2014 IEEE INFOCOM Workshops: 2014 IEEE INFOCOM Workshop on Smart Data Pricing

Optimal Risk-aware Power Procurement for Data Centers in Day-Ahead and Real-Time Electricity Markets

Mahdi Ghamkhari*, Hamed Mohsenian-Rad*, and Adam Wierman**
*Department of Electrical Engineering, University of California, Riverside, CA, USA
**Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA
E-mails: {ghamkhari, hamed}@ee.ucr.edu and adamw@caltech.edu

***Abstract*— With the growing trend in the amount of power consumed by data centers, finding ways to cut their electricity bills has become an important and challenging problem. In this paper, our focus is on the cost reduction that data centers may achieve by exploiting the diversity in the price of electricity in day-ahead and real-time electricity markets. Based on a stochastic optimization framework, we propose to jointly select a data center's service rate and its power demand bids to the day-ahead and real-time electricity markets. In our analysis, we take into account service-level-agreements, risk management constraints, and statistical characteristics of workload and electricity prices. Using empirical electricity price and Internet workload data and through computer simulations, we show that by directly participating in the day-ahead and real-time electricity markets, data centers can significantly reduce their energy expenditure.**

***Keywords*: Data Center, Day-ahead Market, Real-time Market, Price Diversity, Power Procurement, Service-level Agreement.**

I. Introduction

The energy demands of data centers have significantly increased over the past years. Accordingly, the cost of electricity to operate data centers have been skyrocketing. For example, it is estimated that Microsoft and Google each spent over $36 million on annual electricity bills for their data centers in 2007 [1]. The total annual electricity cost of servers and data centers in the United States is estimated at $7.4 billion [2].

The growing energy cost of data centers has motivated various studies to lower data centers' electricity bills. The prior work can be classified into at least five different categories. First, there have been studies to reduce the amount of power that computing and memory devices consume, e.g., see [3]. Second, different methods have been proposed to optimize the operation of hardware and software systems in data centers in response to changes in the workload, e.g., by conducting *dynamic cluster server configuration* [4]. Third, there have been efforts to make the best use of *local energy recourses* at data centers, such as solar and wind generators [5], [6], battery banks [7], and backup diesel generators [8]. Fourth, some recent studies have focused on *workload redistribution* across data centers to benefit from *geographical diversity* in both electricity prices [1], [8] and renewable generation [9], [10]. Finally, there have been studies to manage the operation and available resources of data centers to better respond to the changes in the price of electricity, whether by lowering power consumption or by increasing the use of local energy resources. Examples include [11] for the case of time-of-use prices, [12] for the case of day-ahead prices when hedging is applied, [13] for the case of coincidental peak prices, [14] for the case of prediction-based prices and [15] for the case of a data center offering ancillary services.

This work was supported in part by the U.S. NSF through grants CNS 1319798 and CNS 1319820. H. Mohsenian-Rad is the corresponding author.

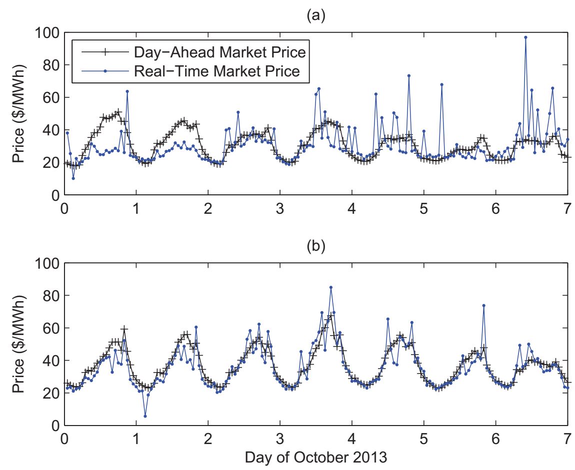

Fig. 1. Two sample empirical price trends for day-ahead and real-time electricity markets during the first week of October 2013: (a) The Ameren retail price trends [18]. (b) The PJM wholesale market price trends [19].

In this paper, our approach is related to the fifth category above. Our focus is on procuring power for data centers in a *deregulated electricity market*, i.e., a market where prices are set by running bidding mechanisms among the electricity suppliers and consumers, c.f., [16]. Compared to the prior work, our study is unique in the sense that we consider a scenario where data centers can buy electricity from both the *day-ahead market* and the *real-time market*. The day-ahead market is usually settled several hours or even a day in advance while the real-time market is settled only one hour or sometimes 15 minutes in advance [17, Chapter 2].

Our goal is to understand the cost reductions that data centers can achieve by exploiting the *price diversity across day-ahead and real-time markets*. To see the potential for such cost reductions, consider the price data in Fig. 1. Here, we show the prices in the day-ahead market versus the real-time market in the Ameren retail market in Illinois [18] and the Pennsylvania-Jersey-Maryland (PJM) wholesale market [19]. Price diversity is evident: at some hours the price is lower in the real-time market while at some other hours the price is lower in the day-ahead market. For the price data in Fig. 1, and compared to buying electricity only from the day-ahead market, procuring electricity from *both* day-ahead and real-time electricity markets may result in a saving up to 13.2%

and 7.5% in the Ameren and PJM markets, respectively.

The significant power consumption of data centers make them *eligible* to directly participate in the day-ahead and real-time electricity markets that currently exist in the U.S., instead of purchasing electricity from regional utilities who charge "insurance premiums" to handle the variations in the wholesale price of electricity. However, a key challenge in procuring power directly from the wholesale markets is that, although the statistical characteristics of electricity prices and Internet workload can be predicted, the actual values of such variables are revealed only at the operating time and not in advance. This can expose data centers to the risk of facing volatile electricity expenditure. To tackle this challenge, we propose to maximize the data center's expected profit, i.e., revenue minus cost, subject to the data center operator's risk management constraints as well as the constraints with respect to power consumption and service-level-agreements (SLAs).

II. System Model

A. Power Market and Cost of Electricity

In most deregulated electricity markets, electricity can be purchased both at the Day-Ahead Market (DAM) and the Real-Time Market (RTM). This is done by submitting demand bids L_{DAM} and L_{RTM} in megawatt hours to the DAM and RTM, respectively. The total amount of purchased electric energy from the two markets combined is obtained as

$$\text{Electric Energy Purchase} = L_{DAM} + L_{RTM}. \tag{1}$$

The day-ahead market is usually settled several hours or a day in advance while the real-time market is settled one hour down to 15 minutes before the operation time [17, Chapter 2]. Let ω_{DAM} and ω_{RTM} denote the *market clearing prices* at the day-ahead and real-time markets, respectively. The total cost of power purchase for each bidding period is obtained as

$$\text{Cost} = L_{DAM}\omega_{DAM} + L_{RTM}\omega_{RTM}. \tag{2}$$

We note that, since L_{DAM} and L_{RTM} are selected *before* the market is settled, the market clearing prices ω_{DAM} and ω_{RTM} are *not* known at the time of choosing the demand bids. Therefore, ω_{DAM} and ω_{RTM} are modeled as two random variables with statistical characteristics $E\{\omega_{DAM}\}$, $Var\{\omega_{DAM}\}$, $E\{\omega_{RTM}\}$, $Var\{\omega_{RTM}\}$, and $Cov\{\omega_{DAM}, \omega_{RTM}\}$. Note that, in this study, we assume that the data center is price taker. That is, its demand bids are not large enough to have noticeable impact on the price of electricity.

B. Power Consumption

The total amount of power consumption in a data center is obtained by adding the total power consumption at the computer servers to the total power consumption at the facility, e.g., for cooling, lighting, etc. For a data center, *power usage effectiveness* (PUE), denoted by E_{usage}, is defined as the ratio of the data center's total power consumption to the power consumption at the servers [2]. The PUE values reported in the literature range from state-of-the-art 1.05 to 3.0 for the common practice [20, Section 12.3.3]. Let P_{server} denote the average power when a switched on server handles a service request. Also let $M \leq M_{max}$ denote the number of servers that are switched on at the data center. Assuming full CPU utilization for all switched on servers, the total power consumption of a data center can be calculated as [10], [21]:

$$\text{Power Consumption} = E_{usage} M P_{server}. \tag{3}$$

Clearly, the power consumption at a data center increases as more servers are switched on to handle more service requests.

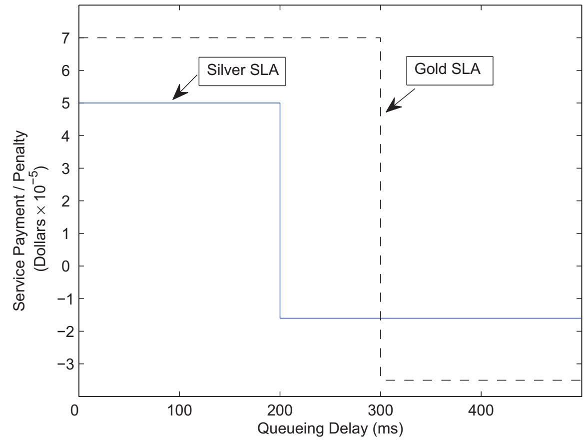

Fig. 2. Two sample service-level agreements (SLAs) in data centers [22].

C. Quality-of-Service, SLAs, and Service Rate

Because of the limited computing capacity of data centers and the stochastic nature of workload, the service requests that are sent to a data center are first placed in a queue until they can be handled by an available computer. To satisfy quality-of-service (QoS) requirements, the waiting time / queuing delay for each incoming service request should be limited to a level that is determined by the *Service Level Agreement* (SLA). The exact SLA depends on the type of service offered which may range from cloud-based computational tasks to video streaming and web services. Two example SLAs based on the study in [22] are shown in Fig. 2, where each SLA is identified by three parameters D, δ, and γ. Parameter D indicates the maximum waiting time that a service request can tolerate. Parameter δ indicates the service money that the data center receives when it handles a single service request *before* deadline D. Parameter γ indicates the *penalty* that the data center must pay to its customers every time it *cannot* handle a service request before deadline D.

Let $\mu \geq 0$ denote the rate at which service requests are removed from the queue and handled by a server. Also let κ denote the number of service requests that a computer server in the data center can handle per second. We have [11]:

$$\mu = \kappa M \quad \Rightarrow \quad M = \frac{\mu}{\kappa}. \tag{4}$$

As we switch on more servers and accordingly increase the service rate μ, more service requests can be handled *before* the SLA deadline D, which in turn increases the payments that the data center receives based on the SLAs. However, increasing

μ will also increase the amount of power consumption at the data center. In fact, from (3) and (4), we have

$$\text{Power Consumption} = \phi\mu, \tag{5}$$

where

$$\phi = E_{usage} P_{server}/\kappa. \tag{6}$$

D. Revenue of Data Center

Consider a power purchase bidding period T. For example, in an hourly market, we have $T = 3600$ seconds. Let $N \gg 1$ denote the number of service requests that arrive at the data center. Based on the SLA model that we discussed in Section II-C, the revenue of the data center within the time period of interest can be calculated as

$$\text{Revenue} = \sum_{i=1}^{N} I_{D,i}\,\delta - (1 - I_{D,i})\,\gamma, \tag{7}$$

where $I_{D,i} = 1$ indicates that the i^{th} service request *was* handled before the SLA deadline D and $I_{D,i} = 0$ indicates that the i^{th} service request was *not* handled before the SLA deadline D. If T and N are large enough, then we can write

$$\text{Revenue} \approx \lambda T((1 - q(\mu))\delta - q(\mu)\gamma), \tag{8}$$

where λ denotes the average arrival rate of service requests and $q(\mu)$ is the probability that a service request is *not* handled within deadline D. A model for $q(\mu)$ can be obtained through a G/D/1 queuing analysis that is already done in [23]:

$$q(\mu) = \begin{cases} q_I(\mu) & \mu_I \le \mu \\ q'_{I+}(\mu_I)(\mu - \mu_I) + q(\mu_I) & \mu_{II} \le \mu < \mu_I \\ q_{II}(\mu) & \mu < \mu_{II}, \end{cases} \tag{9}$$

where

$$q_I(\mu) = \alpha(\mu)\ \exp\left(-\frac{1}{2}\min_{n\ge 1}\ m_n(\mu)\right), \tag{10}$$

and

$$q_{II}(\mu) = \frac{T\lambda - T\mu}{T\lambda} = \frac{\lambda - \mu}{\lambda}. \tag{11}$$

Parameters μ_I and μ_{II} are obtained using [23, Algorithm 1] and the notations in (10) are defined as follows:

$$\alpha(\mu) = \frac{1}{\lambda\sqrt{2\pi}\sigma} e^{\frac{(\mu-\lambda)^2}{2\sigma^2}} \int_{\mu}^{\infty} (r - \mu) e^{-\frac{(r-\lambda)^2}{2\sigma^2}}\, dr, \tag{12}$$

and for each integer number $n \ge 1$ we have

$$m_n(\mu) = \frac{(D\mu + n(\mu - \lambda))^2}{n\sigma^2 + 2\sum\limits_{l=1}^{n-1} \rho(l)(n - l)}. \tag{13}$$

It is worth emphasizing that the *general* service request arrival rate in (9)-(13) is modeled based on its various statistical characteristics, i.e., not only its mean λ, but also its variance σ^2 and its auto-covariance function $\rho(l)$, where lag time $l = 1, 2, \ldots$. Thus, the $q(\mu)$ model in (9) is significantly more elaborate and more accurate than the simplified M/M/1 queuing models that are typically used in most data center power consumption studies, e.g., in [8], [9], [12], [24], [25].

III. Stochastic Profit Maximization

From the results in Section II, there is a *trade-off* when it comes to selecting a data center's service rate: increasing service rate increases the revenue while it also increases the cost. Addressing this trade-off is challenging due to the complexity of the queuing models and also because of the *stochastic nature* of the workload and the day-ahead and real-time electricity market prices. Hence, in this section, we propose a decision making process based on a stochastic optimization framework. We propose two variations for the problem, with and without risk constraints, as we explain next.

A. Design without Risk Constraint

We can model the *profit* for a data center as

$$\text{Profit} = \text{Revenue} - \text{Cost}. \tag{14}$$

When it comes to operating a data center, it is natural to seek to maximize the data center's profit. However, due to the stochastic nature of workload and electricity price, such maximization must be in an average / statistical sense, i.e., in terms of the expected value of the profit. Therefore, we need to solve the following optimization problem to choose both the electricity purchase bidding parameters L_{DAM} and L_{RTM} as well as the service rate μ, one day in advance, i.e., at the time when the day-ahead market bid needs to be submitted:

$$\begin{aligned} &\underset{\substack{L_{DAM},\,L_{RTM} \\ \mu\,\le\,\mu_{max}}}{\textbf{Maximize}} && E\{\text{Profit}\} \\ &\textbf{Subject to} && \text{Energy Consumption} = \text{Energy Purchase}, \end{aligned} \tag{15}$$

where $\mu_{max} = \kappa M_{max}$. From (14), we can write

$$E\{\text{Profit}\} = E\{\text{Revenue}\} - E\{\text{Cost}\}. \tag{16}$$

By substituting (2) and (8) in (16), we have

$$\begin{aligned} E\{\text{Profit}\} =& \lambda T((1 - q(\mu))\delta - q(\mu)\gamma) - \\ & L_{DAM}E\{\omega_{DAM}\} - L_{RTM}E\{\omega_{RTM}\}. \end{aligned} \tag{17}$$

From (1), (5), and (17), we can rewrite problem (15) as

$$\begin{aligned} &\underset{L_{DAM},L_{RTM},\,\mu\le\mu_{max}}{\textbf{Maximize}} && \lambda T\big((1 - q(\mu))\delta - q(\mu)\gamma\big) - \\ &&& L_{DAM}E\{\omega_{DAM}\} - \\ &&& L_{RTM}E\{\omega_{RTM}\} \\ &\textbf{Subject to} && L_{DAM} + L_{RTM} = \phi\mu, \end{aligned} \tag{18}$$

where the probability term $q(\mu)$ is as in (9). The following theorem helps characterising the above optimization problem.

***Theorem 1**: The following results hold for problem (18):*
(a) If $E\{\omega_{DAM}\} > E\{\omega_{RTM}\}$, then, at optimality, we have

$$L^\star_{DAM} = 0, \tag{19}$$

$$L^\star_{RTM} = \phi\mu^\star, \tag{20}$$

$$\mu^\star = \underset{\mu\le\mu_{max}}{\arg\max}\ \lambda T\big((1-q(\mu))\delta - q(\mu)\gamma\big) - \phi\mu E\{\omega_{RTM}\}. \tag{21}$$

(b) If $E\{\omega_{DAM}\} < E\{\omega_{RTM}\}$, then, at optimality, we have

$$L^\star_{RTM} = 0, \tag{22}$$

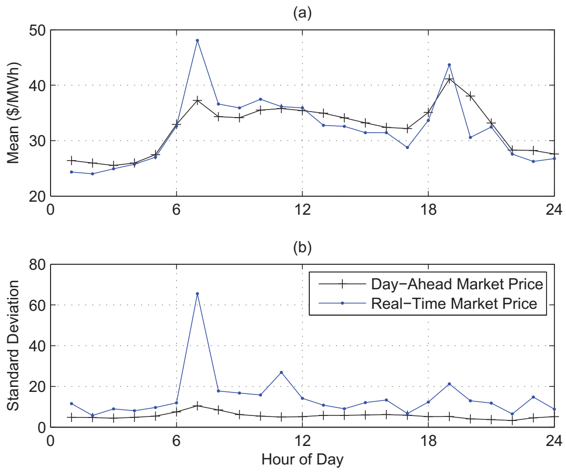

Fig. 3. Statistical characteristics of the day-ahead market and the real-time market prices during the month of October 2013 based on the Ameren electricity price trends in [18]: (a) Mean; (b) Standard deviation.

$$L^{\star}_{DAM} = \phi\mu^{\star}, \tag{23}$$

$$\mu^{\star} = \underset{\mu \le \mu_{max}}{\arg\max}\ \lambda T\big((1-q(\mu))\delta - q(\mu)\gamma\big) - \phi\mu E\{\omega_{DAM}\}. \tag{24}$$

The proof of Theorem 1 is given in Appendix A. From Theorem 1, if we follow the optimal solution of problem (18), then the entire power needs of the data center must be purchased from the market with lower average prices. In practice, it is usually the real-time market that has the lower average prices. For example, for the hourly price data in Fig. 3(a), in 17 out of 24 hours, the average price is lower in the real-time market than the day-ahead market. Consequently, the data center should procure all its energy needs from the real-time market during most of the day. The drawback is that the prices in the real-time market are more volatile, as shown in Fig. 3(b). In fact, due to the high variations in the real-time prices, load entities usually avoid purchasing the entire or even the majority of their electricity needs from the real-time market [26], making the solution of the expected energy cost minimization problem (18) less desirable in practice. This shortcoming can be resolved by including *risk constraints* in the decision making process, as we explain next.

B. Design with Risk Constraint

To address the concerns about volatile electricity prices, we propose to extend problem (15) by adding a risk constraint that limits the variance in the cost of procured energy:

$$\begin{aligned} &\underset{\substack{L_{DAM},\,L_{RTM}\\ \mu\le\mu_{max}}}{\textbf{Maximize}} && E\{\text{Profit}\} \\ &\textbf{Subject to} && Var\{\text{Cost}\} \le \Gamma \\ & && \text{Energy Consumption} = \text{Energy Purchase}, \end{aligned} \tag{25}$$

where Γ is a risk management design parameter. The choice of parameter Γ depends on whether the data center operator is *risk averse* (lower Γ) or *risk seeking* (higher Γ). Note that, problem (15) is a special case of problem (25) where $\Gamma \to \infty$.

From (2), the variance of the cost can be calculated as

$$\begin{aligned} Var\{\text{Cost}\} =& L^2_{DAM} Var\{\omega_{DAM}\} + \\ & L^2_{RTM} Var\{\omega_{RTM}\} + \\ & 2L_{DAM}L_{RTM}Cov\{\omega_{DAM},\omega_{RTM}\}. \end{aligned} \tag{26}$$

From (1), (5), (17), and (26), we can rewrite problem (25) as

$$\begin{aligned} &\underset{L_{DAM},L_{RTM},\,\mu\le\mu_{max}}{\textbf{Maximize}} && \lambda T\big((1-q(\mu))\delta - q(\mu)\gamma\big) - \\ & && L_{DAM}E\{\omega_{DAM}\} - \\ & && L_{RTM}E\{\omega_{RTM}\} \\ &\textbf{Subject to} && 2L_{DAM}L_{RTM}Cov\{\omega_{DAM},\omega_{RTM}\} + \\ & && L^2_{DAM}Var\{\omega_{DAM}\} + \\ & && L^2_{RTM}Var\{\omega_{RTM}\} \le \Gamma \\ & && L_{DAM} + L_{RTM} = \phi\mu. \end{aligned} \tag{27}$$

The following theorem shows that the above nonlinear optimization problem is computationally tractable.

Theorem 2: *For any workload and electricity price parameters λ, σ^2, $\rho(l)$, $Cov\{\omega_{DAM},\omega_{RTM}\}$, $Var\{\omega_{DAM}\}$, and $Var\{\omega_{RTM}\}$, the optimization problem in (27) is convex.*

The proof of Theorem 2 is given in Appendix B. From Theorem 2, problem (27) can be solved using standard convex programming techniques, c.f. [27]. Therefore, solving problem (27) can be considered as a practical yet optimal way to adjust the operation of the data center and to select its demand bids to the day-ahead and real-time electricity markets.

Before we end this section, we shall point out some remarks about problem (27). First, the risk model based on variance is only one option to cope with price uncertainty. Another option is to use the conditional value at risk (CVaR) models from [8], [28], [29] and revise them for risk management across *both* day-ahead and real-time markets. Second, it can be beneficial to also use price prediction, e.g., see [30]–[32], especially for day-ahead market prices which are less volatile. Third, although the optimization (27) is for a scenario where service rate and power purchase bids are updated on an hourly basis, we can easily extend our design to other update intervals. For example, service rates can be updated every 15 minutes [11], [23], while the power procurement bids are updated every hour. Problem (27) can still be used in such scenarios by breaking down the revenue and power consumption terms into multiple terms, each corresponding to a smaller interval at which the service rate is adjusted. Finally, optimization problem (27) can be extended to the case where the data center offers different types of services with different quality-of-service requirements and SLAs. In this regard, the problem formulation can be extended similar to analysis in [33].

IV. Numerical Examples

Consider a data center with $M_{max} = 50,000$ servers. The number of switched on servers M, and accordingly the service rate μ is updated periodically at the beginning of each time slot of length $T = 15$ minutes. We assume that $\kappa = 0.1$ and $P_{server} = 150$ watts. The data center's power usage effectiveness is $E_{usage} = 1.5$. The SLA parameters are set

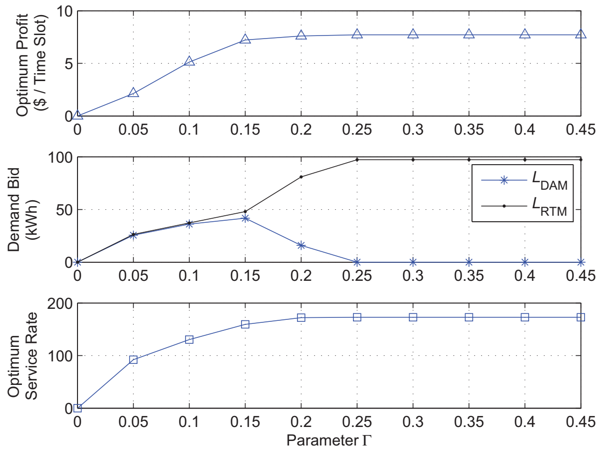

Fig. 4. Numerical results for a single time slot for different values of Γ.

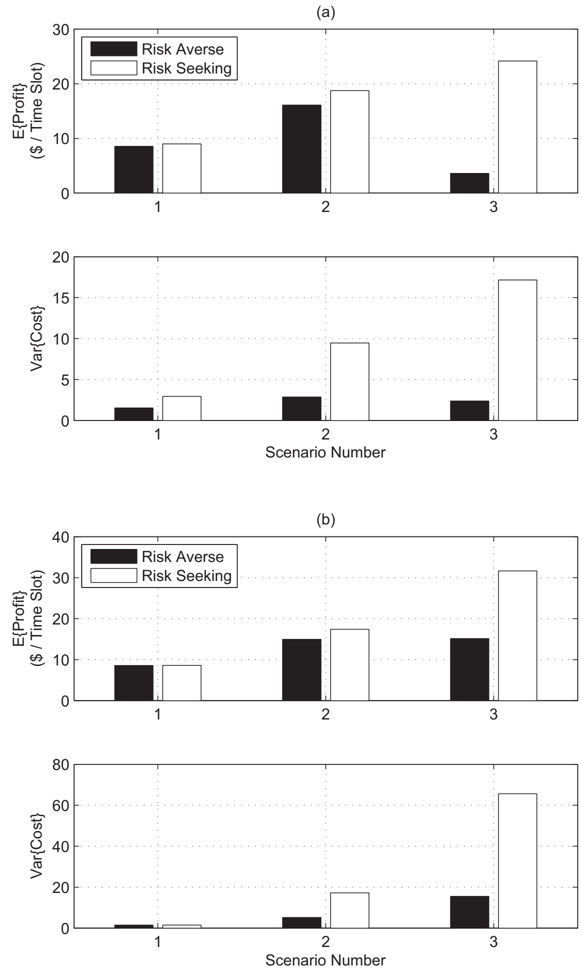

Fig. 5. The expected value of profit and the variance of cost for three workload scenarios over one month: (a) Using Ameren prices. (b) Using PJM prices.

based on the Gold service model in Fig. 2. To simulate the total workload, we use the World Cup 98 web hits data [34]. The electricity price information is based on the hourly day-ahead and real-time prices that are set by Ameren and PJM during December 2012 [18], [19]. In all cases, the revenue and cost are calculated using an event-based simulation, where an event is the arrival of a new service request, c.f. [11].

To gain insights, first we look at the detailed results for the case of solving problem (27) for a single time slot. Here, we use the prices from Ameren. The results are shown in Fig. 4 for different values of parameter Γ. When $\Gamma = 0$, the only feasible solution is $\mu^\star = L^\star_{DAM} = L^\star_{RAM} = 0$, i.e., shutting down the data center. As we become more risk seeking by increasing Γ, both service rate and profit increase. We can see that for the lower values of Γ, it is optimal to procure a large portion of the power needs from the day-ahead market as the prices in the day-ahead market are less volatile. However, as we increase Γ, such portion gradually degrades and the entire demand is eventually procured from the real-time market.

Next, we consider three different time slots / scenarios over one month. The results are shown in Fig. 5. Scenario 1 is for a time slot where the data center workload is *low*, with average $\lambda = 197.64$. Scenario 2 is for a time slot where the workload is *medium*, with average $\lambda = 397.18$. Scenario 3 is for a time slot where the workload is *high*, with average $\lambda = 879.29$. For each scenario, we compare a *risk averse* design with $\Gamma = 1$ and a *risk seeking* design with $\Gamma = 4$. In all three scenarios, while a risk seeking design can increase profit by more aggressively bidding in the real-time market, it makes the cost more volatile. Note that, since the optimal choices of L_{DAM} and L_{RTM} change from one day to another, the variance of cost that is calculated across 30 different days may not be limited to Γ; nevertheless, we can see that by changing Γ we can achieve our intended design goal on controlling the trade-off between maximizing the expected value of the profit and minimizing the risk in power procurement cost.

V. Conclusions

In this paper, we took the first steps towards exploiting the price diversity across the day-ahead and real-time electricity markets to lower data center energy expenditure. Based upon our observations of empirical workload and electricity price data, we proposed a novel stochastic and provably convex profit maximization problem to select data centers' service rates and bids to the day-ahead and real-time electricity markets. Our design incorporates the trade-off between increasing the profit and decreasing the risk in energy procurement cost, for a data center that bids into both day ahead and real time electricity markets. Through simulation results based on different workload intensities, we showed that procuring power from both the day-ahead and real-time electricity markets can significantly help data centers in lowering their energy cost.

Appendix

A. Proof of Theorem 1

We prove part (a) by contradiction. The proof for part (b) is similar. Assume that $(L^\star_{DAM}, L^\star_{RTM}, \mu^\star)$ is an optimal solution of problem (18), where $L^\star_{DAM} = \Delta > 0$. Next, consider $(L^*_{DAM}, L^*_{RTM}, \mu^*)$ such that $L^*_{DAM} = 0$, $L^*_{RTM} = L^\star_{RTM} + \Delta$, and $\mu^* = \mu^\star$. Clearly, $(L^*_{DAM}, L^*_{RTM}, \mu^*)$ is a feasible solution for problem (18). We can show that

$$\begin{aligned}&\text{Profit}(L^*_{DAM}, L^*_{RTM}, \mu^*) - \text{Profit}(L^\star_{DAM}, L^\star_{RTM}, \mu^\star) \\ &= \Delta\left(E\{\omega_{DAM}\} - E\{\omega_{RTM}\}\right) > 0,\end{aligned} \tag{28}$$

where the inequality is due to $E\{\omega_{DAM}\} > E\{\omega_{RTM}\}$. Thus, $(L^\star_{DAM}, L^\star_{RTM}, \mu^\star)$ *cannot* be an optimal solution of (18). ■

B. Proof of Theorem 2

From [23, Theorems 1 and 3], for any choices of workload statistical parameters λ, σ^2, and $\rho(l)$, the probability model $q(\mu)$ is a convex and non-increasing function of service rate μ. Therefore, the objective function in (27) is concave in μ and linear in L_{DAM} and L_{RTM}. From this, and since the equality constraint in (27) is linear, problem (27) is convex as long as the left-hand side in the non-linear inequality constraint in (27) is a convex function. To show this, first, we note that by definition of the *correlation coefficient*, we have [35]:

$$-1 \leq \frac{Cov\{\omega_{DAM}, \omega_{RTM}\}}{\sqrt{Var\{\omega_{DAM}\}}\sqrt{Var\{\omega_{RTM}\}}} \leq 1. \qquad (29)$$

From (29), we can further show that

$$Cov^2\{\omega_{DAM}, \omega_{RTM}\} \leq Var\{\omega_{DAM}\} Var\{\omega_{RTM}\}. \quad (30)$$

Next, we notice that the Hessian matrix of the non-linear function in the inequality constraint of (27) is

$$2\begin{bmatrix} Var\{\omega_{DAM}\} & Cov\{\omega_{DAM}, \omega_{RTM}\} & 0 \\ Cov\{\omega_{DAM}, \omega_{RTM} & Var\{\omega_{RTM}\} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (31)$$

where the rows and columns correspond to the optimization variables L_{DAM}, L_{RTM} and μ, respectively. For the inequality constraint in (27) to be convex, the above matrix must be positive definite. Using the Schur complement and because the variance of electricity price is always positive and also because of the block diagonal structure of the matrix in (31), this matrix is positive definite if and only if

$$Var\{\omega_{RTM}\} - \frac{Cov^2\{\omega_{DAM}, \omega_{RTM}\}}{Var\{\omega_{DAM}\}} > 0. \qquad (32)$$

However, the above inequality always holds due to (30). Therefore, the Hessian matrix in (31) is positive semi definite and accordingly the problem in (27) is convex. ■

References

[1] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs, "Cutting the electric bill for internet-scale systems," in *Proc. of the ACM SIGCOMM*, Barcelona, Spain, 2009.

[2] U.S. Environmental Protection Agency, "EPA report on server and data center energy efficiency," Final Report to Congress, August 2007.

[3] S. Steinke, N. Grunwald, L. Wehmeyer, R. Banakar, M. Balakrishnan, and P. Marwedel, "Reducing energy consumption by dynamic copying of instructions onto onchip memory," in *Proc. of the International Symposium on System Synthesis*, Kyoto, Japan, October 2002.

[4] E. Pinheiro, R. Bianchini, E. V. Carrera, and T. Heath, "Dynamic cluster reconfiguration for power and performance," in *Compilers and Operating Systems for Low Power*, M. K. L. Benini and J. Ramanujam, Eds. Kluwer Academic Publishers, 2003.

[5] R. Bianchini, "Leveraging renewable energy in data centers: present and future," in *Proc. of ACM International symposium on High-Performance Parallel and Distributed Computing*, Delft, Netherlands, June 2012.

[6] K. K. Nguyen, M. Cheriet, M. Lemay, M. Savoie, and B. Ho, "Powering a data center network via renewable energy: A green testbed," *IEEE Internet Computing*, vol. 17, no. 1, pp. 40–49, January 2013.

[7] B. Aksanli, T. Rosing, and E. Pettis, "Distributed battery control for peak power shaving in datacenters," in *Proc. of IEEE International Green Computing Conference*, Arlington, VA, June 2013.

[8] L. Yu, T. Jiang, Y. Cao, and J. Wu, "Risk-constrained operation for internet data centers under smart grid environment," in *Proc. of IEEE WCSP*, Hangzhou, China, October 2013.

[9] Z. Liu, M. Lin, A. Wierman, S. Low, and L. L. H. Andrew, "Geographical load balancing with renewables," in *Proc. of the ACM GreenMetrics Workshop*, San Jose, CA, April 2011.

[10] M. Ghamkhari and H. Mohsenian-Rad, "Optimal integration of renewable energy resources in data centers with behind-the-meter renewable generators," in *Proc. of the IEEE ICC*, Ottawa, Canada, June 2012.

[11] ——, "Energy and performance management of green data centers: A profit maximization approach," *IEEE Trans. on Smart Grid*, vol. 4, no. 2, pp. 1017–1025, June 2013.

[12] L. Rao, X. Liuand, L. Xie, and Z. Pang, "Hedging against uncertainty: A tale of internet data center operations under smart grid environment," *IEEE Trans. on Smart Grid*, vol. 2, no. 3, pp. 555–563, 2011.

[13] 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," in *Proc. of ACM Sigmetrics*, Pittsburgh, PA, 2013.

[14] Z. Liu, I. Liu, S. Low, and A. Wierman, "Pricing data center demand response," (Under Submission) [Online] users.cms.caltech.edu/∼adamw/papers/DCDRpricing.pdf, November 2013.

[15] M. Ghamkhari and H. Mohsenian-Rad, "Data centers to offer ancillary services," in *Proc. of the IEEE Conference on Smart Grid Communications*, Tainan City, Taiwan, October 2012.

[16] M. Shahidehpour, H. Yamin, and Z. Li, *Market Operations in Electric Power Systems*. New York, NY: IEEE Press, 2002.

[17] X. P. Zhang, *Restructured Electric Power Systems: Analysis of Electricity Markets with Equilibrium Models*. Wiley IEEE Press, July 2010.

[18] https://www2.ameren.com/RetailEnergy/realtimeprices.aspx.

[19] http://www.pjm.com/markets-and-operations/energy/real-time/monthlylmp.aspx.

[20] J. H. Kim and M. J. Lee, *Green IT: Technologies and Applications*. Springer, 2011.

[21] X. Fan, W. D. Weber, and L. A. Barroso, "Power provisioning for a warehouse-sized computer," in *Proc. of the ACM International Symposium on Computer Architecture*, San Diego, CA, June 2007.

[22] D. Kusic, J. Kephart, J. Hanson, N. Kandasamy, and G. Jiang, "Power and performance management of virtualized computing environments via lookahead control," in *Proc. of IEEE ICAC*, Nice, France, June 2008.

[23] M. Ghamkhari and H. Mohsenian-Rad, "A Convex Optimization Framework for Service Rate Allocation in Finite-buffer G/D/1 Queuing Systems," (Under Submission), [Online] http://www.ee.ucr.edu/∼hamed/GD12013.pdf, November 2013.

[24] H. Mohsenian-Rad and A. Leon-Garcia, "Coordination of cloud computing and smart power grids," in *Proc. of IEEE International Conference on Smart Grid Communications*, Gaithersburg, MD, Oct. 2010.

[25] ——, "Energy-information transmission tradeoff in green cloud computing," in *Proc. of IEEE Conference on Global Communications (Globecom)*, Miami, FL, December 2010.

[26] R. C. Leou and J. H. Teng, "The optimal portfolio of the day-ahead market and real-time market for the load serving entities," in *Proc. of IEEE INDIN*, Osaka, Japan, July 2010.

[27] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[28] S. J. Deng and S. S. Oren, "Electricity derivatives and risk management," *Energy*, vol. 31, no. 6, pp. 940–953, May 2006.

[29] D. Das and B. F. Wollenberg, "Risk assessment of generators bidding in day-ahead market," *IEEE Trans. on Power Systems*, vol. 20, no. 1, pp. 416–424, February 2005.

[30] D. Bunn, "Forecasting loads and prices in competitive power markets," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 163–169, February 2002.

[31] H. Mohsenian-Rad and A. Leon-Garcia, "Optimal residential load control with price prediction in real-time electricity pricing environments," *IEEE Trans. on Smart Grid*, vol. 1, no. 2, pp. 120–133, 2010.

[32] J. Nair, S. Adlakha, and A. Wierman, "Energy procurement strategies in the presence of intermittent sources," (Under Submission) [Online] http://users.cms.caltech.edu/∼adamw/papers/Wind-preprint.pdf, November 2013.

[33] M. Ghamkhari and H. Mohsenian-Rad, "Profit maximization and power management of green data centers supporting multiple SLAs," in *Proc. of the IEEE International Conference on Computing, Networking and Communications (ICNC'13)*, San Diego, CA, Jan. 2013.

[34] Http://ita.ee.lbl.gov/html/contrib/WorldCup.html.

[35] A. Leon-Garcia, *Probability, Statistics, and Random Processes For Electrical Engineering*, 3rd ed. Prentice Hall, January 2011.