

Optimal Integration of Renewable Energy Resources in Data Centers with Behind-the-Meter Renewable Generator

Mahdi Ghamkhari and Hamed Mohsenian-Rad
 Department of Electrical and Computer Engineering

Texas Tech University, Lubbock, TX, USA

e-mails: {mahdi.ghamkhari, hamed.mohsenian-rad}@ttu.edu

Abstract—Renewable energy resources, such as wind and solar power, are rapidly becoming generation technologies of significance in the United States and around the world. However, because of their *intermittency* and *inter-temporal variations*, the integration of renewable energy resources is usually very challenging. Some of the previously proposed solutions, such as the use of energy storage units and fast-responding generators, are costly and may not be desirable in practice. Instead, in this paper, we argue that due to the *workload flexibility* in the Internet and Cloud Computing data centers they can offer a unique opportunity to tackle the challenges in integrating renewable energy resources. Moreover, by running data centers with behind-the-meter renewable generators, we can significantly reduce not only data centers' *energy cost* but also their Carbon Footprint. To achieve these objectives, we propose to optimally distribute the workload among geographically dispersed data centers such that we can benefit from the *location diversity* of different types of available renewable energy resources. The quality-of-service is guaranteed by real-time monitoring of the queue lengths and effectively *stabilizing* short queue lengths in each data center. Our simulation results, based on various *experimental data*, show that the proposed design can significantly reduce data centers' cost of electricity while maintaining quality-of-service.

I. INTRODUCTION

The growing demand for the Internet services has significantly increased the energy consumed by the Internet and Cloud Computing data centers over the past few years. A typical data center includes hundreds of thousands of computer servers, cooling equipment, and substation transformers. For example, Microsoft's data center in Quincy, WA consumes 48 megawatts electricity which is enough to power 40,000 homes [1]. It is estimated that the total electricity cost of servers and data centers in the United States is \$7.4 Billion annually [2].

The increasing energy cost in data centers has motivated a growing number of research projects not only to reduce the amount of energy consumed by computer servers [3], [4], but also to design resource management algorithms to run data centers more efficiently. One approach in this line of research is *dynamic cluster server configuration* to reduce the total power consumption by consolidating load on only a subset of machines and turning off the rest, during low workload periods when the system is under-utilized. Additional servers are powered on only when the workload increases [5], [6]. Another approach is to *dynamically scale the CPU clock frequency* of the computer servers in response to the varying workload [7], [8]. A high frequency, imposing higher energy consumption, is chosen only at peak workload hours. This allows the energy consumption at data centers to be *elastic*.

Another thread of research, that has emerged only recently, aims to take the advantage of *price-diversity* in the deregulated electricity markets to develop algorithms that distribute the workload among data centers in multiple locations to minimize the total cost of electricity of the data centers [9], [10]. The idea is to constantly monitor the price of electricity at different regions and forward the workload towards data centers that are located in regions with the lowest price of electricity. Migrating an existing workload from one data center to another to benefit from the lower price of electricity in the new location has also been recently examined in [11].

In this paper, we take a very different approach and investigate reducing the energy cost of data centers, where each data center is equipped with a local *behind-the-meter renewable power generator*. In general, the main challenges in integrating renewable energy resources are their *intermittency* and *inter-temporal variations* [12]. However, we will show that due to the elasticity of data centers' energy consumption, they can be among the most appropriate types of load to be run by renewable power. In this regard, our focus is on developing optimal workload distribution algorithms that distribute the incoming service requests among geographically dispersed data centers based on renewable power availability. To the best of our knowledge, this paper is the first to investigate the integration of behind-the-meter renewable power generators into the Internet and cloud computing data centers using workload distribution with explicit quality-of-service provisioning. The results in this paper can complement those in [3]–[11] not only by further reducing data center's total energy cost but also by advancing the design and development of *green* data centers. The contributions in this paper can be summarized as follows.

- We develop a mathematical framework to investigate optimal workload distribution among geographically dispersed data centers with different types of behind-the-meter renewable generators. In this regard, we take into account computer servers' power consumption profiles, data centers' power usage effectiveness, total workload in terms of the number of service requests received at each time of day, instantaneous queue lengths at each data center, availability of renewable power at different locations, and price of electricity at different locations.
- We propose an *optimization-based workload distribution* algorithm that allocates the workload among data centers such that for each data center its total power consumption can follow the time-varying trend of power generated by

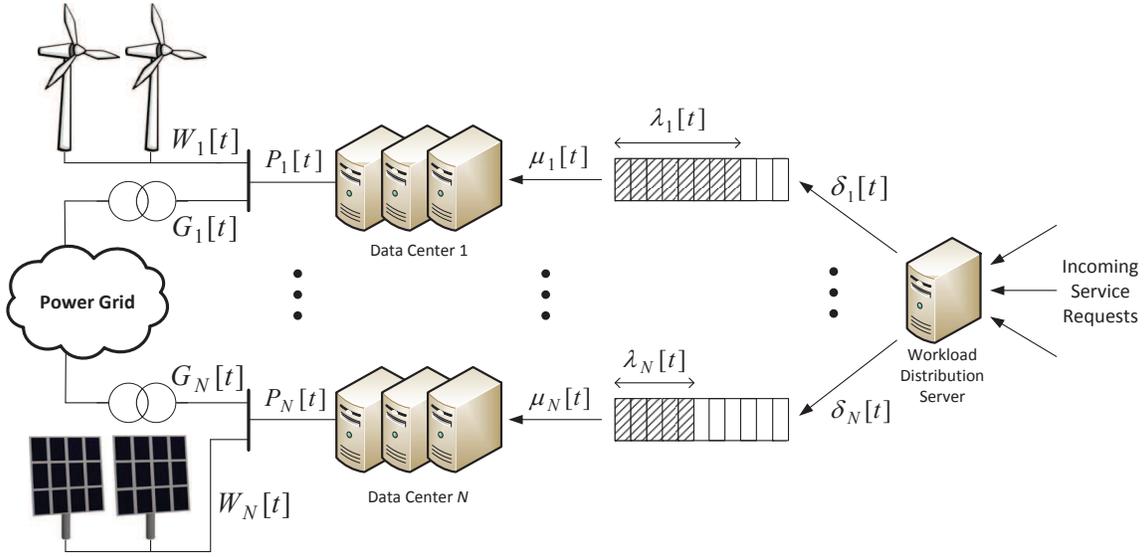


Fig. 1. The system model studied in this paper. The load assignment server distributes service requests among N geographically dispersed green data centers.

the data center's behind-the-meter renewable generators. Moreover, our design constantly monitors the number of service requests waiting at data centers' queues and seeks to stabilize short queues to assure quality-of-service.

- We use various *experimental data*, e.g., for wind power generation, time-of-day price of electricity, and workload trends, to assess the performance of the proposed optimal workload distribution scheme via computer simulations. Simulation results over a 10 days service period show that our design can significantly reduce data centers' cost of electricity while maintaining the quality-of-service.

The rest of this paper is organized as follows. The system model is introduced in Section II. It provides mathematical models for power consumption at data centers, workload distribution, and renewable power generation. The proposed optimal workload distribution algorithm is presented in Section III. The performance of the proposed algorithm is evaluated using experimental data and computer simulations in Section IV. Conclusions and future work are discussed in Section V.

II. SYSTEM MODEL

Consider a group of N data centers that are geographically dispersed at different locations to offer the *same types of service* such as content distribution [13], [14] or cloud computation [15]. We are interested in a scenario where each data center is equipped with a *behind-the-meter renewable power generator*. Such generator can be, for example, a wind farm or an array of solar panels, which are the most common types of commercial renewable generators [12], [16]. The system model to be investigated in this paper is shown in Fig. 1.

A. Workload Distribution

As a key element in the system model, a *workload distribution server* is used in order to distribute the incoming service requests from all users among the data centers as shown in Fig. 1. We assume that time is divided into several equal-length time slots. For example, each time slot can indicate one

second. Let $L[t]$ denote the total number of service requests received by the workload distribution server from all users at time slot t . Also let $\delta_i[t]$ denote the number of service requests that are forwarded to each data center i . In order to assure that all service requests will be handled, we should always have

$$\sum_{i=1}^N \delta_i[t] = L[t]. \quad (1)$$

The service requests forwarded to each data center are enqueued until they are handled in a first-in-first-served order. The rate at which service requests are pulled from the queue to be handled in data center i at time slot t is denoted by $\mu_i[t]$. Such rate can be adjusted by changing the CPU frequency of the data center's computer servers or by changing the utilization level of the computer clusters. Depending on the type and number of service requests that each computer server can handle and also based on the total number of computer servers at each data center, for each data center i and at each time slot t , it is required that we select $\mu_i[t]$ such that

$$\mu_i^{\min} \leq \mu_i[t] \leq \mu_i^{\max}, \quad (2)$$

where μ_i^{\min} and μ_i^{\max} are parameters specific to data center i .

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 each 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 data center's power consumption at the computer servers [2]. The PUE is considered as a measure for data center's energy efficiency. Currently, the typical value for most enterprise data centers is 2.0 or more. However, recent studies have suggested that many data centers can soon reach a PUE of 1.7. A few state-of-the art facilities have reached a PUE of 1.2 [2].

Let P_{idle} denote the average *idle power* draw of a single computer server in a data center and P_{peak} denote the average peak power when a computer server is handling a service request. The ratio $P_{\text{peak}}/P_{\text{idle}}$ denotes the power elasticity of the servers. A higher value of this ratio indicates greater elasticity, leading to less power consumption when the computer server is idle not handling any service request. Let m_i denote the number of computer servers that are ‘on’ at data center i . Also let κ denote the total number of service requests that each computer server can handle at each time slot. We can obtain the *total power consumption* at data center i as a function of the rate of handling service requests $\mu_i[t]$ [17] as follows:

$$P_i(\mu_i[t]) = a_i \mu_i[t] + b_i, \quad (3)$$

where

$$a_i \triangleq (P_{\text{peak}} - P_{\text{idle}})/\kappa, \quad (4)$$

$$b_i \triangleq m_i (P_{\text{idle}} + (E_{\text{usage}} - 1) P_{\text{peak}}). \quad (5)$$

Here, b_i is the *base load*, i.e., the amount of power consumption even if some of the turned ‘on’ computer servers are idle. On the other hand, $a_i \mu_i[t]$ is an *added load* which indicates the extra power consumption depending on the rate at which service requests are handled at data center i at time slot t . The faster service requests are handled at a data center the higher the data center’s total power consumption will become.

We note that, in addition to carefully adjusting the rate of handling service request, i.e., μ_i , the power consumption at data center i can be controlled by also adjusting the number of ‘on’ servers, i.e., m_i . This can be done easily, for example, by turning on only as many servers as are needed to handle the expected number of incoming service requests at each data center, given a reasonable margin to keep some additional computer servers on stand-by mode at each time of day [18].

C. Renewable Power Generation

Due to the *intermittency* of most renewable energy resources, the amount of renewable power available at each data center can significantly *fluctuate* at different times of day. Therefore, in addition to a local renewable generator, each data center is also equipped with a sub-station that is connected to the electric grid in order to compensate lack of local generation when the amount of renewable power is not enough to run data centers. Let $W_i[t]$ denote the renewable power generated by each data center i ’s renewable generator. Also let $G_i[t]$ denote the amount of power exchange with the power grid. We have

$$G_i[t] = P_i(\mu_i[t]) - W_i[t]. \quad (6)$$

If local renewable generation is lower than local consumption, i.e., $P_i(\mu_i[t]) > W_i[t]$, then $G_i[t]$ is positive and the power flow is in the direction from the power grid to the data center. If $P_i(\mu_i[t]) = W_i[t]$, then the data center operates as a *zero-net energy facility* [19]. Since the behind-the-meter renewable power markets have not been established in most regions yet, for the special case when local renewable generation is greater than local consumption, we make the assumption that the amount of renewable power generated will be instantly reduced to match local consumption. This can be done, for

example, by braking the rotors for a sub-set of wind turbines or by applying dynamic *pitch* and *yaw* control [20]. The alternative scenario where data centers *sell* their excessive renewable power generation is beyond the scope of this paper and can be considered as an interesting extension in a future study. Therefore, for the rest of this paper, we assume that $G_i[t] \geq 0$ for all $i = 1, \dots, N$ and at any time slot t .

D. Renewable Power Prediction

The amount of renewable generation at each data center can be predicted for the next few minutes (short term predictions) based on *past measurements*. In this regard, at the beginning of each time slot t and for each data center i , the predicted renewable power can be denoted as [21]:

$$\hat{W}_i[t] = \mathbb{E} \{W_i[t] | W_i[t-1], W_i[t-1], \dots\}, \quad (7)$$

where \mathbb{E} denotes conditional expectation. For example, when it comes to wind power, the prediction in (7) can be done using Markov Chain models based on monthly wind speed statistics [22]. As we will explain in Section III, the predictions on available renewable power in (7) can be used to adjust the operation of the load assignment server and the data centers such that the total cost of electricity is minimized.

E. Locational Price of Electricity

The price of electricity may vary at *different locations*. For example, the average price of electricity is higher in Texas compared to its neighboring states of Oklahoma and Louisiana [23]. Moreover, the price of electricity may vary at *different times of day* when *real-time* or *time-of-use* pricing tariffs are implemented [24]. The prices are usually higher at *peak load hours* in the afternoon [25]. We denote the price of electricity at data center i and at time slot t by $\omega_i[t]$.

III. OPTIMAL WORKLOAD DISTRIBUTION ALGORITHM

In this section, we propose an optimization-based workload distribution algorithm that allocates the workload among data centers such that for each data center its power consumption follow the time-varying trend of power generated while taking into account the price of electricity. As we will see in Section IV, the proposed design can significantly reduce data centers’ electricity cost while maintaining quality-of-service.

A. Quality-of-Service

Let $\lambda_i[t]$ denote the length of the queue in data center i at time slot t . If $\mu_i[t] < \delta_i[t]$ then the queue length increases and at the next time slot $\lambda_i[t+1]$ will be greater than $\lambda_i[t]$. Similarly, if $\mu_i[t] > \delta_i[t]$ then the queue length decreases and at the next time slot $\lambda_i[t+1]$ will be less than $\lambda_i[t]$. In fact, given $\lambda_i[t]$ and based on the choice of $\mu_i[t]$ and $\delta_i[t]$ we can predict $\lambda_i[t+1]$ at the beginning of time slot $t+1$ as

$$\hat{\lambda}_i[t+1] = \lambda_i[t] - \mu_i[t] + \delta_i[t]. \quad (8)$$

Note that, as the queue length increases, it takes longer to respond to the incoming service requests. However, most service-level agreements, e.g., for web-based services, require

data centers to limit the delay in responding to service requests they receive. For example, based on some commonly used step-wise service-level agreements, a data center / service provider is paid only if the response time is less than 200 ms [15]. If the response time goes beyond this limit, the data center will be penalized. Therefore, to assure quality-of-service and to limit the response time within an acceptable range, we can impose an upper bound D on queue lengths. Therefore, we should control the operation of data centers such that we always have $\hat{\lambda}_i[t+1] \leq D$ for all $i = 1, \dots, N$.

B. Optimization Problem

At the beginning of each time slot t , we propose to solve the following problem at the workload distribution server:

$$\begin{aligned} & \underset{\delta[t], \mu[t]}{\text{minimize}} && \sum_{i=1}^N \omega_i[t] \left(a_i \mu_i[t] + b_i - \hat{W}_i[t] \right)^2 \\ & && - \sum_{i=1}^N \psi_i \lambda_i[t] \mu_i[t] \\ & \text{subject to} && \sum_{i=1}^N \delta_i[t] = L[t], \\ & && \lambda_i[t] + \delta_i[t] - \mu_i[t] \leq D, \quad i = 1, \dots, N, \end{aligned} \quad (9)$$

where $\delta_i[t] \geq 0$ and $0 \leq \mu_i[t] \leq \mu_i^{\max}$ for all $i = 1, \dots, N$. Here, there are two set of variables: $\delta[t] = \{\delta_1[t], \dots, \delta_N[t]\}$ and $\mu[t] = \{\mu_1[t], \dots, \mu_N[t]\}$. They both affect the queue lengths and quality-of-service. The latter also affects total power consumption across data centers. Note that, there two components in the objective function in problem (9). The first one, i.e., $\sum_{i=1}^N \omega_i[t] (a_i \mu_i[t] + b_i - \hat{W}_i[t])^2$, is a *weighted summation* of N quadratic terms, each one indicating the difference between the predicted amount of renewable power to be generated and the total power to be consumed at each data center at the next time slot, where the weights are the locational electricity prices. By minimizing the first component in the objective function, we seek to minimize the difference between local renewable generation and local power consumption (this minimizes the total electricity to be purchased from the power grid), particularly in those regions where the price of electricity is higher. The second component, i.e., $\sum_{i=1}^N \psi_i \lambda_i[t] \mu_i[t]$, is added to the objective function for the purpose of *stabilizing* the queues at all data centers in order to enforce quality-of-service, together with constraint $\lambda_i[t] + \delta_i[t] - \mu_i[t] \leq D$ for all $i = 1, \dots, N$. Note that, ψ_i is a design parameter for each data center and is used to adjust the trade-off between cost minimization and improving the quality-of-service. As the number of service requests waiting in data center i 's queue, i.e., queue length $\lambda_i[t]$, increases, we push the optimal solution towards increasing μ_i ; thus, handling service requests faster and also tackling incoming bursts.

C. Implementation

Problem (9) is a quadratic convex optimization problem and can be solved effectively at the workload distribution server using convex programming techniques such as the *interior*

point method (IPM) [26]. The optimal solution can then be implemented by properly modifying some of the existing workload distribution software such as *Apache Hadoop* [27] which is commonly used, e.g., by Google and Yahoo! for Web search query. In this regard, the workload distribution server will periodically collect data center's queue length and other local information and will constantly update the workload distribution plan according to the optimal solution of the optimization problem (9), e.g., once every few seconds.

IV. PERFORMANCE EVALUATION

A. Simulation Setting

Consider $N = 3$ data centers that are located at three geographical regions and all offer the same type of Internet or cloud computing service. We assume that there is a local wind farm associated with each data center, each one comprising of 20 General Electric (GM) 1.5 megawatts wind turbines. The trends of wind power in our simulations are based on the measurements at three wind farms in Texas [22]. A daily example for the considered wind power trends are shown in Fig. 2(a). The electricity price information in our simulations are based on the real-time pricing tariffs currently practiced in three different areas in Chicago, IL [24]. An example for daily prices are shown in Fig. 2(b). Finally, to simulate the total workload, we used the World Cup 98 web requests which are available in [28]. As an example, the number of service requests received at the World Cup 98 servers on June 2, 1998 are shown in Fig. 2(c). Each time slot is assumed to take one second. We assume that each data center i has up to 400,000 servers. The exact number of turned on servers varies and is set at each time of day based on the mechanism explained at the end of Section II-B. For each server, $P_{\text{peak}} = 200$ watts and $P_{\text{peak}} = 100$ watts. We also have $E_{\text{usage}} = 1.2$ which is the reported state of the art power usage effectiveness [2]. The type of service to be offered by data centers is assumed to be video streaming. Each service takes 45 minutes which is equal to the time it takes for each soccer game in each half time. Therefore, we have $\kappa = \frac{4}{3} \times \frac{1}{3600}$ requests per second.

B. Cost Reduction and Quality-of-Service

The optimal workload distribution for a single day simulation is shown in Fig. 2(d) where the simulation data are as in Fig. 2(a)-(c). We can see that the optimal workload distribution is *not* uniform and some data centers receive more workload at different times of day. Let us examine the two scenarios marked in this figure. For Scenario 1 between hours 7 and 8, the total incoming workload in the system is *low*, as we can see in Fig. 2(c). Therefore, it is possible to handle most of the workload using *only* wind power, making wind power availability the key factor in decision making. As a result, the majority of the workload is forwarded to Data Center 1 which has significantly higher wind power availability at this time of day, as we can see in Fig. 2(a). However, for Scenario 2 between hours 20 and 21, the total incoming workload in the system is *high*. Therefore, the wind power available is *not* enough to run data centers and we need to purchase electricity from the grid, making the price of electricity the key factor for

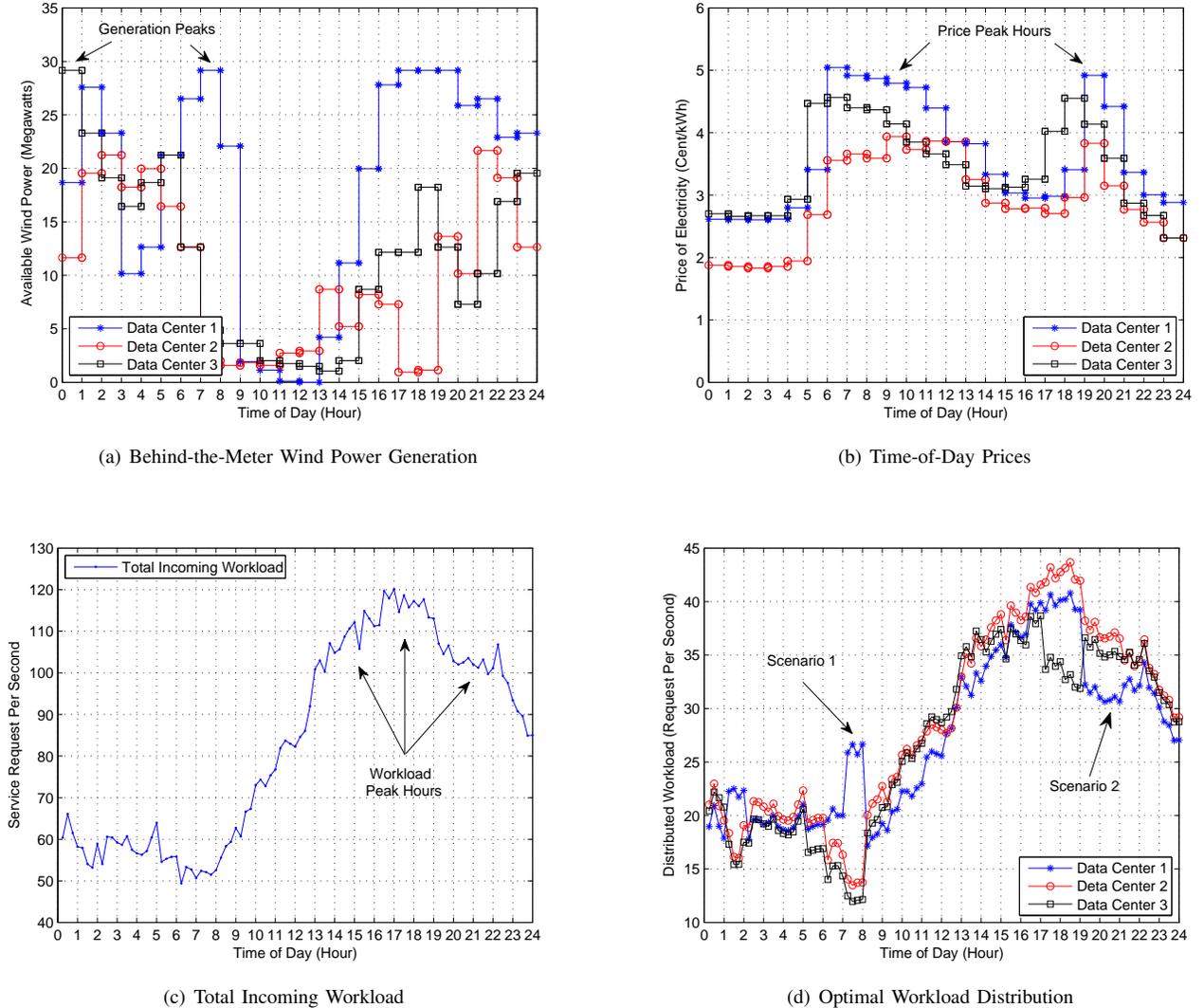


Fig. 2. An example daily simulation results: (a)-(c) The experimental data used for simulations. (d) The resulting optimal workload distribution.

decision making in this case. As a result, the majority of the workload is forwarded to Data Center 2 which has the lowest price at this time of day, as we can see in Fig. 2(b). Finally, we can see from the results in Fig. 3 that the optimal workload distribution scheme can stabilize queues and maintain short queue lengths in all data centers to *assure quality-of-service*.

Comparing to a uniform workload distribution plan that allocates workload among data centers *equally*, the optimal workload distribution scheme can significantly reduce the cost of running data centers as shown in Fig. 4(a). We can see that the proposed design outperforms uniform workload distribution at any time of day. Repeating the simulations for 10 different days, based on the World Cup 98 workload data from May 28, 1998 to June 6, 1998, we can see that the cost of running data centers will reduce by 29.2%, i.e., \$8,797 on average, leading to a total saving of \$87,970 over 10 days.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed an optimization-based workload distribution framework for Internet and cloud computing data

centers with behind-the-meter renewable generators. In this regard, we took into account *computer servers' power consumption profiles, data centers' power usage effectiveness, total workload in terms of the number of service requests received at each time of day, instantaneous queue lengths at each data center, availability of renewable power at different locations, and price of electricity at different locations*. Each data center is designed to adapt its power consumption to constantly follow the time-varying trend of renewable power generated while the number of service requests waiting at the data centers' queues are monitored to assure quality-of-service. Using various experimental data, simulation results show that the proposed design can significantly reduce data centers' cost of electricity while maintaining quality-of-service.

This paper can be extended in several directions. For example, we are currently working on also incorporating service differentiation, statistical whether forecasting information, and propagation delays between user and data center locations into the problem formulation to achieve more accurate designs.

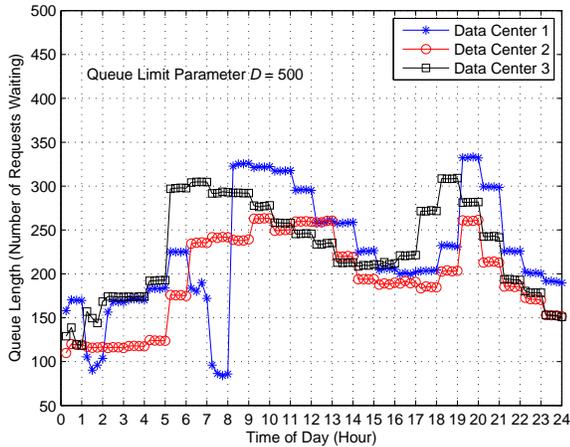
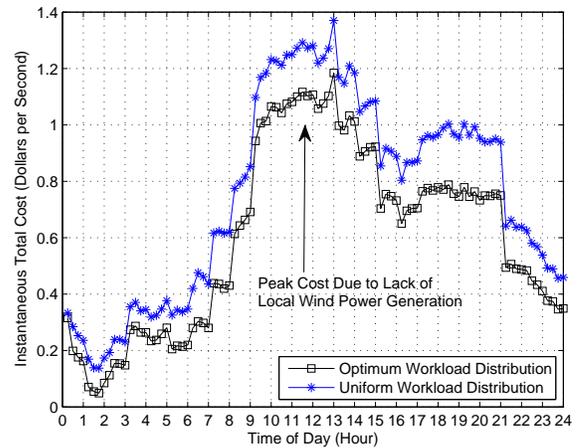


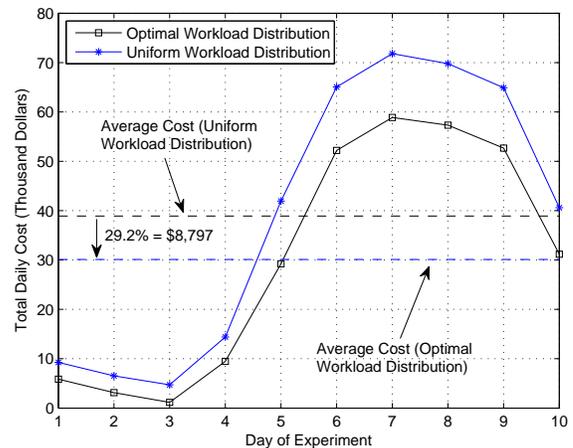
Fig. 3. The queue lengths are stabilized for the data in Fig. 2.

REFERENCES

- [1] R. H. Katz, "Tech titans building boom," *IEEE Spectrum*, pp. 40–54, Feb. 2009.
- [2] United States Environmental Protection Agency, "EPA report on server and data center energy efficiency," Final Report to Congress, Aug. 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, Oct. 2002.
- [4] S. Gurumurthi, A. Sivasubramaniam, M. Kandemir, and H. Franke, "Reducing disk power consumption in servers with DRPM," *Computer*, vol. 36, no. 12, pp. 59–66, Dec. 2003.
- [5] 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.
- [6] J. Heo, D. Henriksson, X. Liu, and T. Abdelzaher, "Integrating adaptive components: An emerging challenge in performance-adaptive systems and a server farm case-study," in *Proc. of the IEEE International Real-Time Systems Symposium*, Tucson, AZ, Dec. 2007.
- [7] 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.
- [8] J. S. Chase, D. Anderson, P. Thakar, A. Vahdat, and R. Doyle, "Managing energy and server resources in hosting centers," in *Proc. of ACM Symp. on Operating systems principles*, Banff, Canada, Oct. 2001.
- [9] J. Li, Z. Li, K. Ren, X. Liu, and H. Su, "Towards optimal electric demand management for internet data centers," *IEEE Transactions on Smart Grid (to appear)*, 2011.
- [10] X. Zheng and Y. Cai, "Energy-aware load dispatching in geographically located internet data centers," *Sustainable Computing (to appear)*, 2011.
- [11] N. Buchbinder, N. Jain, and I. Menache, "Online job-migration for reducing the electricity bill in the cloud," in *Proc. of International IFIP TC 6 Networking Conference*, Valencia, Spain, May 2011.
- [12] S. Rahman, "Going green - the growth of renewable energy," *IEEE Power and Energy Magazine*, vol. 1, pp. 16–18, Nov. 2003.
- [13] 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.
- [14] C. Cranor, M. Green, C. Kalmanek, D. Shur, S. Sibal, J. V. der Merwe, and C. Sreenan, "Enhanced streaming services in a content distribution network," *IEEE Internet Computing*, vol. 5, pp. 66–75, July 2001.
- [15] H. Mohsenian-Rad and A. Leon-Garcia, "Coordination of cloud computing and smart power grids," in *Proc. of the IEEE Smart Grid Communications Conference*, Gaithersburg, MD, Oct. 2010.
- [16] A. Ipakchi and F. Albuyeh, "Grid of the future," *IEEE Power and Energy Magazine*, vol. 7, no. 2, pp. 52–62, Mar. 2009.
- [17] 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.



(a) Instantaneous Total Cost for a Sample Day



(b) Total Daily Cost Over 10 Days

Fig. 4. Total cost of running all three data centers when the optimal workload distribution is compared with a uniform workload distribution: Instantaneous cost for the daily data in Fig. 2; (b) Total daily cost for simulating 10 days.

- [18] Z. Liu, M. Lin, A. Wierman, S. Low, and L. L. H. Andrew, "Geographical load balancing with renewables," in *Proc. of ACM GreenMetrics Workshop*, San Jose, CA, Apr. 2011.
- [19] P. Torcellini, S. Pless, M. Deru, and D. Crawley, "Zero energy buildings: A critical look at the definition," in *Proc. of the ACEEE Summer Study on Energy Efficiency in Buildings*, Aug. Pacific Grove, CA.
- [20] E. Muljadi and C. P. Butterfield, "Pitch-controlled variable-speed wind turbine generation," *IEEE Transactions on Industry Applications*, vol. 37, no. 1, pp. 240–246, Jan. 2001.
- [21] M. Lange and U. Focken, *Physical Approach to Short-Term Wind Power Prediction*. Berlin, Germany: Springer-Verlag, 2005.
- [22] C. Wu, H. Mohsenian-Rad, J. Huang, and A. Y. Wang, "Demand side management for wind power integration in microgrid using dynamic potential game theory," in *Proc. of IEEE GLOBECOM Workshop on Smart Grid Communications and Networking*, Houston, TX, Dec. 2011.
- [23] U.S. Energy Information Administration, "Independent statistics and analysis," http://www.eia.gov/cneaf/electricity/epm/table5_6_a.html.
- [24] H. Mohsenian-Rad and A. Leon-Garcia, "Optimal residential load control with price prediction in real-time electricity pricing environments," *IEEE Transactions on Smart Grid*, vol. 1, no. 2, pp. 120–133, 2010.
- [25] H. Allcott, "Real time pricing and electricity markets," Working Paper, Harvard University, Feb. 2009.
- [26] S. Boyd and L. Vandenbergher, *Convex Optimization*. Cambridge University Press, 2004.
- [27] Apache Hadoop, <http://hadoop.apache.org>, 2011.
- [28] <http://ita.ee.lbl.gov/html/contrib/WorldCup.html>.