

Joint Bidding and Geographical Load Balancing for Datacenters: Is Uncertainty a Blessing or a Curse?

Ying Zhang, Lei Deng, Minghua Chen, and Peijian Wang

Abstract—We consider the scenario where a cloud service provider (CSP) operates multiple geo-distributed datacenters to provide Internet-scale service. Our objective is to minimize the total electricity and bandwidth cost by jointly optimizing electricity procurement from wholesale markets and geographical load balancing (GLB), *i.e.*, dynamically routing workloads to locations with cheaper electricity. Under the ideal setting where exact values of market prices and workloads are given, this problem reduces to a simple LP and is easy to solve. However, under the realistic setting where only distributions of these variables are available, the problem unfolds into a non-convex infinite-dimensional one and is challenging to solve. One of our main contributions is to develop an algorithm that is proven to solve the challenging problem optimally, by exploring the full design space of strategic bidding. Trace-driven evaluations corroborate our theoretical results, demonstrate fast convergence of our algorithm, and show that it can reduce the cost for the CSP by up to 20% as compared to baseline alternatives. Our study highlights the intriguing role of uncertainty in workloads and market prices, measured by their variances. While uncertainty in workloads deteriorates the cost-saving performance of joint electricity procurement and GLB, counter-intuitively, uncertainty in market prices can be exploited to achieve a cost reduction even larger than the setting without price uncertainty.

Index Terms—Geographic Load Balancing, Deregulated Electricity Market, Data Centers, Bidding Curve

I. INTRODUCTION

As cloud computing services become prevalent, the electricity cost of worldwide datacenters hosting these services has skyrocketed, reaching \$16B in 2010 [22]. Electricity cost represents a large fraction of the datacenter operating expense [43], and it is increasing at an alarming rate of 12% annually [5]. Consequently, reducing electricity cost has become a critical concern for datacenter operators [34].

There have been substantial research on reducing power consumption and the related cost of datacenters [17], [19], [40], [42]. Among them, geographical load balancing (GLB) is a promising technique [34], [35], [39]. By *dynamically* routing workloads to locations with cheaper electricity, GLB

Ying Zhang and Minghua Chen are with Department of Information Engineering, the Chinese University of Hong Kong, Hong Kong. Lei Deng is with School of Electrical Engineering & Intelligentization, Dongguan University of Technology, Guangdong, China. Peijian Wang is with School of Electrical and Information Engineering, Xi'an Jiaotong University, Xi'an, China.

The work presented in this paper was supported in part by the University Grants Committee of the Hong Kong Special Administrative Region, China (Theme-based Research Scheme Project No. T23-407/13-N and Collaborative Research Fund No. C7036-15G). The work of Peijian Wang was supported in part by the National Natural Science Foundation of China (61402358) and the China Postdoctoral Science Foundation (2014M562418, 2015T81033).

Part of this paper was presented at IEEE INFOCOM 17 [49].

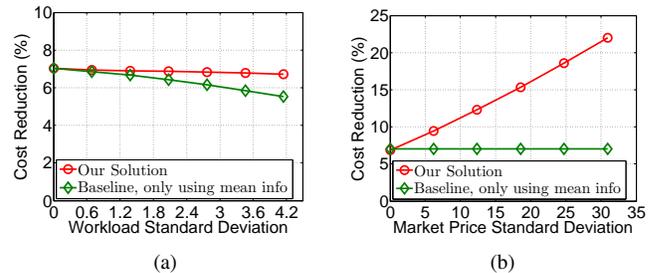


Fig. 1. (a) We fix market prices to their means and increase standard deviations of workloads. Cost reductions of our solution and baseline decrease as the standard deviations increase. (b) We fix workloads to their means and increase standard deviations of prices. Cost reduction of our solution increases as the standard deviations increase, while that of baseline stays constant. More details are in Sec. V and Sec.VII-C.

has been shown to be effective in reducing electricity cost (*e.g.*, by 2–13% [34]) of geo-distributed datacenters operated by a CSP. Many existing works explore price *diversity across geographical locations* to reduce electricity cost [34], [35], [45]. Some recent studies also advocate additional price *diversity across time* at a location by, for example, using electricity storage system and demand response for arbitrage [40] or opportunistically optimizing various electricity procurement options [8], [16], [44].

Inspired by these advances and recent practices that CSPs moving into electricity markets (*e.g.*, Google Energy LLC [18]), we consider the scenario where a CSP jointly performing GLB and electricity procurement from deregulated markets. The market prices are set by running *auction* mechanisms among the electricity suppliers and consumers, cf., [47]. The goal is to minimize the total electricity and bandwidth cost, by exploiting price diversity in both geographical locations (by GLB) and time (by procurement in local sequential markets).

Under the ideal setting where exact values of market prices and workloads are given, the problem reduces to a simple LP and is easy to solve by, for example, the solution in [34]. In practice, however, the actual values of these variables are revealed only at the operating time, and only their distributions are available when procuring electricity by submitting bids to markets (bidding). Under such realistic settings, the problem unfolds into a non-convex infinite-dimensional one. Our focus in this paper is to develop an algorithm to solve the problem optimally.

Our study highlights the intriguing role of uncertainty in workloads and market prices, measured by their variances. On one hand, workload uncertainty undermines the efficiency of balancing supply and demand (proportional to workload) on

electricity markets. As a result, the cost-saving performance of joint bidding and GLB deteriorates as workload uncertainty increases, as illustrated in Fig. 1(a). On the other hand, counter-intuitively, higher uncertainty in market prices allows us to extract larger *coordination* gain in sequential procurement in day-ahead and real-time markets [4], [15], [26]. As shown in Fig. 1(b), capitalizing such gain leads to a cost reduction even *larger* than the setting without price uncertainty.

In our solution, we explore the full design space of strategic bidding to *simultaneously* exploit the price uncertainty and combat the workload uncertainty, so as to maximize the cost saving. We make the following contributions.

▷ We present necessary backgrounds on electricity markets in Sec. II. Then in Sec. III, we formulate the problem of cost minimization by joint bidding and GLB, under the realistic setting where only distributions of market prices and workloads are available. The problem is a non-convex infinite-dimensional one and is in general challenging to solve.

▷ To address the non-convexity challenge, in Sec. IV, we leverage problem structures to characterize a subregion of the feasible set so that (i) it contains the optimal solution, and (ii) the problem over this subregion becomes a convex one. We then solve the reduced problem by a nested-loop solution.

▷ In the inner loop, we fix the GLB decision and optimize bidding strategies for local sequential markets. We derive an easy-to-compute closed-form optimal solution in Sec. IV-B. The optimal bidding strategies not only address the infinite-dimension challenge, but also allow the CSP to simultaneously exploit price uncertainty and combat workload uncertainty.

▷ In the outer loop, we solve the remaining GLB problem given optimal bidding strategies. While the problem is convex and of finite dimension, its objective function does not admit an explicit-form expression. Consequently, its gradient cannot be computed explicitly, and gradient/subgradient-based algorithms cannot be directly applied. In Sec. IV-C, we tackle this issue by adapting a zero-order optimization algorithm, named General Pattern Search (GPS) [24], to solve the problem without knowing the explicit-form expression of the objective function. Finally, we prove that our nested-loop algorithm solves the joint bidding and GLB problem optimally. We discuss the computational complexity and issues related to practical implementation in Sec. IV-D.

▷ We analyze the impact of demand and price uncertainties on the cost-saving performance in Sec. V. Realizing our optimal bidding curve may require CSP to place an infinite number of bids in each deregulated electricity market. In practice, however, market operator may only accept a finite number of bids from the CSP. In Sec. VI, we carefully quantize the optimal bidding curve so that it can be realized by using a finite number of bids. We bound the performance loss due to such quantization.

▷ By evaluations based on real-world traces in Sec. VII, we show that our solution converges fast and reduces the CSP cost by up to 20% as compared to baseline alternatives.

Our study also adds understanding to energy cost management for entities other than datacenters. For example, [33] and [26] considered similar problems for utilities and microgrids, without fully exploring the bidding design space or pursuing

optimal solution. Results of our study thus can help to optimize bidding strategy designs under such settings.

We discuss related works in Sec. VIII and conclude the paper in Sec. IX. Due to space limitation, some proofs are deferred to our supplementary materials.

II. ELECTRICITY MARKET PRELIMINARY

In a region, there are two electricity wholesale markets, *day-ahead* market and *real-time* market, to balance the electricity supply and demand in two timescales. We show the critical operations in Fig. 2 and explain the details in the following.

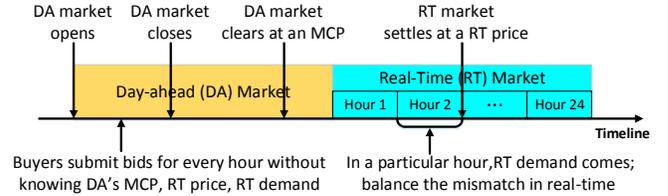


Fig. 2. Operation of day-ahead market and real-time market.

Day-Ahead Market. The day-ahead market is a forward market to trade the electricity one day before dispatching. The electricity supply is auctioned in the day-ahead market. The sellers, *i.e.*, generation companies, submit (hourly) generation offers, and the buyers, *i.e.*, utilities or CSPs, submit (hourly) demand bids, all in the format of $\langle \text{price}, \text{quantity} \rangle$, to the *auctioneer*, *i.e.*, the Independent System Operator (ISO).

In the offers (resp. bids), the generation companies (resp. utilities and CSPs) specify the amount of electricity they want to sell (resp. buy) and at which price. Each seller (resp. buyer) is allowed to submit *multiple* offers (resp. bids) [4] in the same auction with different prices and quantities. The ISO matches the offers with the bids, typically using a well-established double auction mechanism [47]. The outcome of the auction is that it determines a *market clearing price* (MCP) for all the traded units. The bids with prices higher than MCP and the offers with prices lower than MCP will be accepted, and the electricity will be traded at MCP. Upon day-ahead market settlement, the generation companies (resp. utilities and CSPs) will be notified the MCP and quantity of electricity that they commit to generate (resp. consume).

The actual value of MCP is revealed only after the day-ahead market is settled/cleared, and they are unknown to market participants at the time of submitting bids/offers.

We show an example in Fig. 3 from the perspective of our CSP. Suppose that the CSP submits three bids to the day-ahead market: $\langle 30\$/\text{MWh}, 3\text{MWh} \rangle$, $\langle 51\$/\text{MWh}, 4\text{MWh} \rangle$, $\langle 70\$/\text{MWh}, 5\text{MWh} \rangle$. If ISO announces that the MCP is 40\$/MWh after the auction, then the second and the third bid will be accepted since their bidding prices are higher than MCP. Thus the CSP gets $4 + 5 = 9\text{MWh}$ of day-ahead committed supply at the price of MCP, *i.e.*, 40\$/MWh. The day-ahead trading cost is thus $9 \times 40 = 360\$$.

Real-Time Market. The mismatch between day-ahead committed supply (as discussed above) and real-time demand

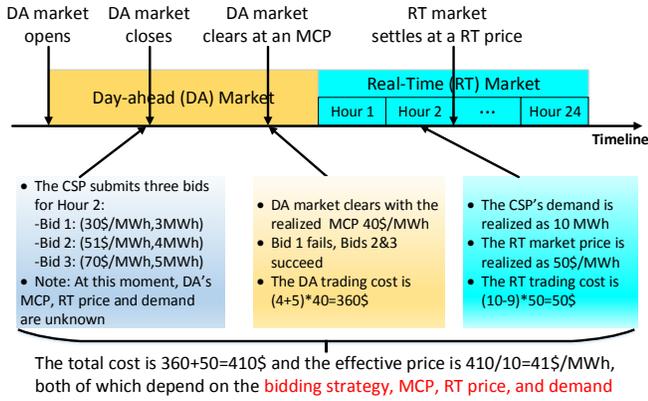


Fig. 3. An illustrating example for the CSP to participate in markets.

is balanced on the real-time market, in a pay-as-you-go fashion. In particular, the system calls the short-start fast-responding generating units, which is usually more expensive, to standby and meet the instantaneous power shortage if any. The real-time price is set after the real-time dispatching and is not exactly known a priori.

- In case that the day-ahead committed supply matches exactly the actual demand, there is no real-time cost.
- In case of under-supply, (*i.e.*, the committed supply is less than the real-time demand), the CSP will pay for extra supply at the real-time price.
- In case of over-supply, the system needs to reduce the power generation output or pay to schedule elastic load [30] to balance the supply, both incurring operational overhead and consequently economic loss. In this case, the CSP will receive a rebate at price $\beta \cdot \text{MCP}$ for the unused electricity (recall that the planned supply is purchased from the day-ahead market at price MCP). Here $\beta \in [0, 1]$ is a discounting factor capturing the overhead-induced cost in handling over-supply situation.

The overall electricity cost for the CSP is the sum of day-ahead procurement cost and the real-time settlement cost, which can be in the form of extra payment or rebate. This pricing model can be viewed as a generalization of the classic Newsvendor problem in operational research [21] (See more discussions in Sec. IV-B), and a real-world market example fitting this model can be found in [20]. We remark that the framework developed in this paper can also be extended to different pricing models described in [15], [30], [33] (see our discussions in [48]).

Back to our example for the CSP in Fig. 3, suppose that the CSP's real-time demand is 10MWh. Since the day-ahead committed supply is only 9MWh, *i.e.*, the under-supply case happens, the CSP needs to buy 1MWh extra electricity from the real-time market. If the real time price is 50\$/MWh, the real-time trading cost of the CSP will be $1 \times 50 = 50\$$. The total cost is the sum of day-ahead trading cost and real-time trading cost, which is $360+50=410\$$.

Cost Structure. An important observation is that the overall cost depends on not only the actual demand, the day-ahead MCP and the real-time price, but also the mismatch between

the day-ahead committed supply and the actual demand. As the day-ahead committed supply depends on day-ahead market bidding strategy of the CSP, the overall cost is thus also a function of the bidding strategy. We remark that such cost structure is unique to electricity procurement in electricity markets and motivates the bidding strategy design [33].

III. SYSTEM MODEL AND PROBLEM FORMULATION

We consider the scenario of a CSP providing computing-intensive services (*e.g.*, Internet search) to users in N regions by operating N geo-distributed datacenters, one in each region, as exemplified in Fig. 4. Service workloads from a region can be served either by the local datacenter or possibly by datacenters in other regions through GLB. The CSP directly participates in wholesale electricity markets in each region, to obtain electricity to serve the local datacenter. Based on (i) distributions of hourly service workloads and (ii) distributions of market settlement prices, the CSP aims at minimizing the expected total operating cost by optimizing GLB and bidding strategies in the markets. The hourly timescale aligns with both the settlement timescale in wholesale markets [39] and the suggested time granularity for performing GLB [34]. Without loss of generality, we focus on minimizing cost of a particular operation hour of the CSP, as shown in Fig. 3.

We summarize the key notations of this paper in Table I.

TABLE I
Key notations.

Notation	Definition
P_j , r.v.	The MCP of day-ahead market j
$f_{P_j}(p)$	PDF of P_j
μ_j^{RT}	Expectation of the price of real-time market j
β	Discounting factor of selling back electricity
z_{ij}	Unit bandwidth cost of routing the workload in region i to the datacenter in region j
U_i , r.v.	User workload in region i with upper bound \bar{u}_i and lower bound \underline{u}_i
$f_{U_i}(u)$	PDF of U_i
V_i , r.v.	Workload to be served by the datacenter in region i
$f_{V_i}(v)$	PDF of V_i
λ_i	Percentage of the workload in region i that has to be served locally
\bar{v}_j	The upper bound of the workload served by datacenter j
C_j	Capacity of the datacenter in region j
α_{ij}	Percentage of the workload in region i served by the datacenter in region j
α	GLB strategy formed by $[\alpha_{ij} : i, j = 1, \dots, N]$
$q_j(p)$	Bidding curve for the day-ahead electricity market in region j

A. Workload and Geographical Load Balancing

Workload and Electricity Demand. We assume that each datacenter is power-proportional, which means that its electricity consumption is proportional to its workload [34]. For example, Google reports that each search requires about 0.28Wh electricity for its datacenters [34]. Without loss of generality, we assume that the workload-to-electricity coefficients are one for all datacenters and thus use the workload served by a datacenter to represent its electricity demand. Our results can

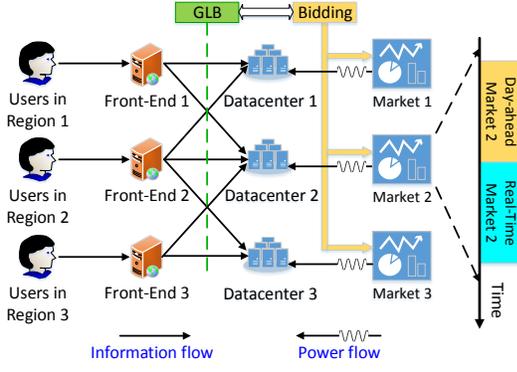


Fig. 4. The scenario that we consider in this paper.

be easily generalized to the case where the coefficients are different for different datacenters.

We model the workload originated from region i as a random variable U_i in the range $[\underline{u}_i, \bar{u}_i]$, with a continuous probability density function (PDF) $f_{U_i}(u)$ that can be empirically estimated from historical data. We assume that all U_i 's are independent.

Geographical Load Balancing.¹ We denote the GLB decision by $\alpha = [\alpha_{ij} : i, j = 1, \dots, N] \in \mathbb{R}_{N \times N}$ which satisfies

$$\sum_j \alpha_{ij} \geq 1, \quad \forall i = 1, \dots, N, \quad (1)$$

$$\alpha_{ii} \geq \lambda_i, \quad \forall i = 1, \dots, N, \quad (2)$$

$$\bar{v}_j \triangleq \sum_{i=1}^N \alpha_{ij} \bar{u}_i \leq C_j, \quad \forall j = 1, \dots, N. \quad (3)$$

$$0 \leq \alpha_{ij} \leq 1, \quad \forall i, j = 1, \dots, N, \quad (4)$$

$$\alpha_{ij} = 0, \quad \forall (i, j) \in \mathcal{G}, \quad (5)$$

where $\mathcal{G} \triangleq \{(i, j) \mid \text{workloads from region } i \text{ cannot be served by datacenter } j \text{ due to quality of service constraints}\}$.

Here α_{ij} represents the fraction of the workload originated from region i that will be routed to datacenter j . Constraints in (1) mean that all workloads must be served. Constraints in (2) capture that λ_i fraction of the workload originated from region i can only be served locally due to various reasons such as delay requirements. Constraints in (3) ensure that the total workload coming into datacenter j can be served even when the workloads in all regions are realized as their maximum values. Constraints in (5) describe that the workload cannot be served by a datacenter if the resulting quality of service, e.g., the delay incurred by serving the workload by a remote datacenter, is not satisfactory. We define the set of all feasible GLB decisions as

$$\mathcal{A} \triangleq \{\alpha \in \mathbb{R}^{N \times N} \mid \alpha \text{ satisfies (1) - (5)}\}. \quad (6)$$

¹Under the conventional setting where datacenters obtain electricity from utilities, GLB is performed in CSP's real-time operation. Under the setting we consider in this paper, CSP needs to bid for electricity in the day-ahead market, where the amount of electricity to bid is a function of GLB decisions. As such, we consider doing joint GLB and electricity bidding in CSP's day-ahead operation, in order to fully explore the new design space enabled by the setting considered in this paper. It is conceivable to perform GLB in both day-ahead and real-time operations of CSP to further minimize the energy cost, which we leave for future study.

Given the GLB decision α , the total workload routed to datacenter j is given by $V_j = \sum_i \alpha_{ij} U_i$. Since $U_i, \forall i$ are random variables, V_j is also a random variable with a continuous PDF

$$f_{V_j}(v) = f_{U_{1j}} \otimes f_{U_{2j}} \otimes \dots \otimes f_{U_{Nj}}(v), \quad (7)$$

where \otimes is the convolution operator and the distribution functions in the convolution are given by

$$f_{U_{ij}}(u) = \begin{cases} \frac{1}{\alpha_{ij}} f_{U_i}\left(\frac{u}{\alpha_{ij}}\right), & \text{if } \alpha_{ij} > 0, \\ \delta(u), & \text{if } \alpha_{ij} = 0, \end{cases} \quad (8)$$

where $\delta(\cdot)$ denotes Dirac delta function.

Bandwidth Cost. Let $z_{ij} \geq 0$ be the unit bandwidth cost from region i to datacenter j . The expected network cost of routing the workload to different datacenters is given by

$$\text{BCost}(\alpha) = \sum_{i=1}^N \sum_{j=1}^N z_{ij} \cdot \alpha_{ij} \cdot \mathbb{E}(U_i). \quad (9)$$

B. Electricity Market Price and Bidding Curve

Day-ahead MCP and Real-time Market Price. At the time of making joint bidding and GLB decisions, MCPs of day-ahead markets in N regions are unknown. We model them as N random variables P_j ($j \in \{1, 2, \dots, N\}$), not necessarily independent to each other. Each P_j has a probability distribution $f_{P_j}(p)$ that can be empirically estimated from historical data [8]. Here we assume that the CSP has negligible market power and its bidding and GLB behavior will not affect the dynamics of electricity markets.²

Similarly, the real-time market prices in N regions are also unknown when making bidding and GLB decisions. We model the price of real-time market j as a random variable P_j^{RT} whose probability distribution can also be empirically estimated from historical data [8]. We define $\mu_j^{\text{RT}} \triangleq \mathbb{E}[P_j^{\text{RT}}]$ as the expectation of P_j^{RT} . We assume that the real-time market prices P_j^{RT} 's are independent to the day-ahead MCPs P_j ($j \in \{1, 2, \dots, N\}$).

Bidding Curve. We explore the full design space of bidding strategy via *bidding curve*, which is a well-accepted concept in the power system community [15], [26]. Bidding curve, denoted as $q_j(p)$, is a function that maps the (realized) day-ahead market MCP to the amount of electricity the CSP wishes to obtain from day-ahead market j , by placing multiple bids. We remark that it is a common practice for one entity (e.g., a utility company) to submit multiple bids to one electricity market.

Bidding curve is useful in designing bidding strategies in the following sense. First, any set of bids can be mapped to a bidding curve. Suppose the CSP submits K bids, namely $\langle b_j^k, q_j^k \rangle, k = 1, \dots, K$, to the day-ahead market of region j , where b_j^k is the bidding price and q_j^k is the bidding quantity of the k -th bid. The corresponding bidding curve is a step-wise decreasing function as

$$q_j(p) = \sum_{k: b_j^k \geq p} q_j^k, \quad \forall p \in \mathbb{R}^+. \quad (11)$$

²The assumption is reasonable as, e.g., datacenters in the US only consume 2% of total electricity [14], and it is usually used in the literature such as [39].

$$\text{ECost}_j(q_j(p), \alpha) = \int_0^{+\infty} f_{P_j}(p) \left[\underbrace{pq_j(p)}_{\text{Day-ahead trading cost}} - \underbrace{\beta p \int_0^{q_j(p)} (q_j(p) - v) f_{V_j}(v) dv}_{\text{Rebate of over-supply}} + \underbrace{\mu_j^{\text{RT}} \int_{q_j(p)}^{\bar{v}_j} (v - q_j(p)) f_{V_j}(v) dv}_{\text{Cost of under-supply}} \right] dp. \quad (10)$$

Real-time trading cost

Expected electricity cost of datacenter j conditioning on day-ahead market j 's MCP $P_j = p$

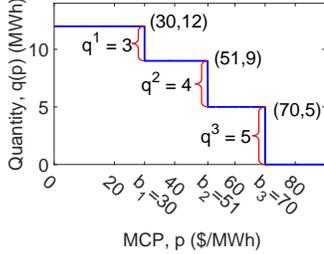


Fig. 5. An illustrating example for the (step-wise) bidding curve constructed from the submitted three bids in Fig. 3.

For example, considering the three bids in Fig. 3, we can construct the corresponding bidding curve as shown in Fig. 5.

Recall that if day-ahead market MCP is p , then all bids whose bidding prices are higher than p will be accepted. Thus, the right hand side of (11) represents the total amount of electricity obtained from the day-ahead market when the MCP is p . Clearly, the purchased amount will be non-increasing in MCP p . Thus, a valid bidding curve $q_j(p)$ must be a non-increasing function.

Second, any non-increasing function is a valid bidding curve and can be realized by placing a set of bids. For example, the bidding curve in (11) can be realized by placing the K bids $\langle b_j^k, q_j^k \rangle, k = 1, \dots, K$ stated above.

Based on the above two observations, we design bidding strategy by choosing a bidding curve from the feasible set

$$\mathcal{Q} \triangleq \{q(p) \mid q(p_1) \leq q(p_2), \forall p_1 \geq p_2, p_1, p_2 \in \mathbb{R}^+\}. \quad (12)$$

Remark. In this paper, we assume that the CSP is allowed to submit any number, possibly infinite number, of bids, or the CSP can directly submit a bidding curve. This assumption allows us to significantly simplify the derivation of optimal solution to the joint bidding and GLB problem in Sec. IV. In Sec. VI, we discuss how to approximately realize a continuous bidding curve with a limited number of bids in the practical implementation. Our simulation results in Sec. VII (Tab. II) suggest that the performance loss due to the approximation error is minor.

Electricity Cost. Given the bidding curve $q_j(p)$ and the GLB decision α , we denote the *expected* electricity procurement cost of the CSP in electricity market j as $\text{ECost}_j(q_j(p), \alpha)$, which consists of settlement in both day-ahead trading and real-time trading.

- In day-ahead trading, suppose that the MCP in the day-ahead market j is p , the committed supply will be $q_j(p)$ and the day-ahead trading cost is $p \cdot q_j(p)$.
- In real-time trading, the day-ahead committed supply $q_j(p)$ may not exactly match the real-time demand V_j .

Suppose that the real-time demand is realized as $V_j = v$. If $v > q_j(p)$, *under-supply* happens and we need to buy $v - q_j(p)$ amount of electricity at expected price μ_j^{RT} . It is not difficult to compute the expected cost, due to under-supply, as $\mu_j^{\text{RT}} \int_{q_j(p)}^{\bar{v}_j} (v - q_j(p)) f_{V_j}(v) dv$. Since $f_{V_j}(v)$ is a continuous PDF, we can replace the lower limit of the integral by $q_j(p)$ without affecting the result of the integral. Similarly, if *over-supply* happens, the unused electricity $(q_j(p) - v)$ will be sold back at a discounted price βp and the expected rebate due to over-supply is $\beta p \int_0^{q_j(p)} (q_j(p) - v) f_{V_j}(v) dv$. The expected real-time trading cost is simply the under-supply cost minus the over-supply rebate.

Based on the above analysis, we obtain the expression of $\text{ECost}_j(q_j(p), \alpha)$ in (10) by applying the total expectation theorem. Note that $\text{ECost}_j(q_j(p), \alpha)$ is related to the GLB decision α through the distribution of V_j (the workload of datacenter j), which is computed by (7) and (8).

C. Problem Formulation

According to our investigations of some real-world pricing schemes, the bandwidth cost can be comparable to the electricity cost [48], so our objective is to minimize the summation of electricity cost and bandwidth cost. We now formulate the problem of joint bidding and GLB:

$$\begin{aligned} \mathbf{P1:} \quad & \min \sum_{j=1}^N \text{ECost}_j(q_j(p), \alpha) + \text{BCost}(\alpha) \\ & \text{var. } \alpha \in \mathcal{A}, q_j(p) \in \mathcal{Q}, j = 1, \dots, N. \end{aligned}$$

where \mathcal{A} is the set of all feasible GLB decisions, defined in (6) and \mathcal{Q} is the set of all feasible bidding curves, defined in (12). It is straightforward to see both \mathcal{A} and \mathcal{Q} are convex sets. The objective is to minimize the summation of electricity cost of N datacenters and network cost, by optimizing bidding strategies and GLB decisions. The consideration of joint bidding and GLB as well as the market and demand uncertainty differentiates our work from existing works, *e.g.*, [8], [34], [35], [43]. We emphasize that it is important to consider input uncertainty to fully capitalize the economic benefit of joint bidding and GLB under real-world market mechanisms.

Challenges. There are two challenges in solving problem **P1**. First, it can be shown that the objective function of **P1** is non-convex with respect to $q_j(p)$ (see our technical report [48]). Second, the optimization variable $q_j(p)$ is a functional variable with infinite dimension. Thus it is highly non-trivial to solve this non-convex infinite-dimensional problem optimally by existing solvers, without incurring forbidden complexity.

IV. AN OPTIMAL JOINT BIDDING AND GLB SOLUTION

In this section, we design an algorithm to solve the challenging problem **P1** optimally.

A. Reducing **P1** to a Convex Problem and Approach Sketch

To begin with, we define a sub-region of \mathcal{Q} as follows

$$\hat{\mathcal{Q}}_j = \{q_j(p) | q_j(p) \in \mathcal{Q}, \text{ and } q_j(p) = 0, \forall p \geq \mu_j^{\text{RT}}\}. \quad (13)$$

As compared to \mathcal{Q} defined in (12), the new constraint in the definition of $\hat{\mathcal{Q}}_j$, i.e., $q_j(p) = 0, \forall p \geq \mu_j^{\text{RT}}$, means that we do not submit any bid to day-ahead market j with bidding price higher than μ_j^{RT} , i.e., the expected price of real-time market j . It is easy to verify that both \mathcal{Q} and $\hat{\mathcal{Q}}_j$ are convex sets.

Theorem 1: The following problem **P2** is *convex* and has the same optimal solution as **P1**:

$$\begin{aligned} \mathbf{P2:} \quad & \min \sum_{j=1}^N \text{ECost}_j(q_j(p), \alpha) + \text{BCost}(\alpha) \\ \text{var.} \quad & \alpha \in \mathcal{A}, q_j(p) \in \hat{\mathcal{Q}}_j, j = 1, \dots, N. \end{aligned}$$

Proof: See Appendix A in our supplementary materials. ■

Remarks. (i) Problems **P1** and **P2** differ only in the feasible set of bidding curve $q_j(p)$. It is \mathcal{Q} in **P1** but $\hat{\mathcal{Q}}_j$ in **P2**. The objective function is nonconvex over \mathcal{Q} but convex over $\hat{\mathcal{Q}}_j$, as shown in the proof of Theorem 1 in Appendix A; hence, **P1** is a nonconvex problem but **P2** now is a convex one. (ii) Intuitively, the optimal bidding curve for day-ahead market j must be in $\hat{\mathcal{Q}}_j$. This is because the CSP can always buy electricity from real-time market j at an expected price μ_j^{RT} ; thus it is not economic to submit bids with bidding price higher than μ_j^{RT} to day-ahead market j . Such bidding strategies must be in set $\hat{\mathcal{Q}}_j$, defined in (13).

Theorem 1 allows us to solve **P1** by solving the convex problem **P2**. However, **P2** still suffers the infinite-dimension challenge, since optimizing bidding curves in general requires us to specify the value of $q_j(p)$ for every $p \in [0, \mu_j^{\text{RT}}]$. To illustrate our design, we first rewrite problem **P2**,

$$\begin{aligned} & \min_{\alpha \in \mathcal{A}} \min_{q_j(p) \in \hat{\mathcal{Q}}_j, \forall j} \left\{ \sum_{j=1}^N \text{ECost}_j(q_j(p), \alpha) + \text{BCost}(\alpha) \right\} \\ = & \min_{\alpha \in \mathcal{A}} \left\{ \underbrace{\sum_{j=1}^N \left[\min_{q_j(p) \in \hat{\mathcal{Q}}_j} \text{ECost}_j(q_j(p), \alpha) \right]}_{\text{Problem EP}_j(\alpha), \text{ solved in Sec. IV-B}} + \text{BCost}(\alpha) \right\} \\ & \underbrace{\hspace{10em}}_{\text{Problem P3, solved in Sec. IV-C}} \end{aligned} \quad (14)$$

The structure of the expression in (14) suggests a nested-loop approach to solve problem **P2**.

- *Inner Loop:* The CSP optimizes its bidding strategies for each regional day-ahead market with given GLB decision α , by solving the following problems:

$$\mathbf{EP}_j(\alpha) : \min_{q_j(p) \in \hat{\mathcal{Q}}_j} \text{ECost}_j(q_j(p), \alpha), \quad j = 1, \dots, N.$$

- *Outer Loop:* After solving the inner-loop problems $\mathbf{EP}_j(\alpha)$ and obtaining the optimal bidding curves, denoted by $q_j^*(p; \alpha), \forall j = 1, \dots, N$, the CSP optimizes the (finite-dimensional) GLB decision α by solving the following problem:

$$\mathbf{P3:} \quad \min_{\alpha \in \mathcal{A}} \sum_{j=1}^N \text{ECost}_j(q_j^*(p; \alpha), \alpha) + \text{BCost}(\alpha).$$

In our technical report [48], we prove that the inner-loop problem $\mathbf{EP}_j(\alpha)$ and outer-loop problem **P3** are both convex, which is perhaps not surprising. In the next two subsections, we solve $\mathbf{EP}_j(\alpha)$ and **P3** to obtain an optimal joint bidding and GLB solution to **P2**, which is also optimal for **P1**.

B. Inner Loop: Optimal Bidding Given GLB Decision

The inner-loop problem $\mathbf{EP}_j(\alpha)$ is concerned about designing optimal bidding strategy for day-ahead market in region j (by choosing $q_j(p) \in \hat{\mathcal{Q}}_j$) with GLB decision α given, in face of demand and price uncertainty.

1) *Comparison with the Newsvendor Problem:* Problem $\mathbf{EP}_j(\alpha)$ can be considered as a generalized version of the Newsvendor problem [21]. In the classic Newsvendor problem, with only the demand distribution in hand, the vendor decides how many newspapers to stock at a fixed wholesale price, knowing that both over-supply (loss due to discounted selling-back price) and under-supply (losing some profit) will lead to economic loss. The objective is to optimize the vendor's expected profit or cost performance, subject to (only) the demand uncertainty.

Problem $\mathbf{EP}_j(\alpha)$ shares the same objective and demand uncertainty model as the Newsvendor problem. Meanwhile, it is built upon an auction based model for the newspaper procurement, in which the wholesale price is no longer fixed. It is determined by the auction on the wholesale market and only its distribution is available when the vendor bids into the market to purchase the newspapers. Naturally, the ideal quantity to purchase should be a function of the realized wholesale price, so as to optimize the vendor's expected profit or loss. This setting hence naturally suggests the vendor to explore the "bidding curve" design, a new design space that is not available in the original Newsvendor problem. In summary, the wholesale price uncertainty and the need of exploring the bidding curve design differentiate Problem $\mathbf{EP}_j(\alpha)$ from the classic Newsvendor problem.

2) *An Optimal Solution:* Let the cumulative distribution function (CDF) of V_j , i.e., the demand of datacenter j , be $F_{V_j}(x) \triangleq \int_0^x f_{V_j}(v)dv$, where $f_{V_j}(v)$ is PDF of V_j given in (7). The following theorem shows that $\mathbf{EP}_j(\alpha)$ admits a closed-form solution $q_j^*(p; \alpha)$, addressing the infinite-dimension challenge.

Theorem 2: Assume that the CDF $F_{V_j}(\cdot)$ is strictly increasing; its inverse exists and is denoted by $F_{V_j}^{-1}(\cdot)$. Given a GLB decision α , the optimal bidding curve for solving $\mathbf{EP}_j(\alpha)$ is given by, for $j = 1, \dots, N$,

$$q_j^*(p; \alpha) = \begin{cases} F_{V_j}^{-1} \left(\frac{\mu_j^{\text{RT}} - p}{\mu_j^{\text{RT}} - \beta p} \right), & \text{if } p \in [0, \mu_j^{\text{RT}}]; \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

Proof: See Appendix B in our supplementary materials. ■

The proof of Theorem 2 is delegated in Appendix B. The optimal bidding curve $q_j^*(p; \alpha)$ is *universal* in that it does not depend on the distribution of day-ahead MCP P_j . This is because $q_j^*(p; \alpha)$ actually minimizes the expected electricity procurement cost for any p . It is easy to extend this result to the general non-decreasing $F_{V_j}(x)$ (See Appendix B in our supplementary materials).

Remarks. The optimal bidding curve $q_j^*(p; \alpha)$ in (15) is a decreasing function in the day-ahead MCP p . This meets our intuition that a good bidding strategy should purchase more electricity when p is low and less electricity when p is high. Similarly, it also meets our intuition that $q_j^*(p; \alpha)$ increases in the sell-back discounting factor β . A larger β means less penalty of over-supply, thus the CSP should bid for more electricity from the day-ahead market to balance the penalty of under-supply. By taking into account both the demand statistics ($F_{V_j}(\cdot)$) and market condition (p and β), $q_j^*(p; \alpha)$ achieves the best balance between over-supply and under-supply to minimize the expected electricity cost, and allows our overall solution to be robust to demand uncertainty, as shown in our case study in Fig. 1(a) as well as simulation results in Sec. V. We provide more insightful discussions in Sec. V.

C. Outer Loop: Optimal GLB with Optimal Bidding Curve as a Function of GLB Decision

After obtaining the optimal bidding strategy $q_j^*(p; \alpha)$ as a function of GLB decision α , we now solve the outer-loop problem **P3** for optimizing GLB. While **P3** is convex and of finite dimension, its objective function does not admit an explicit-form expression since we do not have an explicit expression of the optimal objective value of **EP_j**(α). Thus, gradient-based algorithms cannot be directly applied.

We tackle this issue by adapting a zero-order optimization algorithm, named General Pattern Search (GPS) [24], to solve the out-loop problem without knowing explicit expression of the objective function. *Zero-order optimization algorithms* are widely used to solve optimization problems without directly accessing the derivative information. The GPS algorithm in [24] is a popular zero-order optimization algorithm for solving problems with linear constraints, which is suitable for **P3**.

Our adapted GPS algorithm is an iterative algorithm. In each iteration, the algorithm first creates a set of searching directions, named “patterns”, which *positively spans* the entire feasible set. It then searches the directions one by one in order to find a direction, along which the objective value decreases. And we will update to a better solution if we find one. In each search, the algorithm needs to evaluate the objective value of **EP_j**(α) given a GLB decision α , which can be obtained by plugging the optimal solution $q_j^*(p; \alpha)$ into the objective function of **EP_j**(α). In this manner, our adapted GPS algorithm works like gradient-based algorithms, but without the need to compute gradient/subgradient. We summarize our proposed nested-loop algorithm in Algorithm 1.

In general, GPS algorithm is not guaranteed to converge to the globally optimal solution [24]. In the following theorem,

Algorithm 1 An Algorithm for Solving **P3** Optimally

```

1: initialize  $\alpha^0 \leftarrow \mathbf{I}_{N \times N}$ ,  $t \leftarrow 0$ 
2: while not converge do
3:   current_value  $\leftarrow \mathbf{P3-Obj}(\alpha^t)$ 
4:   Get  $\alpha^{t+1}$  by invoking P3-Obj and comparing with
     current_value at most  $2N^2$  times (see [24, Fig. 3.4])
5:    $t \leftarrow t + 1$ 
6: end while
7:  $\alpha^* \leftarrow \alpha^t$ 
8: Compute  $q_j^*(p; \alpha^*)$  by (15) for all  $j \in \{1, 2, \dots, N\}$ 
9: return  $\alpha^*$ ,  $q_j^*(p; \alpha^*)$  for all  $j \in \{1, 2, \dots, N\}$ 

```

A subroutine to compute the objective value of **P3**

```

10: function P3-OBJ( $\alpha$ )
11:   initialize  $j \leftarrow 1$ , val  $\leftarrow \mathbf{BCost}(\alpha)$  by (9)
12:   while  $j \leq N$  do
13:     Compute  $q_j^*(p; \alpha)$  by (15)
14:     val  $\leftarrow \text{val} + \mathbf{ECost}_j(q_j^*(p; \alpha), \alpha)$  by (10)
15:      $j \leftarrow j + 1$ 
16:   end while
17:   return val
18: end function

```

we prove that our Algorithm 1 actually converges to the optimal solution to the convex problem **P3**, under proper conditions.

Theorem 3: Assume that $f_{U_j}(u)$, $j = 1, \dots, N$, are differentiable and their derivatives are continuous. Algorithm 1 converges to a globally optimal solution to **P3**, which is also an optimal solution to **P1** and **P2**.

Proof: The proof follows the facts that **P3** is convex and GPS algorithm converges to a point satisfying the KKT condition [24]. For full proof, see Appendix C in our supplementary materials. ■

D. Complexity and Practical Considerations

In this section, we discuss the computational complexity and some practical considerations for our solution.

1) *Computational Complexity:* In our model and analysis, we assume that both MCP P_j and the demand U_j are continuous random variables. When applying them to practice, we need to sample a PDF (which is a continuous function) into a probability mass function (which is a discrete sequence). So we assume that we sample both the PDF of P_j , *i.e.*, $f_{P_j}(p)$, and the PDF of U_j , *i.e.*, $f_{U_j}(v)$, into sequences with length m . The value of m depends on both the ranges of MCP and demand and the accuracy we aim to achieve. Based on such sampling, we show the computational complexity of our proposed solution, *i.e.*, Algorithm 1.

Theorem 4: If Algorithm 1 converges in n_{iter} iterations, its time complexity is $O(n_{\text{iter}}((N^3 m \log(Nm) + N^3 m^2)))$.

The proof of Theorem 4 is in Appendix D of our supplementary materials. The complexity is linear with the number of iterations until convergence. However, characterizing the exact convergence rate of GPS algorithm is still an open problem, despite of some recent progress for the convex and unconstrained cases [11], [12]. It remains challenging to

obtain a tight bound for the number of iterations. Instead, we empirically evaluate the convergence rate of Algorithm 1 under a realistic setting in Sec. VII-D. The results show that Algorithm 1 converges within 20 iterations, see e.g., Fig. 9, and it is much faster than an existing algorithm that numerically estimates gradients.

The highest-order parameter for the complexity is N , *i.e.*, the number of datacenters of the CSP. But in reality N is usually small: For example, there are only 10 deregulated electricity markets in US. Thus, Theorem 4 shows that the complexity of our Algorithm 1 is affordable in practice.

2) *Imperfect Probability Distributions.*: In our model and solution, we require perfect probability distributions of day-ahead MCP P_j and the regional demand U_j . However, in practice, learning distributions from historical data inevitably introduces certain estimation error. Thus it is important to evaluate the robustness of our solution to the estimation error. In Sec.V, we empirically show that our solution works pretty well for *imperfect* probability distributions of the demand U_j which only use the first-order (expectation) and second-order (variance) statistic information.

V. IMPACTS OF DEMAND AND PRICE UNCERTAINTY

In this section, we study the impacts of demand and price uncertainties, to better understand the observations in Fig. 1(a) and 1(b). We will use the variance of a random variable to measure its uncertainty.³ Taking normal distribution as an example, the distribution of a random variable with a larger variance will be more “stretched” and it is more likely to take very large or small values.

Unless otherwise specified, our discussions in this section involve a single datacenter.

A. Impact of Demand Uncertainty

Demand uncertainty is one of the main challenges handled by this work and it is interesting to ask how the performance will change with different levels of demand uncertainty. Given any purchased amount of electricity from the day-ahead market, a larger demand uncertainty will increase the possibility of real-time mismatch. As elaborated in Sec. II, both over-supply and under-supply will introduce inefficiency to the market and incur additional cost. Thus, the demand uncertainty is always an unwished curse to increase the electricity cost, even for our carefully designed bidding strategy.

Now, we formalize our statement in Lemma 1.

Lemma 1: Assume that the day-ahead MCP is positive and follows an arbitrary distribution, and that the electricity demand (proportional to workload) follows Truncated Normal, Gamma, or Uniform distribution, with a variance σ_D^2 . The optimal expected electricity cost, achievable by using the strategy in (15), is non-decreasing in σ_D^2 .

The proof of Lemma 1 is in Appendix E. As shown in Lemma 1, the bidding strategy by $q_j^*(p; \alpha)$ in (15) cannot fully

eliminate this curse. However, it can handle the demand variability carefully such that the performance will not deteriorate too much, as illustrated in the empirical studies in Fig. 1(a) and Fig. 11.

B. Impact of Price Uncertainty

The price uncertainty in the day-ahead market is the fundamental reason to motivate the continuous bidding curve design and differentiate $\mathbf{EP}_j(\alpha)$ in this paper from the classic Newsvendor problem [21]. Different from demand uncertainty, uncertainty in MCP of day-ahead market allows the optimal bidding curve $q_j^*(p; \alpha)$ to save cost. In particular, the unique two-sequential-market structure where the real-time market serves as a backup for the day-ahead market allows our bidding strategy $q_j^*(p; \alpha)$ to fully explore the benefit of low MCP values but control the risk of high MCP values. We elaborate as follows. When MCP fluctuates, its value, denoted by p , takes small and large values. When p is small, we can purchase cheap electricity from the day-ahead market and thus enjoy “gain”. When p is large, we have to purchase expensive electricity from the day-ahead market and thus suffer “loss”. However, when $p \geq \mu_j^{\text{RT}}$, our optimal bidding strategy $q_j^*(p; \alpha)$ will not purchase any electricity from the day-ahead market but purchase all electricity from the real-time market at the expected price μ_j^{RT} , bounding the “loss” due to high MCP values. Overall, the gain out-weights the loss and we achieve cost saving by leveraging MCP uncertainty. In fact, the larger the MCP uncertainty, the more significant the saving, as illustrated in our case study in Fig. 1(b).

Now, we make the above intuitive explanations more rigorous in Lemma 2.

Lemma 2: Assume that the electricity demand (proportional to workload) is positive and follows an arbitrary distribution, and that the day-ahead MCP follows Truncated Normal, Gamma, or Uniform distribution, with a variance σ_p^2 . The optimal expected electricity cost, achievable by using the strategy in (15), is non-increasing in σ_p^2 .

The proof for Lemma 2 is in Appendix F. It implies that a larger price uncertainty in the day-ahead market will bring more benefit of the two-stage market structure and decrease the cost expectation.

VI. BIDDING WITH FINITE BIDS

We remark that the previously demonstrated advantages can only be realized when submitting an infinite number of bids or a continuous bidding curve is allowed. If not, its feasibility to solve practical problems can be questioned. In this part, we want to adapt our previous design to tackle the problem when only K bids (b^k, q^k) , $k = 1, \dots, K$ can be submitted.

Recall that the bid (b^k, q^k) succeeds only when the MCP of the day-ahead market is lower than or equal to the bidding price b^k . Implicitly, submitting K bids (b^k, q^k) , $k = 1, \dots, K$ can be viewed as proposing a step-wise bidding curve

$$\bar{q}(p) = \sum_{k: b^k \geq p} q^k.$$

Our task in this section is to optimize $\bar{q}(p)$, *i.e.*, the values of $b^k, q^k, \forall k$, to minimize the electricity cost expectation.

³We remark that Lemma 1 and 2 developed in this section still hold without the assumption of specific distributions but using the “increasing convex ordering” or “variability ordering” defined in [41] to measure the uncertainties. See our discussions in [48].

A. Performance Loss Characterization

Firstly, we characterize the cost difference of two different bidding curves by the following lemma.

Lemma 3: When the day-ahead MCP distribution for electricity market is given as $f_{P_j}(p)$ and we denote the costs of two bidding curves $q^1(p), q^2(p)$ ($q^1(p) = q^2(p) = 0$, for $p \geq \mu_j^{\text{RT}}$) as $\text{ECost}_j(q^1(p)), \text{ECost}_j(q^2(p))$,⁴ respectively, we can have

$$\begin{aligned} & |\text{ECost}_j(q^1(p)) - \text{ECost}_j(q^2(p))|^2 \\ & \leq \mathcal{M} \cdot \int_0^{\mu_j^{\text{RT}}} |q^1(p) - q^2(p)|^2 dp, \end{aligned}$$

where $\mathcal{M} = \int_0^{\mu_j^{\text{RT}}} [f_{P_j}(p)(2\mu_j^{\text{RT}} - \beta p - p)]^2 dp$ is a constant determined by the market condition and irrelevant to the bidding curves.

The proof of Lemma 3 is in Appendix G of our supplementary materials. Essentially Lemma 3 shows that if two bidding curves are close in terms of the distance $\int_0^{\mu_j^{\text{RT}}} |q^1(p) - q^2(p)|^2 dp$, their expected costs are also close, which is quite intuitive.

We denote the optimal bidding curve in (15) and its cost by $q^*(p)$ and C^* . Obviously, C^* serves as a lower bound for $\text{ECost}(\bar{q}(p))$.⁵ By applying Lemma 3, we can have

$$\text{ECost}_j(\bar{q}(p)) - C^* \leq \sqrt{\mathcal{M} \cdot \int_0^{\mu_j^{\text{RT}}} |q^*(p) - \bar{q}(p)|^2 dp}. \quad (16)$$

Remarks. (a) This result guarantees that the performance loss compared with the optimal bidding curve by submitting only K bids is upper bounded. And the upper bound is jointly determined by the market condition (\mathcal{M}) and the bidding strategy design ($\int_0^{\mu_j^{\text{RT}}} |q^*(p) - \bar{q}(p)|^2 dp$). (b) (16) also suggests a guideline for designing a “good” step-wise bidding curve: we seek a $\bar{q}(p)$ with a small value of $\int_0^{\mu_j^{\text{RT}}} |q^*(p) - \bar{q}(p)|^2 dp$. Alternatively speaking, we need to find a stepwise function to approximate the continuous bidding curve.

B. Step-wise Bidding Curve Design

Without loss of generality, we assume that the bidding prices are indexed increasingly with $b^k \leq b^{k+1}$ and $b^0 = 0, b^{K+1} = \mu_j^{\text{RT}}$. Thus when $p \in (b^k, b^{k+1}]$, we have $\bar{q}(p) = \sum_{l=k+1}^K q^l$. To have a good step-wise bidding curve, it is natural to find a $\bar{q}(p)$ to minimize $\int_0^{\mu_j^{\text{RT}}} |q^*(p) - \bar{q}(p)|^2 dp$, i.e., to solve the following problem,

$$\text{FB: } \min \sum_{k=0}^K \int_{b^k}^{b^{k+1}} |q^*(p) - \sum_{l=k+1}^K q^l|^2 dp \quad (17a)$$

$$\text{s.t. } b^k \leq b^{k+1} \quad (17b)$$

$$q^k \geq 0 \quad (17c)$$

$$\text{var. } b^k, q^k, k = 1 \dots, K. \quad (17d)$$

⁴Since we consider only bidding strategy here, the GLB-related parameter α is ignored to simplify the notation.

⁵ C^* can be viewed as the optimal value of the cost minimization problem without the “finite-bid” constraint.

Algorithm 2 A Heuristic Algorithm for Solving FB

Input: Optimal bidding curve $q^*(p)$, number of bids K .

Output: $(b^k, q^k), k = 1, \dots, K$.

1: **initialize** $(b^k, q^k), k = 1, \dots, K$.

2: **while** not converge **do**

3: **for** $k = 1, \dots, K$ **do**

4: Find a value \tilde{b}^k that satisfies

$$2q^*(\tilde{b}^k) - \frac{1}{b^{k+1} - \tilde{b}^k} \int_{\tilde{b}^k}^{b^{k+1}} q^*(p) dp - \frac{1}{\tilde{b}^k - b^{k-1}} \int_{\tilde{b}^k}^{b^{k+1}} q^*(p) dp = 0$$

by binary search.

5: Update $b^k = \tilde{b}^k$ if \tilde{b}^k decreases the objective value of (17a). Otherwise, keep b^k as it is.

6: **end for**

7: **end while**

8: $s^{k+1} = \frac{1}{b^{k+1} - b^k} \int_{b^k}^{b^{k+1}} q^*(p) dp, \forall k$.

9: $q^k = s^k - s^{k+1}, \forall k$

10: **return** $(b^k, q^k), k = 1, \dots, K$.

According to (16), any algorithm to produce a solution with a small objective value of **FB** can produce a set of bids with low cost expectation. But unfortunately, **FB** is nonconvex and the global optimal solution is difficult to obtain. Apart from the off-the-shelf solvers, we provide a heuristic algorithm in Alg. 2 to solve **FB** iteratively. In each iteration, we update one bidding price b^k to decrease the objective value of **FB** while keeping the other bidding prices fixed, the idea of which is motivated by the coordinated gradient descent method [29]. The detailed derivation and analysis of Alg. 2 is relegated in [48] due to the space limitation and we only provide the following proposition on its convergence property.

Proposition 1: The objective value of **FB** is non-increasing in each iteration and thus Alg. 2 will converge.

The correctness of Proposition 1 is guaranteed by the facts that the objective value is non-increasing (Line 5 of Alg. 2) and that the optimal objective value of **FB** is lower bounded by 0. The proof is omitted.

Back to our joint optimization framework to meet the finite-bid constraint, we can firstly ignore it and adopt Alg. 1 to produce the optimal, yet possibly continuous, bidding curves $q_j^*(p; \alpha^*), \forall j$. After that, we use Alg. 2 to produce step-wise bidding curves $\bar{q}_j(p)$ to approximate $q_j^*(p; \alpha^*)$ for different datacenters. Obviously the objective value by $q_j^*(p; \alpha^*)$ is a lower bound for the optimal value of the problem with the finite-bid constraint. Then, according to (16), if $\bar{q}_j(p)$ is close to $q_j^*(p; \alpha^*)$ for all j , the objective value in terms of $\bar{q}_j(p), \forall j$ is also close to the optimal value.

VII. EMPIRICAL EVALUATIONS

In this section, we use trace-driven simulations to evaluate the performance of our proposed solution.

A. Dataset and Settings

Network Settings. We consider a CSP operating 3 datacenters in San Diego, Houston, and New York City. We assume that due to quality of experience consideration, the CSP cannot balance workloads between datacenters in San Diego and New

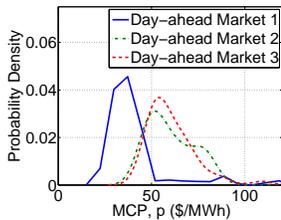


Fig. 6. Empirical distributions of MCPs, 2pm.

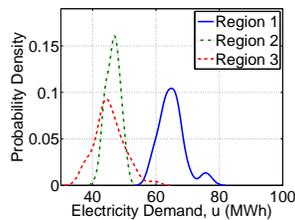


Fig. 7. Empirical distributions of electricity demands, 2pm.

York City. We set the unit bandwidth cost of routing workloads across datacenters as $z_{ij} = \kappa \cdot (\mu_1^{\text{RT}} + \mu_2^{\text{RT}} + \mu_3^{\text{RT}}) / 3$ if $i \neq j$, and $z_{ii} = 0$, $i = 1, 2, 3$. We let $\kappa = 0.1$ as a default setting, and we vary the values of κ to evaluate the overall cost-saving performance under different bandwidth-cost settings.

Workload and Electricity Demand. We get the numbers of service requests per hour against the Akamai CDN in North America for 48 days from Akamai’s Internet Observatory website [1]. By using the conversion ratio claimed by Google for its datacenters [34], we scale up the request information to create an electricity demand series with averaged hourly demand of 125 MWh. The total demand is divided into three regions according to regional electricity consumptions of the three locations [48]. We set the ratio of demand of region i to be served locally, *i.e.*, λ_i , to be 0.7. We also set datacenter j ’s capacity C_j to be 30% larger than region j ’s peak demand, since it was reported that on average 30% or more of the capacity of datacenters is idling in operation [8], [14].

Electricity Prices in Day-ahead and Real-time Markets. We obtain the electricity prices (MCPs of day-ahead market and real-time market prices) from the websites of three regional ISOs which serve the customers in San Diego, Houston, and New York, respectively [7] [13] [32]. The discounting factor β of selling back unused electricity is set as 0.5, which means that the CSP suffers half loss in case of over-supply.

Evaluation and Comparison. We test our design on 24 instances, each corresponding to one hour of the day. For each hour, the distributions of electricity demand, day-ahead MCPs and real-time prices are learned from our dataset, and the real-time price expectation is computed from the distribution accordingly. For illustration purpose, we plot the empirical distributions of MCPs and demands for 2pm in Fig. 6 and Fig. 7, respectively. We denote our solution as OptBidding-OptGLB, in which the GPS part is based on an implementation used in [9], [10]. We test the following four baseline alternatives. (i) NoBidding-NoGLB: it represents the strategy of buying all electricity in real-time markets without doing GLB. It serves as the *benchmark* to compute cost reduction for other algorithms. (ii) OptBidding-NoGLB: it represents the strategy of optimally bidding in day-ahead markets but without doing GLB. (iii) NoBidding-OptGLB: it represents the strategy of doing no bidding in the day-ahead markets but purchasing all electricity in real-time markets and doing optimal GLB (adapted from the solution in [39]). (iv) SimpleBidding-OptGLB: it represents a joint bidding and GLB strategy proposed in [8], in which the CSP only

TABLE II
Cost-saving performance of different schemes.

Solution	Daily Cost (k\$)	Reduction (%)
NoBidding-NoGLB	161.9	-
NoBidding-OptGLB (adapted from [39])	154.5	4.6
SimpleBidding-OptGLB [8]	155.8	3.8
OptBidding-NoGLB	135.4	16.4
OptBidding-OptGLB (Our solution)	128.2	20.8
OptBidding-OptGLB (1 bid)	133.3	17.7
OptBidding-OptGLB (3 bids)	128.6	20.5

submits one bid to each day-ahead market j with bidding price being μ_j^{RT} and the GLB strategy is optimized by a Matlab solver `fmincon`, which implements a gradient-based interior-point algorithm [28]. The gradient of the objective function can be specified as an input or is numerically estimated otherwise [28].

B. Performance Comparison and Impact of Finite Bids

We compare the performance of different solutions in terms of the expected daily cost in Tab. II. Further, we also evaluate the performance loss due to that we approximate the optimal bidding curve (which may require the CSP to submit infinite number of bids) by using only 1 and 3 bids in our solution. We show the cost reduction of using infinite number of bids, 1 bid, and 3 bids in the last three rows of Tab. II, respectively.

We have the following observations. First of all, as seen from Tab. II, we can see that our proposed solution outperforms all the other alternatives and reduces the CSP’s operating cost by 20.8% as compared to the benchmark NoBidding-NoGLB. Meanwhile, we observe that SimpleBidding-OptGLB only reduces the cost by 3.8%, which is much less than that achieved by our solution OptBidding-OptGLB. Moreover, the cost reduction (3.8%) is even less than NoBidding-OptGLB (4.6%), which does not perform bidding in the day-ahead markets but purchases all electricity from the real-time markets. This highlights the importance of designing intelligent strategies for bidding on the day-ahead markets.

In addition to intelligent bidding strategy design, we observe that GLB also brings extra cost saving for CSP. For example, NoBidding-OptGLB reduces the cost by 4.6% as compared to NoBidding-NoGLB, and OptBidding-OptGLB achieves 4.4% extra reduction as compared to OptBidding-NoGLB.

Here, we use the simple method explained in Sec. VI to approximate the optimal bidding curve with a finite number of bids (in particular, 1 and 3 bids in this experiment). From the last two rows in Tab. II, we observe that submitting 1 bid can achieve reasonably good performance (17.7% vs 20.8%). Submitting 3 bids can almost achieve the same performance as submitting infinite number of bids (20.5% vs 20.8%). This observation suggests that our solution performs well in practice even if the CSP is only allowed to submit a small number of bids to a day-ahead market. To understand this observation, we visualize the optimal bidding curves of three datacenters for one optimization instance (4pm) in Fig. 8. We can see that all three bidding curves are “flat” and thus can be

accurately approximated by step-wise functions corresponding to submitting only a small number of bids.

C. Impact of Market Price Uncertainty and Demand Uncertainty

In Sec. I, we provide two experiments related to the electricity demand uncertainty and market price uncertainty (Fig. 1(a) and Fig. 1(b)) to motivate the study in this paper and we describe the details here. Our Solution denotes the strategy OptBidding-OptGLB and **Baseline** denotes a simple strategy: in each region, we pick only one market with cheaper electricity, day-ahead market or real-time market depending on the price expectations, and buy the expected amount of electricity demand in the picked market (If picking the real-time market, we submit no bid in the day-ahead market; if picking the day-ahead market, we submit one bid with the bidding price infinity and the bidding quantity as the expected electricity demand). To understand their individual impact separately, we construct two experiments.

In Fig. 1(a), we set the day-ahead MCP and real-time price to be constant (their sample means), and test the performance of our solution and the baseline with different levels of demand uncertainty (we manipulate the data such that the demand expectations stay the same and their sample STDs increase from 0 to 4.2, where 0 STD represents the scenario without demand uncertainty.). As we can observe, the cost reduction ratio of our solution decreases from 7% to 6.7% while that of the baseline solution decreases from 7% to 5.5%, which fits our analysis in Sec. V-A. It also means that even though the demand uncertainty curses the performance of both two schemes, our solution behaves more robustly. In Fig. 1(b), we set the electricity demand to be constant (its sample mean), and test the performance with different levels of market price uncertainty (similarly, we keep the day-ahead MCP expectation the same and increase its sample STD from 0 to 30.). In this case, the performance of the baseline solution stays the same. This observation is not surprising because the baseline's decision will be the same for any level of market price uncertainty and we also only care about the expected cost. On the other hand, the cost reduction ratio of our solution increases from 7% to 21%. Also, this result should not be surprising based on our analysis in Sec. V-B.

D. Convergence Rate of the Joint Bidding and GLB Algorithm

In this subsection, we empirically evaluate the convergence rate of Algorithm 1. We run our algorithm for two instances with workload/price distribution of 10am and 2pm, respectively. From Fig. 9, we can see that our algorithm converges rather fast, within 20 iterations. The computational complexity of each iteration is polynomial in the problem size (Theorem 4). The main efforts in each iteration are just put to evaluate the objective values by a given set of candidate solutions, and the number of such candidate solutions is less than 18 (2 times the dimensions of α).

We also compare the convergence time of Algorithm 1, which is based on the GPS algorithm, with a Matlab nonlinear optimization solver `fmincon` [28], which was used to solve

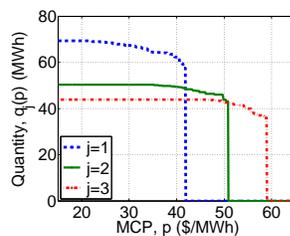


Fig. 8. Optimal bidding curves for three day-ahead markets, 4pm.

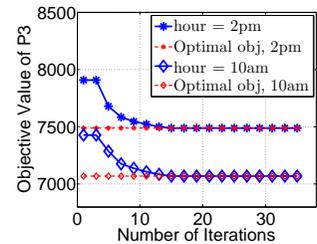


Fig. 9. Objective values in each iteration of our Algorithm 1.

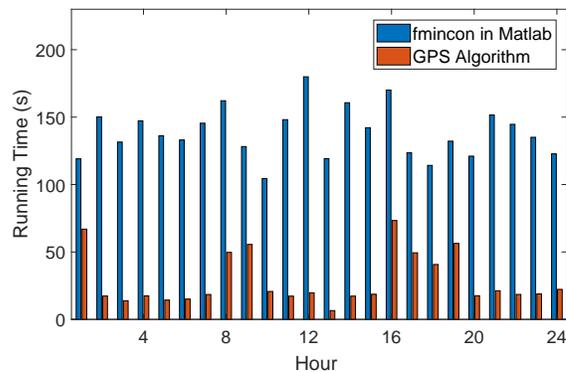


Fig. 10. Running time comparison of our algorithm and a gradient-based solver.

P3. We remark that `fmincon` is a gradient-based solver, and it estimates gradients of the objective function even if the function has no closed form; such gradient estimation can be expensive in time. We numerically solve 24 problem instances, which represent 24 hours of one day, for 100 times by Algorithm 1 (GPS Algorithm) and `fmincon`, respectively. The averaged running times are shown in Fig. 10. The two algorithms produce solutions with the similar objective values. However, as we can observe, Algorithm 1 is 2-9 times faster than `fmincon` for all the 24 cases, which highlights the advantage of the derivative-free design of Algorithm 1.

E. Impact of Inaccurate Demand Distribution Estimation

We properly scale the electricity demand of all three regions such that the demand expectations stay the same and the average of the normalized sample standard deviations among all three regions changes from 0.02 to 0.13, to mimic low to high uncertainty in workloads and hence electricity demand. Here normalized sample standard deviation is defined as the ratio of the sample standard deviation to the sample mean. We apply our solution OptBidding-OptGLB to the set of scaled demands and plot the cost reduction in Fig. 11. From Fig. 11, similar to the observations obtained in Sec. VII-C but using real-world data traces, we can see that the cost reduction decreases as the demand uncertainty increases, but the performance loss is minor, suggesting that our solution OptBidding-OptGLB is robust to demand uncertainty.

The main purpose of the experiments in this subsection is to study the impact of imperfect distribution estimation. In our solution OptBidding-OptGLB, we use the distribution of

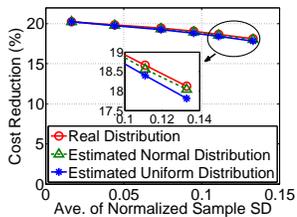


Fig. 11. Cost reductions with different levels of demand uncertainty and different estimated distributions.

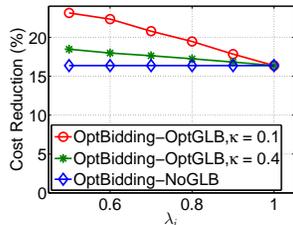


Fig. 12. Cost reductions when more workloads must be locally served, under different bandwidth cost.

the demand U_j for region j as input. In practice, however, the CSP may not have the exact demand distributions, but just their estimations based on historical data. It is common for these estimated distributions to have the same mean and variance as the actual demand distributions, but it is difficult, if not impossible, for the estimated distribution to match the actual distribution exactly. A central question is then how sensitive is the performance of our solution OptBidding-OptGLB to the accuracy of the distribution estimation, given that we have obtained an accurate estimate of the mean and variance?

We explore answers to this question by comparing the performance achieved by our solution OptBidding-OptGLB based on the following distributions for demand with the same mean and variance: actual distribution, *normal distribution*, and *uniform distribution*. We compare their cost reductions in Fig. 11. As seen, the performance loss is minor, implying that accurate first and second order statistics of the demand distribution may be enough to determine the performance of our solution OptBidding-OptGLB. This observation also suggests an interesting direction for future work.

F. Impact of Local Service Requirement and Bandwidth Cost

We investigate the impact of local service requirement, where we changes the percentage of demand that must be served locally, *i.e.*, λ_i , from 0.5 to 1.0. The simulation results are in Fig. 12, where we can see the cost reduction of our solution OptBidding-OptGLB decreases as λ_i increases. This matches our intuition that larger λ_i means that the CSP has less room to do GLB. When $\lambda_i = 1$, *i.e.*, all demand should be served locally, our solution OptBidding-OptGLB coincides with OptBidding-NoGLB.

We also study the impact of bandwidth cost, where we choose two different values (0.1 and 0.4) for the bandwidth cost factor κ . We show the cost reduction in Fig. 12. As seen, a larger κ , meaning higher bandwidth cost, leads to smaller reduction, which matches out intuition.

VIII. RELATED WORK

The seminal works [34], [35] propose the idea of GLB to effectively reduce electricity cost of datacenter operators. Later on many works [25], [27], [39], [45] have broadened the landscape of GLB with more practical considerations and design spaces. Instead of studying the benefit of GLB, several works [8], [43] study the impact of GLB on the electricity supply chain and electricity markets. Several works investigate how

GLB should be operated in the presence of demand uncertainty and/or market price uncertainty. Both [36] and [46] utilize the long-term forward contracts to reduce operation risk. Similar to our work, [8], [16], [17], [44] deal with uncertainty by bidding in the day-ahead market. However, [16], [17] do not fully exploit the bidding design space, and [8], [44] do not consider demand uncertainty. Instead, our work fully exploits the bidding design space and simultaneously considers the demand and market uncertainty. Note that the problem of designing the optimal bids/offers has also been considered in [3], [4], [15], [26], [33] *etc.*, under various pricing models, but for a single-regional-market scenario. In particular, the authors in [3] design the optimal offer strategies for renewable generation company with given day-ahead market prices but uncertain wind generation.

A number of techniques, including stochastic programming, online algorithm design [25], [42], [50], and robust optimization [2], [23], [31], [51], have been applied to study the operation of energy systems and electricity markets, in the presence of the future uncertainty. The online algorithm and robust optimization approaches assume no or only partial distributional information, and are thus appealing when full distributional information is not available. Meanwhile, with distributional information available, the stochastic optimization approach, *e.g.*, the one discussed in this paper, usually can achieve better average-case performance as compared to the online algorithm and robust optimization approaches, as they can be too pessimistic and focus on the worst-case performance rather than the average-case performance.

IX. CONCLUDING REMARKS

We develop an algorithm that is proven to minimize the total electricity and bandwidth cost of a CSP in face of workload and price uncertainties, by jointly optimizing strategic bidding in wholesale markets and GLB. Evaluations based on real-world traces show that our algorithm can reduce the CSP's cost by up to 20%. We show that, interestingly, while uncertainty in workloads deteriorates cost saving, uncertainty in day-ahead market prices allows us to achieve a cost reduction even *larger* than the setting without price uncertainty. Our work has several limitations. First, we assume that the day-ahead market's MCPs and real-time market prices are mutually independent, which may limit the applicability of our results in the scenarios where they are correlated. Second, we assume that the CSP has negligible market power, which may not hold for local markets even though globally datacenters today only consume less than 2% of the total electricity [14], and we do not model possible strategic behaviours of other market participants. Finally, the convergence rate of the GPS-like algorithm in Algorithm 1 is still open. Addressing these limitations is an interesting research direction.

REFERENCES

- [1] "Akamai Internet observatory," available at <https://www.akamai.com>.
- [2] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng, "Adaptive robust optimization for the security constrained unit commitment problem," *IEEE Trans. on Power Syst.*, vol. 28, no. 1, pp. 52–63, 2013.

- [3] E. Y. Bitar, R. Rajagopal, P. P. Khargonekar, K. Poolla, and P. Varaiya, "Bringing wind energy to market," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1225–1235, 2012.
- [4] T. K. Boomsma, N. Juul, and S.-E. Fleten, "Bidding in sequential electricity markets: The Nordic case," *European Journal of Operational Research*, vol. 238, no. 3, pp. 797–809, 2014.
- [5] R. Brown, "Report to congress on server and data center energy efficiency: Public law 109-431," *Lawrence Berkeley National Lab*, 2008.
- [6] C. Burrus and T. W. Parks, *DFT/FFT and Convolution Algorithms: Theory and Implementation*. John Wiley & Sons, Inc., 1991.
- [7] "CAISO archive," available at <http://www.caiso.com>.
- [8] 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 e-Energy*, 2014.
- [9] A. L. Custódio, H. Rocha, and L. N. Vicente, "Incorporating minimum frobenius norm models in direct search," *Computational Optimization and Applications*, vol. 46, no. 2, pp. 265–278, 2010.
- [10] A. L. Custódio and L. N. Vicente, "Using sampling and simplex derivatives in pattern search methods," *SIAM Journal on Optimization*, vol. 18, no. 2, pp. 537–555, 2007.
- [11] M. Dodangeh and L. N. Vicente, "Worst case complexity of direct search under convexity," *Mathematical Programming*, vol. 155, no. 1-2, pp. 307–332, 2016.
- [12] E. D. Dolan, R. M. Lewis, and V. Torczon, "On the local convergence of pattern search," *SIAM Journal on Optimization*, vol. 14, no. 2, pp. 567–583, 2003.
- [13] "ERCOT archive," available at <http://www.ercot.com>.
- [14] "Facts about data centers," available at <http://energy.gov>.
- [15] S.-E. Fleten and E. Pettersen, "Constructing bidding curves for a price-taking retailer in the Norwegian electricity market," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 701–708, 2005.
- [16] M. Ghamkhari, H. Mohsenian-Rad, and A. Wierman, "Optimal risk-aware power procurement for data centers in day-ahead and real-time electricity markets," in *Proc. INFOCOM Workshop on SDP*, 2014.
- [17] M. Ghamkhari, A. Wierman, and H. Mohsenian-Rad, "Energy portfolio optimization of data centers," *IEEE Trans. Smart Grid*, 2016.
- [18] "Google energy wiki," http://en.wikipedia.org/wiki/Google_Energy.
- [19] Y. Guo and Y. Fang, "Electricity cost saving strategy in data centers by using energy storage," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 6, pp. 1149–1160, 2013.
- [20] R. Herranz, A. M. San Roque, J. Villar, and F. A. Campos, "Optimal demand-side bidding strategies in electricity spot markets," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1204–1213, 2012.
- [21] M. Khouja, "The single-period (news-vendor) problem: Literature review and suggestions for future research," *Omega*, vol. 27, no. 5, pp. 537–553, 1999.
- [22] J. Koomey, "Growth in data center electricity use 2005 to 2010," *A report by Analytical Press, The New York Times*, 2011.
- [23] E. Kuznetsova, Y.-F. Li, C. Ruiz, and E. Zio, "An integrated framework of agent-based modelling and robust optimization for microgrid energy management," *Applied Energy*, vol. 129, pp. 70–88, 2014.
- [24] R. M. Lewis and V. Torczon, "Pattern search methods for linearly constrained minimization," *SIAM Journal on Optimization*, vol. 10, no. 3, pp. 917–941, 2000.
- [25] M. Lin, Z. Liu, A. Wierman, and L. L. Andrew, "Online algorithms for geographical load balancing," in *Proc. IGCC*, 2012.
- [26] G. Liu, Y. Xu, and K. Tomsovic, "Bidding strategy for microgrid in day-ahead market based on hybrid stochastic/robust optimization," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 227–237, 2016.
- [27] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. Andrew, "Greening geographical load balancing," in *Proc. ACM SIGMETRICS*, 2011.
- [28] "Find minimum of constrained nonlinear multivariable function," <https://www.mathworks.com/help/optim/ug/fmincon.html>.
- [29] L. Moreau, R. Bachmayer, and N. E. Leonard, "Coordinated gradient descent: A case study of lagrangian dynamics with projected gradient information," *IFAC Proceedings Volumes*, vol. 36, no. 2, pp. 57–62, 2003.
- [30] B. Neupane, T. B. Pedersen, and B. Thiesson, "Evaluating the value of flexibility in energy regulation markets," in *Proc. ACM e-Energy*, 2015.
- [31] S. Nojavan, A. Najafi-Ghalelou, M. Majidi, and K. Zare, "Optimal bidding and offering strategies of merchant compressed air energy storage in deregulated electricity market using robust optimization approach," *Energy*, vol. 142, pp. 250–257, 2018.
- [32] "NYISO archive," available at <http://www.nyiso.com>.
- [33] F. Paganini, P. Belzarena, and P. Monzón, "Decision making in forward power markets with supply and demand uncertainty," in *Proc. CISS*, 2014.
- [34] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs, "Cutting the electric bill for Internet-scale systems," in *Proc. ACM SIGCOMM*, 2009.
- [35] L. Rao, X. Liu, and W. Liu, "Minimizing electricity cost: Optimization of distributed Internet data centers in a multi-electricity-market environment," in *Proc. IEEE INFOCOM*, 2010.
- [36] L. Rao, X. Liu, L. Xie, and Z. Pang, "Hedging against uncertainty: A tale of Internet data center operations under smart grid environment," *IEEE Trans. Smart Grid*, vol. 2, no. 3, pp. 555–563, 2011.
- [37] S. M. Ross, *Stochastic processes*. Wiley, New York, 1996.
- [38] W. Rudin *et al.*, *Principles of mathematical analysis*. McGraw-hill New York, 1964, vol. 3.
- [39] H. Shao, L. Rao, Z. Wang, X. Liu, Z. Wang, and K. Ren, "Optimal load balancing and energy cost management for Internet data centers in deregulated electricity markets," *IEEE Trans. Parallel and Distrib. Syst.*, vol. 25, no. 10, pp. 2659–2669, 2014.
- [40] Y. Shi, B. Xu, B. Zhang, and D. Wang, "Leveraging energy storage to optimize data center electricity cost in emerging power markets," in *Proc. ACM E-Energy*, 2016.
- [41] J.-S. Song, "The effect of leadtime uncertainty in a simple stochastic inventory model," *Management Science*, vol. 40, no. 5, pp. 603–613, 1994.
- [42] Q. Sun, S. Ren, C. Wu, and Z. Li, "An online incentive mechanism for emergency demand response in geo-distributed colocation data centers," in *Proc. ACM e-Energy*, 2016.
- [43] 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, 2012.
- [44] 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. CISS*, 2016.
- [45] R. Wang, N. Kandasamy, C. Nwankpa, and D. R. Kaeli, "Datacenters as controllable load resources in the electricity market," in *Proc. IEEE ICDCS*, 2013.
- [46] L. Yu, T. Jiang, Y. Cao, and Q. Zhang, "Risk-constrained operation for Internet data centers in deregulated electricity markets," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 5, pp. 1306–1316, 2014.
- [47] X.-P. Zhang, *Restructured Electric Power Systems: Analysis of Electricity Markets with Equilibrium Models*. John Wiley & Sons, 2010.
- [48] Y. Zhang, L. Deng, M. Chen, and P. Wang, "Joint bidding and geographical load balancing for datacenters: Is uncertainty a blessing or a curse?" *CUHK, Tech. Rep. 2016*, <https://staff.ie.cuhk.edu.hk/~%7Emhchen/papers/BGLBInfocomTech.pdf>.
- [49] —, "Joint bidding and geographical load balancing for datacenters: Is uncertainty a blessing or a curse?" in *Proc. IEEE INFOCOM*, 2017.
- [50] Y. Zhang, M. H. Hajiesmaili, and M. Chen, "Peak-aware online economic dispatching for microgrids," in *Proc. ACM e-Energy*, 2015.
- [51] M. Zugno and A. J. Conejo, "A robust optimization approach to energy and reserve dispatch in electricity markets," *European Journal of Operational Research*, vol. 247, no. 2, pp. 659–671, 2015.



Ying Zhang received his B.Eng. degree from the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2013. He received his Ph.D. degree from the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, in 2017. His research interests include energy system operation and optimization (e.g., microgrids and energy-efficient data centers), statistical arbitrage and algorithmic trading.



Lei Deng is currently an assistant professor in School of Electrical Engineering & Intelligentization, Dongguan University of Technology. He received the B.Eng. degree from the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2012, and the Ph.D. degree from the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, in 2017. In 2015, he was a Visiting Scholar with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. His

research interests are timely network communications, energy-efficient timely transportation, and spectral-energy efficiency in wireless networks.



Minghua Chen (S04 M06 SM 13) received his B.Eng. and M.S. degrees from the Dept. of Electronic Engineering at Tsinghua University in 1999 and 2001, respectively. He received his Ph.D. degree from the Dept. of Electrical Engineering and Computer Sciences at University of California at Berkeley in 2006. He spent one year visiting Microsoft Research Redmond as a Postdoc Researcher. He joined the Dept. of Information Engineering, the Chinese University of Hong Kong in 2007, where he is currently an Associate Professor. He is also an

Adjunct Associate Professor in Institute of Interdisciplinary Information Sciences, Tsinghua University. 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) and The Chinese University of Hong Kong Young Researcher Award in 2013. 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 is currently an Associate Editor of the IEEE/ACM Transactions on Networking. He serves as TPC Co-Chair of ACM e-Energy 2016 and General Chair of ACM e-Energy 2017. He receives the ACM Recognition of Service Award in 2017 for service contribution to the research community. His current research interests include smart societal systems (e.g., smart power grids, energy-efficient data centers, and intelligent transportation system), networked system, online competitive optimization, distributed optimization, and delay-constrained network coding.



Peijian Wang received the Ph.D. degree from Xi'an Jiaotong University, China, in 2012. He is currently an Assistant Professor with the Department of Computer Science and Technology, Xi'an Jiaotong University. His research interests are in the fields of cloud computing, green computing and big data.

Supplementary Materials

APPENDIX A PROOF OF THEOREM 1

To prove Theorem 1, we firstly provide Proposition 2 and 3 to aid our analysis.

Proposition 2: Given two feasible⁶ demands \tilde{V} and V with $\tilde{V} = \delta V$, where $\delta \in (0, 1)$ is a constant, we have

$$\text{Cost}_j(\delta q(p), f_{\tilde{V}}(v)) = \delta \text{Cost}_j(q(p), f_V(v)), \quad (18)$$

for any $q(p) \in \mathcal{Q}$.

Proposition 3: Given two feasible demands V^1 and V^2 with PDF $f_{V^1}(v)$ and $f_{V^2}(v)$, if $q^1(p), q^2(p) \in \hat{\mathcal{Q}}_j$ and $V^1 + V^2$ is also feasible, we have

$$\begin{aligned} \text{Cost}_j(q^1(p) + q^2(p), f_{V^1+V^2}(v)) &\leq \\ \text{Cost}_j(q^1(p), f_{V^1}(v)) + \text{Cost}_j(q^2(p), f_{V^2}(v)). \end{aligned} \quad (19)$$

A. Proof of Proposition 2

Firstly we can have $f_{\tilde{V}}(\delta v) = \frac{1}{\delta} f_V(v)$ by $\tilde{V} = \delta V$. Then,

$$\begin{aligned} &\int_0^{\delta q(p)} (\delta q(p) - \tilde{v}) f_{\tilde{V}}(\tilde{v}) d\tilde{v} \\ &= \int_0^{q(p)} (\delta q(p) - \delta v) f_{\tilde{V}}(\delta v) d(\delta v), \\ &= \int_0^{q(p)} (\delta q(p) - \delta v) \frac{1}{\delta} f_V(v) d(\delta v), \\ &= \delta \int_0^{q(p)} (q(p) - v) f_V(v) dv. \end{aligned}$$

By similar arguments we have

$$\int_{\delta q(p)}^{\delta \tilde{v}} (\tilde{v} - \delta q(p)) f_{\tilde{V}}(\tilde{v}) d\tilde{v} = \delta \int_{q(p)}^{\tilde{v}} (v - q(p)) f_V(v) dv.$$

According to the cost function (10), we have

$$\begin{aligned} &\text{Cost}_j(\delta q(p), f_{\tilde{V}}(\tilde{v})) \\ &= \int_0^{+\infty} f_{P_j}(p) [p\delta q(p) - \beta p \int_0^{\delta q(p)} (\delta q(p) - \tilde{v}) f_{\tilde{V}}(\tilde{v}) d\tilde{v} \\ &\quad + \mu_j^{\text{RT}} \int_{\delta q(p)}^{\delta \tilde{v}} (\tilde{v} - \delta q(p)) f_{\tilde{V}}(\tilde{v}) d\tilde{v}] dp \\ &= \int_0^{+\infty} f_{P_j}(p) [p\delta q(p) - \delta\beta p \int_0^{q(p)} (q(p) - v) f_V(v) dv \\ &\quad + \mu_j^{\text{RT}} \delta \int_{q(p)}^{\tilde{v}} (v - q(p)) f_V(v) dv] dp \\ &= \delta \text{Cost}_j(q(p), f_V(v)). \end{aligned}$$

The proof is completed.

⁶Recall that demand V is feasible means that the maximum value of V is less than or equal to the datacenter's capacity.

B. Proof of Proposition 3

We rewrite the cost function as

$$\begin{aligned} &\text{Cost}_j(q(p), f_V(v)) \\ &= \mu_j^{\text{RT}} E[V] + \int_0^{+\infty} f_{P_j}(p) \left[(p - \mu_j^{\text{RT}}) q(p) \right] dp \\ &\quad + \int_0^{+\infty} f_{P_j}(p) \left[(\mu_j^{\text{RT}} - \beta p) \int_0^{q(p)} (q(p) - v) f_V(v) dv \right] dx. \end{aligned}$$

Note the first two terms are linear in $f_V(v)$ and $q(p)$ respectively, and $(\mu_j^{\text{RT}} - \beta p) \geq 0$ for all p such that $q^i(p) > 0$. By letting $V = V^1 + V^2$, we only need to prove that

$$\begin{aligned} &\int_0^{q^1(p)+q^2(p)} (q^1(p) + q^2(p) - v) f_V(v) dv \leq \\ &\int_0^{q^1(p)} (q^1(p) - v^1) f_{V^1}(v^1) dv^1 + \int_0^{q^2(p)} (q^2(p) - v^2) f_{V^2}(v^2) dv^2. \end{aligned} \quad (20)$$

(20) can be rewritten as

$$\begin{aligned} &\mathbb{E} [(q^1(p) + q^2(p) - V^1 - V^2)^+] \\ &\leq \mathbb{E} [(q^1(p) - V^1)^+] + \mathbb{E} [(q^2(p) - V^2)^+] \\ &= \mathbb{E} [(q^1(p) - V^1)^+ + (q^2(p) - V^2)^+]. \end{aligned}$$

This inequality is obviously true because for any realization v^1, v^2 we can have

$$(q^1(p) + q^2(p) - v^1 - v^2)^+ \leq (q^1(p) - v^1)^+ + (q^2(p) - v^2)^+.$$

Then we establish the inequality of (20) and the proof for Lemma 3 is completed.

The discussions in Proposition 2 and 3 only involve one datacenter, so we hide the GLB decision α and abuse the notations a little bit to lighten the formula. We will denote $\text{Cost}_j(q(p), f_V(v))$ as the electricity cost of datacenter j when its demand follows $f_V(v)$ and it submits a bidding curve $q(p)$.

Now we are ready to prove Theorem 1 by the following steps,

To prove **P2** is convex, it is enough to show that its objective function is convex over its feasible region. Hence, we only need to show that $\text{ECost}_j(q_j(p), \alpha)$ is convex in $(q_j(p), \alpha)$.

We denote the vector of the electricity demand for each location before GLB by \mathbf{U} . Let α^1, α^2 be two feasible GLB solutions and $\alpha = \delta\alpha^1 + (1-\delta)\alpha^2, \delta \in [0, 1]$ be their convex combination. Then, the electricity demands for datacenter j by $\alpha^1, \alpha^2, \alpha$ can be expressed as $V^1 = (\alpha^1 \mathbf{U})_j, V^2 = (\alpha^2 \mathbf{U})_j, V = (\alpha \mathbf{U})_j = \delta V^1 + (1-\delta)V^2$, respectively. Let the distributions of V^1 and V^2 be $f^1(y)$ and $f^2(y)$, respectively. The distribution for V is then given by $\tilde{f}^1 \odot \tilde{f}^2(y)$, where $\tilde{f}^1(y)$ and $\tilde{f}^2(y)$ are the distributions for δV^1 and $(1-\delta)V^2$. We have

$$\begin{aligned} &\delta \text{ECost}_j(q_j^1(p), \alpha^1) + (1-\delta) \text{ECost}_j(q_j^2(p), \alpha^2) \\ &= \delta \text{Cost}_j(q^1(p), f^1(v)) + (1-\delta) \text{Cost}_j(q^2(p), f^2(v)), \\ &\stackrel{(E_a)}{=} \text{Cost}_j(\delta q_j^1(p), \tilde{f}^1(v)) + \text{Cost}_j(q_j^2(p), \tilde{f}^2(v)), \\ &\stackrel{(E_b)}{\geq} \text{Cost}_j(\delta q^1(p) + (1-\delta)q^2(p), f_{\delta V^1 + (1-\delta)V^2}(v)), \\ &= \text{ECost}_j(\delta q_j^1(p) + (1-\delta)q_j^2(p), \delta\alpha^1 + (1-\delta)\alpha^2). \end{aligned}$$

(E_a) and (E_b) are established by Proposition 2 and Proposition 3, respectively.

Moreover, to prove that **P1** and **P2** have the same optimal solution, we only need to show that, with any α , the optimal bidding curve of datacenter j belongs to $\hat{\mathcal{Q}}_j$, which is true by Theorem 2. The proof is complete.

APPENDIX B PROOF OF THEOREM 2

To solve $\mathbf{EP}_j(\alpha)$, we need to assign a value $q_j(p)$ for each p , to specify how much electricity to buy for any realization of MCP.

The sketch of the proof is as follows: We note that there is a constraint that $q_j(p) \in \hat{\mathcal{Q}}_j$. In the following, we first ignore this constraint and solve the relaxed problem optimally. Then we will show that the optimal solution of the relaxed problem actually satisfies this constraint and thus is optimal to the original problem $\mathbf{EP}_j(\alpha)$. We minimize the objective of unconstrained $\mathbf{EP}_j(\alpha)$ by minimizing the function value inside the integral for each p .

Now, let $c(q) = pq - \beta p \int_0^q (q-v)f_{V_j}(v)dv + \mu_j^{\text{RT}} \int_q^{\bar{v}_j} (v-q)f_{V_j}(v)dv$, we have

$$\begin{aligned} \frac{dc(q)}{dq} &= p - \mu_j^{\text{RT}} \int_q^{\bar{v}_j} f_{V_j}(v)dv - \beta p \int_0^q f_{V_j}(v)dv \\ &= p - \mu_j^{\text{RT}} + (\mu_j^{\text{RT}} - \beta p) \int_0^q f_{V_j}(v)dv. \end{aligned}$$

We discuss the form of the optimal solution as follows,

- If $p \in [0, \mu_j^{\text{RT}}]$, we have $p - \mu_j^{\text{RT}} \leq 0$ and $\mu_j^{\text{RT}} - \beta p \geq 0$. Then $\frac{dc(q)}{dq}$ increases with q . The optimal solution can be obtained by solving $\frac{dc(q)}{dq} = 0$ and the solution is $q_j^*(p) = F_{V_j}^{-1} \left(\frac{\mu_j^{\text{RT}} - p}{\mu_j^{\text{RT}} - \beta p} \right)$.
- If $p \in (\mu_j^{\text{RT}}, \mu_j^{\text{RT}}/\beta)$, we have $p - \mu_j^{\text{RT}} \geq 0$ and $\mu_j^{\text{RT}} - \beta p \geq 0$. Then $\frac{dc(q)}{dq} \geq 0, \forall q$. The optimal solution is $q_j^*(p) = 0$.
- If $p \in [\mu_j^{\text{RT}}/\beta, +\infty]$, we have $\mu_j^{\text{RT}} - \beta p \leq 0$. With the fact that $\int_0^q f_{V_j}(v)dv \leq 1$, we can have $\frac{dc(q)}{dq} \geq p - \mu_j^{\text{RT}} + (\mu_j^{\text{RT}} - \beta p) = (1 - \beta)p \geq 0$. Then the optimal solution is $q_j^*(p) = 0$.

The we can get that the optimal solution to the relaxed problem is

$$q_j^*(p; \alpha) = \begin{cases} F_{V_j}^{-1} \left(\frac{\mu_j^{\text{RT}} - p}{\mu_j^{\text{RT}} - \beta p} \right), & \text{if } p \in [0, \mu_j^{\text{RT}}]; \\ 0, & \text{otherwise.} \end{cases}$$

Note that $\frac{\mu_j^{\text{RT}} - p}{\mu_j^{\text{RT}} - \beta p} \in (0, 1)$ decreases with p and $F_{V_j}^{-1}(\cdot)$ is an increasing function, so $q_j^*(p; \alpha) \in \hat{\mathcal{Q}}_j$. Also, in the processing of obtaining $q_j^*(p; \alpha)$, we do not restrict our attention to $\hat{\mathcal{Q}}_j$, instead we search the entire bidding curve design space \mathcal{Q} , which means that $q_j^*(p; \alpha)$ is also the optimal bidding curve for **P1**. The proof is complete.

Extensions. The above proof can be extended to the case where $F_{V_j}(v)$ is increasing but not strictly increasing. Recall that we want to find a $q_j(p)$ to minimize $pq - \beta p \int_0^q (q-v)f_{V_j}(v)dv + \mu_j^{\text{RT}} \int_q^{\bar{v}_j} (v-q)f_{V_j}(v)dv$, whose derivative is

$p - \mu_j^{\text{RT}} + (\mu_j^{\text{RT}} - \beta p)F_{V_j}(q)$. Note that the derivative of this function is non-decreasing and is negative when $q = 0$ and positive when $q = \bar{v}_j$, where \bar{v}_j is the upper bound of the demand for datacenter j . The optimal solution can be found by investigating the following cases. The first case is discussed in Theorem 2, where there is only a unique solution so that the derivative is zero. This unique solution is the optimal solution. The second case is that there are multiple solutions for the derivative being zero, and any of these solutions is optimal. The last case is where there is no solution for the derivative being zero, i.e., the derivative is not continuous. Then there must exist at least one critical point at which the derivative ‘jumps’ from negative to positive. Any of these critical points is an optimal solution and can be found numerically by binary search.

APPENDIX C PROOF OF THEOREM 3

Again, to prove that **P3** is convex, we only need to prove that $\mathbf{ECost}_j(q_j^*(p; \alpha), \alpha)$ is convex in α .

$$\begin{aligned} & \delta \mathbf{ECost}_j(q_j^*(p; \alpha^1), \alpha^1) + (1 - \delta) \mathbf{ECost}_j(q_j^*(p; \alpha^2), \alpha^2) \\ (E_a) & \mathbf{ECost}_j(\delta q_j^*(p; \alpha^1), \delta \alpha^1) + \mathbf{ECost}_j((1 - \delta)q_j^*(p; \alpha^2), (1 - \delta)\alpha^2), \\ (E_b) & \mathbf{ECost}_j(\delta q_j^*(p; \alpha^1) + (1 - \delta)q_j^*(p; \alpha^2), \delta \alpha^1 + (1 - \delta)\alpha^2), \\ & \geq \mathbf{ECost}_j(q_j^*(p; \delta \alpha^1 + (1 - \delta)\alpha^2), \delta \alpha^1 + (1 - \delta)\alpha^2) \end{aligned}$$

(E_a) is by Proposition 2 and (E_b) is by Proposition 3. The last step is due to the fact that $q_j^*(p; \delta \alpha^1 + (1 - \delta)\alpha^2)$ is the optimal bidding curve when the GLB decision is $\delta \alpha^1 + (1 - \delta)\alpha^2$, so its electricity cost should not be higher than that of $\delta q_j^*(p; \alpha^1) + (1 - \delta)q_j^*(p; \alpha^2)$.

According to Theorem 4.2 of [24], GPS algorithm is guaranteed to converge to a solution satisfying the KKT condition with four hypotheses. And the solution would be an optimal solution if **P3** is convex, which has already been proved above. We examine these hypotheses one by one as follows.

- Hypothesis 0 requires that proper patterns and step lengths are used, which is satisfied by the implementations of GPS algorithm [9], [10].
- Hypothesis 1 requires that the matrix in the constraint is rational, which is automatically satisfied by (1)-(5).
- Denote the objective function of **P3** by $F_3(\alpha)$. Hypothesis 2 requires that, given the initial point α^0 , the set $\{\alpha \in \mathcal{A} | F_3(\alpha) \leq F_3(\alpha^0)\}$ is compact. It can be guaranteed by the continuity of the objective function of **P3** and the compactness of \mathcal{A} .
- Hypothesis 3 requires that the objective function of **P3** is continuously differentiable on an open neighbourhood of $\{\alpha \in \mathcal{A} | F_3(\alpha) \leq F_3(\alpha^0)\}$, which we prove in the following.

To prove that Hypothesis 3 holds for **P3**, we only need to show that $\mathbf{ECost}_j(q_j^*(p; \alpha), \alpha)$ is continuously differentiable with respect to $\alpha_{ij}, \forall i, j$. A sufficient condition is that both $f_{V_j}(v)$ and $q_j^*(p; \alpha)$ are continuously differentiable with respect to α_{ij} .

For $f_{V_j}(v)$, remember that $f_{V_j}(v)$ is a convolution of several functions. We denote $\bar{f}_{U_{ij}}(v)$ be the convolution of $f_{U_{kj}}(u), k \neq i$, and then

$$f_{V_j}(v) = \frac{1}{\alpha_{ij}} f_{U_i} \left(\frac{v}{\alpha_{ij}} \right) \otimes \bar{f}_{U_{ij}}(v).$$

Note that $\bar{f}_{U_{ij}}(v)$ is not related with α_{ij} , and the condition in Theorem 2 provides that $f_{U_i}(u)$ is continuously differentiable, then $f_{V_j}(v)$ is continuously differentiable with respect to α_{ij} .

For $q_j^*(p; \alpha)$, remember that it is derived from the inverse function of $F_{V_j}(v)$ and $F_{V_j}(v)$ is continuously differentiable (since its derivative $f_{V_j}(v)$ is continuously differentiable.). By the Inverse function theorem [38], $q_j^*(p; \alpha)$ is also continuously differentiable.

APPENDIX D PROOF OF THEOREM 4

We first describe the complexity to solve our inner-loop problem $\mathbf{EP}_j(\alpha)$, i.e., compute the optimal datacenter j 's bidding curve $q_j^*(p; \alpha)$ through (15) when the GLB decision is given by α . We need five steps to obtain $q_j^*(p; \alpha)$. (i) We obtain the PDF of U_{ij} , i.e., $f_{U_{ij}}(v)$, for all $i \in \{1, 2, \dots, N\}$. Through (8), we can obtain $f_{U_{ij}}(v)$ in $O(m)$ for each i , and thus get all $f_{U_{ij}}(u)$'s ($\forall i \in \{1, 2, \dots, N\}$) in $O(Nm)$. (ii) We obtain the PDF of datacenter j 's allocated demand V_i , i.e., $f_{V_j}(v)$. We can obtain $f_{V_j}(v)$ through (7) by doing convolution $N - 1$ times in $O(N^2 m \log(Nm))$ [6]. Note that $f_{V_j}(v)$ could take values at Nm different points. (iii) We obtain the CDF of V_j , i.e., $F_{V_j}(v)$. We can iteratively do summation to obtain $F_{V_j}(v)$ in $O(Nm)$. (iv) We obtain the inverse function of the CDF of V_j , i.e., $F_{V_j}^{-1}(v)$. We only need to inverse all Nm points of $F_{V_j}(v)$, which requires $O(Nm)$ complexity. (v) We obtain the optimal bidding curve $q_j^*(p; \alpha)$. Since we have sampled $f_{P_j}(p)$ into a length- m sequence, we only need to get $q_j^*(p; \alpha)$ for at most m different values for p . Thus we can construct $q_j^*(p; \alpha)$ in $O(m)$ steps. Therefore, the total complexity is the sum of (i)-(v), i.e., $O(Nm) + O(N^2 m \log(Nm)) + O(Nm) + O(Nm) + O(Nm) + O(m) = O(N^2 m \log(Nm))$.

We then analyze the computational complexity of the subroutine $\mathbf{P3-OBJ}(\alpha)$, i.e., evaluating the objective value of $\mathbf{P3}$ for any given GLB decision α . Step 13 needs $O(N^2)$ from (9). Steps 15 is the complexity to compute $q_j^*(p; \alpha)$, which requires $O(N^2 m \log(Nm))$. Step 16 is the complexity to compute $\mathbf{ECost}_j(q_j^*(p; \alpha), \alpha)$ by (10). For any $P_j = p$, the day-ahead trading cost part can be computed in $O(1)$; the rebate of over-supply can be computed in $O(Nm)$; the cost of under-supply can be computed in $O(Nm)$; thus the total complexity for given $P_j = p$ is $O(Nm)$. Since we have sampled $f_{P_j}(p)$ into a length- m sequence, the total complexity to compute $\mathbf{ECost}_j(q_j^*(p; \alpha), \alpha)$ will be $O(Nm^2)$. Since $\mathbf{P3-OBJ}(\alpha)$ should do N iterations for all datacenters, the total complexity to evaluate $\mathbf{P3-OBJ}(\alpha)$ is $O(N^2 + N(N^2 m \log(Nm) + Nm^2)) = O(N^3 m \log(Nm) + N^2 m^2)$.

Finally we come to analyze the computational complexity of our global solution, i.e., Algorithm 1. During the **while** loop, each iteration requires at most $(2N + 1)$ invokes for

the subroutine $\mathbf{P3-OBJ}(\alpha)$, and thus incurs $O((2N + 1) \times (N^3 m \log(Nm) + N^2 m^2)) = O((N^4 m \log(Nm) + N^3 m^2))$. Suppose that our Algorithm 1 converges in n_{iter} iterations. Then the computational complexity of our Algorithm 1 is $O(n_{\text{iter}}((N^4 m \log(Nm) + N^3 m^2)))$.

APPENDIX E PROOF OF LEMMA 1

To aid our analysis, we introduce two stochastic orderings called ‘‘increasing convex ordering’’ (\geq_{ic}) and ‘‘variability ordering’’ (\geq_{var}), the definitions of which are presented below.

Definition 1: ([41, Definition 4.1]) For two random variables X and Y , $X \geq_{\text{ic}} Y$ if and only if $\mathbb{E}[f(X)] \geq \mathbb{E}[f(Y)]$ for **all** nondecreasing convex functions f .

Definition 2: ([41, Definition 4.8]) Consider two random variables X and Y with the same mean $\mathbb{E}[X] = \mathbb{E}[Y]$, having distribution functions f and g . Suppose X and Y are either both continuous or discrete. We say X is more variable than Y , denoted as $X \geq_{\text{var}} Y$, if the sign of $f - g$ changes exactly twice with sign sequence $+, -, +$.

A useful relationship between them is presented in Proposition 4.

Proposition 4: ([41, Lemma 4.9]) $X \geq_{\text{var}} Y$ implies that $X \geq_{\text{ic}} Y$.

We consider two electricity demands V_1 and V_2 with the same expectations and V_1 has a larger variance. According to the definition of ‘‘variability ordering’’ and the properties of involved unimodal distributions, $V_1 \geq_{\text{var}} V_2$. We denote C_1 and C_2 as the cost of V_1 and V_2 by the optimal bidding curve in (15). Our purpose is to show that $C_1 \geq C_2$.

Let $C_1(p)$ and $C_2(p)$ be the cost expectation conditioning on that the day-ahead MCP is realized as p , and $C_1 = \int_0^{+\infty} C_1(p) f_{P_1}(p) dp$, $C_2 = \int_0^{+\infty} C_2(p) f_{P_2}(p) dp$. It would be sufficient if we can show that $C_1(p) \geq C_2(p), \forall p$.

Also, note that when the day-ahead MCP is fixed as p , the problem $\mathbf{EP}_j(\alpha)$ will reduce to the classic Newsvendor problem and (15) is the corresponding optimal solution. According to Proposition 4, we have $V_1 \geq_{\text{ic}} V_2$. By the following proposition, we can immediately have $C_1(p) \geq C_2(p), \forall p$.

Proposition 5: [41, Proposition 4.3] For the Newsvendor problem, given two future demands D_1, D_2 , if $D_1 \geq_{\text{ic}} D_2$, $\mathbb{E}[D_1] = \mathbb{E}[D_2]$, then the optimal cost of D_1 is not less than that of D_2 .

The proof is complete.

APPENDIX F PROOF OF LEMMA 2

We first define $C_{\text{opt}}(p)$ as the expected cost *under the optimal bidding strategy* when the day-ahead MCP is realized as p . The total cost expectation by (15) can be expressed as $\mathbb{E}_P[C_{\text{opt}}(p)]$, where the expectation is taken with respect to the distribution of day-ahead MCP. We consider two stochastic day-ahead MCP denoted by P^1 and P^2 with $\mathbb{E}[P^1] = \mathbb{E}[P^2]$ and P^1 having a larger variance. According to the definition of ‘‘variability ordering’’ and the properties of involved unimodal distributions, $P^1 \geq_{\text{var}} P^2$. Our goal is to show that $\mathbb{E}_{P^1}[C_{\text{opt}}(p)] \leq \mathbb{E}_{P^2}[C_{\text{opt}}(p)]$.

Since $P^1 \geq_{\text{var}} P^2$ implies $P^1 \geq_{\text{ic}} P^2$ (by Proposition 4), according to the following lemma, it will be sufficient to show that $C_{\text{opt}}(p)$ is a **concave** function of p . (A more direct result is that $\mathbb{E}_{P^1}[-C_{\text{opt}}(p)] \geq \mathbb{E}_{P^2}[-C_{\text{opt}}(p)]$ if $-C_{\text{opt}}(p)$ is convex.)

Lemma 4: ([37]) If X and Y are nonnegative random variables with $\mathbb{E}[X] = \mathbb{E}[Y]$, then $X \geq_{\text{ic}} Y$ if and only if $\mathbb{E}[f(X)] \geq \mathbb{E}[f(Y)]$ for all convex functions f .

Let $\delta \in (0, 1)$ and $p^0 = \delta p^1 + (1 - \delta)p^2$. We will show that $C_{\text{opt}}(p^0) \geq \delta C_{\text{opt}}(p^1) + (1 - \delta)C_{\text{opt}}(p^2)$.

Recall that $C_{\text{opt}}(p) = pq_j^*(p) - \beta p \int_0^{q_j^*(p)} (q_j^*(p) - v)f_{V_j}(v)dv + \mu_j^{\text{RT}} \int_{q_j^*(p)}^{v_j} (v - q_j^*(p))f_{V_j}(v)dv$. To lighten the formula, we further denote $Q_{\text{over}}(q_j^*(p)) = \int_0^{q_j^*(p)} (q_j^*(p) - v)f_{V_j}(v)dv$ and $Q_{\text{under}}(q_j^*(p)) = \int_{q_j^*(p)}^{v_j} (v - q_j^*(p))f_{V_j}(v)dv$ as the expected over-supply and under-supply, respectively. Then our proof will proceed as follows,

$$\begin{aligned} & C_{\text{opt}}(p^0) \\ &= p^0 q_j^*(p^0) - \beta p^0 Q_{\text{over}}(q_j^*(p^0)) + \mu_j^{\text{RT}} Q_{\text{under}}(q_j^*(p^0)) \\ &\stackrel{(E_a)}{=} \delta \left(p^1 q_j^*(p^0) - \beta p^1 Q_{\text{over}}(q_j^*(p^0)) + \mu_j^{\text{RT}} Q_{\text{under}}(q_j^*(p^0)) \right) + \\ &\quad (1 - \delta) \left(p^2 q_j^*(p^0) - \beta p^2 Q_{\text{over}}(q_j^*(p^0)) + \mu_j^{\text{RT}} Q_{\text{under}}(q_j^*(p^0)) \right) \\ &\stackrel{(E_b)}{\geq} \delta \left(p^1 q_j^*(p^1) - \beta p^1 Q_{\text{over}}(q_j^*(p^1)) + \mu_j^{\text{RT}} Q_{\text{under}}(q_j^*(p^1)) \right) + \\ &\quad (1 - \delta) \left(p^2 q_j^*(p^2) - \beta p^2 Q_{\text{over}}(q_j^*(p^2)) + \mu_j^{\text{RT}} Q_{\text{under}}(q_j^*(p^2)) \right) \\ &= \delta C_{\text{opt}}(p^1) + (1 - \delta)C_{\text{opt}}(p^2). \end{aligned}$$

We get step (E_a) by replacing the p^0 outside $q_j^*(\cdot)$ with $\alpha p^1 + (1 - \alpha)p^2$ and rearranging the terms. (E_b) is due to the fact that $q_j^*(p^1)$ and $q_j^*(p^2)$ are the optimal electricity procurement. (Remember that we obtain $q_j^*(p^1), q_j^*(p^2)$ by minimizing $pq_j^*(p) - \beta p Q_{\text{over}}(q_j^*(p)) + \mu_j^{\text{RT}} Q_{\text{under}}(q_j^*(p))$ for p^1, p^2). The proof is completed.

APPENDIX G PROOF OF LEMMA 3

We firstly reformulate the cost function from (10) to the following one,

$$\begin{aligned} & \text{ECost}_j(q(p)) \\ &= \int_0^{+\infty} f_P(p) \left[(\mu_j^{\text{RT}} - \beta p) \int_0^{q(p)} (q(p) - v)f_V(v)dv - (\mu_j^{\text{RT}} - p)q(p) \right] dp \\ &\quad + \mu_j^{\text{RT}} \mathbb{E}[V] \\ &\stackrel{(E_a)}{=} \int_0^{\mu_j^{\text{RT}}} f_P(p) \left[(\mu_j^{\text{RT}} - \beta p) \int_0^{q(p)} F_V(v)dv - (\mu_j^{\text{RT}} - p)q(p) \right] dp \\ &\quad + \mu_j^{\text{RT}} \mathbb{E}[V]. \end{aligned}$$

(E_a) comes from the facts that $q(p) = 0$ for $p \geq \mu_j^{\text{RT}}$ and

$$\begin{aligned} & \int_0^{q(p)} (q(p) - v)f_V(v)dv = \int_0^{q(p)} (q(p) - v)dF_V(v) \\ &= (q(p) - v)F_V(v)|_0^{q(p)} - \int_0^{q(p)} F_V(v)d(q(p) - v) \\ &= \int_0^{q(p)} F_V(v)dv. \end{aligned}$$

We will prove the inequality in the following steps.

$$\begin{aligned} & |\text{ECost}_j(q^1(p)) - \text{ECost}_j(q^2(p))|^2 \\ &= \left| \int_0^{\mu_j^{\text{RT}}} f_P(p) \left[(\mu_j^{\text{RT}} - \beta p) \int_{q^2(p)}^{q^1(p)} F_V(v)dv - (\mu_j^{\text{RT}} - p)(q^1(p) - q^2(p)) \right] dp \right|^2 \\ &\stackrel{(E_b)}{\leq} \left| \int_0^{\mu_j^{\text{RT}}} f_P(p) \left[(\mu_j^{\text{RT}} - \beta p) \left| \int_{q^2(p)}^{q^1(p)} F_V(v)dv \right| + (\mu_j^{\text{RT}} - p)|q^1(p) - q^2(p)| \right] dp \right|^2 \\ &\stackrel{(E_c)}{\leq} \left| \int_0^{\mu_j^{\text{RT}}} f_P(p) \left[(\mu_j^{\text{RT}} - \beta p)|q^1(p) - q^2(p)| + (\mu_j^{\text{RT}} - p)|q^1(p) - q^2(p)| \right] dp \right|^2 \\ &= \left| \int_0^{\mu_j^{\text{RT}}} f_P(p) \left[(2\mu_j^{\text{RT}} - \beta p - p)|q^1(p) - q^2(p)| \right] dp \right|^2 \\ &\stackrel{(E_d)}{\leq} \int_0^{\mu_j^{\text{RT}}} [f_P(p)(2\mu_j^{\text{RT}} - \beta p - p)]^2 dp \int_0^{\mu_j^{\text{RT}}} |q^1(p) - q^2(p)|^2 dp. \end{aligned}$$

Step (E_b) is obtained by replacing the two terms in the integral by their absolute values, which is similar to $|a + b| \leq ||a| + |b||$. Step (E_c) is due to the fact that $F(v) \leq 1$ and (E_d) is the direct application of Cauchy-Schwarz inequality.