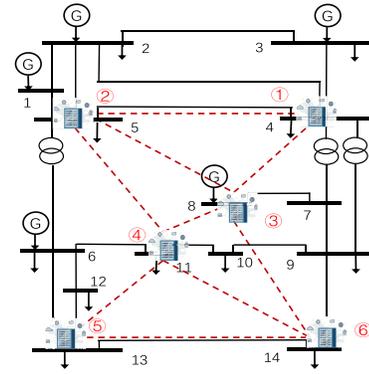


Fig. 7. 6 distributed DCs in an IEEE 14 bus test system.

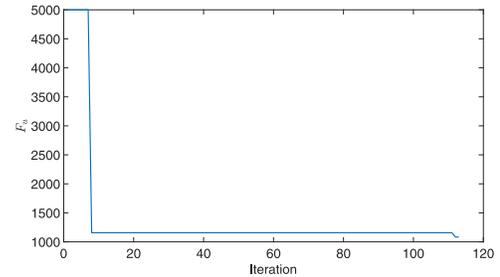


Fig. 8. The iteration process of the proposed algorithm.

not increase any more, because the cost reduction from workload transfer cannot compensate transfer cost. In other words, this communication line will not be used when the transfer price exceeds a certain threshold. Provided the mean of the price is fixed, the price with larger variance is more likely to be high (at the same time more likely to be low) compared with the price with smaller variance. Therefore, when the fluctuation is significant, more benefit of the low price is enjoyed and the poison of the high price is circumvented.

VI. SIMULATION

To validate the effectiveness of the proposed algorithm, we carry out the simulations on two standard IEEE test feeders, IEEE 14 bus case and IEEE 118 bus case [37], of which the details are presented in next two subsections.

A. IEEE 14 Bus Test Case

1) *Simulation Setup:* As illustrated in Fig. 7, we consider a CSP with 6 DCs and 9 communication lines. In our simulation we scale the demand and line capacity to make the LMP in the network different. We set $\mathbf{c}_1 = [20, 30, 40, 30, 45]^T$ (vector \mathbf{c}_1 in our model should be of length N_b , but here we only give the cost of 5 generators, other elements of \mathbf{c}_1 are 0. Below we give the parameters in the same manner.) $\mathbf{z}^0 = [4.4, 5.5, 3.3, 6.6, 5.5, 3.85]^T$, $\mathbf{Z} = [6, 8, 5.5, 9, 7, 5]^T$, and $\mathbf{M} = [3.6, 4.8, 2.4, 1.2, 6.0, 2.4, 1.2, 4.8, 3.6]^T$, $\mathbf{c}_2 = [9, 13.5, 18.0, 10.5, 15.9, 21.0, 18.3, 13.5, 18.0]^T$.

2) *Price Maker v.s. Price Taker:* First we show the proposed algorithm is effective in finding the global optimal solution. As demonstrated in Fig. 8, the algorithm stops after 112 iterations and obtains optimal solution. Note that in our

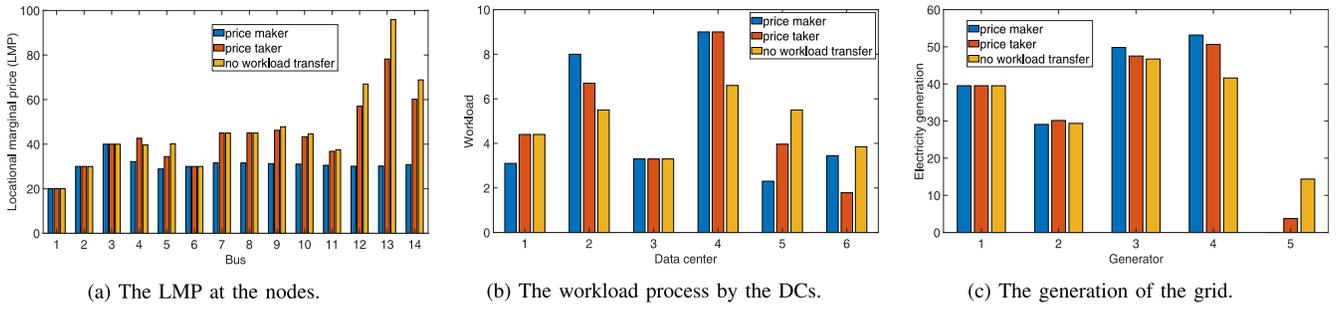


Fig. 9. The different impact of price taker and price maker behavior.

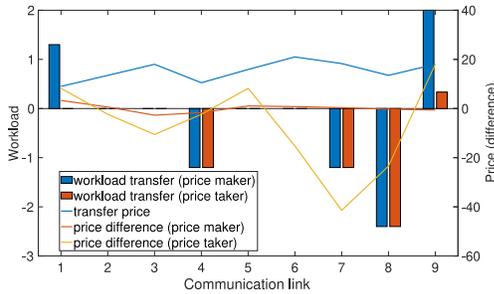


Fig. 10. The price difference and workload transfer of the links.

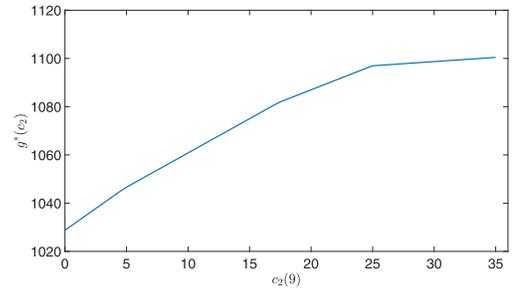


Fig. 11. The influence of transfer price on total cost of CSP.

simulation we initialize $F_u = 5000$, which is large enough for this simulation setting.

The impact of the workload transfer of DCs on electricity market is compared under three different settings in which CSP participates in the electricity market: 1) the price maker model, 2) the price taker model as in work [12], and 3) no workload transfer. Note that in our simulations the price is terms on \$/MWh. Fig. 14(a) shows that LMP in both price maker model and price taker model is flattened compared with the case without workload transfer, because workload transfer relieves the power network congestion. The price maker model influences the price on behalf of the CSP, therefore it results in a relatively lower price than the price taker model. Illustrated in Fig. 14(b) is the workload processed at each DC after workload transfer. As a price maker, CSP can anticipate how its strategy changes the price, therefore it transfers the workload at DC 5 and 6 (bus 13 and 14) to other DCs without increasing the prices of other nodes too much but significantly reducing the prices at bus 13 and 14.

The workload transfer will influence the generation scheduling of the grid as shown in Fig. 14(c). When no workload transfer is conducted, the grid has to use the most expensive generator, generator 5, to balance the supply and demand. The output of generator 5 is reduced when the workload transfer is allowed, and further reduced to zero when CSP is a price maker. Note that even the price taker model is essentially minimizing the social cost (the total cost of the generation and workload transfer), it obtains a social cost of 5357.3 which is only slightly lower than that of the price maker model, 5365.5. While the latter yields a much smaller CSP cost (1083.0) than the former (1348.0). This is because that CSP, as a price maker, can significantly reduce LMP at expense of a slightly higher transfer cost.

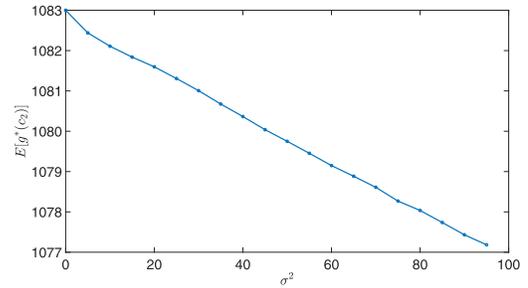


Fig. 12. The expected total cost decreases with the increasing transfer price uncertainty.

Furthermore, it is worth mentioning that the workload transfer strategy of CSP can result in a win-win situation, i.e., mutually beneficial for both CSP and ISO. As seen in this example, without workload transfer the costs of CSP and ISO are 1583.7 and 5435.6, respectively, which are reduced to 1083.0 and 5251.0 when the price maker workload transfer strategy is implemented. Further discussions about the conditions on which the win-win situation is guaranteed to appear remain to be our future work.

Fig. 10 shows workload transfer, transfer price and electricity price difference between two end nodes of a communication link. When the signs of workload transfer and price difference are the same, the workload is transferred from the higher price node to the lower one. For the price taker model, DCs with net workload injection (extraction) are two DCs located in the two buses with lowest (highest) LMP. It is also notable that even when the electricity price difference is not larger than the transfer price, there may still be workload transferred through a link, for example link 4. This is because a DC can play as a relay to transfer workload for two DCs with electricity price difference larger than the transfer price. In the price taker model, the price difference between the source DC and

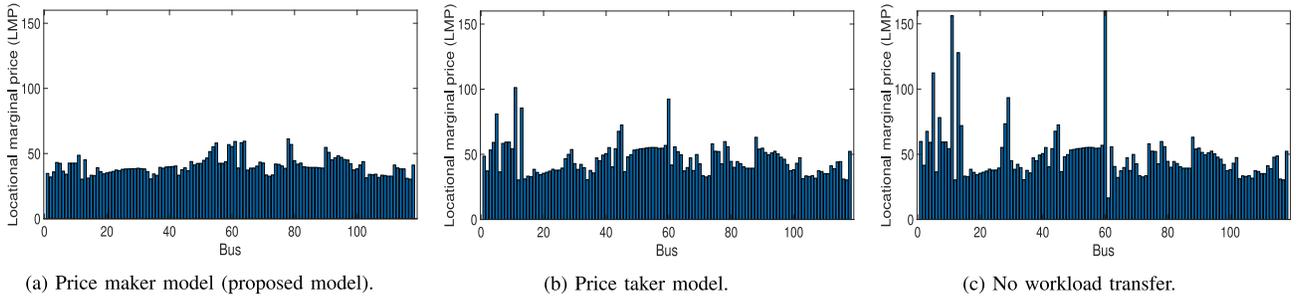


Fig. 13. Electricity prices across all buses.

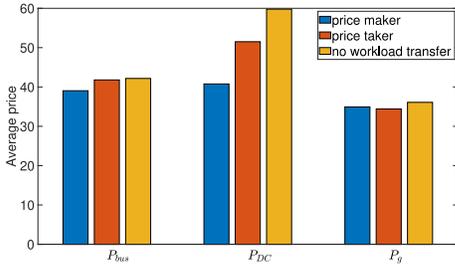


Fig. 14. Average electricity price over buses, data centers and generators.

destination DC must be larger than the transfer price between the two DCs, which is the incentive for workload transfer. However, the situation is quite different in the price maker model. There can be workload transfer even electricity price difference between the source DC and the destination DC is less than transfer price. Because the price maker understands how its behavior influences the prices, its behavior does not totally determined by the prices as that of the price taker does. This is demonstrated by the result that there is workload transfer in price maker model through link 1, where the price difference of the two DCs is less than the transfer price.

3) *The Transfer Price Uncertainty*: Fig. 11 demonstrates that the optimal total cost of CSP is an increasing and concave function of c_2 . In our simulation, we vary the transfer price of the link 9 (connecting DC 5 and 6) and keep all the other transfer prices fixed. When the mean of transfer price of link 9 is 18 and the variance increases from 0 to 95, the total cost decreases from 1083.0 to 1077.2, which coincides with Theorem 3 as shown in Fig. 12.

B. IEEE 118 Bus Test Case

In this case, we enumerate 34 DCs, each of which is connected to 3 other DCs on average. Fig. 13 shows the electricity prices over all buses under different models. Compared with Fig. 13(c), Fig. 13(a) and Fig. 13(b) show that workload transfer can smooth the price diversity over buses, which verifies that the links among DCs can be regarded as power transmission lines to relieve the power network congestion.

Then we show the average electricity price from three perspectives. Let define $P_{bus} = \frac{\lambda^T(\mathbf{z}+\mathbf{d}^0)}{\mathbf{1}^T(\mathbf{z}+\mathbf{d}^0)}$, $P_{DC} = \frac{\lambda^T\mathbf{z}}{\mathbf{1}^T\mathbf{z}}$, and $P_g = \frac{\mathbf{c}_1^T\mathbf{x}}{\mathbf{1}^T\mathbf{x}}$, as the average price over all buses, data centers, and generators. Note that the average price of generators is the electricity generation price. It is shown by Fig. 14 that the price maker model yields the lowest selling price no matter for all the buses or for the buses where there are DCs,

which means that the price model benefits all the consumers. When there is a DC at a bus the price is significantly lower than other two models. For the electricity generation price, the price difference is not significant, and the price maker model generates slightly higher price than the price taker model does. It is obvious that P_{bus} is higher than P_g , because that there is congestion which leads to transmission cost.

VII. CONCLUSION

The growing electricity demand of data centers has driven cloud service providers to pursue effective and efficient data center operation scheme to reduce their operational expenditure. In this context, the CSP that operates urban neighboring data centers can transfer workload to the data centers where electricity is cheaper. However, the enormous electricity consumption endows the data centers market power, which means the electricity prices change with the workload redistribution. In order to model the market power of data center, the optimization problem for CSP is formulated as a bilevel problem, where the upper level deals with the goal of CSP and the lower level is the economic dispatch problem that yields the electricity prices. Compared with previous works, the market power is clearly characterized in this bilevel model. Although being mathematically elegant, this bilevel problem is nonconvex, and thus to solve it directly is challenging. By exploiting the structure and property of the problem, we reformulate it and develop a polytope cutting algorithm to obtain the global optimal solution. The simulation results verify that the workload transfer has significant impact on the electricity prices and in some certain settings, both CSP and ISO can observe cost reduction from the workload transfer. However, it is not clear under what condition such win-win situation is guaranteed to exist, which deserves further effort.

APPENDIX A PROOF OF THEOREM 1

Proof: Let $(\mathbf{x}^*, \Theta^*, \lambda^*, \mathbf{m}^*)$ be the equilibrium decision of both players. At the equilibrium, both players are playing the best response strategy. Therefore, given \mathbf{m}^* , $(\mathbf{x}^*, \Theta^*, \lambda^*)$ is the minimizer of **ED**, and given $(\mathbf{x}^*, \Theta^*, \lambda^*)$, \mathbf{m}^* is the solution of **CSP**.

In other words, $(\mathbf{x}^*, \Theta^*, \lambda^*)$ satisfies the following KKT conditions of problem **ED** when $\mathbf{m} = \mathbf{m}^*$.

$$\mathbf{c}_1^{*T} - \lambda^{*T} - \alpha^{*T} + \bar{\alpha}^{*T} = 0; \quad (19a)$$

$$\mathbf{J}\Theta^* = \mathbf{x}^* - (\mathbf{z}^0 - \mathbf{B}\mathbf{m}^*) - \mathbf{d}^0; \quad (19b)$$

TABLE I
SUMMARY OF WORKS ON THE PARTICIPATION OF DC IN ELECTRICITY MARKET

References	Scenario				Objective	Method / Mathematical Tool
	Workload transfer: intra DC/ inter DC	electricity purchase in single/ multiple markets	Power flow model	Price taker/ Price maker		
Lei Rao [11]	Inter DC	Multi-market	No	Price taker	Electricity cost minimization with QoS guaranteed	Geographical load balance / Mixed integer quadratic program
Tianyi Chen [12], [39], Yuan Yao [13]	Inter DC	Multi-market	No	Price taker	Operational cost minimization	Online resource allocation / Stochastic optimization & Lyapunov optimization
Niangjun Chen [4], Yuanxiong Guo [5]	Intra DC	Single market	No	Price taker	Tenant: payoff maximization & DC: social cost minimization	Tenant supply function bidding for load curtailment / pricing mechanism design
Peijian Wang [14]	Intra DC	Multi-market	No	Price taker	CSP operational cost minimization	Second chance bidding / Markov decision process based optimization
Ying Zhang [10]	Intra DC	Multi-market	No	Price taker	CSP operational cost minimization	Geographical load balance & day ahead and real time trading / convex optimization
Hao Wang [19], [20]	Inter DC	Single market	No	Price maker	DC: Energy Cost Minimization & Utility: load balance (minimizing electric load index)	Two stage problem to model the interaction between DC and utility / Mixed integer linear program
Peijian Wang [21]	Inter DC	-	No	Price maker	Electricity Cost Minimization	Supply function to model the market power / Mixed integer nonlinear program
Nguyen H. Tran [22] S. Bahrami [23]	Inter DC	Single market	No	Price maker	DC operational cost minimization & Utility: profit maximization	Workload shifting & dynamic server allocation / Stackelberg game & matching game
Zhenhua Liu [24]	Intra DC	Single market	No	Price maker	DC: payoff maximization & Utility: Social cost minimization	Prediction-based pricing / Market design
This paper	Inter DC	Single market (LMP)	dc power flow	Price maker	DC operational cost minimization	Price anticipating workload transfer strategy / Two-level problem

$$\lambda^{*T} \mathbf{J} + \mu^{*T} \mathbf{K} = 0; \quad (19c)$$

$$\mathbf{K} \Theta^* \leq \mathbf{L}; \quad (19d)$$

$$0 \leq \mathbf{x}^* \leq \mathbf{X}; \quad (19e)$$

$$\mu^{*T} (\mathbf{K} \Theta^* - \mathbf{L}) = 0; \quad (19f)$$

$$\underline{\alpha}^* \mathbf{x}^* = 0; \quad (19g)$$

$$\bar{\alpha}^{*T} (\mathbf{x}^* - \mathbf{X}) = 0; \quad (19h)$$

$$\mu^*, \underline{\alpha}^*, \bar{\alpha}^* \geq 0. \quad (19i)$$

Similarly, when $(\mathbf{x}, \Theta, \lambda) = (\mathbf{x}^*, \Theta^*, \lambda^*)$, \mathbf{m}^* satisfies the following KKT conditions.

$$\mathbf{c}_2^{*T} - \lambda^{*T} \mathbf{B} + \underline{\gamma}^{*T} \mathbf{B} - \bar{\gamma}^{*T} \mathbf{B} - \underline{\beta}^{*T} + \bar{\beta}^{*T} = 0; \quad (20a)$$

$$0 \leq \mathbf{m}^* \leq \mathbf{M}; \quad (20b)$$

$$0 \leq \mathbf{z}^0 - \mathbf{Bm}^* \leq \mathbf{Z}; \quad (20c)$$

$$\underline{\gamma}^{*T} (\mathbf{z}^0 - \mathbf{Bm}^*) = 0; \quad (20d)$$

$$\bar{\gamma}^{*T} (\mathbf{z}^0 - \mathbf{Bm}^* - \mathbf{Z}) = 0; \quad (20e)$$

$$\underline{\beta}^{*T} \mathbf{m}^* = 0; \quad (20f)$$

$$\bar{\beta}^{*T} (\mathbf{m}^* - \mathbf{M}) = 0; \quad (20g)$$

$$\underline{\gamma}^*, \bar{\gamma}^*, \underline{\beta}^*, \bar{\beta}^* \geq 0. \quad (20h)$$

It can be verified that (19) and (20) is exactly the KKT conditions of problem **SCM**. Since problem **SCM** is a convex optimization problem with Slater conditions satisfied (This is usually true for practical engineering problems), $(\mathbf{x}^*, \Theta^*, \lambda^*, \mathbf{m}^*)$ satisfying its KKT conditions is the optimal solution. ■

APPENDIX B PROOF OF THEOREM 2

Proof: (The proof is inspired by the work in [33] and [38].) **CSPb** is equivalent to the following two level problem:

$$\begin{aligned} \mathbf{CSPc} : \min_{\mathbf{m}} & g(\mathbf{m}, \mathbf{x}, \Theta, \Lambda) \\ \text{s.t.} & (4), (5) \\ & (\mathbf{x}, \Theta, \Lambda) \text{ solves} \\ & \min_{\mathbf{x}, \Theta, \Lambda} h(\mathbf{m}, \mathbf{x}, \Lambda) \\ \text{s.t.} & (9a)-(9f), \end{aligned} \quad (21)$$

where $h(\mathbf{m}, \mathbf{x}, \Lambda) = \mathbf{c}_1^T \mathbf{x} - [(\mathbf{z}^0 - \mathbf{Bm})^T \lambda + \mathbf{d}^{0T} \lambda - \mathbf{L}^T \mu - \mathbf{X}^T \bar{\alpha}]$. The KKT conditions in (9) are equivalent to the **CSPc**'s lower level problem which combines both the primal and dual problem of **ED**.

Now we prove that \mathbb{S}_2 is the face of \mathbb{S}_1 by contradiction.

According to the definition of \mathbb{S}_1 and \mathbb{S}_2 , for any $y \in \mathbb{S}_2$, there exist $y_i \in \mathbb{S}_1$ and $r_i > 0$, $i = 1, 2, \dots, p$ and $\sum_{i=1}^p r_i = 1$, such that $y = \sum_{i=1}^p r_i y_i$. Suppose that there exists a $q \in \{1, 2, \dots, p\}$ such that $y_q = (\mathbf{m}_q, \mathbf{x}_q, \Lambda_q) \notin \mathbb{S}_2$. Then there exists a \hat{y}_q such that $\hat{y}_q = (\mathbf{m}_q, \mathbf{x}_q, \hat{\Lambda}_q) \in \mathbb{S}_2$ and $h(y_q) = h(\mathbf{m}_q, \mathbf{x}_q, \Lambda_q) > h(\mathbf{m}_q, \mathbf{x}_q, \hat{\Lambda}_q) = h(\hat{y}_q)$. Thus,

$$\hat{y} = r_q \hat{y}_q + \sum_{i=1, i \neq q}^p r_i y_i \in \mathbb{S}_1. \quad (22)$$

Therefore,

$$h(y) = r_q h(y_q) + \sum_{i=1, i \neq q}^p r_i h(y_i) > r_q h(\hat{y}_q) + \sum_{i=1, i \neq q}^p r_i h(y_i) \quad (23)$$

which contradicts with the assumption $y \in \mathbb{S}_2$.

Therefore, \mathbb{S}_2 as well as the optimal solution must be on the face of \mathbb{S}_1 . ■

APPENDIX C PROOF OF THEOREM 3

Proof: First we prove that $g^*(\mathbf{c}_2)$ is a concave function of \mathbf{c}_2 :

From (13), it can be obtained $g^*(\mathbf{c}_2) = \mathbf{c}_2^T \mathbf{m}^*(\mathbf{c}_2) + \mathbf{c}_1^T \mathbf{x}^*(\mathbf{c}_2) + L^T \mu^*(\mathbf{c}_2) + \mathbf{X}^T \bar{\alpha}^*(\mathbf{c}_2) - \mathbf{d}^{0T} \lambda^*(\mathbf{c}_2)$. Let \mathbf{c}_2^0 and $\bar{\mathbf{c}}_2$ be any values in $[\underline{\mathbf{c}}_2, \bar{\mathbf{c}}_2]$, and $\delta \in (0, 1)$, then $\mathbf{c}_2^0 = \delta \mathbf{c}_2^1 + (1 - \delta) \bar{\mathbf{c}}_2 \in [\underline{\mathbf{c}}_2, \bar{\mathbf{c}}_2]$. Therefore,

$$\begin{aligned} g^*(\mathbf{c}_2^0) &= \mathbf{c}_2^T \mathbf{m}^*(\mathbf{c}_2^0) + \mathbf{c}_1^T \mathbf{x}^*(\mathbf{c}_2^0) \\ &\quad + L^T \mu^*(\mathbf{c}_2^0) + \mathbf{X}^T \bar{\alpha}^*(\mathbf{c}_2^0) - \mathbf{d}^{0T} \lambda^*(\mathbf{c}_2^0) \\ &= \delta \left[\mathbf{c}_2^T \mathbf{m}^*(\mathbf{c}_2^0) + \mathbf{c}_1^T \mathbf{x}^*(\mathbf{c}_2^0) \right. \\ &\quad \left. + L^T \mu^*(\mathbf{c}_2^0) + \mathbf{X}^T \bar{\alpha}^*(\mathbf{c}_2^0) - \mathbf{d}^{0T} \lambda^*(\mathbf{c}_2^0) \right] \\ &\quad + (1 - \delta) \left[\mathbf{c}_2^T \mathbf{m}^*(\mathbf{c}_2^0) + \mathbf{c}_1^T \mathbf{x}^*(\mathbf{c}_2^0) + L^T \mu^*(\mathbf{c}_2^0) \right. \\ &\quad \left. + \mathbf{X}^T \bar{\alpha}^*(\mathbf{c}_2^0) - \mathbf{d}^{0T} \lambda^*(\mathbf{c}_2^0) \right] \quad (24) \end{aligned}$$

Since $(\mathbf{m}^*(\mathbf{c}_2^1), \mathbf{x}^*(\mathbf{c}_2^1), \Lambda^*(\mathbf{c}_2^1))$ and $(\mathbf{m}^*(\mathbf{c}_2^0), \mathbf{x}^*(\mathbf{c}_2^0), \Lambda^*(\mathbf{c}_2^0))$ are the optimal solutions that minimize $g(\mathbf{c}_2)$ given $\mathbf{c}_2 = \mathbf{c}_2^1$ and $\mathbf{c}_2 = \mathbf{c}_2^0$, respectively.

$$\begin{aligned} g^*(\mathbf{c}_2^0) &\geq \delta \left[\mathbf{c}_2^T \mathbf{m}^*(\mathbf{c}_2^1) + \mathbf{c}_1^T \mathbf{x}^*(\mathbf{c}_2^1) \right. \\ &\quad \left. + L^T \mu^*(\mathbf{c}_2^1) + \mathbf{X}^T \bar{\alpha}^*(\mathbf{c}_2^1) - \mathbf{d}^{0T} \lambda^*(\mathbf{c}_2^1) \right] \\ &\quad + (1 - \delta) \left[\mathbf{c}_2^T \mathbf{m}^*(\mathbf{c}_2^0) + \mathbf{c}_1^T \mathbf{x}^*(\mathbf{c}_2^0) + L^T \mu^*(\mathbf{c}_2^0) \right. \\ &\quad \left. + \mathbf{X}^T \bar{\alpha}^*(\mathbf{c}_2^0) - \mathbf{d}^{0T} \lambda^*(\mathbf{c}_2^0) \right] \\ &= \delta g^*(\mathbf{c}_2^1) + (1 - \delta) g^*(\mathbf{c}_2^0). \quad (25) \end{aligned}$$

Above inequality implies that $g^*(\mathbf{c}_2)$ is a concave function of \mathbf{c}_2 . For a concave function $g^*(\cdot)$, and \mathbf{c}_2^a and \mathbf{c}_2^b satisfying $\mathbb{E}[\mathbf{c}_2^a] = \mathbb{E}[\mathbf{c}_2^b]$ and $\sigma_a^2 \geq \sigma_b^2$, it directly follows that $\mathbb{E}[g^*(\mathbf{c}_2^a)] \leq \mathbb{E}[g^*(\mathbf{c}_2^b)]$ according to [10, Proposition 4]. ■

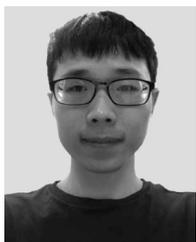
APPENDIX D SUMMARY OF RELATED WORKS

See Table I.

REFERENCES

- [1] "Equinix customer case study CDM Smith," Equinix, Redwood City, CA, USA, Rep. v050218 CS-EN CC-DD 0616, 2018.
- [2] Z. Yang *et al.*, "LMP revisited: A linear model for the loss-embedded LMP," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 4080–4090, Sep. 2017.
- [3] A. Wierman, Z. Liu, I. Liu, and H. Mohsenian-Rad, "Opportunities and challenges for data center demand response," in *Proc. Int. Green Comput. Conf. (IGCC)*, 2014, pp. 1–10.
- [4] N. Chen, X. Ren, S. Ren, and A. Wierman, "Greening multi-tenant data center demand response," *Perform. Eval.*, vol. 91, pp. 229–254, Sep. 2015.
- [5] Y. Guo, H. Li, and M. Pan, "Colocation data center demand response using Nash bargaining theory," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4017–4026, Sep. 2018.
- [6] *Average Household Electricity Consumption*. Accessed: Dec. 2019. [Online]. Available: <http://shrinkthatfootprint.com/average-household-electricity-consumption>
- [7] Y. Feng, B. Li, and B. Li, "Price competition in an oligopoly market with multiple IaaS cloud providers," *IEEE Trans. Comput.*, vol. 63, no. 1, pp. 59–73, Jan. 2014.
- [8] J. Sun, Z. Yang, Z. Guo, and Q. Yang, "Long term operational optimization of data center network under uncertainty," in *Proc. IEEE 14th Int. Conf. Control Autom. (ICCA)*, 2018, pp. 522–527.
- [9] J. Sun, Q. Yang, and Z. Yang, "Probability based online algorithm for switch operation of energy efficient data center," *IEEE Trans. Cloud Comput.*, to be published.
- [10] Y. Zhang, L. Deng, M. Chen, and P. Wang, "Joint bidding and geographical load balancing for datacenters: Is uncertainty a blessing or a curse?" *IEEE/ACM Trans. Netw.*, vol. 26, no. 3, pp. 1049–1062, Jun. 2018.
- [11] L. Rao, X. Liu, L. Xie, and W. Liu, "Minimizing electricity cost: Optimization of distributed Internet data centers in a multi-electricity-market environment," in *Proc. IEEE INFOCOM*, 2010, pp. 1–9.
- [12] T. Chen, Q. Ling, and G. B. Giannakis, "An online convex optimization approach to proactive network resource allocation," *IEEE Trans. Signal Process.*, vol. 65, no. 24, pp. 6350–6364, Dec. 2017.
- [13] Y. Yao, L. Huang, A. B. Sharma, L. Golubchik, and M. J. Neely, "Power cost reduction in distributed data centers: A two-time-scale approach for delay tolerant workloads," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 1, pp. 200–211, Jan. 2014.
- [14] P. Wang, Y. Zhang, L. Deng, M. Chen, and X. Liu, "Second chance works out better: Saving more for data center operator in open energy market," in *Proc. Annu. Conf. Inf. Syst. (CISS)*, 2016, pp. 378–383.
- [15] J. Camacho, Y. Zhang, M. Chen, and D. M. Chiu, "Balance your bids before your bits: The economics of geographic load-balancing," in *Proc. ACM 5th Int. Conf. Future Energy Syst.*, 2014, pp. 75–85.
- [16] *Data Centers in United States*. Accessed: Aug. 2019. [Online]. Available: <https://cloudscene.com/market/data-centers-in-united-states/washington-dc>
- [17] A. Vafamehr, M. E. Khodayar, and K. Abdelghany, "Oligopolistic competition among cloud providers in electricity and data networks," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1801–1812, Mar. 2019.
- [18] A. Vafamehr, M. E. Khodayar, S. D. Manshadi, I. Ahmad, and J. Lin, "A framework for expansion planning of data centers in electricity and data networks under uncertainty," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 305–316, Jan. 2019.
- [19] H. Wang, J. Huang, X. Lin, and H. Mohsenian-Rad, "Exploring smart grid and data center interactions for electric power load balancing," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 41, no. 3, pp. 89–94, 2014.
- [20] H. Wang, J. Huang, X. Lin, and H. Mohsenian-Rad, "Proactive demand response for data centers: A win-win solution," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1584–1596, May 2016.
- [21] P. Wang, L. Rao, X. Liu, and Y. Qi, "D-pro: Dynamic data center operations with demand-responsive electricity prices in smart grid," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1743–1754, Dec. 2012.
- [22] N. H. Tran, S. Ren, Z. Han, S. Jang, S.-I. Moon, and C. S. Hong, "Demand response of data centers: A real-time pricing game between utilities in smart grid," in *Proc. Feedback Comput.*, 2014.

- [23] S. Bahrami, V. W. S. Wong, and J. Huang, "Data center demand response in deregulated electricity markets," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2820–2832, May 2019.
- [24] Z. Liu, I. Liu, S. Low, and A. Wierman, "Pricing data center demand response," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 42, no. 1, pp. 111–123, 2014.
- [25] T. Li and M. Shahidehpour, "Strategic bidding of transmission-constrained GENCOs with incomplete information," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 437–447, Feb. 2005.
- [26] M. Zugno, J. M. Morales, P. Pinson, and H. Madsen, "Pool strategy of a price-maker wind power producer," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3440–3450, Aug. 2013.
- [27] R. Baldick. (2018). *Course Notes for EE394V Restructured Electricity Markets: Locational Marginal Pricing*. [Online]. Available: <http://users.ece.utexas.edu/~baldick/classes/394V/Localional.pdf>
- [28] D. Deka, S. Backhaus, and M. Chertkov, "Estimating distribution grid topologies: A graphical learning based approach," in *Proc. Power Syst. Comput. Conf. (PSCC)*, 2016, pp. 1–7.
- [29] S. Park, D. Deka, and M. Chertkov, "Exact topology and parameter estimation in distribution grids with minimal observability," in *Proc. Power Syst. Comput. Conf. (PSCC)*, 2018, pp. 1–6.
- [30] D. Deka, S. Backhaus, and M. Chertkov, "Structure learning in power distribution networks," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 1061–1074, Sep. 2018.
- [31] G. Cavraro, R. Arghandeh, K. Poolla, and A. Von Meier, "Data-driven approach for distribution network topology detection," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2015, pp. 1–5.
- [32] A. Sinha, P. Malo, and K. Deb, "A review on bilevel optimization: From classical to evolutionary approaches and applications," *IEEE Trans. Evol. Comput.*, vol. 22, no. 2, pp. 276–295, Apr. 2018.
- [33] S. Dempe, *Foundations of Bilevel Programming*. Dordrecht, The Netherlands: Springer, 2002.
- [34] J. Fortuny-Amat and B. McCarl, "A representation and economic interpretation of a two-level programming problem," *J. Oper. Res. Soc.*, vol. 32, no. 9, pp. 783–792, 1981.
- [35] S. Binato, M. V. F. Pereira, and S. Granville, "A new benders decomposition approach to solve power transmission network design problems," *IEEE Trans. Power Syst.*, vol. 16, no. 2, pp. 235–240, May 2001.
- [36] A. M. Geoffrion, "Generalized benders decomposition," *J. Optim. Theory Appl.*, vol. 10, no. 4, pp. 237–260, 1972.
- [37] *steady State Power Flow Simulation and Optimization for MATLAB and Octave*. Accessed: Aug. 2019. [Online]. Available: <https://github.com/MATPOWER/matpower>
- [38] M. Inuiguchi and P. Saridichainunta, "Bilevel linear programming with ambiguous objective function of the follower," *Fuzzy Optim. Decis. Making*, vol. 15, no. 4, pp. 415–434, 2016.
- [39] T. Chen, A. G. Marques, and G. B. Giannakis, "DGLB: Distributed stochastic geographical load balancing over cloud networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 28, no. 7, pp. 1866–1880, Jul. 2017.



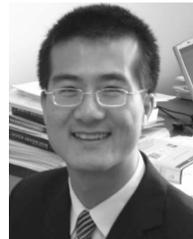
Jun Sun (Student Member, IEEE) received the B.S. degree from the College of Astronautics, Nanjing University of Aeronautics and Astronautics, China, in 2015. He is currently pursuing the Ph.D. degree with the College of Control Science and Engineering, Zhejiang University, where he is a member of the Group of Networked Sensing and Control with the State Key Laboratory of Industrial Control Technology. His research interests include game theory and optimization with applications in electricity market, data centers, and machine learning.



Minghua Chen (Senior Member, IEEE) received the B.Eng. and M.S. degrees from the Department of Electronic Engineering, Tsinghua University in 1999 and 2001, respectively, and the Ph.D. degree from the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley in 2006. In 2019, he joined the School of Data Science, City University of Hong Kong as a Full Professor. He worked with Microsoft Research Redmond and the Chinese University of Hong Kong. His current research interests include online optimization and algorithms, energy systems (e.g., smart power grids and energy-efficient data centers), intelligent transportation systems, distributed optimization, delay-constrained network coding, machine learning in networked and societal systems, and capitalizing the benefit of data-driven prediction in algorithm/system design. He received the Eli Jury Award from UC Berkeley in 2007 (presented to a graduate student or recent alumnus for outstanding achievement in the area of Systems, Communications, Control, or Signal Processing), the Chinese University of Hong Kong Young Researcher Award in 2013, and the ACM Recognition of Service Award in 2017 for the service contribution to the research community. He also received several best paper awards, including the IEEE ICME Best Paper Award in 2009, the IEEE TRANSACTIONS ON MULTIMEDIA Prize Paper Award in 2009, and the ACM Multimedia Best Paper Award in 2012. He also coauthors several papers that are Best Paper Award Runner-up/Finalist/Candidate for conferences, including ACM MobiHoc in 2014 and ACM e-Energy in 2015, 2016, 2018, and 2019. He serves as the TPC Co-Chair, the General Chair, and the Steering Committee Chair for ACM e-Energy in 2016, 2017, and 2018—present, respectively. He also serves as an Associate Editor for the IEEE/ACM TRANSACTIONS ON NETWORKING from 2014 to 2018. He is currently serving as the TPC Co-Chair for ACM MobiHoc 2020.



Haoyang Liu received the B.S. and M.S. degrees from the Department of Structural Engineering, Tongji University, Shanghai, China, in 1997 and 1999, respectively, and the Ph.D. degree from the Department of Civil Engineering, Johns Hopkins University, in 2003. He is the Co-Founder and a CEO with Beijing Noitom Technologies Inc. He has been a Professor with Beijing Sport University since 2017. His current research interests include motion capture, information visualization, and mixed reality.



Qinmin Yang (Member, IEEE) received the bachelor's degree in electrical engineering from the Civil Aviation University of China, Tianjin, China, in 2001, the M.S. degree in control science and engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2004, and the Ph.D. degree in electrical engineering from the University of Missouri–Rolla, Rolla, MO, USA, in 2007, where he was a Postdoctoral Research Associate from 2007 to 2008. From 2008 to 2009, he was a System Engineer with Caterpillar Inc. From 2009 to 2010, he was a Postdoctoral Research Associate with the University of Connecticut. Since 2010, he has been with the State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, China, where he is currently a Professor. He has also held visiting positions with the University of Toronto and Lehigh University. His research interests include intelligent control, renewable energy systems, smart grid, and industrial big data. He has been serving as an Associate Editor for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS and *Automatica Sinica*.



Zaiyue Yang (Member, IEEE) received the B.S. and M.S. degrees from the Department of Automation, University of Science and Technology of China, Hefei, China, in 2001 and 2004, respectively, and the Ph.D. degree from the Department of Mechanical Engineering, University of Hong Kong in 2008. In 2010, he joined the College of Control Science and Engineering, Zhejiang University, Hangzhou, China. He was a Postdoctoral Fellow and a Research Associate with the Department of Applied Mathematics, Hong Kong Polytechnic University. Then, he joined the Department of Mechanical and Energy Engineering, Southern University of Science and Technology, Shenzhen, China, in 2017, where he is currently a Professor. His current research interests include smart grid, signal processing, and control theory. He is an Associate Editor of the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS.