Joint Geographic Load Balancing and Electricity Procurement for Datacenters in Deregulated Electricity Markets

ZHANG, Ying

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy
in
Information Engineering

The Chinese University of Hong Kong
June 2017

Abstract

The flourishing Internet-scale cloud services are revolutionizing the landscape of human activity. The rapid growth of such services has triggered an increasing deployment of massive energy-hungry geo-distributed datacenters worldwide. In this thesis, 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 cost and bandwidth cost by dynamically routing workloads to datacenters with cheaper electricity, *i.e.*, geographic load balancing (GLB).

Most existing studies on GLB assume that the use of GLB has no impact on electricity prices, even though GLB increases local electricity demand variation. In practice, however, electricity retail prices are determined by how supply and demand are dynamically balanced by local electricity utilities. Firstly, in order to understand GLS's economic potential and impact, we carry out a comprehensive study on how GLB interacts with electricity supply chains. In particular, we show that a separate GLB solution, which relies on utility companies for electricity procurement (EP), will make the electricity supply chains less efficient. Then, utility companies have to increase electricity retail prices to ensure certain profit margin. Consequently, CSP doing GLB may end up getting minor cost reduction or even paying higher electricity cost than not doing GLB, as shown in our case study based on real-world traces.

Secondly, motivated by the recent practice of large CSPs moving into electricity markets, we allow CSPs to join the deregulated market directly and propose a joint GLB and EP solution. By considering the real-world market mechanisms and exploring the full design space of strategic bidding,

we formulate a stochastic optimization problem to minimize the total cost expectation. Under the ideal setting where exact values of market prices and workloads are given, this problem reduces to a simple linear programming and is easy to solve. However, under the realistic setting where only distributional information of these variables is available when making decisions, the problem unfolds into a non-convex infinite-dimensional one and is challenging. One of our main contributions is to develop a nested-loop algorithm that is proven to solve the challenging problem optimally. Our study also highlights the intriguing role of uncertainty in demands and prices, measured by their variances. While uncertainty in electricity demands deteriorates the cost-saving performance of joint GLB and EP, counter-intuitively, uncertainty in market prices can be exploited to achieve a cost reduction even *larger* than the setting without price uncertainty.

Finally, our 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.

This thesis demonstrates the necessity and benefit of the joint optimization framework when performing GLB. We believe that our study provides an important guideline for the CSP to cut its electricity bills by taking advantage of its presence in multiple deregulated markets.

摘要

随着信息科技的发展，云服务深入我们生活的方方面面。与此同时，人们在世界范围内不同地区兴建了大规模的数据中心以支持此类云服务。本论文的研究对象即为一个同时运营多个不同地区数据中心的云服务提供商。我们研究一种“区域负载均衡”的方法，其目标是通过合理地将工作量分配到电价较低的数据中心以降低云服务提供商的用电成本和带宽成本。

首先，为了深入理解“区域负载均衡”，我们详细地研究了一种比较简单的方法。在这种简单的方法中，我们仅考虑不同区域供电公司提供的电价高低，而具体的购电环节则由供电公司独立完成。我们发现这种简单的方法会增加每个数据中心用电量的不确定性，此种不确定性会降低不同地区电力供应链的效率，从而会给供电公司带来额外的损失，因此每个地方的电力公司会提高电价以弥补此损失。由此可见，区域负载均衡的优势很难由此简单的方法充分实现。

然后，我们允许云服务提供商直接参与到开放的电力市场中购电，并把区域负载均衡问题与购电问题结合起来，提出了一个联合优化的框架。但是考虑到现实中的电力市场结构，我们在日前电力市场中做决策时并没有电价和用电量的确切信息，而只有统计信息，我们面临的是一个非凸和无线维的随机优化问题。本论文设计了一个具有嵌套结构的算法去迭代地解决此优化问题，并且我们证明我们的算法在一定条件下可以收敛到全局最优解。与此同时，我们的研究也彰显了一个关于“不确定性”在电力市场中的有趣现象：*尽管用电量的不确定性会增加我们的用电成本，我们的联合优化框架可以利用日前电力市场中电价的不确定性进一步降低用电成本.*

最后，我们使用现实中的用电量数据和电力市场数据验证我们的理论结果。实验表明，我们的算法可以非常快速地收敛并且为云服务提供商

降低接近20%的用电成本。

本论文验证了联合优化方案在区域负载均衡中的必要性和有效性。我么相信我们的研究成果会增进人们对不同区域电力供应链和电力市场的理解，并且为云服务提供商（这种同时存在于多个区域电力市场的特殊电力用户）的降低用电成本的问题提供重要参考。

Acknowledgement

First of all, I would like to thank my advisor Minghua Chen for his supervision. From the beginning when I knew little about this filed till the end of my PhD study to complete this thesis, he guided me and helped me at each step. From him, I learned how to select interesting topics, how to come up with a model and formulate research problems, how to evaluate the results and possibly gain further improvement, and finally how to write papers to present the findings professionally. Although Professor Minghua Chen has a very busy schedule, he does spend a lot of time and effort to make sure that I am on the track to finish my PhD study on time. His substantial support, attitude on research and enthusiasm towards life have influenced me a lot and will continue to benefit my future career for sure.

Besides my advisor, I would like to express my gratitude to my research collaborators: Lei Deng, Mohammad H. Hajiesmaili, Jose Camacho, Professor Peijian Wang, Professor Dahming Chiu, and Professor Qi Zhu. Their broad knowledge and deep insights significantly improved my research results and it is really a pleasure to work with them.

My friends and labmates at CUHK make my PhD life a really enjoyable journey. They are Xin Tao, Shaoquan Zhang, Lei Deng, Hanling Yi, Jincheng Zhang, Hanxu Hou, Yang Yang, Qiulin Li, Lin Yang and many others whose names are too long to list. I enjoy the numerous time of playing and studying with them and have been inspired by their hard-working spirits. They make CUHK a home away from home.

Last but not least, I am profoundly grateful to my parents for making me who I am and enabling me to do what I am doing, and to Jenny for her accompany in the past several years. Although they could not be

here with me or fully understand my research, they always cheer for every achievement during my PhD process and pay great attention to every little aspect of my life. Their warm encouragement and endless love help me to go through a lot of difficulties and keep me strong forever. This thesis is dedicated to them.

To My Parents

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivations

As cloud computing services become prevalent, the electricity cost of worldwide datacenters hosting these services has skyrocketed, reaching $16B in 2010 [37]. Electricity cost represents a large fraction of the datacenter operating expense [78], and it is increasing at an alarming rate of 12% annually [14]. Consequently, reducing electricity cost has become a critical concern for datacenter operators [60, 2].

There have been substantial research on reducing power consumption and related cost of datacenters [72, 69, 29, 32]. Among them, geographical load balancing (GLB) is a promising technique [60, 61, 67]. By *dynamically* routing workloads to locations with cheaper electricity, GLB has been shown to be effective in reducing electricity cost (*e.g.*, by 2–13% [60]) of geo-distributed datacenters operated by a cloud service providers (CSP). Many existing works explore price *diversity across geographical locations* to reduce electricity cost [60, 61, 81]. 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 [69] or opportunistically optimizing various electricity procurement options [17, 28, 80].

Nevertheless, most existing works related GLB focus on addressing technical feasibility and revealing the abundant benefits of GLB, assuming the electricity prices are not affected by GLB. In practice, however, the electricity prices are determined by how supply and demand are dynamically balanced by local utilities, and thus may as well be affected by GLB. In particular, the fact that the electricity is a non-storable commodity forces the utility to predict the demand and schedule its supply in advance. Since GLB increases demand variation, it may incur extra errors in demand prediction. As we will show in Chapter 4, prediction errors will lead to over-/under- supply and consequently economic loss for utilities and utilities may have to increase electricity retail prices to ensure certain profit margin in face of such extra economic loss caused by GLB.

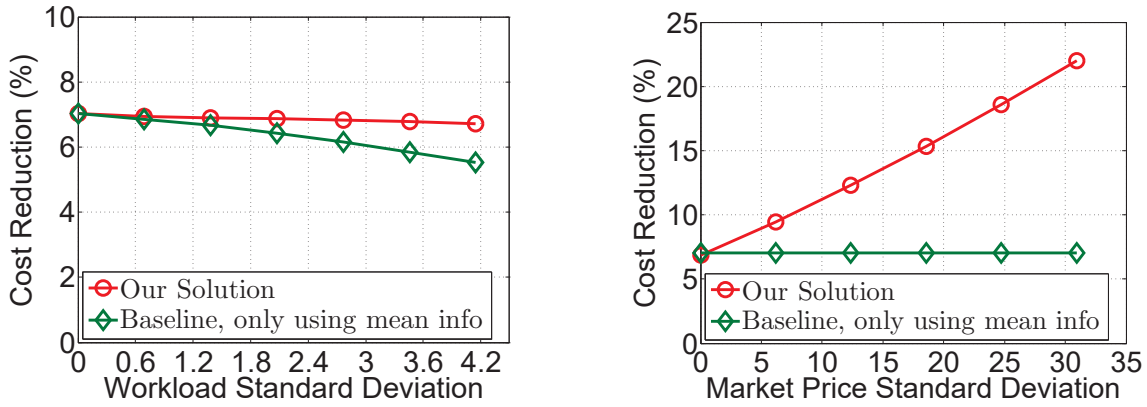
As one of the contributions in this thesis, we note that *GLB can cause non-negligible demand variation for a utility.* For example, Facebook, Apple, Google and Amazon have built or will build large datacenters in Prineville (Oregon, US) to leverage the chilly outdoor air for datacenter cooling at low cost. A fully-operated datacenter (*e.g.*, Google's datacenter in Oregon) is estimated to consume 90 MW power [7]. Power Pacific, a large utility serving Oregon including Prineville, sells 35 GWh daily [1]. Hence, these datacenters once all in full operation could consume 8.6 GWh daily or 22% of Power Pacific sales today, and 33% in 4 years if we aggressively consider datacenter electricity demand grows 15% annually as estimated in [38] while conventional demand remains steady. If datacenters can shift 30% electricity demand away by doing GLB according to the

estimate in [60], then GLB could lead to 10% demand variation for Power Pacific. Therefore, in order to understand and unleash GLB's economic potential, it is critical to understand the interaction between the GLB ability to alter electricity demand patterns, and the impact of its uncertainty on the electricity prices.

Motivated by the above observations, we develop relevant models and carry out a comprehensive study of the impact of GLB on the electricity supply chain. Particularly, we show that as the simple-designed GLB introduces extra local demand uncertainty, which will force utility companies to increase electricity retail prices to ensure certain profit margin. Consequently, CSP doing GLB may end up getting minor cost reduction or even paying *higher* electricity bills than not doing GLB, as shown in our case study based on real-world traces.

Inspired by recent practices that CSPs moving into electricity markets, we consider the scenario where a CSP jointly performing GLB and electricity procurement from deregulated markets. In this new model, the market prices are set by running *auction* mechanisms among the electricity suppliers and consumers, cf, [89]. 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, where exact values of market prices and workloads are given, the optimization problem reduces to a simple linear programming (LP) and is easy to solve, by for an example solution in [60]. In practice, however, the actual values of these variables are re-



(a) (b)

Figure 1.1: (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 Chapter 7 and Chapter 10.2.2.

vealed only at the operating time, and only their distributions are available when procuring electricity by submitting bids to markets (bidding). Such obstacles make it challenging to exploit the benefit of GLB under realistic settings. We show that, to fully exploit the design space, we need to solve a non-convex stochastic optimization problem with infinite dimensions. One of our contributions in this thesis is to develop an algorithm to solve the problem optimally.

The results of our study highlights the intriguing role of uncertainty in the deregulated electricity with a sequential structure. On one hand, workload uncertainty undermines the efficiency of balancing supply and

demand (proportional to workload) on electricity markets. [1] As a result, the cost-saving performance of joint bidding and GLB deteriorates as workload uncertainty increases, as illustrated in Fig. 1.1(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 [46, 26, 12]. As shown in Fig. 1.1(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.

1.2 Thesis Contributions and Organization

The organization and main contributions of this thesis are summarized as follows.

▷ We discuss some related works in Chapter 2 and provide some necessary preliminaries about electricity supply chain and deregulated electricity market in Chapter 3 to bring the readers to the same page. We introduce CSPs doing GLB as a *new* type of customers – they can make their local demand more *elastic* to prices by "shifting" electricity demand among geo-locations. They are very different from conventional electricity customers whose demands are localized and inelastic.

▷ By analysis and case study using real-world traces, we investigate

[1] In this thesis, we assume that the datacenters are power-proportional [45] and we will use electricity demands interchangeably with workloads.

the impact of GLB on the supply chain and its economic consequence in Chapter 4. We show that electricity utilities rely on accurate demand prediction to efficiently balance supply and demand. As GLB will incorporate the price and demand information of remote areas into local demand and make accurate demand prediction harder, it causes trading inefficiency between utilities and CSPs and subsequent economic loss to the utilities. In face of such economic loss, utilities will have to increase retail prices to ensure certain profit margin. Consequently, CSPs doing GLB may end up getting poor cost reduction or even paying higher electricity bills than not doing GLB – 1% higher in our case study.

▷ Then in Chapter. 5, 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 Chapter 6, 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 Chapter 6.2. 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 Chapter 6.3, we tackle this issue by adapting a zero-order optimization algorithm, named General Pattern Search (GPS) [43], 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 Chapter 6.4

▷ We analyze the impact of demand and price uncertainties on the cost-saving performance in Chapter 7. 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 Chapter 8, we carefully quantize the optimal bidding curve so that it can be realized by using a finite number of bids. We also bound the performance loss due to such quantization.

▷ We also discuss how to extend our joint optimization framework when other market pricing models are used to handle the real-time mismatch in Chapter 9.

▷ By evaluations based on real-world traces in Chapter 10, we show that our solution converges fast and achieve satisfactory performance. In particular, the joint optimization approach reduces the CSP cost by up to 20% as compared to baseline alternatives. We test the performance under

different system parameter settings and show that the merit of our design is still remarkable when the distributional information is inexact, or only a finite number of bids can be submitted.

Our study also adds understanding to electricity cost management for entities other than datacenters. For example, [55] and [46] considered similar problems for utilities and microgrids, without fully exploring the bidding design space or pursing optimal solution. Results of our study thus can help to optimize the bidding strategy design under such settings. Part of the results in this thesis have been published in [17, 90] and submitted for journal publication in [91].

□ **End of chapter.**

Chapter 2

Related Work

As energy consumed by datacenters keeps increasing dramatically, reducing power and related cost for IDCs is becoming a very important research topic. A large number of research works can be found in a recent survey [10], and the references therein. This chapter only discusses the most relevant work to this thesis.

Benefits of GLB: The seminal works [60, 61] propose the idea of GLB to effectively reduce electricity cost of datacenter operators. Later on many works [67, 81, 44, 48] have broadened the landscape of GLB with more practical considerations and design spaces. Besides economic benefit, Some other works [48, 42, 71] highlight that GLB can also be applied to efficiently utilize renewable energy with environmental considerations.

GLB with Demand Response: It has been shown promising for the datacenters to participate into demand response (DR) programs in different manners. See [49, 50, 79, 84, 19] and the references therein. Researchers also propose to combine GLB with DR to realize mutual interest of datacenters and electricity providers, in the scenario of regulated or deregulated markets. Particularly, in [77], the authors show that datacenters can help the smart grid operator to balance the load ratio in different locations to make the system more reliable; in [73], the authors use game theory to

study the interactions between the datacenters and different utility companies (monopoly providers), which are modelled as independent players without sharing information; in [48], the authors show that, by properly setting the pricing signals, we can encourage the datacenters to use more renewable energy and reduce the carbon footprint; in [29], the authors show that datacenter can gain economic profits by offering ancillary services to the deregulated market operator.

Impact of Conventional GLB on Electricity Supply Chain: Regarding this optimal procurement, CSPs are completely new players in the electricity markets. Recently, the impacts of geo-distributed datacenters on electricity prices have been studied in [79, 49], in the context of demand response. In particular, [49] analyzed the pricing model only for one datacenter while we consider multiple datacenters instead. [79] showed that the electricity price will be changed when GLB reroute enough amount of workloads so that the energy consumption of individual location is significantly changed, while in this thesis, we show a subtle observation, that the electricity price for CSPs can be increased by its larger demand uncertainty, which is more common in toady's practice. Different from conventional utilities, the CSP is able to bid in different regional markets, and this scenario provides new study cases for the existing literature on strategic bidding [66, 33, 47].

GLB with Market and Demand Uncertainty: Several works [62, 88, 87, 28, 29, 80] study the GLB strategies in the presence of demand uncertainty and/or electricity price uncertainty. Both [62] and [88] utilize

the long-term forward contracts to reduce operation risk. In contrast, our work considers the bidding-based procurement in day-ahead markets. Aligned with this direction, [28] and [29] treat the CSP as a price taker and only optimize the bidding quantity and [17] only considers one bid, which does not fully exploit the design space of bidding strategies. Camacho *et al.* in [17] and Wang *et al.* in [80] fully exploit the design space but they only consider the market uncertainty and do not consider demand uncertainty. Instead, this thesis fully exploits the bidding design space and simultaneously considers the demand and market uncertainty.

Electricity Trading in One Regional Market: Several papers [11, 46, 33, 26, 12] and [55] consider the electricity procurement strategy of the electricity consumer in one electricity market, which is a subproblem considered in this thesis. [55] only optimizes the procurement quantity in the day-ahead market and does not exploit the full design space of bidding strategy. [26] considers a linear-wise bidding curve with the bidding prices at the critical points given and model the future demand as a function of the MCP. [46] and [33] try to optimize the bidding curve but their solutions rely on existing solvers or genetic algorithms, and thus have no optimality guarantee. The authors in [11] design the optimal *offer* strategies for renewable generation company with given day-ahead market prices but uncertain power output.

A brief summarized comparison is provided in Table 2.1.

□ **End of chapter.**

Table 2.1: Summary and comparison of related works and this thesis. N/A: the papers do not consider day-ahead market. ✓✓: the solutions are optimal.

Reference	Day-ahead market uncertainty	Demand uncertainty	Full bidding design space	GLB	Mismatch cost
Ghamkhari *et al.*[28]	✗	✗	✗	✗	✗
Ghamkhari *et al.*[29]	✗	✗	✗	✓	✗
Rao *et al.*[62]	N/A	✓	✗	✓	✗
Liu *et al.*[46]	✓	✓	✓	✗	✓
Paganini *et al.* [55]	✓	✓	✓	✗	✓
Yu *et al.*[88]	N/A	✓	✗	✓	✓
Herranz *et al.*[33]	✓	✗	✓	✗	✓
Bitar *et al.*[11]	✓	✓	✓✓	✗	✓
Wang *et al.*[80]	✓	✗	✓✓	✓	✓
This thesis	✓	✓	✓✓	✓	✓

Chapter 3

Background

In this chapter, we provide some necessary preliminaries on electricity supply chain and deregulated market.

3.1 The Electricity Supply Chain

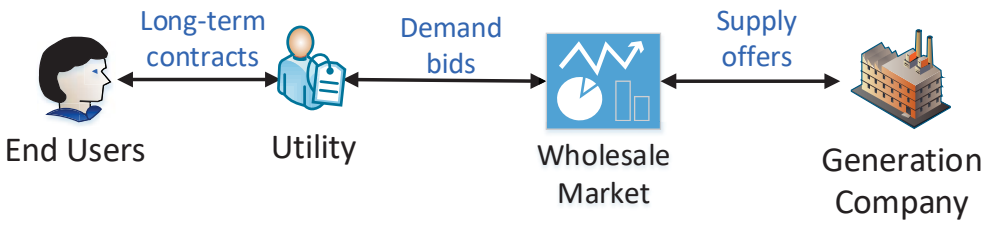


Figure 3.1: An Overview of the Electricity Supply Chain.

Firstly, we provide a high-level introduction of the electricity supply chain, which is in Fig. 3.1. Under the electricity market deregulation, electricity supply chains consist of four components: Generating Companies (GENCOs), Electricity Wholesale Market (Market), Utility Companies (Utilities), Customers (in particular, Cloud Services Providers (CSPs) that owns multiple geo-distributed datacenters). Different from regulated markets, the utilities act as Provider of Last Resort and are not responsible for electricity generation and transmission anymore [58]. Electricity

trading happens between Utilities and GENCOs, by strategic behaviors of two parties, and also between Utilities and Customers, often by long-term contracts (for example, the time-of-use pricing scheme).

3.1.1 Components of Supply Chain

We firstly describe the roles of the three parties in details.

GENCO. GENCOs run the generating units and sell electricity on the wholesale Market. Utilities buy from the Market and sell retail to CSPs. From its generation to its consumption in the data-centers, electric energy flows the entire supply chain. The trading at each step of the chain jointly determines the final prices offered to the customers. Consequently, changes on one side of the chain may propagate to the other extreme. One well-known example is the extremely high prices experienced by customers in 2001 due to inefficiencies in the spot markets in California [35]. For our study, it suffices to consider three components in the supply chain: Market, Utilities, and CSPs.

Utilities. Similar to the retailers in a generic supply chain, utilities buy commodity – electricity – from spot markets and sell to end customers (like CSPs). Utilities make profit by setting a proper retail price, which may be different from MCPs.

Meanwhile, utilities are unique retailers in two senses:

- utilities are trading a non-storable commodity (electricity) with extremely short "expiration time";

- utilities have to schedule electricity supply one day before the demand arrives, by bidding in the day-ahead market.

These two facts incentive the utilities to *predict* precisely both the demand quantity and time-of-arrival, so as to *schedule* the right amount of supply to serve the demands at the right time. For example, a utility that predicts a datacenter needs 30MWh electricity tomorrow at 2-3pm needs to buy today, from the day-ahead market, the exact amount of electricity for its dispatch tomorrow 2-3pm. If there are errors in the prediction, utilities will suffer from over-/under- supply. Over-/under- supply leads to either unused electricity or unmatched demand (to be compensated in more volatile real-time markets), which immediately translates into economic loss for the utility.

Consequently, when setting the retail price, utilities have to take into account the potential economic loss due to demand prediction error. Larger demand uncertainty leads to larger prediction error, and thus higher economic loss. This observation is crucial in understanding the results in Chapter 4 and motivates the joint optimization framework design in this thesis.

Customers (CSPs) In this paper, we consider CSPs that operate energy-hungry geo-distributed datacenters (*e.g.*, Google and Microsoft) to provide *computing-intensive* services (*e.g.*, search) to its users through the Internet. Depending on whether they perform GLB, CSPs' roles as electricity customers differ significantly.

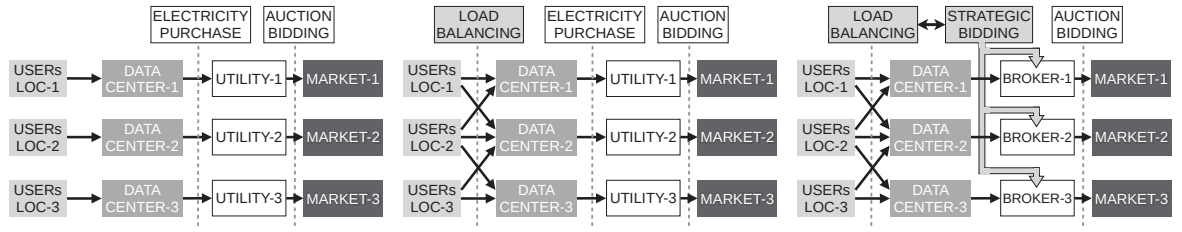
- Without GLB, a CSP manages its geo-distributed datacenters separately. Each datacenter only serves its regional workload, and it purchases electricity from local utilities for its energy needs. In this case, from the utilities' point of view, each datacenter is no different from traditional electricity customers (*e.g.*, commercial buildings).

- However, CSPs can perform GLB for various purposes, including but not limited to reducing the total electricity cost of its geo-distributed datacenters. As long as the quality of service does not degrade, routing service requests to datacenters at locations with cheaper electricity price can provide remarkable cost reduction [60]. According to the widespread estimate in [52], the workload of a datacenter that can be geographically load-balanced corresponds to 20-30% of the datacenter electricity demand. In such a scenario, CSPs represent a completely *new* type of electricity customers to the geo-isolated market infrastructure and to local utilities, in the sense that CSPs' energy demand at one location is *elastic*, caused by CSPs moving their workload around.

3.1.2 Supply Chain Evolutions with GLB

The electricity supply chain will evolve intriguingly under different GLB and EP models. And we briefly describe 3 variants involved in this thesis.

- *No GLB Model*: In this scenario (see Fig. 3.2(a)), electricity utilities purchase electricity from local electricity spot markets. Then, the utilities sell electricity like a commodity to datacenter owners to support



(a) Conventional electricity supply chain.

(b) Electricity supply chain with GLB.

(c) Geo-Distributed electricity supply chain.

Figure 3.2: Three electricity ecosystems related to this thesis. [17]

their operation.

- *Conventional GLB Model*: The scenario evolves to Fig. 3.2(b) if GLB is conducted. The critical change is that different supply chains, which are originally separated, interact with each others.

- *Joint GLB and EP Model*: In this scenario (see Fig. 3.2(c)), data-center owners directly purchase electricity from local spot markets, either by obtaining a valid license[1] or through a broker (*e.g.*, utilities are ideal broker candidates).

3.2 Deregulated Electricity Market

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. 3.3 and explain the details in the following.

[1]As a real-world example, in February 2010 the Federal Energy Regulatory Commission authorized Google to buy and sell energy at market rates [30].

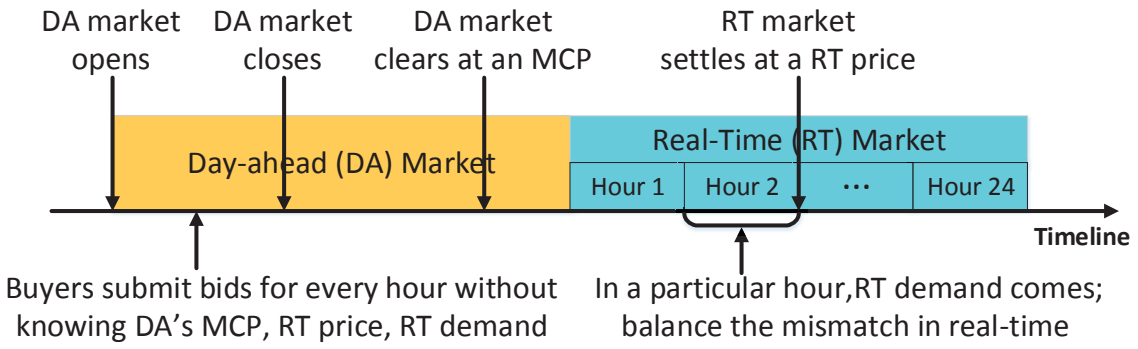


Figure 3.3: Operation of day-ahead market and real-time market.

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 <*marginal price, quantity*>, 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 marginal price. Each seller (resp. buyer) is allowed to submit *multiple* offers (resp. bids) [12] 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 [89]. 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 quantity and MCP 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. However, the statistical information can be learned from historical data.

We show an example in Fig. 3.4 from the perspective of our CSP. Suppose that the CSP submits three bids to the day-ahead market: <30$/MWh, 3MWh>, <51$/MWh, 4MWh>, <70$/MWh, 5MWh>. Now if ISO announces that the MCP is 40$/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 $4 + 5 = 9$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 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 are not exactly known a priori.

- In case that the day-ahead committed supply matches exactly the actual demand, there is no real-time cost.

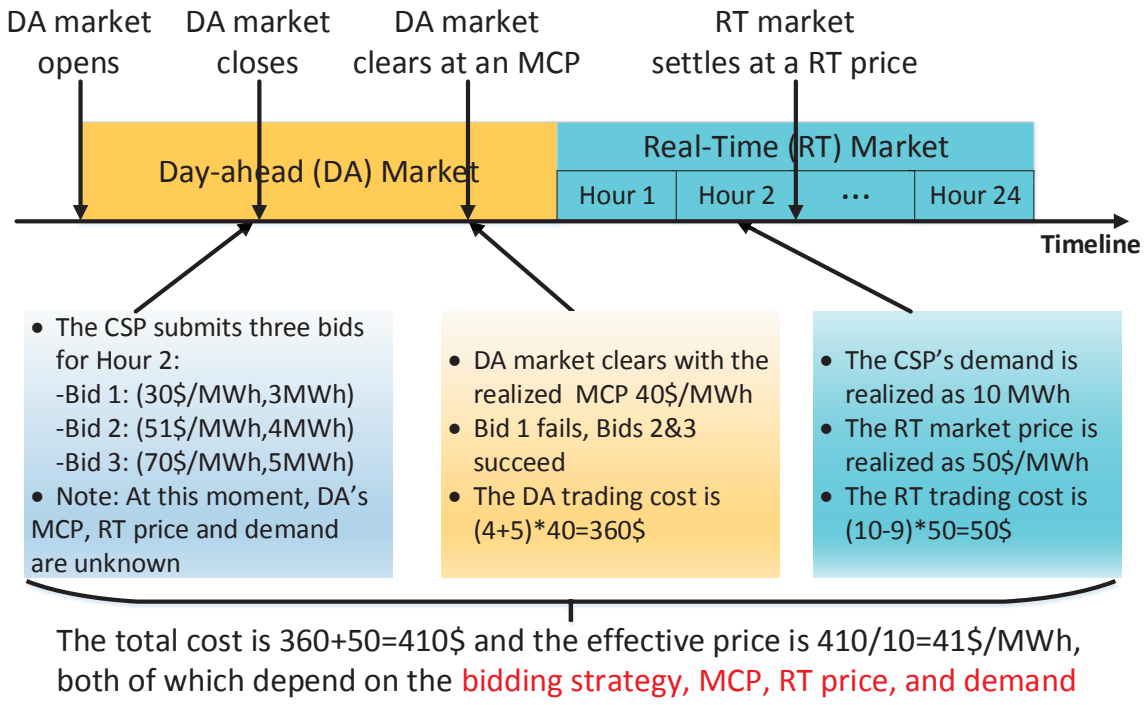


Figure 3.4: An illustrating example for the CSP to participate in markets.

- 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 [54] 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. A concrete real-world example fit the above description could be found in [33] (a Spanish Market). We remark that the real-world pricing mechanisms to handle the real-time mismatch could be different in markets and our description here might be a little bit specific for the purpose of math modelling in Chapter 5. However, the developed framework can also be extended to different pricing models described in [55, 54, 26] (see our discussions in Chapter 9).

Back to our example for the CSP in Fig. 3.4, 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. Now 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 [55].

□ **End of chapter.**

Chapter 4

A Subtle yet Important Issue of Doing GLB and EP Separately

In this chapter we present a model to analyze how GLB will interact with the electricity supply chain. In particular, we show that utility companies have to increase retail prices in order to ensure certain profit margin in face of the economic loss caused by a simple GLB solution. Consequently, CSPs doing "careless"GLB (as in Fig. 3.2(b)) might end up paying *higher* electricity prices than not doing GLB (as in Fig. 3.2(a)).

4.1 A Separate GLB and EP Solution

Here, we briefly describe a simple GLB solution. This seminal work [60] firstly identifies the geographical electricity price diversity and that this diversity can be exploited by the special electricity customer CSP. The basic idea is very simple and intuitive. The CSP owns several datacenters (or computing clusters) in different locations. Each datacenter is an end customer and signs a long-term contract with its local utility company. Naturally, different datacenters will have different retail prices. When some request comes, which represents some amount of electricity consumption [45], we can rout this request to the datacenter with cheaper electricity

while respecting the performance guarantees (like delay constraints). With the electricity prices and demands given as constants, the optimization problem reduces to a simple LP, which can be efficiently solved by off-the-shelf solvers, like CVX [31]. The authors of [60] shows this technique is promising to save millions of US dollars per year even for a relatively small system like Akamai.

4.2 GLB increases Prediction Error of Utilities' Demand

We begin our arguement by showing that this simple GLB approach will increase the demand prediction error of utilities. Before presenting our empirical study, the underlying logic is quite intuitive: with GLB, the local demands not only depend on the local information, (like the local temperature,) but also depend on the remote information, (like electricity consumptions and prices in other locations) and the private information of the CSPs, (like how much remote workload will be routed to this datacenter), both of which are usually either not taken into account by or not revealed to the utilities. Then the utilities' ability to make accurate predictions are inevitably depressed by GLB.

4.2.1 Dataset Characterization

We firstly describe the dataset we use in our empirical study.

Datacenters' demand: We use traces from the Akamai CDN as the

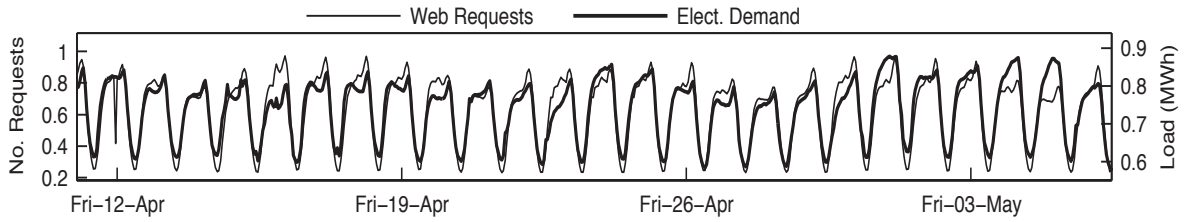


Figure 4.1: Evolution of the (aggregated) electricity demand and web workload between April 12th and May 6th 2013.

user request workload of the (virtual) CSP in its three datacenters. We crawl Akamai's Internet Observatory website [9] to obtain the number of HTTP requests per minute against the Akamai CDN in North America. Akamai CDN relies on co-location datacenters that individually do not represent large electricity consumption. Nevertheless, using the conversion rate of $1kJ$ per query ($0.28\ Watts \cdot h$) claimed by Google for its datacenters [60], the crawled workload aggregately creates a power consumption of 125 MW, which may serve well to approximate the consumption of three Facebook's datacenters at full utilization (according to [7, 6]).

Since Akamai does not dissect the information of its workload per location, we run a preliminary experiment to make an educated approximation of the workload splitting for the three locations by the following method.

We aggregate the electricity demand curves from the three locations into a time series, respecting the time difference between the aggregated time series of each location. We compare this (normalized) electricity demand aggregate with the time series of the (normalized) number of web requests against the Akamai CDN. The two seires are displayed in Fig. 4.1. The

correlation coefficient of these aggregated curves is 0.92. Most differences appear during the morning and more noticeably in some weekends, what we associate with the industrial and commercial activity. Then we split the number web requests (electricity demand) to three subsets according to the ratios of electricity demand among the locations. This method is reasonable assuming that a random sample of the population in these three areas will provide similar results about the usage of electricity and web services (and the ratio between these two) and it should provide us a good estimation.

Utilities's demand and electricity price: To obtain the total electricity demand of each of the three local utilities, we crawl the hourly electricity demand from the spot markets in San Diego [16], CA, Houston [24], TX, and New York [4], NY for 2009-2012, and choose nodal demand so that the datacenter demand represents to 30% of the utility's demand (following the back-of-the-envelope computation presented in the introduction). We also collect the day-ahead MCPs and real-time prices of the three spot markets for the same period.

Finally, to maximize their prediction accuracy, utilities take into account the weather conditions and daily activity patterns. We crawl the hourly weather conditions [5] in the three areas and the official holidays calendar for 2009 - 2012. We omit the weekends in all our experiments, due to the seasonality of the workload and electricity demand during these days.

4.2.2 Prediction Method

In our empirical study, we change the proportion of the allowed GLB workload from 0 to 60% of the total workload (the cases beyond 30% aggressively evaluate a futuristic scenario reflecting the datacenter's increasing capability to conduct GLB). For each hour, the CSP solves a standard GLB cost-minimization problem as the one in [60] to allocate its allowed GLB workload optimally. The evaluation is carried out assuming that utilities use commonly adopted *neural networks* (NN)-based demand forecast algorithms [74] to predict their electricity demand[1]. The inputs of the NNs include the weather forecast, historical demand records, and whether it is a public holiday/weekend or not, while the output is the hourly electricity demand. Utilities use NNs as a black-box, which requires training with training data set. Once they are trained, the NN takes the inputs in the testing data set to predict the demand for each hour, which results in a certain estimation error.

We train the NN with data from 2009-2011 and use the trained model to perform hourly demand prediction during 2012. To this end, we use different training data sets, one for the case without GLB (original workload traces) and one for each GLB eligible ratio that we study (workload traces 'optimized' by GLB). We compare the predicted demand and the actual demand to record the *mean absolute percentage error* (MAPE) in Table 4.1.

[1]For a real-world practice, the readers are referred to `http://www.mathworks.com/matlabcentral/fileexchange/28684-electricity-load-and-price-forecasting-webinar-case-study`.

Table 4.1: MAPE and Prices vs. Balanced Load

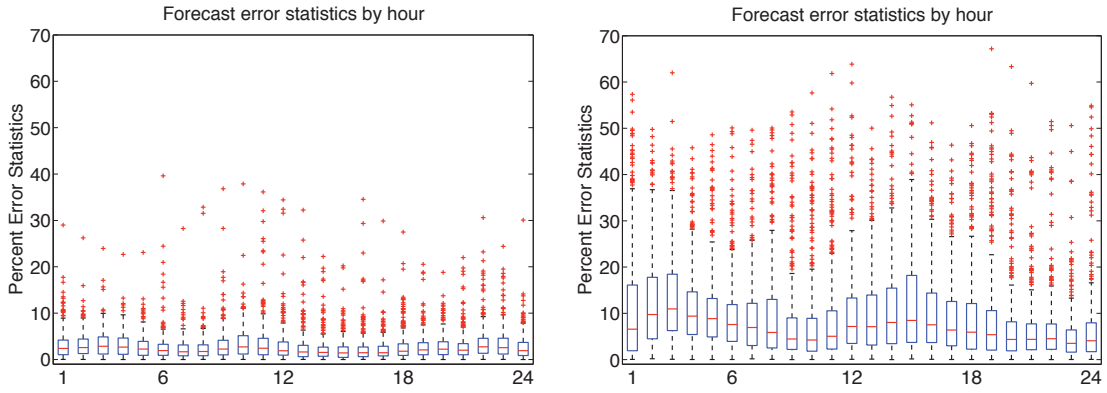
GLB	San Diego		Houston		New York	
(%Load)	MAPE (%) & Avg. Price ($/MWh)					
0	3.0	47.9	2.7	43.9	3.0	70.2
15	6.8	49.3	3.5	45.5	6.4	70.8
30	8.2	49.8	7.3	47.2	7.6	71.0
45	10.7	50.8	10.5	48.7	8.6	71.2
60	14.3	52.2	14.8	50.8	10.7	71.6
MAPE/GLB	0.714		0.921		0.345	

4.2.3 Utilities' Demand Prediction Error

In Table 4.1, each datacenter location has two associated columns. We report the MAPE with varying GLB load (in percentage, increased at 15% resolution) in the first column. The last row shows the ratio between MAPE and proportion of routable workload to other locations. Several interesting observations can be made from this table.

First, without GLB (corresponding to the third row of 0% GLB load), the NN algorithm can predict the actual demand pretty accurately – with a MAPE at most 3%. A closer look into the prediction accuracy of the NN algorithm for the San Diego site shows the hourly MAPE has a mean of 3% and a standard variation of 6%. These results show that without GLB, NNs can predict accurately the real-world electricity demand, justifying its widespread adoption in practice.

Second, as the GLB load percentage increases, MAPE of the NN algo-



(a) (b)

Figure 4.2: (a) Statistics of demand prediction error without GLB; (b) Statistics of demand prediction error with GLB at 10% (*i.e.*, the allowed demand variation caused by the CSP performing GLB is 10%).

rithm also increases remarkably for all three locations. For example, in Table 4.1, when the GLB load increases to 30%, the MAPE for San Diego increases to 8.15%, 2.7 times of that of no GLB. The standard deviation of MAPE is 11.3%, almost twice of that of no GLB. These results are in sharp contrast to the case of no GLB, and confirm our intuition that GLB introduces demand uncertainty and extra errors in the demand prediction.

For a better illustration, we also visualize the hourly forecast error statistics for the case without GLB and the case with 10% GLB in Fig. 4.2(a) and 4.2(b), respectively. As we can see, both the values and variances of prediction errors for all hours are increased evidently.

4.3 Prediction Error Increases Retail Price for CSPs

We proceed to show that larger demand prediction errors will lead to higher retail prices. Let d be the actual demand for a particular hour in the next day, $\tilde{d}$ be the utility's prediction of d, and w^b be the average (MCP) price at which the utility purchased $\tilde{d}$ amount of electricity for that hour from the day-ahead market.

Without prediction error, *i.e.*, $\tilde{d} = d$, given a retail price p_0 [2], the utility obtains a desired profit for the hour as

$$\left(p_0 - w^b\right) d. \tag{4.1}$$

With prediction error, the utility suffers additional economic loss as compared to the error-free case.

- In case of over-prediction, there is $\tilde{d} - d > 0$ amount of electricity surplus (and it cannot be stored). In today's practice, the utility can sell them back to a GENCO at an average marginal price denoted as w^s (usually $w^b > w^s$). The economic loss to the utility is $\left(w^b - w^s\right)\left(\tilde{d} - d\right)$.
- In case of under-prediction, there is $d - \tilde{d} > 0$ amount of unmatched demand to be urgently balanced by the utility to avoid power outage. In today's practice, the utility can purchase supply in the hour-ahead

[2] The process of how a utility determines its retail price can be highly involved (consideration factors include competition from other utilities). A vital requirement that the price has to be high enough to guarantee the (expected) profit is larger than a minimum for the utility to stay in business.

or real-time markets to satisfy urgent demand, but at a expected price higher than in day-ahead markets. Denote the average marginal price of buying electricity in urgency as w^u (usually $w^u > w^b$). The economic loss to the utility is then $\left(w^u - w^b\right)\left(d - \tilde{d}\right)$.

In order to compensate the economic loss of the utility due to prediction errors, and to obtain the same expected profit in (4.1), the utility needs to set a retail price p *higher* than p_0 (the price for the error-free case) according to:

$$p = p_0 + \left(w^b - w^s\right)\mathbb{E}\left[\left(\tilde{d} - d\right)^+ / d\right] + \left(w^u - w^b\right)\mathbb{E}\left[\left(d - \tilde{d}\right)^+ / d\right] > p_0. \tag{4.2}$$

Denote MAPE by Δd, *i.e.*,

$$\Delta d = \mathbb{E}\left[\left|\tilde{d} - d\right| / d\right].$$

To ensure the expected profit is at least the desired one in (4.1), the relationship between the retail price and MAPE Δd can be characterized by

$$p = p_0 + \left(w^u - w^s\right)\Delta d. \tag{4.3}$$

We continue our previous empirical study to compute the retail prices with and without prediction errors according to (4.3) with $p_0 = w^b$ (modeling an altruistic utility targeting zero expected profit). The numerical results are reported in the second column of each datacenter location in Table 4.1. We can observe that the retail prices for all three datacenters

are increased and different locations have different price increment per % GLB, depending on their individual market profiles. As an example, *the retail price for San Diego on average increases by 0.7% for every increment of 1% in the GLB load.*

GLB's Performance Degradation: Next, adding the updated pricing information, we can evaluate how the performance degradation of GLB will be degraded by the introduced demand uncertainty. We do this for the cases where the CSP is able to move 0%, 15%, 30%, and 60% of the total local utility demand, which we denote as NOGLB, GLB@15, GLB@30, and GLB@60 respectively. We study and compare the total electricity cost (sum of the three locations for the year 2012) between the baseline case, NOGLB, and the rest (in percentage).

Results show that in the GLB@15 case *the CSP actually ends up paying a total bill 1% higher than not doing GLB at all.* In the GLB@30 case where the CSP can move up to 30% of its overall workload, the ability of aggressively moving workload to low-price locations improves the results, in spite of the increase in the electricity prices due to higher degrees of uncertainty. However, the savings in the overall electricity bill is still minor, about 3%, while the CSP is already moving the full allowed GLB workload of its datacenters. Finally, higher benefits could be achieved with larger *allowed* GLB load. For the GLB@60 case, the GLB effect provides 9% cost reduction, but note that this case requires the CSP to move a workload that is beyond the *feasible* percentage in datacenters nowadays (20-30% according to [60]).

4.4 Discussions

Based on this (simplified) electricity pricing model, demand predictions are critical for the operation of the utilities. The good news is that, conventionally electricity demand is rather predictable as it follows regular patterns that repeats daily, with seasonality during weekends and holidays.

Although its impact depends on the amount of routed workload, GLB may introduce utterly different demand patterns. As we justified by the previous example, just adapting local demand prediction methods to GLB may not be enough to yield accurate predictions and extra economic loss by GLB is inevitable. In the next chapter, we introduce a cooperative model in which CSPs join the wholesale markets to purchase electricity. In this way, CSPs doing GLB can exploit their appearance in multiple locations, while bypassing such trading inefficiency.

□ **End of chapter.**

Chapter 5

A Joint GLB and EP Solution: Problem Formulation

In this thesis, 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. 5.1. 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 [67] and the suggested time granularity for performing GLB[60].

Without loss of generality, we focus on minimizing cost of a particular operation hour of the CSP, as shown in Fig. 3.4.

5.1 Workload and Geographical Load Balancing

Workload and Electricity Demand. We assume that each datacenter is power-proportional, which means that its electricity demand is proportional to its workload [60]. For example, Google reports that each search requires about 0.28Wh electricity for its datacenters [60]. 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 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 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

[1] Under the conventional setting where datacenters obtain electricity from utilities, GLB is performed in CSP's real-time operation. Under the considered setting, 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 work. It is conceivable to perform GLB in both day-ahead and real-time operations of CSP to further minimize the energy cost, which we discuss in Chapter 11.

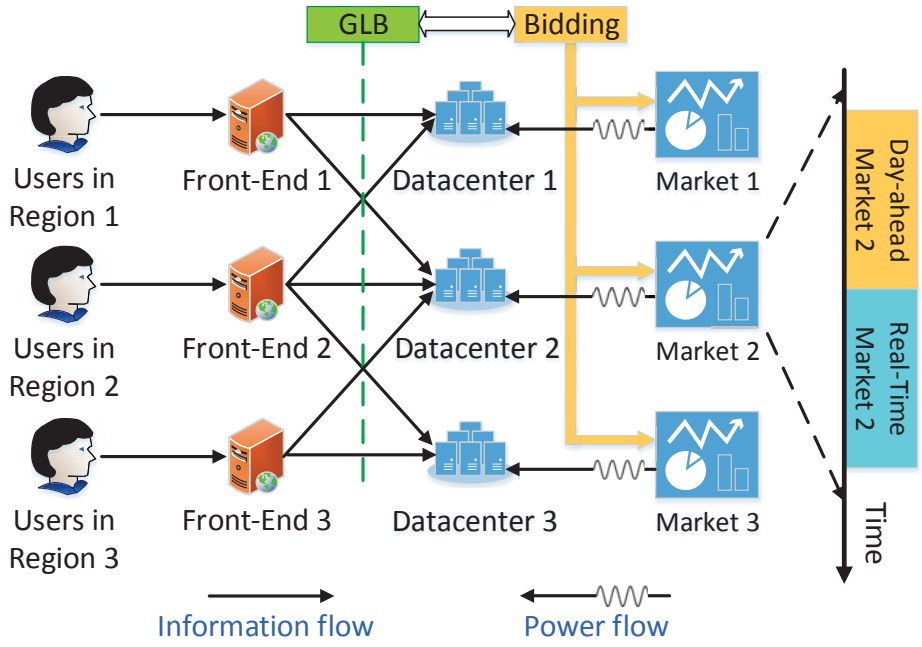


Figure 5.1: The scenario that we consider in this work.

$\boldsymbol{\alpha} = [\alpha_{ij} : i, j = 1, \ldots, N] \in \mathbb{R}_{N \times N}$ which satisfies

$$\sum_j \alpha_{ij} \geq 1, \quad \forall i = 1, \ldots, N, \tag{5.1}$$

$$\alpha_{ii} \geq \lambda_i, \quad \forall i = 1, \ldots, N, \tag{5.2}$$

$$\bar{v}_j \triangleq \sum_{i=1}^{N} \alpha_{ij} \bar{u}_i \leq C_j, \quad \forall j = 1, \ldots, N. \tag{5.3}$$

$$0 \leq \alpha_{ij} \leq 1, \quad \forall i, j = 1, \ldots, N, \tag{5.4}$$

$$\alpha_{ij} = 0, \quad \forall (i, j) \in \mathcal{G}, \tag{5.5}$$

where $\mathcal{G} \triangleq \{(i, j) |$ workloads from region i cannot be routed to datacenter $j\}$ captures the topological constraints.

Here α_{ij} represents the fraction of the workload originated from region

i that will be routed to datacenter j. Constraints in (5.1) mean that all workloads must be served. Constraints in (5.2) 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 (5.3) ensure that the total workload coming into datacenter j can be served even in the largest realization of workload. Constraints in (5.5) describe that the workload cannot be routed to a datacenter that is too far away from its own region. We define the set of all feasible GLB decisions as

$$\mathcal{A} \triangleq \{\boldsymbol{\alpha} \in \mathbb{R}^{N\times N} | \boldsymbol{\alpha} \text{ satisfies } (5.1) - (5.5)\}. \tag{5.6}$$

Given the GLB decision $\boldsymbol{\alpha}$, the total workload for 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 PDF

$$f_{V_j}(v) = f_{U_{1j}} \otimes f_{U_{2j}} \otimes \ldots \otimes f_{U_{Nj}}(v), \tag{5.7}$$

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

$$f_{U_{ij}}(u) = \begin{cases} \frac{1}{\alpha_{ij}} f_{U_i}\left(\frac{u}{\alpha_{ij}}\right), & \text{if } \alpha_{ij} > 0, \\ \delta(u), & \text{if } \alpha_{ij} = 0, \end{cases} \tag{5.8}$$

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

Bandwidth Cost. To understand and compare the scales of electricity and bandwidth cost of serving the internet services, we estimate the bandwidth cost and electricity cost of one google search.[2] We assume that, to

[2]It should be noted that the electricity price and bandwidth prices may vary enormously in different places and time, so the estimation is more like a Fermi problem and we only care about the order.

serve one google search, we need to we need to consume 0.28Wh electricity [30] and deliver the traffic volume of one webpage, which is roughly 300 KB [82].

- For the electricity cost, the electricity price to the end customer is about 0.07 \$/KWh, so the cost of powering one google search is about $0.07 * 0.00028 = 1.96 * 10^{-5}$\$.
- For the bandwidth cost, we assume that the pay-by-traffic charging scheme is used. I check the pricing scheme of ALIYUN, one major CDN service provider in Mainland China. The cost of delivering one GB data is close to 0.05 USD [18], so the cost of google search is like to be $0.05 \times \frac{300}{1024^2} = 1.4 * 10^{-5}$\$.

So according to the data and rough calculation, the two types of cost are of the same order and need to be jointly considered.

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

$$\mathsf{BCost}(\boldsymbol{\alpha}) = \sum_{i=1}^{N} \sum_{j=1}^{N} z_{ij} \cdot \alpha_{ij} \cdot \mathbb{E}(U_i). \tag{5.9}$$

5.2 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 independent random variables

P_j ($j \in [1, N]$), each with probability distribution $f_{P_j}(p)$ that can be empirically estimated from historical data [17]. Here we assume that the CSP has negligible market power and its bidding and GLB behavior will not affect the dynamics of electricity markets[3].

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 [17]. We define $\mu_j^{\mathsf{RT}} \triangleq \mathbb{E}[P_j^{\mathsf{RT}}]$ as the expectation of P_j^{RT}. We assume that all day-ahead MCPs P_j's and real-time market prices P_j^{RT}'s are independent[4].

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 [26, 46]. 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$\left\langle b_j^k, q_j^k \right\rangle, k = 1, \ldots, K$, to the day-ahead

[3]The assumption is reasonable as, *e.g.*, datacenters in the US only consume 2% of total electricity [3], and it is usually used in the literature such as [67].

[4]We remark that this independence assumption may not hold in practice. But it significantly simplifies our analysis and allows us to reveal some important insights. A comprehensive study of considering correlations between day-ahead MCPs and real-time prices would be an interesting future work.

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

$$q_j(p) = \sum_{k:b_j^k \geq p} q_j^k, \quad \forall p \in \mathbb{R}^+. \tag{5.10}$$

For example, considering the three bids in Fig. 3.4, we can construct the corresponding bidding curve as shown in Fig. 5.2.

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 (5.10) represents the total amount of electricity obtained when the day-ahead 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 (5.10) can be realized by placing the K bids $\left\langle b_j^k, q_j^k \right\rangle, k = 1, \ldots, K$ stated above.

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

$$\mathcal{Q} \triangleq \left\{ q(p) \mid q\left(p_1\right) \leq q\left(p_2\right), \forall p_1 \geq p_2, p_1, p_2 \in \mathbb{R}^+ \right\}. \tag{5.11}$$

Remark. Here, we assume that the CSP is allowed to submit any number, possibly infinite number, of bids. This assumption allows us to significantly simplify the derivation of optimal solution to the joint bidding and GLB problem in Chapter 6. In Chapter 8, we relax this assumption

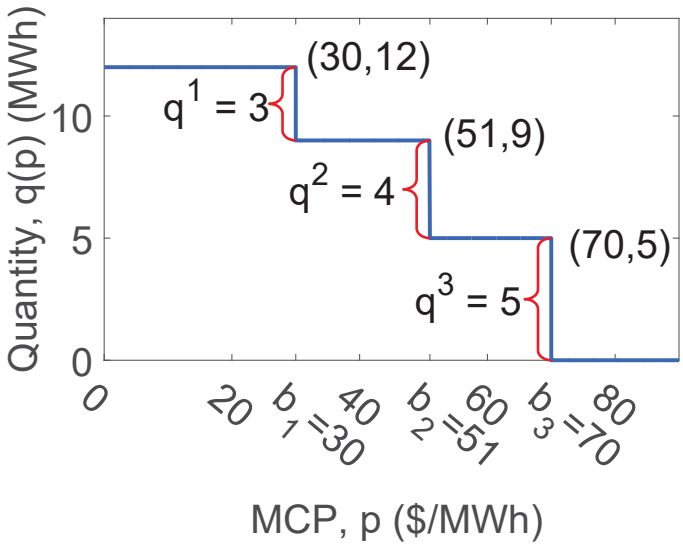


Figure 5.2: An illustrating example for the (step-wise) bidding curve constructed from the submitted three bids in Fig. 3.4.

and discuss how to approximately realize a continuous bidding curve with a limited number of bids in the practical implementation. Our simulation results in Chapter 10 (Tab. 10.2) 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 $\boldsymbol{\alpha}$, we denote the *expected* electricity procurement cost of the CSP in electricity market j as $\mathsf{ECost}_j\left(q_j(p), \boldsymbol{\alpha}\right)$, 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

$$\mathsf{ECost}_j\left(q_j(p), \boldsymbol{\alpha}\right) = \underbrace{\int_0^{+\infty} f_{P_j}(p) \Big[\underbrace{pq_j(p)}_{\substack{\text{Day-ahead}\\ \text{trading cost}}} \underbrace{- \underbrace{\beta p \int_0^{q_j(p)} (q_j(p) - v) f_{V_j}(v) dv}_{\text{Rebate of over-supply}} + \underbrace{\mu_j^{\mathsf{RT}} \int_{q_j(p)}^{\bar{v}_j} (v - q_j(p)) f_{V_j}(v) dv}_{\text{Cost of under-supply}}}_{\text{Real-time trading cost}} \Big] dp}_{\text{Expected electricity cost of datacenter } j \text{ conditioning on day-ahead market } j\text{'s MCP } P_j = p}. \tag{5.12}$$

exactly match the real-time demand V_j. If $V_j = v$ and $v > q_j(p)$, *under-supply* happens and we need to buy $v - q_j(p)$ amount of electricity at expected price μ_j^{RT}, so the expected cost due to under-supply would be $\mu_j^{\mathsf{RT}} \int_{q_j(p)}^{\bar{v}_j} (v - q_j(p)) f_{V_j}(v) dv$. 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 $\mathsf{ECost}_j\left(q_j(p), \boldsymbol{\alpha}\right)$ in (5.12) by applying the total expectation theorem. Note that $\mathsf{ECost}_j\left(q_j(p), \boldsymbol{\alpha}\right)$ is related to the GLB decision $\boldsymbol{\alpha}$ through the distribution of V_j (the workload of datacenter j), which is computed by (5.7) and (5.8).

We provide the following proposition to reveal an important property of (5.12).

Proposition 1. *The cost function* (5.12) *is generally non-convex in* $q_j(p)$.

The proof for Proposition 1 is in Appendix 13.1. Essentially Proposi-

tion 1 indicates that the optimization problem involving (5.12) is nonconvex and requires sophisticated design.

5.3 Problem Formulation

We now formulate the problem of joint bidding and GLB:

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

where $\mathcal{A}$ is the set of all feasible GLB decisions, defined in (5.6) and $\mathcal{Q}$ is the set of all feasible bidding curves, defined in (5.11). 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.*, [60, 61, 78, 17]. 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 respective to $q_j(p)$ (see Proposition 1). Second, the optimization variable $q_j(p)$ is a functional variable with infinite dimensions. Thus it is highly non-trivial to solve this non-convex infinite-dimensional problem optimally by existing solvers, without incurring forbidden complexity.

5.4 An Alternative Two-stage Formulation

In our previous formulation **P1**, we assume that we will decide GLB strategy and the EP strategy simultaneously before the day-head markets are closed. When real-time demands come, we will follow our previous decision and allocate the demand proportionally to different datacenters. However, readers may have already realized that, instead of sticking to our day-ahead decision, we can perform another optimization to optimally route the demand in real-time, with the exact information of the real-time demand and the electricity procurement amount for each datacenter. Under this scheme, GLB in real-time is used not only to exploit the price diversity across different regions, but also to handle the mismatch between day-head procurement and real-time demand. So another natural formulation for the joint optimization framework essentially span two stages: the first stage is day-ahead, when we submit bidding curves to day-ahead markets; the second stage is real-time, when we allocate demand. Since we optimize the GLB strategy for different realizations of $U_i, P_j, \forall i, j$, we can have a larger gain as compared with optimizing with only their statistical information. However, as we will show in Chapter 11, the optimization problem is too complicated and challenging to solve. In the main body of this thesis, we will focus on solving **P1** since it is intellectually interesting and its empirical performance is satisfactory.

□ **End of chapter.**

Chapter 6

A Joint GLB and EP Solution: Algorithm Design

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

6.1 Reducing P1 to a Convex Problem and Approach Sketch

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

$$\hat{\mathcal{Q}}_j = \{q_j(p) | q_j(p) \in \mathcal{Q}, \text{ and } q_j(p) = 0, \forall p \geq \mu_j^{\mathsf{RT}}\}. \tag{6.1}$$

As compared to $\mathcal{Q}$ defined in (5.11), the new constraint in the definition of $\hat{\mathcal{Q}}_j$, *i.e.*, $q_j(p) = 0, \forall p \geq \mu_j^{\mathsf{RT}}$, 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 opti-*

mal solution as ***P1****:*

$$\textbf{\textit{P2:}} \quad \min \quad \sum_{j=1}^{N} \mathsf{ECost}_j\left(q_j(p), \boldsymbol{\alpha}\right) + \mathsf{BCost}(\boldsymbol{\alpha})$$

$$\textit{var.} \quad \boldsymbol{\alpha} \in \mathcal{A},\ q_j(p) \in \hat{\mathcal{Q}}_j, j = 1, \ldots, N.$$

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 13.2; 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 (6.1).

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

$p \in [0, \mu_j^{\mathsf{RT}})$. To illustrate our design, we first rewrite problem **P2**,

$$\min_{\boldsymbol{\alpha}\in\mathcal{A}} \min_{q_j(p)\in\hat{\mathcal{Q}}_j,\forall j} \left\{ \sum_{j=1}^{N} \mathsf{ECost}_j\left(q_j(p), \boldsymbol{\alpha}\right) + \mathsf{BCost}(\boldsymbol{\alpha}) \right\}$$

$$= \underbrace{\min_{\boldsymbol{\alpha}\in\mathcal{A}} \left\{ \sum_{j=1}^{N} \underbrace{\left[\min_{q_j(p)\in\hat{\mathcal{Q}}_j} \mathsf{ECost}_j\left(q_j(p), \boldsymbol{\alpha}\right) \right]}_{\text{Problem } \mathbf{EP}_j(\boldsymbol{\alpha}), \text{ solved in Chapter 6.2}} + \mathsf{BCost}(\boldsymbol{\alpha}) \right\}}_{\text{Problem } \mathbf{P3}, \text{ solved in Chapter 6.3}} \tag{6.2}$$

The structure of the expression in (6.2) 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 $\boldsymbol{\alpha}$, by solving the following problems:

$$\mathbf{EP}_j(\boldsymbol{\alpha}): \quad \min_{q_j(p)\in\hat{\mathcal{Q}}_j} \mathsf{ECost}_j\left(q_j(p), \boldsymbol{\alpha}\right), \quad j = 1, \ldots, N. \tag{6.3}$$

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

$$\textbf{P3:} \quad \min_{\boldsymbol{\alpha}\in\mathcal{A}} \quad \sum_{j=1}^{N} \mathsf{ECost}_j\left(q_j^*(p; \boldsymbol{\alpha}), \boldsymbol{\alpha}\right) + \mathsf{BCost}(\boldsymbol{\alpha}). \tag{6.4}$$

According to Theorem 1, **P2** is convex and then, the inner-loop problem $\mathbf{EP}_j(\boldsymbol{\alpha})$ and outer-loop problem **P3** are both convex, which are perhaps

not surprising. In the following parts, we solve $\mathbf{EP}_j(\boldsymbol{\alpha})$ and **P3** to obtain an optimal joint bidding and GLB solution to **P2**, which is also optimal for **P1**.

6.2 Inner Loop: Optimal Bidding Given GLB Decision

The inner-loop problem $\mathbf{EP}_j(\boldsymbol{\alpha})$ 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 $\boldsymbol{\alpha}$ given, in face of demand and price uncertainty. Note that $\mathbf{EP}_j(\boldsymbol{\alpha})$ is closely related to the classic Newsvendor problem [36]. In the Newsvendor problem, the market prices are given and only the buying quantity should be optimized under demand uncertainty, while in $\mathbf{EP}_j(\boldsymbol{\alpha})$ we need to optimize both the bidding quantities and bidding prices simultaneously under both price and demand uncertainties.

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 (5.7). The following theorem shows that $\mathbf{EP}_j(\boldsymbol{\alpha})$ admits a closed-form solution $q_j^*(p;\boldsymbol{\alpha})$, addressing the infinite-dimension challenge.

Theorem 2. *Given GLB decision $\boldsymbol{\alpha}$, we assume that $F_{V_j}(x)$ is strictly increasing; thus its inverse exists and is denoted as $F_{V_j}^{-1}(x)$. The optimal*

bidding curve for solving $\boldsymbol{EP}_j(\boldsymbol{\alpha})$ *is given by, for* $j = 1, \ldots, N$,

$$q_j^*(p;\boldsymbol{\alpha}) = \begin{cases} F_{V_j}^{-1}\left(\frac{\mu_j^{RT}-p}{\mu_j^{RT}-\beta p}\right), & \text{if } p \in \left[0, \mu_j^{RT}\right); \\ 0, & \text{otherwise.} \end{cases} \tag{6.5}$$

The proof of Theorem 2 is delegated in Appendix 13.3.

Extensions. The extension to the case where $F_{V_j}(v)$ is not strictly increasing should be easy to derive by the proof above. Recall we want to find a $q_j(p)$ to minimize $pq - \beta p \int_0^q (q-v) f_{V_j}(v)dv + \mu_j^{\mathrm{RT}} \int_q^{\bar{v}_j} (v-q) f_{V_j}(v)dv$ with derivative $p - \mu_j^{\mathrm{RT}} + (\mu_j^{\mathrm{RT}} - \beta p) F_{V_j}(q)$. 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. We present a brief discussion here. Other than the case in Theorem 2 (we can find a unique solution to make the derivative equal to 0), we can have another two cases: (i) there are multiple solutions for the derivative to be 0. In this case. any solution is an optimal solution. (ii) there is no solution for the derivative to be 0 (the derivative is not continuous.). Then there is a critical point at which the derivative 'jumps' from negative value to positive value. Both cases can be solved numerically by binary search.

Remarks. The optimal bidding curve $q_j^*(p;\boldsymbol{\alpha})$ is *universal* in that it does not depend on the distribution of day-ahead MCP P_j. This is because $q_j^*(p;\boldsymbol{\alpha})$ actually minimizes the expected electricity procurement cost for any p. This salient feature is appealing as it means that the CSP does not need to re-optimize its bidding strategy upon possible changes in market mechanism or pricing policy. Also, the structure of $q_j^*(p;\boldsymbol{\alpha})$ helps us to

combat the demand uncertainty and leverage the price uncertainty. More insightful discussions can be found in Chapter 7.

6.2.1 Connections with Newsvendor Problem

As a single-period inventory problem, the Newsvendor problem is one of the most classic problems in operation research and has been extensively studied before. Comprehensive reviews are provided in [36] and [59].

In the basic version of Newsvendor problem, the vendor (or retailer) needs to decide the optimal ordering quantity from the suppliers to maximize his expected profit by selling the goods to the customers at a higher price. Usually, the decision should be made before the real-time demand comes, so the vendor needs to optimize his decision based on statistical information of future demand. On the one hand, if he orders too little, he loses some chances of making profit. On the other hand, if he orders too much, the unsold goods will incur some loss. So the scenario is quite similar to that of $\mathbf{EP}_j(\boldsymbol{\alpha})$ studied in this thesis.

The Newsvendor problem can provide key insights for the inventory or supply chain (consisting of supplier, retailer and customer) management problems, especially with the perishable goods like electricity. Due to this reason, multiple variants of the basic version have been studied. For example, other than the expected profit, we can consider alternative objectives. In [23, 41], the authors maximize a general concave utility function, which can capture the vendor's risk-aversion nature. In [41, 40], the authors maximize the probability of reaching certain profit level, which is more

practical in real-world management. Also, the vendor can decide not only the ordering quantity but also the retail price, which can affect the real-time demand. For this reason, ordering quantity and retail price are jointly optimized in [40, 83, 57].

A key factor in the Newsvendor problem is the future demand randomness/uncertainty, to which some papers are devoted. In [27], the authors studied how the optimal preodering quantity and profit will be changed by manipulating demand uncertainty. In [70], a result that a larger demand uncertainty will increase the cost expectation was established with proper definitions of uncertainty (variability) levels, which is similar to Lemma 1 in this thesis. And in [63], the authors presented a more intriguing result, showing how a larger demand uncertainty could decrease the cost expectation, with a different definition of uncertainty.

The main difference between $\mathbf{EP}_j(\boldsymbol{\alpha})$ studied in this thesis and the Newsvendor problems in most literatures is that, in $\mathbf{EP}_j(\boldsymbol{\alpha})$, the vendor (CSP) makes some order from the supplier (electricity day-ahead market) *by bidding*, while in Newsvendor problem, the vendor only needs to tell the supplier how much he wants to order and he can surely buy at *a fixed and known price*. In $\mathbf{EP}_j(\boldsymbol{\alpha})$, the vendor is not only unsure about the future demand, but also unsure about his ordering quantity from the supplier due to the randomness of auction result (MCP in day-ahead markets). Alternatively speaking, the vendor is faced with both demand uncertainty and price (day-ahead MCP) uncertainty. Also, in $\mathbf{EP}_j(\boldsymbol{\alpha})$, the ordering strategy of CSP consists of *bidding prices* and *bidding quantities*. The coupling

Table 6.1: Comparisons with Literatures on Newsvendor Problem

References	Demand Uncertainty Impact	Price Uncertainty Impact	With Supply Uncertainty	With Risk Managment	Multiple Decision Variables
Eeckhoudt *et al.* [23]	✗	✗	✗	✓	✗
Whitin *et al.* [83]	✓	✗	✗	✗	✓
Wu *et al.* [85]	✗	✗	✓	✓	✗
Laul *et al.* [40]	✗	✗	✓	✓	✓
Polatoglu1 *et al.* [57]	✗	✗	✗	✗	✓
Merzifonluoglu *et al.* [53]	✗	✗	✓	✗	✓
Gerchak *et al.* [27]	✓	✗	✗	✗	✗
Song *et al.* [70]	✓	✗	✗	✗	✗
This Thesis	✓	✓	✓	✗	✓

nature of the two variables makes the problem even more challenging. A summary and comparison is provided in Table 6.1.

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

After obtaining the optimal bidding strategy $q_j^*(p; \boldsymbol{\alpha})$ as a function of GLB decision $\boldsymbol{\alpha}$, 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 $\mathbf{EP}_j(\boldsymbol{\alpha})$. Thus, gradient-based algorithms cannot be directly applied.

We tackle this issue by adapting a zero-order optimization algorithm, named General Pattern Search (GPS) [43], 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 [43] 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 $\mathbf{EP}_j(\boldsymbol{\alpha})$ given a GLB decision $\boldsymbol{\alpha}$, which can be obtained by plugging the optimal solution $q_j^*(p;\boldsymbol{\alpha})$ into the objective function of $\mathbf{EP}_j(\boldsymbol{\alpha})$. 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 [43]. In the following theorem, we prove that our Algorithm 1 actually converges to the optimal solution to the convex problem **P3**, under proper conditions.

Theorem 3. *Assume that $f_{U_j}(u)$, $j = 1, \ldots, N$, 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**.*

Remarks. Theorem. 3 follows the facts that **P3** is convex and GPS algorithm converges to a point satisfying the KKT condition [43]. The proof is deferred in Appendix 13.4.

6.4 Complexity and Practical Considerations

In this part, we discuss the computation complexity and some practical considerations for our solution.

6.4.1 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 $O(n_{iter}((N^5 m \log(Nm) + N^3 m^2)))$.*

The proof of Theorem 4 is in Appendix 13.5. The complexity is linear with the number of iterations until convergence. However, exactly char-

acterizing the convergence rate of GPS algorithm is still an open problem [22], and thus it is hard to get sharp bounds for the number of iterations, *i.e.*, n_{iter}. Instead, we empirically evaluate the convergence rate of our Algorithm 1 in Chapter 10.2.3. The results show that our Algorithm 1 converges fast – within 30 iterations – for the practical setting considered (*i.e.*, $n_{\mathsf{iter}} \leq 30$).

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 and less than 20 Datacenters of Google. Thus, Theorem 4 shows that the complexity of our Algorithm 1 is affordable in practice.

6.4.2 Imperfect Knowledge of 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 Chapter 7 and 10.2.5, we empirically show that our solution works pretty well for *imperfect* probability distributions of the demand and market prices, which only use the first-order (expectation) and second-order (variance) statistic information.

□ **End of chapter.**

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

1: **initialize** $\boldsymbol{\alpha}^0 \leftarrow \mathbf{I}_{N\times N}$, $t \leftarrow 0$
2: **while** not converge **do**
3: `current_value` $\leftarrow$ **P3-Obj**$(\boldsymbol{\alpha}^t)$
4: Get $\boldsymbol{\alpha}^{t+1}$ by invoking **P3-Obj** and comparing with `current_value` at most $2N^2$ times (see [43, Fig. 3.4])
5: $t \leftarrow t+1$
6: **end while**
7: $\boldsymbol{\alpha}^* \leftarrow \boldsymbol{\alpha}^t$
8: Compute $q_j^*(p;\boldsymbol{\alpha}^*)$ by (6.5) for all $j \in [1,N]$
9: **return** $\boldsymbol{\alpha}^*, q_j^*(p;\boldsymbol{\alpha}^*)$ for all $j \in [1,N]$

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

10: **function P3-Obj**$(\boldsymbol{\alpha})$
11: **initialize** $j \leftarrow 1$, val $\leftarrow \mathsf{BCost}(\boldsymbol{\alpha})$ by (5.9)
12: **while** $j \leq N$ **do**
13: Compute $q_j^*(p;\boldsymbol{\alpha})$ by (6.5)
14: `val` $\leftarrow$ `val` $+ \mathsf{ECost}_j\left(q_j^*(p;\boldsymbol{\alpha}), \boldsymbol{\alpha}\right)$ by (5.12)
15: $j \leftarrow j+1$
16: **end while**
17: **return** `val`
18: **end function**

Chapter 7

Impacts of Demand and Price Uncertainty

In this chapter, we study the impacts of demand and price uncertainties, to better understand the observations in Fig. 1.1(a) and 1.1(b). We will use the variance of a random variable to measure its uncertainty. 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 chapter involve a single datacenter.

7.1 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 Chapter 3, 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* (6.5)*, is non-decreasing in* σ_D^2*.*

The proof for Lemma 1 is in Appendix 13.7. Though $q_j^*(p;\boldsymbol{\alpha})$ in (6.5) cannot fully eliminate this curse, it can handle the demand uncertainty carefully such that the performance will not deteriorate too much, as illustrated in the empirical studies in Fig. 1.1(a) and Fig. 10.7. And we will provide more discussions immediately.

7.1.1 $q_j^*(p;\boldsymbol{\alpha})$ is Robust to Demand Uncertainty

In this part, we want to provide some theoretical analysis on the robustness of the optimal bidding curve towards demand uncertainty. Specially, we want to understand how the demand uncertainty will degrade the performance and how our proposed "optimal bidding curve" by (6.5) will behave when the demand uncertainty increases.

Before that, we measure the uncertainty of the stochastic demand V_j by its expected "absolute deviation" (AD), which is formally defined as

$$\text{AD} = \int_0^{\bar{v}_j} |v - \mathbb{E}[V_j]| f_{V_j(\boldsymbol{\alpha})}(v) dv.$$

A larger AD means that the real-time demand is likely to deviate more from its expectation and implies that the demand is more uncertain.

With V_j Given, the simplest bidding strategy, which we refer to as NaiveBidding, would be to submit one bid, with bidding quantity $\mathbb{E}[V_j]$ and bidding price μ_j^{RT}.[1] In this way, the bidding curve of NaiveBidding would be a stepwise function

$$\tilde{q}_j(p) = \begin{cases} \mathbb{E}[V_j], \text{if } p \leq \mu_j^{\mathsf{RT}} \\ 0, \text{otherwise.} \end{cases} \tag{7.1}$$

With $\tilde{q}_j(p)$, the cost function (5.12) can be simplified as

$$\mathbb{E}[V_j]\mu_j^{\mathsf{RT}} + \int_0^{\mu_j^{\mathsf{RT}}} f_{P_j}(p) \left[(\mu_j^{\mathsf{RT}} - \beta x)\tfrac{\mathrm{AD}}{2} - (\mu_j^{\mathsf{RT}} - x)\mathbb{E}[V_j]\right] dx.$$

It is saying that the expected cost scales linearly with AD and the performance degradation by demand uncertainty would be quite noticeable, which is validated by our simulation results in Fig. 1.1(a).

Furthermore, AD can be as large as $\mathbb{E}[V_j]$ in the worst case, and the expected cost could be further revealed as

$$\mathbb{E}[V_j]\mu_j^{\mathsf{RT}} + (1 - \frac{\beta}{2}) \int_0^{\mu_j^{\mathsf{RT}}} f_{P_j}(p) \left(x - \frac{\mu_j^{\mathsf{RT}}}{2 - \beta} \right) \mathbb{E}[V_j] dx,$$

which can be larger than $\mathbb{E}[V_j]\mu_j^{\mathsf{RT}}$ [2].

It should be noted that $\mathbb{E}[V_j]\mu_j^{\mathsf{RT}}$ is the expected cost if the datacenter does not bid in the day-ahead market but purchases all the electricity from the real-time market. In other words, *the carelessly-designed bidding strategy will incur even more cost than not bidding*, which is undesirable.

[1] The bidding price here is from [17]

[2] Just consider a simple example that the market clearing price is only distributed from $\frac{\mu_j^{\mathsf{RT}}}{2-\beta}$ to μ_j^{RT}

Next we provide Proposition 2 to show how our carefully-designed bidding curve will behave instead.

Proposition 2. *With $q_j^*(p)$ given by* (6.5)*, the value of the objective function* (5.12) *is always upper bounded by $E[V_j]\mu_j^{RT}$ for any demand distribution $f_{V_j}(v)$.*

The proof of Proposition 2 is in Appendix 13.6. Essentially it tells that, no matter how eccentric the demand is, bidding in the day-ahead market by following (6.5) will always bring benefit as compared to not bidding. So, besides minimizing the expected cost, another advantage of this bidding curve is that it performs "robustly" to future demand uncertainty. The reason is that, when we construct bidding curve by (6.5), the stochastic information of future demand is fully utilized, while for NaiveBidding, only the expectation is used.

7.2 Impact of Price Uncertainty

The price uncertainty in the day-ahead market is the fundamental reason to motivate the continuous bidding curve design and differentiates $\mathbf{EP}_j(\boldsymbol{\alpha})$ in this paper from the classic Newsvendor problem [36]. Different from demand uncertainty, uncertainty in MCP of day-ahead market allows the optimal bidding curve $q_j^*(p;\boldsymbol{\alpha})$ 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 $q_j^*(p;\boldsymbol{\alpha})$ 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 enjoys "gain". When p is large, we have to purchase expensive electricity from the day-ahead market and thus suffers "loss". However, when $p \geq \mu_j^{\mathsf{RT}}$, our optimal bidding strategy $q_j^*(p; \boldsymbol{\alpha})$ 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. 1.1(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* (6.5)*, is non-inceasing in σ_P^2.*

The proof for Lemma 2 is in Appendix 13.8. 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.

7.3 Generalizations

In this part, we generalize our results in Lemma 1 and 2 by relaxing the assumptions of specific distributions.

There are different approaches to measure and compare the uncertainties of random variables [64, 70]. We provide two metrics, "increasing convex ordering" and "variability ordering", in the following two definitions.

Definition 1. *([70, Definition 4.1]) For two random variables X and Y, $X \geq_{ic} Y$ if and only if $\mathbb{E}[f(X)] \geq \mathbb{E}[f(Y)]$ for **all** nondecreasing convex functions f.*

Definition 2. *([70, 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 $X \geq_{var} Y$, if the sign of $f - g$ changes exactly twice with sign sequence $+, -, +$.*

We remark that $X \geq_{\text{var}} Y$ implies that $X \geq_{\text{ic}} Y$, so the "variability ordering" is stronger than "increasing convex ordering". Now, we present our main results in the following two theorems, which are similar to Lemma 1 and 2.

Theorem 5. *Assume that the day-ahead MCP is positive and follows an arbitrary distribution. Consider two types of electricity demands V^1 and V^2 with $\mathbb{E}[V^1] = \mathbb{E}[V^2]$. If $V^1 \geq_{var} V^2$ or $V^1 \geq_{ic} V^2$, the optimal expected*

electricity cost by V^1, which can be achieved by using the strategy in (6.5)*, is* ***not lower*** *than that by V^2.*

Theorem 6. *Assume that the electricity demand is nonnegative and follows an arbitrary distribution. Consider two types of day-ahead MCPs P^1 and P^2 with $\mathbb{E}[P^1] = \mathbb{E}[P^2]$. If $P^1 \geq_{var} P^2$ or $P^1 \geq_{ic} P^2$, the optimal expected electricity cost incurred by P^1, which can be achieved by using the strategy in* (6.5)*, is* ***not higher*** *than that by P^2.*

Theorem 5 says that a demand with a higher uncertainty "ordering" will lead to higher cost expectation while Theorem 6 says that a price with a higher uncertainty "ordering" will lead to lower cost expectation. The proofs of these two theorems are embedded in those of Lemma 1 and 2, and are omitted. We remark that some limitations still exist, because for some random variables, we cannot compare their uncertainties by Definition 1 or 2.

□ **End of chapter.**

Chapter 8

Bidding with Finite Bids

We remark that the previously demonstrated advantages can only be realized when submitting 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 $(b^k, q^k), k = 1, \ldots, K$ can be submitted. Our arguments for this part focus on a single datacenter unless otherwise mentioned.

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 $(b^k, q^k), k = 1, \ldots, K$ 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 part is to optimize $\bar{q}(p)$, *i.e.*, the values of $b^k, q^k, \forall k$, to minimize the electricity cost expectation.

8.1 Performance Loss Characterization

Firstly, we quantize 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 by two bidding curves $q^1(p), q^2(p)$ ($q^1(p) = q^2(p) = 0$, for $p \geq \mu_j^{RT}$) as $\mathsf{ECost}_j\left(q^1(p)\right), \mathsf{ECost}_j\left(q^2(p)\right)$, respectively, we can have*

$$|\mathsf{ECost}_j\left(q^1(p)\right) - \mathsf{ECost}_j\left(q^2(p)\right)|^2 \leq \mathcal{M} \cdot \int_0^{\mu_j^{RT}} |q^1(p) - q^2(p)|^2 dp,$$

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

Essentially Lemma 3 is saying that if two bidding curves are close in terms of the distance measured by $\int_0^{\mu_j^{\mathsf{RT}}} |q^1(p) - q^1(p)|^2 dp$, their expected costs are also close, which is quite intuitive.

We denote the optimal bidding curve in (6.5) and its cost by $q^*(p)$ and C^*, respectively. Obviously C^* serves as a lower bound for $\mathsf{ECost}(\bar{q}(p))$.[1] By applying Lemma 3, we can have

$$\mathsf{ECost}_j(\bar{q}(p)) - C^* \leq \sqrt{\mathcal{M} \cdot \int_0^{\mu_j^{\mathsf{RT}}} |q^*(p) - \bar{q}(p)|^2 dp}. \tag{8.1}$$

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 how the bids are designed ($\int_0^{\mu_j^{\mathsf{RT}}} |q^*(p) - \bar{q}(p)|^2 dp$). (b) It also provides a guideline for designing a "good" step-wise bidding curve: the $\bar{q}(p)$ with a small value of $\int_0^{\mu_j^{\mathsf{RT}}} |q^*(p) - \bar{q}(p)|^2 dp$. Alternatively speaking,

[1] C^* can be viewed as the optimal value of the cost minimization problem without the "stepwise bidding curve" constrain.

we need to find a stepwise function to approximate the continuous bidding curve.

8.2 Step-wise Bidding Curve Design

To have a good step-wise bidding curve, it is natural to find a $\bar{q}(p)$ to minimize $\int_0^{\mu_j^{\mathsf{RT}}} |q^*(p) - \bar{q}(p)|^2 dp$. Without loss of generality, we assume the bidding prices are indexed increasingly with $b^k \leq b^{k+1}$ and $b^0 = 0, b^{K+1} = \mu_j^{\mathsf{RT}}$. We denote by s^k the procurement quantity from the day-ahead market when the MCP is higher than b^{k-1} but not higher than b^k, *i.e.*, $s^k = \bar{q}(p)$ for $p \in (b^{k-1}, b^k]$, and we can have

$$\begin{cases} s^k = \sum_{l=k}^{K} q^l \\ q^k = s^k - s^{k+1}. \end{cases}$$

And the problem to optimize a step-wise bidding curve (**FB**) is cast below.

$$\begin{aligned} \mathbf{FB} \quad \min \quad & \sum_{k=0}^{K} \int_{b_k}^{b^{k+1}} |q^*(p) - s^{k+1}|^2 dp && \text{(8.2a)} \\ \text{s.t.} \quad & b^k \leq b^{k+1} && \text{(8.2b)} \\ & s^{k+1} \leq s^k && \text{(8.2c)} \\ \text{var.} \quad & b^k, s^k, k = 1 \ldots, K. && \text{(8.2d)} \end{aligned}$$

It is easy to see that the above problem is non-convex and the different terms of the objective function are coupled with each other by the optimization variable b^k. So the global optimal solution of **FB** is difficult to

obtain. In the following we will present an algorithm that guarantees to converge to a local optimal solution.

8.2.1 To Optimize the Bidding Quantities

Let us firstly consider a subproblem: how to determine the values s^k of the step-wise function when the bidding prices b^k are given. By changing the optimization variable b^k to input parameter, **FB** reduces to the optimization problem of determining the optimal bidding quantities, which we denote as **Bidding-Q**.

$$\begin{aligned} \textbf{Bidding-Q} \quad \min \quad & \sum_{k=0}^{K} \int_{b_k}^{b^{k+1}} |q^*(p) - s^{k+1}|^2 dp \\ \text{s.t.} \quad & s^{k+1} \leq s^k \\ \text{var.} \quad & s^k, k = 1 \ldots, K. \end{aligned}$$

Note that the objective function of **Bidding-Q** is separable, we can firstly ignore the constraints and solve it by minimizing each term of the objection function individually. The optimal solution is given by [2]

$$\hat{s}^{k+1} = \frac{1}{b^{k+1} - b^k} \int_{b_k}^{b^{k+1}} q^*(p) dp, \forall k. \tag{8.4}$$

This result is very intuitive: the best constant to approximate a function in an interval (b^k, b^{k+1}) is the averaged value of the function in that interval. With the fact that $q^*(p)$ is a nonincreasing function, $\hat{s}^{k+1}$ automatically satisfies Constraint (8.2c) and thus, (8.4) is the optimal solution

[2]The optimal solution is the unique solution making the first-order derivative of the objective function equal to 0.

to **Bidding-Q**. In other words, given the bidding prices, the corresponding optimal bidding quantities can obtained by (8.4).

8.2.2 To Optimize the Bidding Prices

Then, we turn to consider the problem of how to set the bidding prices $b^k, \forall k$ with the bidding quantities given by (8.4), *i.e.*, solving **Bidding-P** below.

$$\begin{aligned} \textbf{Bidding-P} \quad \min \quad & \sum_{k=0}^{K} \int_{b^k}^{b^{k+1}} |q^*(p) - \hat{s}^{k+1}|^2 dp \\ \text{s.t.} \quad & b^k \leq b^{k+1} \\ \text{var.} \quad & b^k, k = 1 \ldots, K. \end{aligned}$$

As compared with **Bidding-Q**, the objective function of **Bidding-P** is not separable, for example, two terms $\int_{b^k}^{b^{k+1}} |q^*(p) - \hat{s}^{k+1}|^2 dp$ and $\int_{b^{k-1}}^{b^k} |q^*(p) - \hat{s}^{k+1}|^2 dp$ are coupled by b^k; thus, the optimization variables are also coupled with each other. Additionally, this problem is still non-convex.

To further understand the problem structure, we firstly try to characterize how to optimize b^k when $b^1, b^2, \cdots, b^{k-1}, b^{k+1}, b^{K-1}, b^K$ are given, *i.e.*, to minimize

$$\begin{aligned} & \text{Obj}(b_k) \\ = \quad & \int_{b^{k-1}}^{b^k} |q^*(p) - \hat{s}^k|^2 dp + \int_{b^k}^{b^{k+1}} |q^*(p) - \hat{s}^{k+1}|^2 dp. \end{aligned} \tag{8.6}$$

A necessary condition for the optimal solution is to satisfy the first-order

optimality condition, *i.e.*,

$$\begin{aligned} & d\text{Obj}(b^k)/db^k \\ = \quad & \left(\hat{s}^{k+1} - \hat{s}^k\right) \cdot \left(2q^*(b^k) - \hat{s}^k - \hat{s}^{k+1}\right) \\ = \quad & 0. \end{aligned}$$

It is easy to see that $\left(\hat{s}^{k+1} - \hat{s}^k\right) \leq 0$. However, the second term $\left(2q^*(b_k) - \hat{s}^k - \hat{s}^{k+1}\right)$ is not monotonic with b_k, [3] which indicates that (8.6) is nonconvex in b_k. So even minimizing only two consecutive terms with a single variable b_k is challenging. Nevertheless, we can find a solution to $d\text{Obj}(b^k)/db^k = 0$ as long as $d\text{Obj}(b^k)/db^k$ is continuous, for example, by gradient descent method.

Based on the above understandings, we propose a heuristic algorithm to solve **Bidding-P** and **FB** iteratively. The basic idea is as follows. In each round, we firstly fix b^0 and b^2 and find a new b^1 that improves the current solution and satisfies $d\text{Obj}(b^1)/db^1 = 0$, then we fix b^1 and b^3 to update b^2, then fix b^3 and b^5 to update b^4, and so on. In this way, we can sequentially update the variables from b^1 to b^K. It is worth emphasizing that when b^{k-1}, b^k, b^{k+1} satisfies the first-order condition, this condition may not hold after we optimize b^{k+1}. So, after we optimize b^K, we still can decrease the objective value of (8.2a) for all k by going through another round of optimization, starting from b^1. Because the objective value is non-increasing in each iteration and lower bounded by 0, this algorithm is guaranteed to converge. We summarize the algorithm in Alg. 2.

[3] Note that $\hat{s}^k$ and $\hat{s}^{k+1}$ are also functions of b_k.

Algorithm 2 A Heuristic Algorithm for Solving **FB**

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

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

1: **initialize** $(b^k, q^k), k = 1, \ldots, K$.
2: **while** not converge **do**
3: **for** $k = 1, \ldots, K$ **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 (8.2a).
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** $(b^k, q^k), k = 1, \ldots, K$.

Back to our joint optimization framework, we can firstly ignore the "finite-bid" constraint and adopt the "continuous-bidding-curve" solution, *i.e.*, Alg. 1 to produce the optimal, yet possibly continuous, bidding curves $q_j^*(p; \boldsymbol{\alpha}^*), \forall j$. After that, we use Alg. 2 to produce step-wise bidding curves $\bar{q}_j(p)$ to approximate $q_j^*(p; \boldsymbol{\alpha}^*), \forall j$. Obviously the objective value by $q_j^*(p; \boldsymbol{\alpha}^*)$ is a lower bound for the optimal value, and according to (8.1), the performance of $\bar{q}_j(p), \forall j$ is close to that of $q_j^*(p; \boldsymbol{\alpha}^*)$, so the objective value by $\bar{q}_j(p), \forall j$ is also close to the optimal.

□ **End of chapter.**

Chapter 9

Extensions to Other Pricing Models

In this chapter, we briefly describe how to extend our joint GLB and EP framework to other market models, which handles the real-time mismatch by different pricing mechanisms.

9.1 Real-time Pricing Model Two

We firstly consider the scenario that when the MCP is p, the real-time buying price is $(1+\epsilon_1)p$ while the real-time selling price is $(1-\epsilon_2)p$, where $\epsilon_1 \in (0,\infty), \epsilon_2 \in (0,1)$. This model is used in [55, 56, 76, 75], *etc.*

Denote the real-time mismatch by Δ. We formally describe the relationship between the day-ahead MCP P^{da} and real-time price P^{rt} in (9.1).

$$P^{\mathrm{rt}} = \begin{cases} (1+\epsilon_1)P^{\mathrm{da}}, & \text{if } \Delta > 0, \\ (1-\epsilon_2)P^{\mathrm{da}}, & \text{if } \Delta < 0. \end{cases} \tag{9.1}$$

This pricing mechanism also incentives the customers to make accurate prediction of their future demand and purchase all electricity they need in the day-ahead markets, since both the over-supply and under-supply will introduce additional cost. We denote the electricity consumption of a particular future hour as a random variable V and the submitted bidding curve as $q_j(p)$; the expected electricity cost is expressed in (9.2).

$$\mathsf{ECost1}_j(q_j(p), \boldsymbol{\alpha}) = \int_0^{+\infty} \left[pq_j(p) + (1+\epsilon_1)p\mathbb{E}\left[(V_j - q_j(p))^+\right] - (1-\epsilon_2)p\mathbb{E}\left[(q_j(p) - V_j)^+\right]\right] f_{P_j}(p)dp. \tag{9.2}$$

9.1.1 Single Datacenter Case

Similar to our previous solution, we first consider the subproblem of how to purchase electricity for one single datacenter, *i.e.*, solving the following problem,

$$\textbf{EP1} \quad \min \quad \mathsf{ECost1}_j(q(p), \boldsymbol{\alpha}) \tag{9.3a}$$

$$\text{s.t.} \quad q(p) \in \mathcal{Q}. \tag{9.3b}$$

We provide the optimal solution of Problem **EP1** in Lemma 4

Lemma 4. *The optimal bidding curve of **EP1** is given by*

$$q1_j^*(p) = F_{V_j}^{-1}\left(\frac{\epsilon_1}{\epsilon_1 + \epsilon_2}\right) \tag{9.4}$$

The proof of Lemma 4 is in Appendix 13.12. We remark that under this pricing model, the optimal bidding curve is a constant for any realization of MCP, which means that we can realize such a bidding curve by submitting one bid with an extremely high bidding price, to ensure that we can successfully buy $F_V^{-1}\left(\frac{\epsilon_1}{\epsilon_1+\epsilon_2}\right)$ amount of electricity. As an example, if $\epsilon_1 = \epsilon_2$, the amount of electricity should be purchased is the median of the electricity demand V. If we have the finite-bid constraint, submitting one bid will be sufficient to realize this optimal bidding curve.

9.1.2 Multiple Datacenter Case

We follow the similar approach to solve the problem involving multiple datacenters, based on our results on the single datacenter scenario. Suppose our GLB decision is $\boldsymbol{\alpha}$, and we denote the optimal electricity cost of datacenter j with $\boldsymbol{\alpha}$ as $\mathsf{ECost1}_j(q1_j^*(p), \boldsymbol{\alpha})$, which can be computed by substituting $q(p)$ in (9.2) by (9.4). The optimal geographic load balancing strategy can thus be obtained by solving the following problem **GLB1**,

$$\textbf{GLB1} \quad \min \quad \sum_{j=1}^{N} \mathsf{ECost1}_j(q1_j^*(p), \boldsymbol{\alpha}) + \mathsf{BCost}(\boldsymbol{\alpha}) \tag{9.5a}$$

$$\text{s.t} \quad \boldsymbol{\alpha} \in \mathcal{A}. \tag{9.5b}$$

Even though **GLB1** has not closed-form objective function, it is a convex optimization problem and can be optimally solved by any algorithm which guarantees at least a local optimal solution, like General Pattern Search [43]. We formally establish this property of **GLB1** in Theorem 7.

Theorem 7. ***GLB1*** *is a convex optimization problem. Provided that the objective function of* ***GLB1*** *is continuously differentiable, General Pattern Search algorithm will converge to its global optimal solution.*

The proof of this theorem exactly follows the logic in the proof of Theorem 1 and is omitted.

9.2 Real-time Pricing Model Three

Next we consider another pricing model, according to which the real-time price is jointly determined by the day-ahead MCP and the total mismatch

(between day-ahead electricity procurement and real-time demand) of all participants in the markets. This model is used to evaluate the value of flexibility for electricity market in [54] and to analyze the impact of renewable penetration for microgrid in [86].

Mathematically, if the day-ahead MCP is P^{da}, then the real-time price is given by $P^{\mathrm{rt}} = P^{\mathrm{da}} + a\sum_i \Delta_i + \epsilon$, where Δ_i is the mismatch by the i^{th} market participant, and ϵ is noise, capturing the factors we ignored. For the purpose of simplicity, we assume that $\epsilon, \Delta_i, \forall i$ are zero-mean and mutually independent random variables. Also, we assume that the datacenter owner cannot impact or predict the consequence of other participants' behaviour, *i.e.*, the datacenter has no incentive or capability to arbitrage the markets, then for one participant, the real-time price can be characterized by (9.6).

$$P_j^{\mathrm{rt}} = P_j^{\mathrm{da}} + a\Delta_j + \epsilon_j. \tag{9.6}$$

On one hand, when $\Delta_j > 0$, meaning that real-time demand is higher than day-ahead procurement and we need to **buy** additional electricity at higher price (the real-time price is higher than the day-ahead MCP statistically); on the other hand, when $\Delta_j < 0$, meaning that real-time demand is lower than day-ahead procurement and we need to **sell** additional electricity at lower price (the real-time price is lower than the day-ahead MCP statistically). This pricing model will transfer the real-time mismatch into economic loss and incentive the customer to plan its demand in day-ahead markets.

According to the pricing model by (9.6), the expected cost by submit-

$$\mathsf{ECost2}(q_j(p), \boldsymbol{\alpha}) = \int_0^{+\infty} \left[pq_j(p) + \int_0^{\bar{V}} (v - q_j(p)) \left(p + a(v - q_j(p)) + \epsilon_j\right) f_{V_j}(v) dv \right] f_{P_j}(p) dp \tag{9.7}$$

---

ting a bidding curve $q_j(p)$ can be expressed in (9.7).

9.2.1 Single Datacenter Case

The optimal electricity procurement (bidding) strategy can be obtained by solving **EP2**, shown below.

$$\begin{aligned} \textbf{EP2} \quad \min \quad & \mathrm{ECost2}(q_j(p), \boldsymbol{\alpha}) && \text{(9.8a)} \\ \text{s.t.} \quad & q_j(p) \in \mathcal{Q}. && \text{(9.8b)} \end{aligned}$$

And we directly present the optimal solution in Lemma 5.

Lemma 5. *The optimal solution of **EP2** is* $q2_j^*(p) = \mathbb{E}[V_j], \forall p$, *and the corresponding optimal cost is*

$$\mathbb{E}[P_j]\,\mathbb{E}[V_j] + a\,Var(V_j).$$

The proof for Lemma 5 exactly follows the logic of those for Theorem 2 and Lemma 4 and omitted.

Remarks: Under this pricing model, the bidding curve is a constant for any MCP p, which means that we can realize this bidding curve by submitting one bid with a bidding quantity $\mathbb{E}[V]$ and an extremely high bidding price, so that the bid will succeed for any realization of MCP. Besides, the

expected cost under the optimal bidding strategy is determined both by the demand expectation and its variance. Under this model, the intuition that a larger demand variance will lead to larger real-time mismatch is more clear than the results in Chapter 7.1.

9.2.2 Multiple Datacenter Case

Now we would like to proceed with the scenario with N datacenters. With workload allocation decision α, we denote the expected electricity cost of datacenter j with the optimal bidding strategy by $\mathsf{ECost2}_j(\boldsymbol{\alpha})$ and the bandwidth cost by $\mathsf{BCost}(\boldsymbol{\alpha})$. The optimal workload allocation strategy can be obtained by solving the following problem **GLB2**.

$$\textbf{GLB2} \quad \min \quad \sum_{j=1}^{N} \mathsf{ECost2}_j(q2_j^*(p), \boldsymbol{\alpha}) + \mathsf{BCost}(\boldsymbol{\alpha}) \tag{9.9a}$$

$$\text{s.t.} \quad \boldsymbol{\alpha} \in \mathcal{A}. \tag{9.9b}$$

By assuming that the original demand from each location $U_i, \forall i$ are mutually independent, the electricity cost expectation can be expressed more explicitly, in the following,

$$\begin{aligned} &\sum_{j=1}^{N} \mathsf{ECost2}_j(q2_j^*(p), \boldsymbol{\alpha}) \\ =&\sum_{j=1}^{N} \left[\sum_{i=1}^{N} \boldsymbol{\alpha}_{i,j} \mathbb{E}[U_i] + a\boldsymbol{\alpha}_{i,j}^2 \mathrm{Var}[U_i] \right], \end{aligned}$$

which is a quadratic function of $\boldsymbol{\alpha}$. With the fact that the other term $\mathsf{BCost}(\boldsymbol{\alpha})$ is linear in $\boldsymbol{\alpha}$, we can conclude that **GLB2** is a convex problem

and can be optimally solved by standard solvers, like [31].

Remarks: Under this model, the optimal workload allocation and bidding strategies only depends on the expectation and variance of future demands, which is easier to get than their exact probability distributions.

□ **End of chapter.**

Chapter 10

Empirical Evaluations

In this chapter, we use trace-driven simulations to evaluate the performance of the joint GLB and EP framework modelled in Chapter 5 and our algorithm designed in Chapter 6.

10.1 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 York City. We set the unit bandwidth cost of routing workloads across datacenters as $z_{ij} = \kappa \cdot \left(\mu_1^{\mathsf{RT}} + \mu_2^{\mathsf{RT}} + \mu_3^{\mathsf{RT}}\right)/3$ 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 [8]. By using the conversion ratio claimed by Google for its datacenters [60], we scale up the request information to create an electricity demand series with averaged hourly

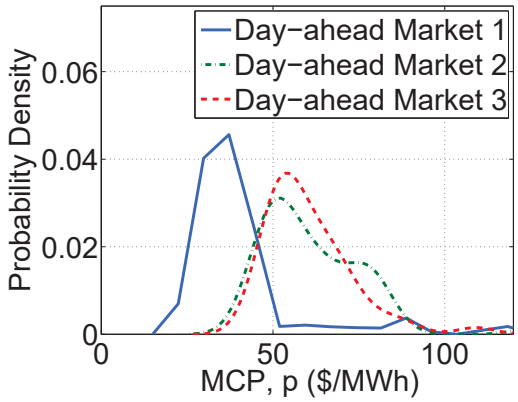


Figure 10.1: Empirical distributions of MCPs, 2pm.

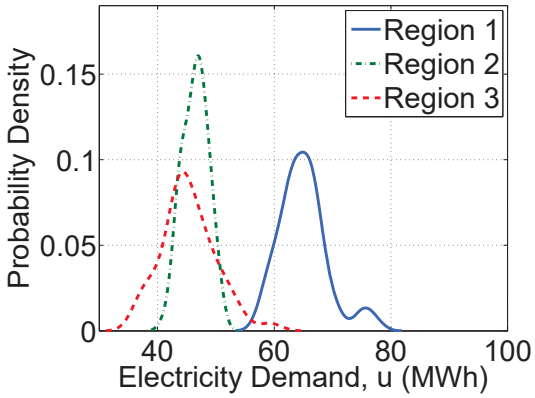


Figure 10.2: Empirical distributions of electricity demands, 2pm.

demand of 125 MWh. The total demand is divided into three regions according to regional electricity consumptions of the three locations (a detailed description is in Chapter 4.2.1). 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 [17, 3, 25].

Electricity Prices in Day-ahead and Real-time Markets. We obtain the electricity prices (MCP of day-ahead market and real-time market price) from three regional ISO websites which serve the customers in San Diego, Houston, and New York, respectively [16] [24] [4]. 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.

We provide the brief statistics of demand and price traces in Table 10.1.

Evaluation and Comparison. We test our design on 24 instances,

Table 10.1: Hourly Electricity Demand and Price Statistics in the Experiments

	Region	Mean	STD
DAM Prices ($)	1	42.7	71.7
	2	55.6	63.3
	3	56.4	17.1
RT Prices ($)	1	42.3	71.5
	2	56.4	64.6
	3	57.4	16.1
Electricity Demands (MWh)	1	52.8	19.2
	2	39.1	12.2
	3	36.4	13.4

each corresponding to one hour of the day. For each hour, the distributions of electricity demand, day-ahead MCP 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. 10.1 and Fig. 10.2, respectively. We denote our solution as OptBidding-OptGLB, in which the GPS part is based on an implementation used in [20, 21]. 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 opti-

Table 10.2: Cost-saving performance of different schemes.

Solution	Daily Cost (k$)	Reduction (%)
NoBidding-NoGLB	161.9	-
NoBidding-OptGLB (adapted from [67])	154.5	4.6
SimpleBidding-OptGLB [17]	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**

mal GLB (adapted from the solution in [67]). (iv) SimpleBidding-OptGLB: it represents a joint bidding and GLB strategy proposed in [17], in which the CSP only 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*.

10.2 Experimental Results

10.2.1 Performance Comparison and Impact of Finite Bids

We compare the performance of different solutions in terms of the expected daily cost in Table 10.2. 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 Table 10.2, respectively.

We have the following observations. First of all, as seen from Table 10.2, we can see that our proposed solution outperforms all 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 Chapter 8 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 Table 10.2, 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. 10.3. 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.

10.2.2 Impact of Market Price Uncertainty and Demand Uncertainty

In Chapter 1, we provide two experiments related to the electricity demand variability and market price variability (Fig. 1.1(a) and Fig. 1.1(b)) to motivate our study and we describe the details here. Our Solution denotes the strategy by 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. 1.1(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%. It means that even though the demand uncertainty curses the performance of both two schemes, our solution behaves more robustly. In Fig. 1.1(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 Chapter 7.2. Because when the market price uncertainty is larger, it is more likely that we can buy cheaper electricity from the day-ahead market while the performance loss due to higher price is always capped by μ_j^{RT}.

10.2.3 Convergence Rate of the Joint Bidding and GLB Algorithm

In this part, we empirically evaluate the convergence rate of our proposed Algorithm 1. We run our algorithm for two instances with workload/price distribution of 10am and 2pm, respectively. From Fig. 10.4, we can see that our algorithm converges rather fast – within 30 iterations – for the practical setting considered. The computation 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 $\boldsymbol{\alpha}$).

In Fig. 10.5, we also report the accumulative statistics of the convergence information for all the 24-hour instances. As we can see, in more than 80% of testing instances, the algorithm will achieve 99.5% optimality within 20 iterations and 99.9% optimality within 40 iterations.

Many modern gradient-based numerical solvers are advanced in the sense that it can estimate the gradient information if not given directly, like the `fmincon` in Matlab [51], which can also be used to solve **P3**. Even though the gradient information provides a searching direction to decrease the objective, estimating such information is also computationally expensive. We show the performance of the solutions produced by Algorithm 1 and `fmincon` in Fig. 10.6. As we can observe, the two algorithms produce solutions with the similar objective values, which may not be too surprising since **P3** is convex. However, in terms of running time to produce such solutions, Algorithm 1 (GPS Algorithm) is 2-9 times fasters than `fmincon`.

10.2.4 Impact of Demand Uncertainty and Distribution Estimation

To study the impact of demand uncertainty, 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 workloaddemand. Here normalized sample standard deviation is

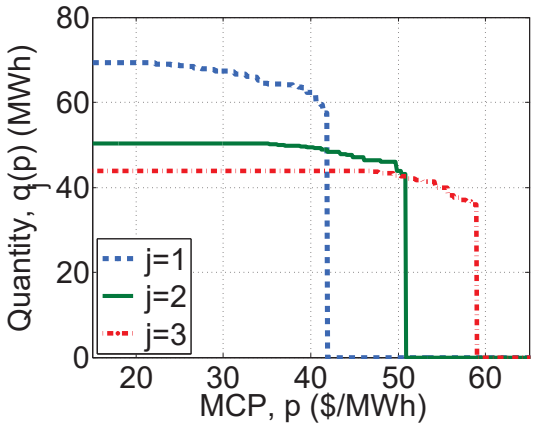


Figure 10.3: Optimal bidding curves for three day-ahead markets, 4pm.

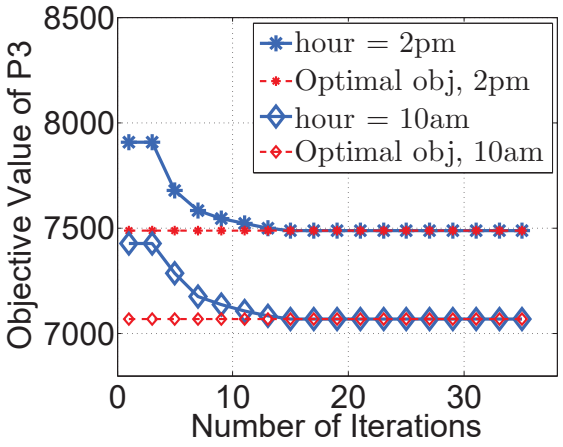


Figure 10.4: Objective values in each iteration of our Algorithm 1.

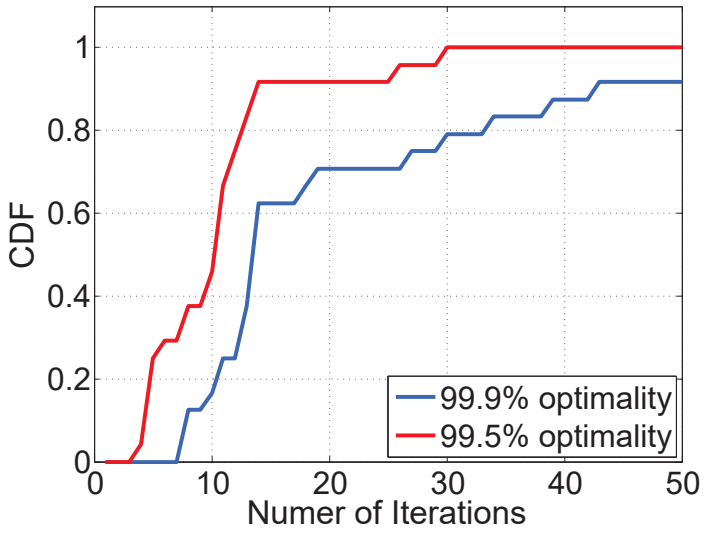


Figure 10.5: Statistics of convergence information for 24 hours

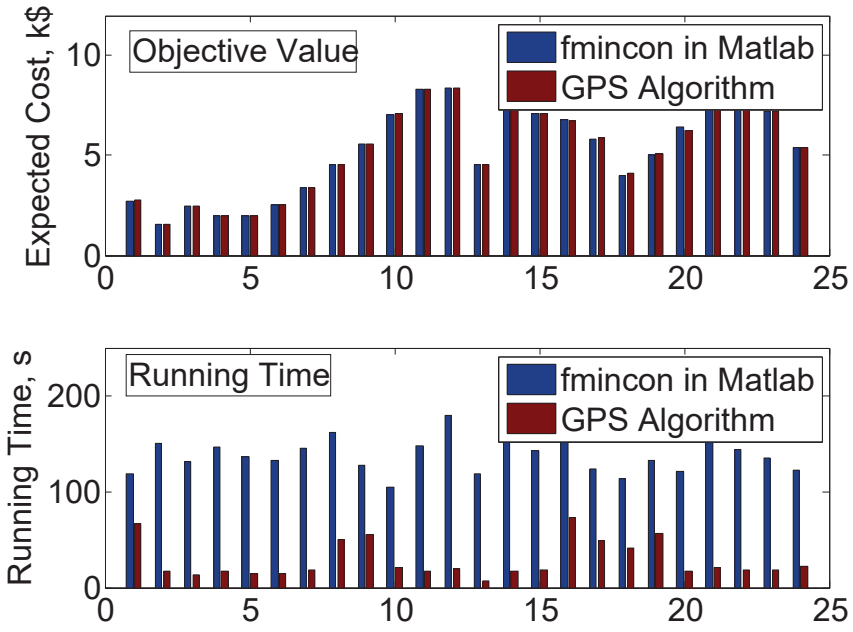


Figure 10.6: Comparisons with gradient-based algorithm

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. 10.7. From Fig. 10.7, we can see that the cost reduction decreases as the demand certainty increases, but the performance loss is minor, suggesting that our solution OptBidding-OptGLB is robust to demand uncertainty.

We also study the impact of distribution estimation. In our solution OptBidding-OptGLB, we use the distribution of the demand U_j for region j as input. In practice, however, the CSP may not have the exact demand distributions, but just their estimates 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. 10.7. 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.

10.2.5 Impact of Market Price Uncertainty and Distribution Estimation

To study the impact of Market Price uncertainty, we use a similar way to manipulate the MCP such that the average of the normalized sample standard deviations changes from 0.3 to 1.08, to mimic low to high uncertainty in market price uncertainty. We apply our solution OptBidding-OptGLB to the set of scaled MCPs and plot the cost reductions in Fig. 10.8.

We also study the impact of distribution estimation. In our solution OptBidding-OptGLB, we use the distribution of the demand P_j for region

j as input. In practice, however, the CSP may not have the exact MCP distributions, but just their estimates 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. 10.8. As seen, the cost reductions of three schemes increase as the market price uncertainty increases, the underlying reason is explained in Chapter10.2.2 and Chapter 6.2. Moreover, the cost reduction due to the distribution estimation error is minor.

We want to remark that the difference of the price profiles in our dataset is significant and it is easy to recognize which market is more economic (The inaccuracy of the MCP distribution will result in no uneconomic decision as long as the inaccuracy is not huge). And, the bidding strategy for each datacenter is not affected by the MCP distributions.

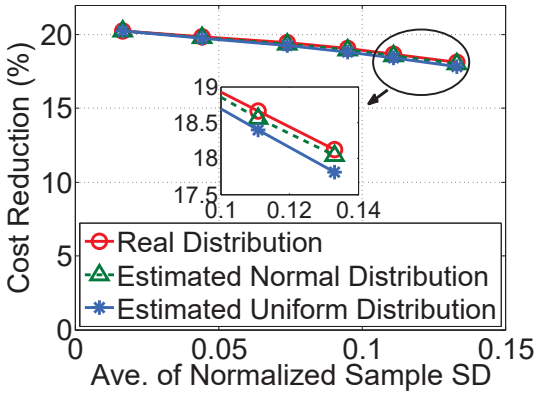


Figure 10.7: Cost reductions with different levels of demand uncertainty and different estimated distributions.

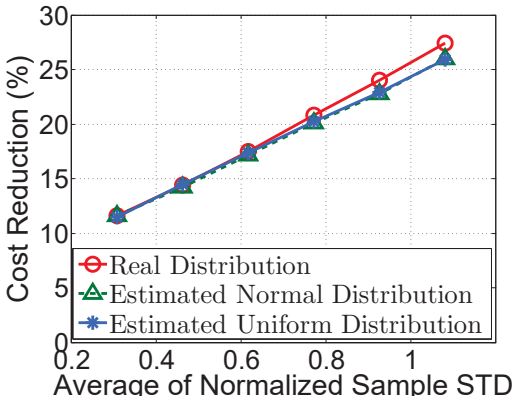


Figure 10.8: Cost reductions with different levels of price uncertainty.

10.2.6 Impact of Local Service Requirement

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. 10.9, 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. 10.9. As seen, a larger κ, meaning higher bandwidth cost, leads to smaller reduction, which matches out intuition.

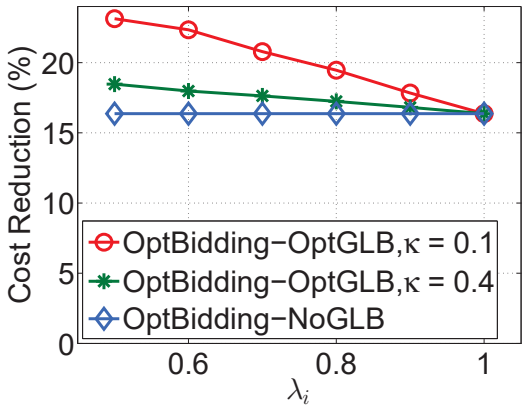


Figure 10.9: Cost reductions when more workloads must be locally served, under different bandwidth cost.

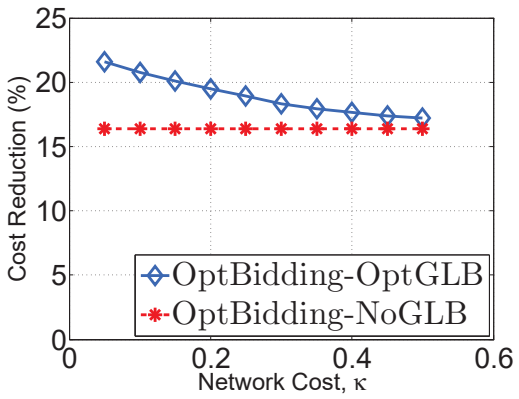


Figure 10.10: Cost reduction ratios with different levels of network cost

10.2.7 Impact of Bandwidth Cost

Moving more workload to the datacenters with lower electricity price will cut the electricity bills, but incur more internet cost. If the internet cost is too large, GLB may not be so economic, which motivates our evaluations in this part. We test the cost reduction by GLB with different network cost by increasing κ from 0.05 to 0.5 and the result is shown in Fig. 10.10.

As we can see, higher network cost will lead to smaller cost reductions; but even when σ is 0.5, the cost reduction by OptBidding-OptGLB is still over 17%, which means the design space of broker-assisted GLB is still much rewarding to exploit.

10.3 Reflections on Experimental Results

In this part, we make some reflections on the previous experimental results to better understand the source of the economic gains and the impact of simulation trace properties. From Table 10.2, we can observe that the cost reduction of our joint optimization framework (OptBidding-OptGLB) is roughly equal to the summation of those of optimizing EP and GLB independently (OptBidding-NoGLB and NoBidding-OptGLB), which are two sources of economic gains.

We firstly make some discussions on EP. According to our statistics in Table 10.1, the day-ahead MCP expectation is almost equal to the real-time price expectation for each market, but the cost reduction by only optimizing EP is over 16%. So, the benefit from joining the day-ahead market is not because the day-ahead market can provide electricity that is always cheaper than real-time market, but because it provides a chance of obtaining cheaper electricity, which comes from the multiplicity between the day-ahead MCP and the real-time price. However, a simple bidding strategy (SimpleBidding-OptGLB) can only reduce the cost expectation by no more than 4%, which means that to fully exploit this chance is non-trivial. We remark that, in our simulation, mutual independence between regional day-ahead MCP and real-time price is assumed, but in practice, positive correlation between them could exist [28], which will degrade the performance. Also, as shown in Chapter 7, the trace with higher price (day-ahead MCP) uncertainty will lead to higher economic gain and the

trace with higher demand uncertainty lower economic gain.

As for GLB, its economic gain comes from the regional "price" diversity and could be affected by many factors, including bandwidth cost, data-center capacities, *etc.* Note that the "price" here means not the price expectation, but the averaged buying cost, which is jointly determined by the market conditions and the demand statistical characteristics. We remark that the design space of GLB is not only seeking electricity with lower price, but also actively manipulating the workload so that it can be satisfied more economically. In our simulation, we assume that the original demands in all locations are mutually independent, but in practice, both positive and negative correlations could exist. Since the demand after GLB is a linear combination of the original demands with positive coefficients, the trace with negative correlation between original demands will lead to higher economic gain which the trace with positive correlation lower economic gain.

□ **End of chapter.**

Chapter 11

An Alternative Formulation

As mentioned in Chapter 5.4, there exists another natural formulation, which adopts the two-stage optimization framework [68]. In this formulation, we can defer the optimization of GLB strategy to the second stage (real-time), at which time we know the exact information of day-ahead MCPs and demands. We briefly discuss this two-stage formulation in this chapter.

11.1 Problem Formulation

In the new formulation, the optimization variables for EP and GLB essentially span two stages, day-ahead and real-time, respectively. To be consistent, we still use the bidding curves for datacenter j $q_j(p)$ to denote the *day-ahead* EP strategy and the matrix $\boldsymbol{\alpha}$ to denote the *real-time* GLB strategy.

In the second stage (real-time), we know the demand u_i from location i, the day-ahead MCP p_j for datacenter j and also the corresponding electricity procurement amount $q_j = q_j(p_j)$. We only need to determine how to route the workload and the mismatch will be automatically balanced by the market. Our objective is to minimize the summation of bandwidth

cost and electricity cost in real-time markets. The optimization problem we need to solve is as follows,

$$\begin{aligned} \textbf{S2:} \quad \min \quad & \sum_{j=1}^{N} \mathsf{ecost}_j\left(\boldsymbol{\alpha}\right) + \mathsf{bcost}(\boldsymbol{\alpha}) \\ \text{var.} \quad & \boldsymbol{\alpha} \in \mathcal{A}_{\mathsf{u}}, \end{aligned}$$

where $\mathsf{ecost}_j\left(\boldsymbol{\alpha}\right) = \mu_j^{\mathsf{RT}}(v_j - q_j)^+ - \beta p_j(q_j - v_j)^+$ with $v_j = \sum_i \alpha_{ij} u_i$, and $\mathsf{bcost}(\boldsymbol{\alpha}) = \sum_{i=1}^{N}\sum_{j=1}^{N} z_{ij}\alpha_{ij}u_i$ denote the electricity cost and bandwidth cost, respectively. The feasible region $\mathcal{A}_{\mathsf{u}}$ is an analogy of $\mathcal{A}$, but is imposed by the exact realization of demand $D_i, \forall i$. [1]

It is easy to see that the optimal solution and objective value of **S2** is determined by the EP strategy $q_j(p), \forall j$ in the first stage (day-ahead). We denote the optimal value of Problem **S2** by $CS2\left(\left[q_j(p)\right]_{j=1:N}\right)$, which is a random variable due to the randomness of $U_i, P_i, \forall i$. When we submit bidding curves in the day-ahead markets, our objective is to minimize the total cost expectation, and the optimization problem is as follows,

$$\begin{aligned} \textbf{S1:} \quad \min \quad & \sum_{j=1}^{N} \mathbb{E}_{P_j}\left[P_j q_j(P_j)\right] + \mathbb{E}_{P_j,V_i,i,j=1:N}\left[CS2\left(\left[q_j(p)\right]_{j=1:N}\right)\right] \\ \text{var.} \quad & q_j(p) \in \mathcal{Q}, j = 1, \ldots, N, \end{aligned}$$

where $\mathbb{E}_{\mathbf{R}}\left[\cdot\right]$ is the expectation taken by the joint distribution of $\mathbf{R}$.

11.2 Problem Properties and Challenges

In this part, we reveal some structures of Problem **S1** and **S2**.

[1] The main difference between $\mathcal{A}_{\mathsf{u}}$ and $\mathcal{A}$ comes from the capacity constraint (5.3) and $\mathcal{A}_{\mathsf{u}} \subseteq \mathcal{A}, \forall \mathsf{u}$.

We firstly provide the following proposition to connect **P1** and **S1**.

Proposition 3. *The optimal value of **S1** is a lower bound for that of **P1**.*

Proposition 3 holds directly by the fact that, given any feasible solution $(\hat{\boldsymbol{\alpha}}, \hat{q}_j(p), \forall j)$ of **P1**, $(\hat{q}_j(p), \forall j)$ is feasible to **S1** and $\hat{\boldsymbol{\alpha}}$ is feasible to **S2**.

In our following discussions, we restrict our attention to the cases that the bidding curves satisfy

$$q_j(p) = 0 \text{ for } p \geq \mu_j^{\mathsf{RT}}, \forall p. \tag{11.1}$$

We claim that we will lose no optimality by this restriction. The intuition is very clear and similar to that of Theorem 1: since we can obtain the electricity in real-time at price μ_j^{RT}, there is no need to buy more expensive electricity in day-ahead market with the risk of additional mismatch cost. We also make the intuition rigorous in the following proposition.

Proposition 4. *There is an optimal solution of **S1** that satisfies* (11.1).

The proof is deferred to Chapter 13.13.

Under the condition of (11.1), it is easy to see that Problem **S2** is to minimize a convex polyhedral with some linear constraints and can be solved by linear programming. We provide a property of **S1**, which is not so obvious, in the following lemma.

Lemma 6. *Under Condition* (11.1)*, Problem **S1** is convex.*

The proof is deferred to Chapter 13.14.

Even though **S1** is convex, several obstacles exist to make the problem challenging, which we list below.

(**C1**) The optimization variable of **S1** are functional, so its dimensionality is infinite and the off-the-shelf numerical solvers are not applicable.

(**C2**) The objective function is an expectation taken by the distributions of several random variables $V_i, P_i, \forall i$. To compute the objective value for each bidding curve design, we need to evaluate the real-time cost $CS2\left(\cdot\right)$ for each possible realization of $U_i, P_i, \forall i$, the number of which could be exponential. [2] So, it would be computationally intensive to only **evaluate** the objective value of **S1**.

(**C3**) $CS2\left(\left[q_j(p)\right]_{j=1:N}\right)$ in the objective function involves another optimization problem. So we cannot have the closed form or derivative information of the objective function.

Several simple heuristics to handle these challenges are suggested in the following.

- To handle Challenge **C1**, for each bidding curve, we can fix several bidding prices and optimize the corresponding bidding quantities. Then, the optimization of a continuous function is transformed to the optimization of a vector. This approach is also adopted in [46, 33].
- To handle Challenge **C2**, we can construct a fixed number of representative scenarios from the dataset using Monte Carlo Method [29, 65] or clustering algorithms [39].

[2] Consider a simple scenario in which we have 3 datacenters and markets ($N = 3$) and each random variable U_i or P_i has 10 different realizations. By assuming mutual independence among the 6 random variables $U_i, P_i, i = 1, 2, 3$, we could have 10^6 possible instances, which means that we need to solve 10^6 optimization problems.

- To handle Challenge **C3**, we can apply GPS algorithm to find a local optima of **S1**, or we can explicitly plug **S2** into **S1** with duplicated variables and nonanticipativity constraint, see Chapter 2.4 of [68] for details.

□ **End of chapter.**

Chapter 12

Conclusion and Future Work

In this thesis, we consider the problem of how a CSP jointly does load balancing and electricity procurement for its geographically located datacenters, with stochastic electricity demand and price information. We show that the joint optimization framework is necessary to realize the full potential of GLB, as a separate solution may increase the demand uncertainty and make electricity supply chains in all regions less efficient. This problem is formulated as a challenging nonconvex optimization problem. And we solve this problem optimally by carefully studying its structure. As part of the solution, we use "bidding curve" to characterize the optimal bidding strategy. By fully utilizing the stochastic information, the optimal bidding curve not only minimizes the cost expectation, but also is shown to be robust to demand uncertainty. The merit of our design was extensively shown by empirical simulations. We believe that this work serves an important guideline for the CSPs to participate in the wholesale electricity market in different locations and allocate their demands geographically.

The current study relies on the distribution of price and demand. It would also be interesting to extend the study to the scenario where we only have first and second moment statistics. Also, we currently assume that the workloads in different locations and prices in day-ahead and real-

time markets are mutually independent, but it is reasonable to believe that the users' activities in different locations are correlated with each other. It deserves effort to study how the correlation can bring additional benefit, for example, how GLB can utilize this correlation to stabilize the demands. Lastly, if the percentage of datacenters' energy consumption increases further, like going beyond 10% of total electricity consumption, how CSP (this new type of customers being able to move their demands geographically) will impact on the electricity supply chain or whether the current market mechanism should be redesigned to improve its efficiency, these topics are also interesting to explore.

□ **End of chapter.**

Chapter 13

Appendix

13.1 Proof of Proposition 1

Proof. We note that (5.12) is an integral over p. A naive but critical observation is that the function inside the integral is separable over p.

We write the inside function (excluding the constants $f_{P_j}(p)$) as follows,

$$
\begin{aligned}
& C(q_j(p)) \\
=& pq_j(p) - \beta p \int_0^{q_j(p)} (q_j(p) - v) f_{V_j}(v) dv + \mu_j^{\mathsf{RT}} \int_{q_j(p)}^{\bar{v}_j} (v - q_j(p)) f_{V_j}(v) dv \\
=& pq_j(p) - \beta p \int_0^{q_j(p)} (q_j(p) - v) f_{V_j}(v) dv + \mu_j^{\mathsf{RT}} \int_0^{q_j(p)} (q_j(p) - v) f_{V_j}(v) dv + \\
& \mu_j^{\mathsf{RT}} \mathbb{E}[V_j] - \mu_j^{\mathsf{RT}} q_j(p) \\
=& \mu_j^{\mathsf{RT}} \mathbb{E}[V_j] + (p - \mu_j^{\mathsf{RT}}) q_j(p) + (\mu_j^{\mathsf{RT}} - \beta p) \int_0^{q_j(p)} (q_j(p) - v) f_{V_j}(v) dv
\end{aligned}
$$

Then the derivative of $C(q_j(p))$ with respect to $q_j(p)$ is as follows,

$$
\frac{dC(q_j(p))}{q_j(p)} = (p - \mu_j^{\mathsf{RT}}) + (\mu_j^{\mathsf{RT}} - \beta p) \int_0^{q_j(p)} f_{V_j}(v) dv. \tag{13.1}
$$

And its second derivative is

$$
(\mu_j^{\mathsf{RT}} - \beta p) f_{V_j}(q_j(p)).
$$

It is obvious that the second derivative is not always non-negative, for example, when $p > \frac{\mu_j^{\mathsf{RT}}}{\beta}$. But this proof also indicates that the objective function is convex in the set

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

Thus the subprolem $\mathbf{EP}_j(\boldsymbol{\alpha})$ solved in Chapter 6.2 is convex.

The proof is completed. □

13.2 Proof of Theorem 1

Proof. To prove Theorem 1, we firstly provide Proposition 5 and 6 to aid our analysis.

The discussions in Proposition 5 and 6 only involve one datacenter, so we hide the GLB decision $\boldsymbol{\alpha}$ and abuse the notations a little bit to lighten the formula. We will denote $\mathsf{Cost}_j(q(p), f_V(v))$ as the electricity cost of datacenter j when its demand follows $f_V(v)$ and it submits a bidding curve $q(p)$.

Proposition 5. *Given two feasible* [1] *demands* $\tilde{V}$ *and* V *with* $\tilde{V} = \delta V$*, where* $\delta \in (0, 1)$ *is a constant, we can have*

$$\mathsf{Cost}_j(\delta q(p), f_{\tilde{V}}(v)) = \delta \mathsf{Cost}_j(q(p), f_V(v)), \tag{13.2}$$

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

[1] Demand V is feasible means that the maximum value of V is less than or equal to the datacenter's capacity.

Proposition 6. *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 can have*

$$\begin{aligned} &\mathsf{Cost}_j(q^1(p) + q^2(p), f_{V^1+V^2}(v)) \leq \\ &\mathsf{Cost}_j(q^1(p), f_{V^1}(v)) + \mathsf{Cost}_j(q^2(p), f_{V^2}(v)). \end{aligned} \tag{13.3}$$

Besides the technical proofs of Proposition 5 and Proposition 6 (Appendix 13.10 and Appendix 13.11), here we try to explain their implications. The implication of Proposition 5 is very clear: if we scale the bidding curve and electricity demand by the same factor, the electricity cost expectation will also scale accordingly. As for Proposition 6, imagine we have two datacenters in one location (the two datacenters are served by the same market). The bidding curves and demand distributions of the two datacenters are $q^1(p), f_{V^1}(v)$ and $q^2(x), f_{V^2}(v)$ respectively; $f_{V^1+V^2}(v)$ is the probability distribution of the demand summation. The right-hand side of the inequality (13.3) is the sum of the two datacenters' cost, while the left-hand side can be viewed as the cost of the datacenters if they can share their electricity procurements and demands. It means that as long as their bids satisfy $q(p) = 0$ for $p > \mu_j^{\mathsf{RT}}$, *cooperation between the datacenters will help to reduce cost.* The fundamental reason can be explained as follows. Remember that the datacenter will suffer more cost due to mismatch (discounted price to sell back for over-supply or more expensive electricity for under-supply); in case of that both datacenters meet over-supply or under-supply, there is no difference, but in case of that one datacenter meets under-supply while the other one meets over-supply, the cooperation

between them will remove part of the mismatch and thus decrease the cost, which is also quite intuitive.

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

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

Let $V^1 = (\alpha^1 \mathbf{D})_i$, $V^2 = (\alpha^2 \mathbf{D})_i$ and $\alpha = \delta\alpha^1 + (1-\delta)\alpha^2$, we have $V = (\alpha D)_i = \delta V^1 + (1-\delta)V^2$. If the distributions for V^1 and V^2 are $f^1(y)$ and $f^2(y)$, the distribution for Y is 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 $(1-\delta)V^2$. Then,

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

(E_a) and (E_b) are established by Proposition 5 and Proposition 6, respectively.

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

13.3 Proof of Theorem 2

Proof. To solve $\mathbf{EP}_j(\boldsymbol{\alpha})$, 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 $\mathbf{EP}_j(\boldsymbol{\alpha})$. We minimize the objective of unconstraint $\mathbf{EP}_j(\boldsymbol{\alpha})$ by minimizing the function value inside the integral for each p.

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

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

We discuss the form of the optimal solution as follows,

- If $p \leq \mu_j^{\mathrm{RT}}$, $\mu_j^{\mathrm{RT}} \geq \beta p$ and $\frac{dc(q)}{dq}$ increases with q. The optimal solution can be obtained by solving $\frac{dc(q)}{dq} = 0$ and the solution is $q_j^*(p) = F_{V_j}^{-1}\left(\frac{\mu_j^{\mathrm{RT}} - p}{\mu_j^{\mathrm{RT}} - \beta p}\right)$.
- If $p \in \left(\mu_j^{\mathrm{RT}}, \mu_j^{\mathrm{RT}}/\beta\right)$, $p - \mu_j^{\mathrm{RT}} \geq 0$ and $\mu_j^{\mathrm{RT}} - \beta p \geq 0$, we have $\frac{dc(q)}{dq} \geq 0$. The optimal solution is $q_j^*(p) = 0$.

- If $p \geq \mu_j^{\mathrm{RT}}/\beta$, $\mu_j^{\mathrm{RT}} - \beta p \leq 0$ and we can observe that $\frac{dc(q)}{dq} \geq p - \mu_j^{\mathrm{RT}} + (\mu_j^{\mathrm{RT}} - \beta p) \geq 0$. Then the optimal solution is $q_j^*(p) = 0$.

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

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

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

13.4 Proof of Theorem 3

Proof. Again, to prove that **P3** is convex, we only need to prove that $\mathsf{ECost}_j\left(q_j^*(p;\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)$ is convex in $\boldsymbol{\alpha}$.

$$\begin{aligned}
&\delta\mathsf{ECost}_j\left(q_j^*(p;\boldsymbol{\alpha}^1),\boldsymbol{\alpha}^1\right) + (1-\delta)\mathsf{ECost}_j\left(q_j^*(p;\boldsymbol{\alpha}^2),\boldsymbol{\alpha}^2\right) \\
=&\mathsf{ECost}_j\left(\delta q_j^*(p;\boldsymbol{\alpha}^1),\delta\boldsymbol{\alpha}^1\right) + \mathsf{ECost}_j\left((1-\delta)q_j^*(p;\boldsymbol{\alpha}^2),(1-\delta)\boldsymbol{\alpha}^2\right), \\
&\qquad \text{by Proposition 5.} \\
\geq&\mathsf{ECost}_j\left(\delta q_j^*(p;\boldsymbol{\alpha}^1) + (1-\delta)q_j^*(p;\boldsymbol{\alpha}^2),\delta\boldsymbol{\alpha}^1 + (1-\delta)\boldsymbol{\alpha}^2\right), \\
&\qquad \text{by Proposition 6} \\
\geq&\mathsf{ECost}_j\left(q_j^*(p;\delta\boldsymbol{\alpha}^1 + (1-\delta)\boldsymbol{\alpha}^2),\delta\boldsymbol{\alpha}^1 + (1-\delta)\boldsymbol{\alpha}^2\right)
\end{aligned}$$

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

According to [43], GPS algorithm is guaranteed to converge to a solution satisfying the KKT condition (which is optimal if the problem is convex) with four hypothesises(Page 9). We examine these conditions one by one as follows.

- Hypothesis 0 is satisfied by the implementations of GPS algorithm [21, 20].
- Hypothesis 1 is saying that the matrix in the constraint is rational, which is automatically satisfied by (5.1)-(5.5).
- Hypothesis 2 can be guaranteed by the convexity of **P3**, which is proved above.
- Hypothesis 3 is guaranteed by the condition that $f_{U_j}(u)$, $j = 1, \ldots, N$, are continuously differentiable.

To prove that Hypothesis 3 holds, we only need to show that $\mathsf{ECost}_j\left(q_j^*(p; \boldsymbol{\alpha}), \boldsymbol{\alpha}\right)$ is continuously differentiable with respect to $\alpha_{ij}, \forall i, j$. A sufficient condition condition is that both $f_{V_j}(v)$ and $q_j^*(p; \boldsymbol{\alpha})$ 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 $\bar{f}_{U_{\bar{ij}}}(v)$ be the convolution of $f_{U_{kj}}(u), k \neq i$, and then

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

Note that $\bar{f}_{\bar{U}_{ij}}(v)$ 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 $q_j^*(p;\boldsymbol{\alpha})$, 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[34], $q_j^*(p;\boldsymbol{\alpha})$ is also continuously differentiable.

Thus GPS algorithm will converge to a point satisfying KKT condition, which is global optimal by the previous convexity argument. □

13.5 Proof of Theorem 4

Proof. We first describe the complexity to solve our inner-loop problem $\mathbf{EP}_j(\boldsymbol{\alpha})$, *i.e.*, compute the optimal datacenter j's bidding curve $q_j^*(p;\boldsymbol{\alpha})$ through (6.5) when the GLB decision is given by $\boldsymbol{\alpha}$. We need five steps to obtain $q_j^*(p;\boldsymbol{\alpha})$. (i) We obtain the PDF of U_{ij}, *i.e.*, $f_{U_{ij}}(v)$, for all $i \in [1, N]$. Through (5.8), we can obtain $f_{U_{ij}}(v)$ in $O(m)$ for each i, and thus get all $f_{U_{ij}}(u)$'s ($\forall i \in [1, N]$) 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 (5.7) by doing convolution $N-1$ times in $O(N^2 m \log(Nm))$ [15]. 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 $q_j^*(p;\boldsymbol{\alpha})$. Since we have sampled $f_{P_j}(p)$ into a length-m sequence, we only need to get $q_j^*(p;\boldsymbol{\alpha})$ for at most m different values for p. Thus we can construct $q_j^*(p;\boldsymbol{\alpha})$ in $O(m)$ steps. Therefore, the total complexity is the sum of (i)-(v), *i.e.*, $O(Nm)+O(N^2m\log(Nm))+O(Nm)+O(Nm)+O(Nm)+O(m) = O(N^2m\log(Nm))$.

We then analyze the computation complexity of the subroutine **P3-OBJ**($\boldsymbol{\alpha}$), *i.e.*, evaluating the objective value of **P3** for any given GLB decision $\boldsymbol{\alpha}$. Step 13 needs $O(N^2)$ from (5.9). Steps 15 is the complexity to compute $q_j^*(p;\boldsymbol{\alpha})$, which requires $O(N^2m\log(Nm)$. Step 16 is the complexity to compute $\mathsf{ECost}_j\left(q_j^*(p;\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)$ by (5.12). 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 $\mathsf{ECost}_j\left(q_j^*(p;\boldsymbol{\alpha}),\boldsymbol{\alpha}\right)$ will be $O(Nm^2)$. Since **P3-OBJ**($\boldsymbol{\alpha}$) should do N iterations for all datacenters, the total complexity to evaluate **P3-OBJ**($\boldsymbol{\alpha}$) is $O(N^2 + N(N^2m\log(Nm) + Nm^2)) = O(N^3m\log(Nm) + N^2m^2)$.

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

$N^3m^2))))$. □

13.6 Proof of Proposition 2

Proof. This result is easily to prove. Since there is one option for the CSP: do not bid in the day-ahead market, *i.e.*, setting $\tilde{q}_j(p) = 0, \forall p \geq 0$. With $\tilde{q}_j(p)$, the objective value of $\mathbf{EP}_j(\boldsymbol{\alpha})$ is $E[V_j]\mu_j^{\mathsf{RT}}$. Since $q_j^*(p)$ is the optimal solution to minimize the objective value, we can have that its objective value is always upper bounded by $E[V_j]\mu_j^{\mathsf{RT}}$. □

13.7 Proof of Lemma 1

Proof. To aid our analysis, we introduce two stochastic orderings called "increasing convex ordering" ($\geq_{\text{ic}}$) and "variability ordering" ($\geq_{\text{var}}$), the definitions of which are presented Chapter 7.3. And an important property is presented in Proposition 7.

Proposition 7. *([70, Lemma 4.9]) $X \geq_{var} Y$ implies that $X \geq_{ic} Y$.*

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, $V_1 \geq_{var} V_2$. We denote C_1 and C_2 as the cost of V_1 and V_2 by the optimal bidding curve in (6.5). 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 $C_1 = \int_0^{+\infty} C_1(p) f_{P_i}(p) dp$, $C_2 =$

$\int_0^{+\infty} C_2(p) f_{P_i}(p) dp$. 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 $\mathbf{EP}_j(\boldsymbol{\alpha})$ will reduce to the classic Newsvendor problem and (6.5) is the corresponding optimal solution. According to Proposition 7, we can 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 8. *[70, Proposition 4.3] For the Newsvendor problem, given two future demands D_1, D_2, if $D_1 \geq_{ic} D_2$, $\mathbb{E}[D_1] = \mathbb{E}[D_2]$, then the optimal cost of D_1 is not less than that of D_2.*

The proof is complete. □

13.8 Proof of Lemma 2

Proof. 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. And the total cost expectation by (6.5) can be expressed as $\mathbb{E}_P[C_{\text{opt}}(p)]$, where the expectation is taken with respected 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 7), 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 7. *([64]) If X and Y are nonnegative random variables with $\mathbb{E}[X] = \mathbb{E}[Y]$, then $X \geq_{ic} Y$ if and only if $\mathbb{E}[f(X)] \geq \mathbb{E}[f(Y)]$ for all convex functions f.*

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

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

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

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. And (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 $pq_j^*(p) - \beta p Q_{\text{over}}(q_j^*(p)) + \mu_j^{\mathsf{RT}} Q_{\text{under}}(q_j^*(p))$ for p^1, p^2). The proof is completed. □

13.9 Proof of Lemma 3

Proof. We firstly reformulate the cost function from (5.12) to the following one,

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

(E_a) comes from the facts that $q(p) = 0$ for $p \geq \mu_j^{\mathsf{RT}}$ 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 proceed as follows,

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

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

13.10 Proof of Proposition 5

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

$$\begin{aligned}
&\int_0^{\delta q(p)} (\delta q(p) - \tilde{v}) f_{\tilde{V}}(\tilde{v}) d\tilde{v} \\
=&\int_0^{q(p)} (\delta q(p) - \delta v) f_{\tilde{V}}(\delta v) d(\delta v), \quad \text{by changing the integral variable} \\
=&\int_0^{q(p)} (\delta q(p) - \delta v) \frac{1}{\delta} f_V(v) d(\delta v), \quad \text{by } \tilde{f}_{\tilde{V}}(\delta v) = \frac{1}{\delta} f_V(v) \\
=&\delta \int_0^{q(p)} (q(p) - v) f_V(v) dv.
\end{aligned}$$

By similar arguments we have

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

According to the cost function (5.12), we have

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

The proof is completed. □

13.11 Proof of Proposition 6

Proof. We first rewrite the cost function as

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

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

$$\begin{aligned}
&\int_0^{q^1(p)+q^2(p)} (q^1(p) + q^2(p) - v) f_V(v) dv \leq \\
&\int_0^{q^1(p)} (q^1(p) - v^1) f_{V^1}(v^1) dv^1 + \int_0^{q^2(p)} (q^2(p) - v^2) f_{V^2}(v^2) dv^2. \qquad (13.4)
\end{aligned}$$

(13.4) can be rewritten as

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

This inequality is obliviously 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 (13.4) and the proof for Proposition 6 is completed. □

13.12 Proof of Lemma 4

Proof. The first-order derivative of the objective function with respect to $q(p)$ is given by

$$\begin{aligned}
&\frac{d\text{ECost1}(q(p),\boldsymbol{\alpha})}{dq(p)}\\
=&\int_0^{+\infty}[p-(1-\epsilon_2)p\int_0^{q(p)}f_{V_j}(v)dv-(1+\epsilon_1)p\int_{q(p)}^{\bar{C}}f_{V_j}(v)dv]f_{P_j}(p)dp,\\
=&\int_0^{+\infty}p\left[\epsilon_2\int_0^{q(p)}f_{V_j}(v)dv-\epsilon_1\int_{q(p)}^{\bar{C}}f_{V_j}(v)dv\right]f_{P_j}(p)dp\\
=&(\epsilon_1+\epsilon_2)\int_0^{q(p)}f_{V_j}(v)dv-\epsilon_1.
\end{aligned}$$

It is easy to see that the first order derivative is nondecreasing with $q(p)$. By solving $\frac{\text{Cost1}(q(p))}{dq(p)}=0$, we can get the optimal solution as $q^*(p)=F_{V_j}^{-1}\left(\frac{\epsilon_1}{\epsilon_1+\epsilon_2}\right)$.

The proof is completed. □

13.13 Proof of Proposition 4

Proof. To prove Proposition 4, we will prove that, given any solution $\hat{q}_j(p), j=1,\cdots,N$ for **S1**, which may violate (11.1), we can construct another solution $\bar{q}_j(p)=\begin{cases}\hat{q}_j(p), \textit{if } p<\mu_j^{\mathsf{RT}},\\ 0, \textit{if } p\geq\mu_j^{\mathsf{RT}}\end{cases}$ and the objective value of $\bar{q}_j(p), j=1,\cdots,N$ cannot be larger than that of $\hat{q}_j(p), j=1,\cdots,N$. In other words, we can construct another solution $\bar{q}_j(p)$, which satisfies (11.1) and has a smaller objective value.

Now, let us consider an alternative cost of $\bar{q}_j(p), \forall j$, which is incurred by the following strategy: we submit bidding curves $\bar{q}_j(p), \forall j$ to the day-ahead markets, but for any realization of U_i, P_i in real-time, we follow the GLB solution that is optimized with respect to $\hat{q}_j(p), \forall j$. We call the cost by this strategy as "fake" cost of $\bar{q}_j(p), \forall j$. Clearly, the fake cost is an upper bound of the objective value of $\bar{q}_j(p), \forall j$, since we do not follow its optimal strategy in the second stage. We will show that the "fake" cost of $\bar{q}_j(p), \forall j$ cannot be larger than the objective value of $\hat{q}_j(p), \forall j$, which will complete our proof.

Note that the strategies of the "fake" cost of $\bar{q}_j(p), \forall j$ and the objective value of $\hat{q}_j(p), \forall j$ share the same GLB strategy. Then both the electricity demands after GLB $v_j, \forall j$ and the bandwidth costs $\mathsf{bcost}(\cdot)$ are the same. It will be sufficient to only compare their electricity costs for each datacenter, as shown in the following two cases.

Case 1. For the MCP realization p_j with $p_j < \mu_j^{\mathsf{RT}}$, the electricity procurements and the day-ahead electricity costs for $\hat{q}_j(p)$ and $\bar{q}_j(p)$ are the same, so as the real-time electricity costs.

Case 2. For the MCP realization p_j with $p_j \geq \mu_j^{\mathsf{RT}}$, the solution $\hat{q}_j(p)$ will purchase $\hat{q}_j = \hat{q}_j(p_j) > 0$ amount of electricity from the day-ahead market, and the $\bar{q}_j(p)$ will not purchase any electricity from the day-ahead market. Then the electricity cost for $\hat{q}_j(p)$ will be $p_j\hat{q}_j + \mu_j^{\mathsf{RT}}(v_j - \hat{q}_j)^+ - \beta p_j(\hat{q}_j - v_j)^+$ and the electricity cost for $\bar{q}_j(p)$ will be $\mu_j^{\mathsf{RT}} v_j$.

– If $v_j \geq \hat{q}_j$,

$$\begin{aligned} & p_j\hat{q}_j + \mu_j^{\mathsf{RT}}(v_j - \hat{q}_j)^+ - \beta p_j(\hat{q}_j - v_j)^+ \\ =& p_j\hat{q}_j + \mu_j^{\mathsf{RT}}(v_j - \hat{q}_j) \\ =& \mu_j^{\mathsf{RT}}v_j + (p_j - \mu_j^{\mathsf{RT}})\hat{q}_j \\ \geq& \mu_j^{\mathsf{RT}}v_j. \end{aligned}$$

The last step is by the fact that $p_j \geq \mu_j^{\mathsf{RT}}$.

– If $v_j < \hat{q}_j$,

$$\begin{aligned} & p_j\hat{q}_j + \mu_j^{\mathsf{RT}}(v_j - \hat{q}_j)^+ - \beta p_j(\hat{q}_j - v_j)^+ \\ =& p_j\hat{q}_j - \beta p_j(\hat{q}_j - v_j) \\ =& p_j(\hat{q}_j - v_j) + p_jv_j - \beta p_j(\hat{q}_j - v_j) \\ =& (1-\beta)p_j(\hat{q}_j - v_j) + p_jv_j \qquad (13.5) \\ \geq& \mu_j^{\mathsf{RT}}v_j. \end{aligned}$$

The last step is by the fact that the first term of (13.5) is positive and $p_j \geq \mu_j^{\mathsf{RT}}$.

□

13.14 Proof of Lemma 6

Proof. It would be sufficient to show that $\mathbb{E}_{P_j,V_i,i,j=1:N}\left[CS2\left([q_j(p)]_{j=1:N}\right)\right]$ is convex in $q_j(p), \forall j$ since $\mathbb{E}_{P_j}\left[P_jq_j(P_j)\right]$ is linear in $q_j(p)$. Towards this end, we will show that, *for any P_i and U_i realization (one scenario)*, $CS2\left([q_j(p)]_{j=1:N}\right)$ is convex in $q_j(p), \forall j$.

Given two solutions $q_j^1(p), \forall j$, $q_j^2(p), \forall j$, and their convex combination $q_j^3(p) = \alpha q_j^1(p) + (1-\alpha) q_j^2(p)$ with $\delta \in [0,1]$, we will show that

$$CS2\left(\left[q_j^3(p)\right]_{j=1:N}\right) \leq \delta CS2\left(\left[q_j^1(p)\right]_{j=1:N}\right) + (1-\delta) CS2\left(\left[q_j^2(p)\right]_{j=1:N}\right).$$

We denote $\mathsf{ec}_j\left(q_j(p), \boldsymbol{\alpha}\right) = \mathsf{ecost}_j\left(\boldsymbol{\alpha}\right)$ as the real-time electricity cost by bidding curve $q_j(p)$. We can have $CS2\left(\left[q_j^3(p)\right]_{j=1:N}\right) = \mathsf{ec}_j\left(q_j(p), \boldsymbol{\alpha}^*\right) + \mathsf{bcost}(\boldsymbol{\alpha}^*)$, where $\boldsymbol{\alpha}^*$ is the corresponding optimal GLB solution in that scenario.

We firstly prove that $\mathsf{ec}_j\left(q_j(p), \boldsymbol{\alpha}\right)$ is **convex** in $\left(q_j(p), \boldsymbol{\alpha}\right), \forall j$, which would be clear if we rewrite it in a composition form. Specifically, let $u(w) = \mu_j^{\mathsf{RT}} w^+ + \beta p_j w^-$ [2] and $A(q_j(p), \boldsymbol{\alpha}) = \sum_i u_i \alpha_{ij} - q_j(p)$. We can have $\mathsf{ec}_j\left(q_j(p), \boldsymbol{\alpha}\right) = u\left(A\left(q_j(p), \boldsymbol{\alpha}\right)\right)$. Note that due to (11.1), $\mu_j^{\mathsf{RT}} \geq \beta p_j$ and the function $u(w)$ is convex in w. Also, $A(q_j(p), \boldsymbol{\alpha})$ is an affine function of $(q_j(p), \boldsymbol{\alpha})$. According to Chapter 3.2.2 of [13] (*Composition with an affine mapping preserves convexity.*), $\mathsf{ec}_j\left(q_j(p), \boldsymbol{\alpha}\right) = u\left(A\left(q_j(p), \boldsymbol{\alpha}\right)\right)$ is convex in $(q_j(p), \boldsymbol{\alpha})$.

Our argument proceeds as follows. Denoting $\boldsymbol{\alpha}^{k*}$ as the corresponding

[2] $w^+ = \begin{cases} w, w \geq 0 \\ 0, w < 0 \end{cases}$ and $w^- = \begin{cases} 0, w \geq 0 \\ w, w < 0 \end{cases}$

optimal solution for $q_j^k(p)$ in that scenario, we can have

$$\begin{aligned}
&\delta CS2\left(\left[q_j^1(p)\right]_{j=1:N}\right)+(1-\delta)CS2\left(\left[q_j^2(p)\right]_{j=1:N}\right)\\
=&\delta\left[\sum_j \mathsf{ec}_j\left(q_j^1(p),\boldsymbol{\alpha}^{1*}\right)+\mathsf{bcost}(\boldsymbol{\alpha}^{1*})\right]+(1-\delta)\left[\sum_j \mathsf{ec}_j\left(q_j^2(p),\boldsymbol{\alpha}^{2*}\right)+\mathsf{bcost}(\boldsymbol{\alpha}^{2*})\right]\\
\overset{(E_a)}{\geq}&\sum_j \mathsf{ec}_j\left(\delta q_j^1(p)+(1-\delta)q_j^2(p),\delta\boldsymbol{\alpha}^{1*}+(1-\delta)\boldsymbol{\alpha}^{2*}\right)+\mathsf{bcost}(\delta\boldsymbol{\alpha}^{1*}+(1-\delta)\boldsymbol{\alpha}^{2*})\\
=&\sum_j \mathsf{ec}_j\left(q_j^3(p),\delta\boldsymbol{\alpha}^{1*}+(1-\delta)\boldsymbol{\alpha}^{2*}\right)+\mathsf{bcost}(\delta\boldsymbol{\alpha}^{1*}+(1-\delta)\boldsymbol{\alpha}^{2*})\\
\overset{(E_b)}{\geq}&\sum_j \mathsf{ec}_j\left(q_j^3(p),\boldsymbol{\alpha}^{3*}\right)+\mathsf{bcost}(\boldsymbol{\alpha}^{3*})\\
=&CS2\left(\left[q_j^3(p)\right]_{j=1:N}\right),
\end{aligned}$$

where (E_a) is from the convexity of $\mathsf{ec}_j\left(q_j^k(p),\boldsymbol{\alpha}\right)$ and (E_b) is from the optimality of $\boldsymbol{\alpha}^{3*}$ for $q_j^3(p)$.

The proof is completed. □

□ **End of chapter.**

Bibliography

[1] 2011 Oregon Utility Statistics.

[2] Data center users group special report: Energy efficiency and capacity concerns increase. Emerson Network Power, White Paper, 2012.

[3] Facts about data centers. available at `http://energy.gov`.

[4] NYISO archive. available at `http://www.nyiso.com`.

[5] Weatherunderground. `http://www.wunderground.com`.

[6] Facebook's new 'cloud'. Technical report, ECONorthWest, 2011.

[7] How clean is your cloud? Technical report, Greenpeace Climate, 2012.

[8] Akamai Internet observatory. available at `https://www.akamai.com`

[9] Akamai Internet Observatory website.

[10] A. Beloglazov, R. Buyya, Y. C. Lee, A. Zomaya, et al. A taxonomy and survey of energy-efficient data centers and cloud computing systems. *Advances in Computers*, 82(2):47–111, 2011.

[11] E. Y. Bitar, R. Rajagopal, P. P. Khargonekar, K. Poolla, and P. Varaiya. Bringing wind energy to market. *IEEE Trans. Power Syst.*, 27(3):1225–1235, 2012.

[12] T. K. Boomsma, N. Juul, and S.-E. Fleten. Bidding in sequential electricity markets: The Nordic case. *European Journal of Operational Research*, 238(3):797–809, 2014.

[13] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[14] R. Brown. Report to congress on server and data center energy efficiency: Public law 109-431. *Lawrence Berkeley National Lab*, 2008.

[15] C. Burrus and T. W. Parks. *DFT/FFT and Convolution Algorithms: Theory and Implementation*. John Wiley & Sons, Inc., 1991.

[16] CAISO archive. available at `http://www.caiso.com`.

[17] 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.

[18] Pricing scheme of aliyun. `https://intl.aliyun.com/en`.

[19] N. Chen, X. Ren, S. Ren, and A. Wierman. Greening multi-tenant data center demand response. *Performance Evaluation*, 91:229–254, 2015.

[20] A. L. Custódio, H. Rocha, and L. N. Vicente. Incorporating minimum frobenius norm models in direct search. *Computational Optimization and Applications*, 46(2):265–278, 2010.

[21] A. L. Custódio and L. N. Vicente. Using sampling and simplex derivatives in pattern search methods. *SIAM Journal on Optimization*, 18(2):537–555, 2007.

[22] E. D. Dolan, R. M. Lewis, and V. Torczon. On the local convergence of pattern search. *SIAM Journal on Optimization*, 14(2):567–583, 2003.

[23] L. Eeckhoudt, C. Gollier, and H. Schlesinger. The risk-averse (and prudent) newsboy. *Management Science*, 41(5):786–794, 1995.

[24] ERCOT archive. available at `http://www.ercot.com`.

[25] A forgotten data center cost: Lost capacity. available at http://www.datacenterknowledge.com.

[26] S.-E. Fleten and E. Pettersen. Constructing bidding curves for a price-taking retailer in the Norwegian electricity market. *IEEE Trans. Power Syst.*, 20(2):701–708, 2005.

[27] Y. Gerchak and D. Mossman. On the effect of demand randomness on inventories and costs. *Operations research*, 40(4):804–807, 1992.

[28] 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.

[29] M. Ghamkhari, A. Wierman, and H. Mohsenian-Rad. Energy portfolio optimization of data centers. *IEEE Trans. Smart Grid*, 2016.

[30] Google energy wiki. `http://en.wikipedia.org/wiki/Google_Energy`.

[31] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, Mar. 2014.

[32] Y. Guo and Y. Fang. Electricity cost saving strategy in data centers by using energy storage. *IEEE Trans. Parallel Distrib. Syst.*, 24(6):1149–1160, 2013.

[33] 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.*, 27(3):1204–1213, 2012.

[34] Inverse function theorem. `https://en.wikipedia.org/wiki/Inverse_function_theorem`.

[35] P. Joskow. California's electricity crisis. *Oxford Review of Economic Policy*, 17(3):365–388, 2001.

[36] M. Khouja. The single-period (news-vendor) problem: Literature review and suggestions for future research. *Omega*, 27(5):537–553, 1999.

[37] J. Koomey. Growth in data center electricity use 2005 to 2010. *A report by Analytical Press, The New York Times*, 2011.

[38] J. Koomey. Growth in data center electricity use 2005 to 2010. *A report by Analytical Press, completed at the request of The New York Times*, 9, 2011.

[39] K. Krishna and M. N. Murty. Genetic k-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 29(3):433–439, 1999.

[40] A. H.-L. Lau and H.-S. Lau. The newsboy problem with price-dependent demand distribution. *IIE transactions*, 20(2):168–175, 1988.

[41] H.-S. Lau. The newsboy problem under alternative optimization objectives. *Journal of the Operational Research Society*, pages 525–535, 1980.

[42] K. Le, O. Bilgir, R. Bianchini, M. Martonosi, and T. D. Nguyen. Managing the cost, energy consumption, and carbon footprint of Internet services. In *Proc. ACM SIGMETRICS*, 2010.

[43] R. M. Lewis and V. Torczon. Pattern search methods for linearly constrained minimization. *SIAM Journal on Optimization*, 10(3):917–941, 2000.

[44] M. Lin, Z. Liu, A. Wierman, and L. L. Andrew. Online algorithms for geographical load balancing. In *Proc. IGCC*, 2012.

[45] M. Lin, A. Wierman, L. L. Andrew, and E. Thereska. Dynamic right-sizing for power-proportional data centers. *IEEE/ACM Trans. on Networking*, 21(5):1378–1391, 2013.

[46] 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*, 7(1):227–237, 2016.

[47] M. Liu and F. F. Wu. Risk management in a competitive electricity market. *International Journal of Electrical Power & Energy Systems*, 29(9):690–697, 2007.

[48] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. Andrew. Greening geographical load balancing. In *Proc. ACM SIGMETRICS*, 2011.

[49] Z. Liu, I. Liu, S. Low, and A. Wierman. Pricing data center demand response. In *Proc. ACM SIGMETRICS*, 2014.

[50] Z. Liu, A. Wierman, Y. Chen, B. Razon, and N. Chen. Data center demand response: Avoiding the coincident peak via workload shifting and local generation. *Performance Evaluation*, 70(10):770–791, 2013.

[51] Find minimum of constrained nonlinear multivariable function. `https://www.mathworks.com/help/optim/ug/fmincon.html`

[52] D. Meisner, B. T. Gold, and T. F. Wenisch. Powernap: eliminating server idle power. *ACM Sigplan Notices*, 44(3):205–216, 2009.

[53] Y. Merzifonluoglu and Y. Feng. Newsvendor problem with multiple unreliable suppliers. *International Journal of Production Research*, 52(1):221–242, 2014.

[54] B. Neupane, T. B. Pedersen, and B. Thiesson. Evaluating the value of flexibility in energy regulation markets. In *Proc. ACM e-Energy*, 2015.

[55] F. Paganini, P. Belzarena, and P. Monzón. Decision making in forward power markets with supply and demand uncertainty. In *Proc. CISS*, 2014.

[56] P. Pinson, C. Chevallier, and G. N. Kariniotakis. Trading wind generation from short-term probabilistic forecasts of wind power. *Trans. Power Sys.*, 22(3):1148–1156, 2007.

[57] L. H. Polatoglu. Optimal order quantity and pricing decisions in single-period inventory systems. *International Journal of Production Economics*, 23(1-3):175–185, 1991.

[58] Public utility commission of texas. `http://www.puc.texas.gov/consumer/electricity/polr.aspx`.

[59] Y. Qin, R. Wang, A. J. Vakharia, Y. Chen, and M. M. Seref. The newsvendor problem: Review and directions for future research. *European Journal of Operational Research*, 213(2):361–374, 2011.

[60] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs. Cutting the electric bill for Internet-scale systems. In *Proc. ACM SIGCOMM*, 2009.

[61] 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.

[62] 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*, 2(3):555–563, 2011.

[63] A. Ridder, E. Van Der Laan, and M. Salomon. How larger demand variability may lead to lower costs in the newsvendor problem. *Operations Research*, 46(6):934–936, 1998.

[64] S. M. Ross. *Stochastic processes*. Wiley, New York, 1996.

[65] R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo method*, volume 10. John Wiley & Sons, 2016.

[66] S. Sethi, H. Yan, J. H. Yan, and H. Zhang. An analysis of staged purchases in deregulated time-sequential electricity markets. *Journal of Industrial and Management Optimization*, 1(4):443–463, 2005.

[67] 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.*, 25(10):2659–2669, 2014.

[68] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming: modeling and theory*. SIAM, 2009.

[69] 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.

[70] J.-S. Song. The effect of leadtime uncertainty in a simple stochastic inventory model. *Management Science*, 40(5):603–613, 1994.

[71] C. Stewart and K. Shen. Some joules are more precious than others: Managing renewable energy in the datacenter. In *Proc. HotPower*, 2009.

[72] 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.

[73] N. H. Tran, D. H. Tran, S. Ren, Z. Han, E.-N. Huh, and C. S. Hong. How geo-distributed data centers do demand response: A game-theoretic approach. *IEEE Trans. Smart Grid*, 7(2):937–947, 2016.

[74] G. K. Tso and K. K. Yau. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, 32(9):1761–1768, 2007.

[75] J. Usaola and J. Angarita. Bidding wind energy under uncertainty. In *2007 International Conference on Clean Electrical Power*, pages 754–759. IEEE, 2007.

[76] J. Usaola and M. A. Moreno. Optimal bidding of wind energy in intra-day markets. In *2009 6th International Conference on the European Energy Market*, pages 1–7. IEEE, 2009.

[77] H. Wang, J. Huang, X. Lin, and H. Mohsenian-Rad. Exploring smart grid and data center interactions for electric power load balancing. *ACM SIGMETRICS Performance Evaluation Review*, 41(3):89–94, 2014.

[78] 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*, 3(4):1743–1754, 2012.

[79] 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*, 3(4):1743–1754, 2012.

[80] 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.

[81] R. Wang, N. Kandasamy, C. Nwankpa, and D. R. Kaeli. Datacenters as controllable load resources in the electricity market. In *Proc. IEEE ICDCS*, 2013.

[82] Web traffic calculator. `http://www.mobilenet.gov.hk/en/consumer_tips/data_usage_calculator/index.html`.

[83] T. M. Whitin. Inventory control and price theory. *Management science*, 2(1):61–68, 1955.

[84] A. Wierman, Z. Liu, I. Liu, and H. Mohsenian-Rad. Opportunities and challenges for data center demand response. In *IGCC*, pages 1–10. IEEE, 2014.

[85] M. Wu, S. X. Zhu, and R. H. Teunter. A risk-averse competitive newsvendor problem under the cvar criterion. *International Journal of Production Economics*, 156:13–23, 2014.

[86] H. Yi, M. Hajiesmaili, Y. Zhang, M. Chen, and X. Lin. Impact of the uncertainty of distributed renewable generation on deregulated electricity supply chain. *submitted for journal publication*, 2017.

[87] L. Yu, T. Jiang, and Y. Cao. Energy cost minimization for distributed Internet data centers in smart microgrids considering power outages. *IEEE Trans. Parallel Distrib. Syst.*, 26(1):120–130, 2015.

[88] 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.*, 25(5):1306–1316, 2014.

[89] X.-P. Zhang. *Restructured Electric Power Systems: Analysis of Electricity Markets with Equilibrium Models*. John Wiley & Sons, 2010.

[90] Y. Zhang, L. Deng, M. Chen, and P. Wang. Joint bidding and geographical load balancing for datacenters: Is uncertainty a blessing or a curse? In *Proc. IEEE INFOCOM*, 2017.

[91] Y. Zhang, L. Deng, M. Chen, and P. Wang. Joint bidding and geographical load balancing for datacenters: Is uncertainty a blessing or a curse? *submitted for journal publication*, 2017.