

Propagating Electricity Bill onto Cloud Tenants: Using a Novel Pricing Mechanism

Yong Zhan^{*§}, Mahdi Ghamkhari[§], Du Xu^{*}, Hamed Mohsenian-Rad[§]

^{*}Key Laboratory of Optical Fiber Sensing and Communications, Ministry of Education,

University of Electronic Science and Technology of China, Chengdu, P. R. China

[§]Department of Electrical and Computer Engineering, University of California, Riverside, CA, USA

Abstract—Data centers spend millions of dollars on their electricity bills annually. Therefore, there is an interest among data center operators to control the electricity usage so as to minimize energy expenditure. However, when it comes to cloud data centers, the electricity usage is mainly controlled by the tenants. Yet, since most cloud data centers charge their tenants with flat rates, the tenants do not have incentives to change their electricity usage to contribute in cutting electricity bills by participation in demand response. Accordingly, in this paper, we propose a game-theoretic framework together with a time varying pricing (TVP) mechanism for cloud data centers to charge their tenants. In this approach, the TVP propagates the actual energy bill, which comprises both *demand* and *energy* charges, onto tenants' service costs. As an extension to this core idea, the time-varying amount of renewable energy generated by data center's on-site renewable generators is also taken into account to affect the payments. Our proposed pricing method is evaluated under various experimental data and simulations. We show that TVP can boost data center's profit by 8.2% and reduce the energy bill by 33.0% and improve tenants' aggregate surplus by 12.3%, when comparing it with a flat rates model that uses a widely employed billing method in today's cloud data centers.

I. INTRODUCTION

In 2011, data centers (DCs) consumed about 1.5% of the total generated electricity worldwide [1]–[3]. In 2012, this portion increased by 63% [4]. This growing trend is expected to continue over the next decade [5]. Hence, the cost of electricity will continue to be a major factor for data centers.

The electricity utilities usually charge their large consumers, such as DCs, via Peak Pricing (PP), which comprises *energy charge* and *demand charge* [6]–[8]. The energy charge is the cost of total kWh electricity consumed by the DC. The demand charge is about the average peak load in kW, e.g., within 1 hour intervals, during a billing cycle, e.g., a month. The demand charge is deployed in order to shed the electricity utilities' peak loads. The demand charge of a DC can be equal to or even exceed the energy charge [9], [10].

Two practical methods have been discussed in the literature for reducing electricity bill of a data center. The first method is to generate renewable energy (e.g., wind power and solar power) to power their own facilities [11]. However, renewable generation is intermittent and its amount depends on external factors, e.g. wind speed or solar irradiance. The second method is to better management power consumption of computation

servers by turning off the idle servers. Of course, this has to be done by in a careful and coordinated way in order to assure quality-of-service (QoS) requirements. Otherwise, the performance of DC may suffer a sudden degradation, violating the QoS terms in service-level-agreements (SLAs).

Most prior work on cloud data center energy management focuses on how a DC should respond to the energy price to reduce its energy bill [12]–[17]. However, the fact that *these are the tenants but not the data center itself that should respond to such price changes* is often neglected. For example, today's DCs charge their tenants via flat rates [18], i.e., tenants' payments depend on the number of machines they rent and the time of renting the machines, but not how the tenants utilize their rented machines. Accordingly, the tenants have no incentive to optimize the use of their computation resources to minimize their payments. This issue can be resolved if the DCs *propagate the energy bill onto their tenants service payments*. In this case, tenants can be aware of their contributions to the overall energy cost in the data center system and can receive some monetary incentives to adjust their workloads properly.

Interestingly, the DC operators have already pointed out their interests in using pricing mechanisms that provide tenants with right incentives to proactively manage their workloads: For instance, according to a Microsoft whitepaper, *“To further smooth demand, sophisticated pricing can be employed. For example, similar to the electricity market, customers can be incentivated to shift their demand from high utilization periods to low utilization periods. Demand management will further increase the economic benefits of cloud”* [19]. However, the technical challenges are yet to be addressed.

In this paper, we design a pricing method to encourage tenants to response to time-varying and complex electricity rates. Tenants are allowed to periodically adjust their purchases of machines based on the current price of machine and their real demands. Data centers can then turn on/off the machines accordingly to maximize their profit, i.e., their revenue minus energy cost. In summary, our contributions are as follows:

- We model the aggregate utility function of tenants in terms of the number of active machines. We also model the profit of DC by taking into account the bursty nature of tenants' workloads, QoS, DC's power usage effectiveness, the PP for energy bill, the time-varying amount of generation of renewable power.

- We propose a game-based framework for a DC to deduce the Time Varying Pricing (TVP) to charge its tenants. Specifically, we assume that tenants decide their aggregate purchase of machine towards maximizing their aggregate surplus (i.e., aggregate utility minus cost). Further, DC can deduce the optimal price of machine that can optimize its profit based on the knowledge of its tenants' aggregate utility and the real time energy price.
- We evaluate our proposed pricing with various experimental data, such as real workload data, various electricity rates at different locations and real renewable energy generation data. Meanwhile, we compare our proposed pricing with a representative of the pricing that is widely adopted in today's DCs. We demonstrate that our proposed pricing can boost DC's profit and reduce its energy bill while improving tenants' aggregate surplus.

II. RELATED WORK

We first sketch the pricing policies that are widely employed by today's DCs. Traditionally, cloud providers prefer simple pricing methods which are easier to understand by users, such as flat rates and usage-based pricing [20]. For instance, Amazon provides its EC2 tenants with flat rates, e.g., \$0.140 per hour per `m3.large` instance [18]. Clearly, such flat rates do not offer tenants any incentive to shed their peak demands or reduce their demands when the price of energy is relatively high. To ensure that the scarce DC resources are consumed by tenants who value them most, Amazon has also implemented a novel machine purchase option: Spot Instances, where tenants must bid for the spare instances and the DC will then allocate resources to the tenant with the highest bid. But this still does not propagate the energy bill onto tenants' payments.

Next, we briefly discuss demand response policies in DCs. In [16], a mathematical framework is developed for DCs to decide the optimal number of machines that are needed to be switched on in order to maximize DC's profit. In [9], the authors explained that the data center is more likely to be charged with PP than any other pricing tariff. Therefore, they focus on responding to PP scenarios. The authors also adopted a partial execution of service requests to reduce the DC's energy bill, based on the assumption that many DC services can tolerate. On a related work, in [21], the authors designed a hierarchical framework to minimize DC's energy bill via both partial execution and workload deferring mechanisms, where the workload deferring can be adopted if the workload is latency-tolerable, e.g., for batch workload. Note that, to be effective in a cloud data center, the above methods need to adjust tenant's workload. However, they neglect that the tenants, who actually own the workload, may not allow DC to drop or defer their workload if they are not well-incentivized.

Another branch of prior work aims at providing tenants with incentives to cooperate with DCs to reduce the energy cost. In [22], the authors proposed Virtual Electric Utility (VEU) that can fairly propagate the energy cost onto tenants' costs. However, it is assumed that the DC needs to inform the tenants

of the whole information of energy bill and their contributions to the bill. As a result, the scheme in [22] is quite complex and may not be viable to be implemented by DCs [20]. Also, in [23], the authors presented an auction-based pricing method, in which tenants bid for the machines they need by submitting a bid function that shows the tenants' willingness to pay if the tenant can obtain those machines before period t . However, the proposed method only takes into account the time-varying electricity rates, which is not likely to be used for charging large electricity consumers such as DCs. Next, we address this shortcoming by considering peak hour pricing as a more common pricing tariff that is deployed for large DCs [6]–[8].

III. PROBLEM FORMULATION

A. Data Center's Energy Bill and Profit

We divide the operating time of DC into time slots of length T and assume that a billing cycle contains τ time slots. Next, we provide a model for the DC's electricity bill and profit at each time slot $t \in \{1, \dots, \tau\}$. Consider a DC that procures energy from both a local renewable energy generator such as wind turbines by itself and also an electricity utility via a pre-assigned wholesale contract. We assume that the pricing tariff between the DC and electricity utility is peak hour pricing and denote the Energy charge and Demand charge at time slot t by $\alpha(t)$ per KWh and β per KW, respectively. We denote the amount of renewable power that is generated at time slot t by G_t and the amount of renewable generation that is not used by DC is injected to the power grid or will be dropped. However, DC is not paid for the power it injects to the power grid. Thus, the energy bill of DC over a billing cycle is obtained as

$$\begin{aligned} \text{Energy Bill} = & \sum_t \alpha(t)T \max\{P_t - G_t, 0\} \\ & + \beta \max_t \{\max\{P_t - G_t, 0\}\}, \end{aligned} \quad (1)$$

where P_t is the average power usage of DC at time slot t .

Let m_t denote the total amount of machines that are purchased by the tenants of DC at time slot t . In this paper, TVP is adopted to charge DC's tenants, i.e., the price of machines can be changed over the time. Let δ_t denote the price of a machine at time slot t . The total revenue of DC over a billing cycle is obtained as

$$\text{Revenue} = \sum_t \delta_t m_t. \quad (2)$$

We assume that the power consumption of DC is a linear function of the number of switched on machines. The average power consumption at time slot t is defined as:

$$P_t = E_{pue} \rho m_t, \quad (3)$$

where E_{pue} is the Power Usage Effectiveness (PUE) and ρ is a machine's average power draw. From Equation (1) (2) and (3), the total profit of data center over a billing cycle is obtained as

$$\begin{aligned} \text{Profit} = & \sum_t (\delta_t m_t - \alpha(t)T \max\{E_{pue} \rho m_t - G_t, 0\}) \\ & - \beta \max_t \{\max\{E_{pue} \rho m_t - G_t, 0\}\}. \end{aligned} \quad (4)$$

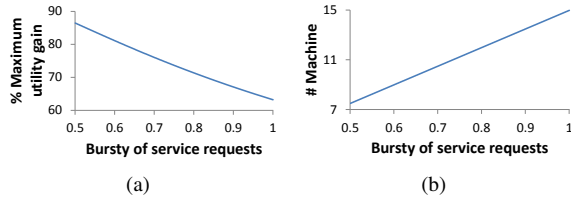


Fig. 1. Impact of γ_t : (a) % Maximum utility gained by tenants when #machine=5 under different γ_t . (b) #machine needed to gain 95th-percentile of maximum utility under different γ_t

B. Tenants' Aggregate Utility and Surplus

We assume that the aggregate utility of tenants is a non-decreasing and concave function of the number of their purchased machines. Meanwhile, we assume that an average of λ_t service requests are generated by the tenants at time slot t . Also, let κ denote the number of service requests that can be handled by a switched on machine at DC. If the requests arrive uniformly, $\lambda_t/(\kappa T)$ machines are sufficient to handle the service requests. However, considering the bursty nature of service requests arrival rates, tenants need more machines than $\lambda_t/(\kappa T)$ to process their service requests without degradation of QoS. Therefore, we model the aggregate utility of tenants by

$$U(m_t) = V_t \lambda_t \left(1 - \exp\left(-m_t \frac{\kappa T}{\gamma_t \lambda_t}\right) \right), \quad (5)$$

where $V_t \lambda_t > 0$ is the maximum utility that can be gained by the tenants at time slot t and $\gamma_t > 0$ is a parameter to model the burstiness of the service requests. As an illustration example, Figure 1 shows the impact of γ_t on tenants' aggregate utility, where $\lambda_t/(\kappa T)$ is set to be 5. Specifically, Figure 1(a) shows how many percent of utility can be gained by the tenants when they purchase 5 machines. Meanwhile, Figure 1(b) shows the number of machines that are needed by the tenants if they want to gain 95th-percentile of maximum utility. From (5), the aggregate surplus gained by the tenants at time slot t is

$$S(m_t) = V_t \lambda_t \left(1 - \exp\left(-m_t \frac{\kappa T}{\gamma_t \lambda_t}\right) \right) - \delta_t m_t. \quad (6)$$

IV. DESIGN OF GAME-THEORETIC FRAMEWORK

We aim at maximizing DC's profit in this paper. To achieve this goal, DC needs to provide its tenants with right incentives (i.e., monetary incentives). Namely, DC will propagate its energy bill onto the price of machine. In this case, tenants of DC may reduce their demands at the peak hour or/and when the energy price is relatively high. By increasing the price of machines, DC gains more profit per each purchased machine by tenants. However, with the increasing of price, tenants may purchase less machines, which decreases the total profit of DC. To capture this trade-off, in this section, we design a game-based price framework for DC to optimize its profit.

A. Tenants' Optimal Machine Purchase

In this section, we analyse tenants' aggregate response to the price of machine. To ensure the quality of service, we assume that the number of machines that are purchased by tenants at time slot t is lower bounded by M_t , i.e., $m_t \in [M_t, \infty)$. We assume that the lower bound is proportional to tenants' aggregate service requests λ_t :

$$M_t = l \lambda_t, \quad (7)$$

where $l \geq 0$ is a parameter that is set by the tenants. The optimum number of machines that should be purchased to maximize the tenants' aggregate surplus is obtained by solving the following optimization problem

$$\begin{aligned} & \underset{m_t}{\text{maximize}} && V_t \lambda_t \left(1 - \exp\left(-m_t \frac{\kappa T}{\gamma_t \lambda_t}\right) \right) - \delta_t m_t \\ & \text{subject to} && m_t \geq M_t. \end{aligned} \quad (8)$$

The above optimization problem is a convex program [24], as the objective function is concave with respect to m_t , when $m_t \geq M_t$. The proof is omitted due to space limitation.

Lemma 1: The solution of program (8) at each time t is:

$$m_t^* = \begin{cases} -\frac{\gamma_t \lambda_t}{\kappa T} \left(\log \delta_t + \log \frac{\gamma_t}{V_t \kappa T} \right), & \text{if } \delta_t \leq \frac{V_t \kappa T}{\gamma_t} \exp\left(\frac{-l \kappa T}{\gamma_t}\right); \\ M_t, & \text{otherwise.} \end{cases} \quad (9)$$

We assume that tenants temporarily quit from the DC if their optimal aggregate surplus is negative.

Theorem 1: The tenants' aggregate surplus is non-negative if and only if

$$\delta_t \leq \frac{V_t}{l} \left(1 - \exp\left(\frac{-l \kappa T}{\gamma_t}\right) \right). \quad (10)$$

Proof 1: Firstly, when

$$\delta_t \leq \frac{V_t \kappa T}{\gamma_t} \exp\left(\frac{-l \kappa T}{\gamma_t}\right),$$

we have

$$\begin{aligned} m_t^* &= -\frac{\gamma_t \lambda_t}{\kappa T} \left(\log \delta_t + \log \frac{\gamma_t}{V_t \kappa T} \right) \text{ and} \\ S(m_t^*) &= V_t \lambda_t - \frac{\delta_t \gamma_t \lambda_t}{\kappa T} \left(1 - \log \frac{\delta_t \gamma_t}{V_t \kappa T} \right). \end{aligned}$$

Since $\partial S(m_t^*)/\partial \delta_t \leq 0$, the minimal $S(m_t^*)$ is reached when

$$\delta_t = \frac{V_t \kappa T}{\gamma_t} \exp\left(\frac{-l \kappa T}{\gamma_t}\right). \quad (11)$$

Thus, in this case, we can ensure that $S(m_t^*) \geq 0$. Secondly, when

$$\frac{V_t \kappa T}{\gamma_t} \exp\left(\frac{-l \kappa T}{\gamma_t}\right) < \delta_t \leq \frac{V_t}{l} \left(1 - \exp\left(\frac{-l \kappa T}{\gamma_t}\right) \right),$$

we have $m_t^* = M_t$ and $S(M_t) \geq 0$. Thirdly, if

$$\delta_t > \frac{V_t}{l} \left(1 - \exp\left(\frac{-l \kappa T}{\gamma_t}\right) \right),$$

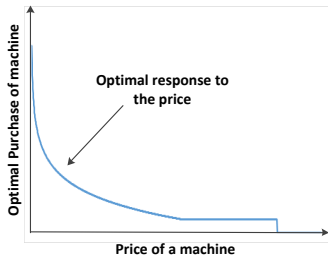


Fig. 2. Tenants' optimal aggregate purchase: a simple example

$m_t^* = M_t$. In this case, tenants' aggregate surplus is always less than 0.

Figure 2 shows an example of tenants' optimal aggregate purchase under different prices of machine. Apparently, if the price is unreasonably low, tenants purchase extremely large amount of machines, which leads to the waste of DC resource. By increasing the price of machine, tenants' aggregate purchase is convexly decreased. Further, when the price reaches $\frac{V_t \kappa T}{\gamma_t} \exp(\frac{-l \kappa T}{\gamma_t})$, tenants keep purchasing M_t machines till the price exceeds $\frac{V_t}{l} \left(1 - \exp(\frac{-l \kappa T}{\gamma_t})\right)$. Then, tenants purchase no machine, i.e., tenants temporarily quit from the DC.

B. DC's Optimal Control

We assume that the DC has perfect knowledge of its tenants' optimal machine purchase i.e., m_t^* in response to machine price δ_t . To maximize the DC's profit, the following optimization problem should be solved to obtain the optimum machine price at each time slot $\delta_1, \dots, \delta_\tau$.

$$\begin{aligned}
& \underset{\delta_t}{\text{maximize}} && \sum_t (\delta_t m_t^* - \alpha(t)T \max\{E_{pue} \rho m_t^* - G_t, 0\}) \\
& && - \beta \max_t \{\max\{E_{pue} \rho m_t^* - G_t, 0\}\} \\
& \text{subject to} && 0 \leq \delta_t \leq \frac{V_t}{l} \left(1 - \exp(\frac{-l \kappa T}{\gamma_t})\right), \\
& && \forall t \in \{1, \dots, \tau\}.
\end{aligned} \tag{12}$$

According to Theorem 1, if $\delta_t > \frac{V_t}{l} \left(1 - \exp(\frac{-l \kappa T}{\gamma_t})\right)$, tenants purchase no machine. Thus, the only constraint of optimization problem (12) ensures profitability of DC. Meanwhile, m_t^* in the objective function can be replaced by

$$m_t^* = \max \left\{ -\frac{\gamma_t \lambda_t}{\kappa T} \left(\log \delta_t + \log \frac{\gamma_t}{V_t \kappa T} \right), M_t \right\}. \tag{13}$$

Apparently, the objective function of the optimization problem (12) is not convex or concave. However, problem (12) becomes a convex program for the special case of $l = 0$:

Theorem 2: When $l = 0$, then problem (12) is convex.

Algorithm 1 Calculation of optimal price

Require:

- δ_t^* (optimal price of machine at time slot t)
 - l (factor of lower bound of purchase of machine)
- 1: Get δ_t^* by solving convex program (14).
 - 2: **if** $l = 0$ **then**
 - 3: **return** δ_t^* .
 - 4: **else if** $l > 0$ **then**
 - 5: **for all** $\forall t \in \{1, \dots, \tau\}$ **do**
 - 6: **if** $\delta_t^* \geq \frac{V_t \kappa T}{\gamma_t} \exp(\frac{-l \kappa T}{\gamma_t})$ **then**
 - 7: $\delta_t^* = \frac{V_t}{l} \left(1 - \exp(\frac{-l \kappa T}{\gamma_t})\right)$.
 - 8: **end if**
 - 9: **end for**
 - 10: **end if**
 - 11: **return** δ_t^* .
-

Proof 2: $\forall t \in \{1, \dots, \tau\}$, when $l = 0$, the optimization problem (12) can be transformed into

$$\begin{aligned}
& \underset{\delta_t}{\text{maximize}} && \sum_t (\delta_t m_t^* - \alpha(t)T \max\{E_{pue} \rho m_t^* - G_t, 0\}) \\
& && - \beta \max_t \{\max\{E_{pue} \rho m_t^* - G_t, 0\}\} \\
& \text{subject to} && 0 \leq \delta_t \leq \frac{V_t \kappa T}{\gamma_t}, \quad \forall t \in \{1, \dots, \tau\},
\end{aligned} \tag{14}$$

where

$$m_t^* = -\frac{\gamma_t \lambda_t}{\kappa T} \left(\log \delta_t + \log \frac{\gamma_t}{V_t \kappa T} \right). \tag{15}$$

The feasible set of the optimization problem (14) is convex and m_t^* is convex and nonincreasing in δ_t . Thus, $\max\{E_{pue} \rho m_t^* - G_t, 0\}$ and $\max_t \{\max\{E_{pue} \rho m_t^* - G_t, 0\}\}$ are both convex with respect to δ_t . Further, let $f(\delta_t) = \delta_t m_t^*$. Since

$$\nabla^2 f(\delta_t) = -\frac{\gamma_t \lambda_t}{\kappa T \delta_t} \leq 0,$$

$f(\delta_t)$ is concave over δ_t . Therefore, the objective function of problem (14) is concave.

Now, we explain Algorithm 1 in two steps, to solve the optimization problem (12) for the general case of $l \geq 0$:

Step 1: Let $l = 0$ to relax the lower bound on tenants' aggregate machine purchase. Then, solve the relaxed problem via a convex programming tool, e.g., CVX [25] (line 1).

Step 2: Check and adjust the price of machine to maximize DC's profit. Firstly, if tenants' lower bound equals to 0, i.e., $l = 0$, according to Theorem 2, the optimal price equals to the optimal solution of the convex program (14) (line 1 – 3). Secondly, $\forall t \in \{1, \dots, \tau\}$, if $l > 0$ and

$$\delta_t^* \geq \frac{V_t \kappa T}{\gamma_t} \exp(\frac{-l \kappa T}{\gamma_t}),$$

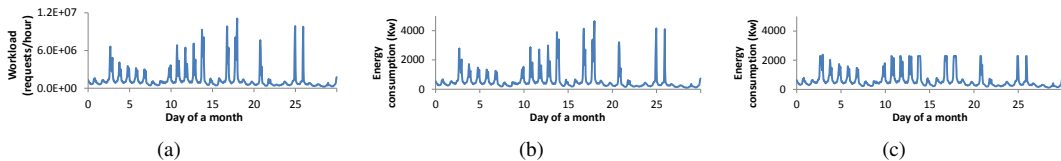


Fig. 3. Comparison of energy consumption between our proposed pricing and Baseline: (a) The aggregate workload [26]; (b) Energy consumption of Baseline; (c) Energy consumption of DC with our proposed pricing.

TABLE I
ELECTRICITY RATES

| Location | Demand Charge (β) | Energy Charge ($\alpha(t)$) |
|-------------------------------|--|-------------------------------|
| Wasco County Schedule-250 [6] | \$ 4.25 per KW | \$ 0.0372 per KWh |
| South California LGS-30 [8] | \$ 10.94 per kW for the first 5,000 kW | \$ 0.0547 per KWh |
| North California LGS-32 [8] | \$ 11.23 per kW for the first 5,000 kW | \$ 0.05566 per KWh |

where δ_t^* is the optimal solution of the convex program (14) at time slot t , $m_t^* = M_t$. Note, in this case, with the increasing of the price of machine till

$$\delta_t^* = \frac{V_t}{l} \left(1 - \exp\left(\frac{-l\kappa T}{\gamma_t}\right) \right),$$

tenants' aggregate purchase will remain to be M_t . Namely, DC can gain more revenue while keeping the energy bill unchanged by setting δ_t^* to be (line 4 – 11):

$$\frac{V_t}{l} \left(1 - \exp\left(\frac{-l\kappa T}{\gamma_t}\right) \right). \quad (16)$$

V. PERFORMANCE EVALUATION

A. Simulation setting

We simulate a DC with sufficient number of machines to support any peak workload. For each switched on machine, $\rho = 200$ watts. As in [16], we set $E_{pue} = 1.2$ and $\kappa = 0.1$. Meanwhile, we assume that a billing cycle is set to repeat on a monthly basis (i.e., 30 days), $T = 1$ hour, $l = 0$ and $\forall t \in \{1, \dots, \tau\}$, $V_t = 0.001$, $\gamma_t = 0.8$. Further, the aggregate workload of World Cup 98 web hits data [26] spanning from June 13, 1998 to July 12, 1998 is used in our simulation, which is shown in Figure 3(a). Obviously, the workload is not smooth. Instead, it has remarkable peak hour and off-peak hour.

Besides, we build a Baseline, which is a representative of the pricing policy that is widely employed by today's DCs. Baseline charges its tenants via flat rates, i.e., the price of each machine will remain unchanged during an relatively long period. For obtaining the optimal flat rates of Baseline, we solve the optimization problem (12) with an additional constraint that the price of machines are same for all the time slots, that is, $\forall t \in \{1, \dots, \tau\}$ and $\forall t' \in \{1, \dots, \tau\}$, we have $\delta_t = \delta_{t'}$.

B. Simulation Results without Renewable Generation

Firstly, we evaluate our proposed pricing without taking into account the generation of renewable energy, i.e., $\forall t \in \{1, \dots, \tau\}$, $G_t = 0$. We adopt several electricity rates in different locations to charge DC for energy consumption, which

are listed in Table I. Figure 3 shows the energy consumption of DC, which is charged by the electricity rates of North California LGS-32 [8]. Apparently, tenants of Baseline do not adjust their demands in response to the electricity rates since they do not receive any incentives to do so (comparing Figure 3(b) with Figure 3(a)). On the contrary, DC with our proposed pricing has provided its tenants with right incentives. In this case, the tenants will proactively "smooth" and reduce their peak demands (comparing Figure 3(c) with Figure 3(b)).

Next, Figure 4 shows the DC's profit and energy bill and tenants' aggregate surplus under different electricity rates. Obviously, with our proposed pricing, DC can gain more profit and reduce its energy bill while improving its tenants' aggregate surplus when comparing with Baseline. For instance, with our proposed pricing, the profit of DC that is charged by the electricity rates of North California LGS-32 [8] can be boosted by 8.2% and its energy bill is reduced by 33.0%. Meanwhile, its tenants' aggregate surplus is increased by 12.3%.

C. Impact of Renewable Generation

Figure 5 shows the results of DC's profit and energy bill and tenants' aggregate surplus, where DC is equipped with 20 local wind turbines. The power generated by the these wind turbines is shown in Figure 6. The wind speed data is gathered from January 1, 2012 to January 30, 2012 and is available in [27]. By comparing Figure 5 with 4, we can find that the local generation of renewable power can further increase DC's profit and tenants' aggregate surplus while reducing the energy bill. Meanwhile, with renewable generation, our proposed pricing can also enhance DC's profit and greatly reduce its energy bill while improving the tenants' aggregate surplus when comparing with Baseline.

VI. CONCLUSION

We proposed a game-theoretic framework for cloud data centers to deduce a time varying pricing (TVP) mechanism to charge their tenants. The advantage of TVP is to propagate the energy bill onto tenants' costs, which will provide tenants with right incentives to cooperate with DC in response to

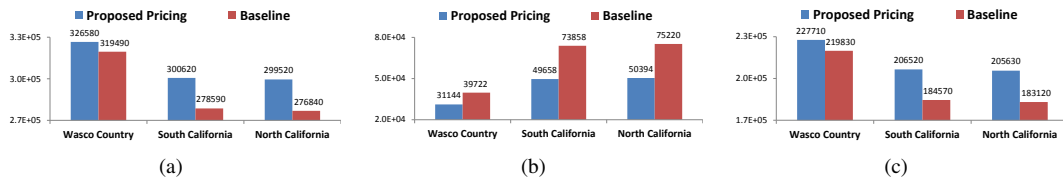


Fig. 4. Performance comparison between our proposed pricing and Baseline without renewable energy: (a) DC's profit over a month. (b) DC's energy bill over a month. (c) Tenants' aggregate surplus over a month

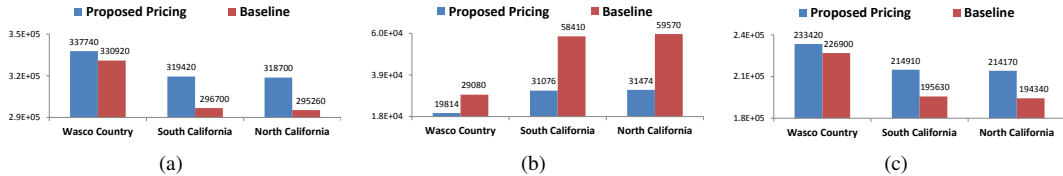


Fig. 5. Performance comparison between our proposed pricing and Baseline with renewable energy: (a) DC's profit over a month. (b) DC's energy bill over a month. (c) Tenants' aggregate surplus over a month

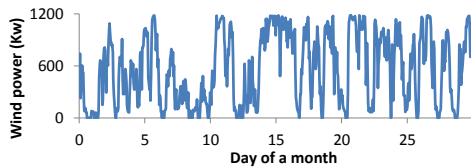


Fig. 6. Wind power generation [27]

the electricity rates. This resolves an important shortcoming in the existing DC energy management literature that often ignores the central role of tenants in shaping the electricity load profile of cloud data centers. Our model is general and takes into account the options to charge tenants also based on local renewable power generation. We demonstrate that our proposed pricing can increase DC's profit and reduce its energy bill while improving tenants' aggregate surplus when comparing with a representative of the pricing that is widely employed by today's DCs. Therefore, our proposed pricing method helps both the cloud data center and the tenants.

REFERENCES

- [1] J. Glanz, "Power, pollution and the internet," *The New York Times*, vol. 22, 2012.
- [2] G. Ghatikar, V. Ganti, N. Matson, and M. A. Piette, "Demand response opportunities and enabling technologies for data centers: Findings from field studies," tech. rep., PG&E/SDG&E/CEC/LBNL, 2012.
- [3] J. Koomey, "Growth in data center electricity use 2005 to 2010," tech. rep., Analytics Press, 2011.
- [4] A. Venkatraman, "Global census shows datacentre power demand grew 63% in 2012." <http://goo.gl/bwFGsA>, 2012.
- [5] R. Brown *et al.*, "Report to congress on server and data center energy efficiency: Public law 109-431," tech. rep., Lawrence Berkeley National Laboratory, 2008.
- [6] "Northern Wasco PUD - Commercial Rates." <http://www.nwasco.com/commercial-rates.cfm/>.
- [7] "GRDA: Wholesale Power Rate for Resale." <http://www.grda.com/electric/customer-service/wholesale-sales/power-rate-for-resale/>.
- [8] "Duke Energy Electric Rates." <http://www.duke-energy.com/rates/>.
- [9] H. Xu and B. Li, "Reducing electricity demand charge for data centers with partial execution," in *Proceedings of the 5th ACM International Conference on Future Energy Systems*, pp. 51–61, 2014.
- [10] M. A. Islam, H. Mahmud, S. Ren, and X. Wang, "Paying to save: Reducing cost of colocation data center via rewards," in *High Performance Computer Architecture, IEEE 21st International Symposium on*, pp. 235–245, 2015.
- [11] "Apple now gets 75 percent of its total power needs from renewable energy." <http://macdailynews.com/2013/03/21/apple-now-gets-75-percent-of-its-total-power-needs-from-renewable-energy/>, 2013.
- [12] H. Chen, C. Hankendi, M. C. Caramanis, and A. K. Coskun, "Dynamic server power capping for enabling data center participation in power markets," in *Proceedings of the IEEE International Conference on Computer-Aided Design*, pp. 122–129, 2013.
- [13] T. Lu, M. Chen, and L. L. Andrew, "Simple and effective dynamic provisioning for power-proportional data centers," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 24, no. 6, pp. 1161–1171, 2013.
- [14] M. Ghamkhari and H. Mohsenian-Rad, "Optimal integration of renewable energy resources in data centers with behind-the-meter renewable generator," in *Proc. of IEEE ICC*, Ottawa, Canada, June 2012.
- [15] J. Camacho, Y. Zhang, M. Chen, and D. M. Chiu, "Balance your bids before your bits: The economics of geographic load-balancing," in *Proceedings of the 5th ACM International Conference on Future Energy Systems*, pp. 75–85, 2014.
- [16] M. Ghamkhari and H. Mohsenian-Rad, "Energy and performance management of green data centers: a profit maximization approach," *Smart Grid, IEEE Transactions on*, vol. 4, no. 2, pp. 1017–1025, 2013.
- [17] M. Ghamkhari and H. Mohsenian-Rad, "Profit maximization and power management of green data centers supporting multiple slas," in *Proc. of IEEE ICNC*, San Diego, CA, 2013.
- [18] "Amazon EC2 Pricing." <http://aws.amazon.com/ec2/pricing/>.
- [19] R. Harms and M. Yamartino, "The economics of the cloud," tech. rep., Microsoft whitepaper, Microsoft Corporation, 2010.
- [20] S. Shakkottai, R. Srikant, A. Ozdaglar, and D. Acemoglu, "The price of simplicity," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 7, pp. 1269–1276, 2008.
- [21] C. Wang, B. Urgaonkar, Q. Wang, and G. Kesidis, "A hierarchical demand response framework for data center power cost optimization under real-world electricity pricing," in *Modelling, Analysis & Simulation of Computer and Telecommunication Systems, IEEE 22nd International Symposium on*, pp. 305–314, 2014.
- [22] C. Wang, B. Urgaonkar, G. Kesidis, U. V. Shanbhag, and Q. Wang, "A case for virtualizing the electric utility in cloud data centers," in *Proceedings of the 6th USENIX Conference on Hot Topics in Cloud Computing*, 2014.
- [23] Y. Xiang and T. Lan, "Smart pricing of cloud resources," *Smart Data Pricing*, pp. 453–476.
- [24] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2009.
- [25] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab Software for Disciplined Convex Programming." <http://cvxr.com/cvx/>, 2015.
- [26] <http://ita.ee.lbl.gov/html/contrib/WorldCup.html>.
- [27] <http://www.windenergy.org/>.