International Journal of Innovative Computing, Information and Control
Volume **12**, Number **5**, October **2016**

ICIC International ©2016 ISSN 1349-4198
pp. 1691–1702

A PROVABLE EFFICIENT CONTROL ALGORITHM TO REDUCE DATA CENTERS POWER COST WITH ON-SITE RENEWABLE ENERGY GENERATION

Ran Zhang[1] and Jianxiong Wan[2]

[1]School of Mathematics and Statistics
Central South University
No. 172, Tongzipo Road, Changsha 410013, P. R. China
seran_zhang@126.com

[2]School of Information Engineering
Inner Mongolia University of Technology
Power City of Jinchuan Development Zone, Hohhot 010051, P. R. China
jxwan@imut.edu.cn

Received April 2016; revised August 2016

Abstract. *Cloud computing and big data era introduces a new opportunity for enterprises to commission their massive computing tasks to Internet data centers (IDC) with relatively low cost. However, the explosive growth of massive data processing has imposed a heavy burden on computation, storage, and communication in IDCs, which hence incurs considerable operational expenditure to data center providers. Data in the cloud computing need to be transferred and processed with high efficiency and cost effectiveness while meeting the demand on reliability. This paper formulates the cost minimization of IDCs servers and renewable power problem to minimize time-averaged expected power cost subject to the constraints of data queue stability and server allocation policy with the consideration of renewable energy utilization. We design an algorithm leveraging the Lyapunov optimization technique to approximately solve the CMISRP problem. All data are originally from Ordos Uni-Cloud Technology Co., Ltd. to simulate the proposed algorithm. Numerical results show that our algorithm can reduce total energy cost while guaranteeing QoS, queue stability and server allocation policy.*
Keywords: Internet data center, CMISRP, Data queue stability, Lyapunov optimization, Cost minimization

1. Introduction.

The 21st century announced the arrival of cloud computing. With the growing amount of industry data, common enterprise computer rooms have been no longer suitable for the needs of high-tech development, so large and medium-sized Internet data centers (IDC) have sprung up like mushrooms after rain. In research field, topics about IDC are inevitably brushing professors' and professionals' brains. S. Zaman and D. Grosu [1] and P. Bonacquisto et al. [2] focused on fixed power price scenarios and J. Zhao et al. [3] extended their works to environment with geo-distributed IDCs and time-varying power price. However, the global problem faced by IDCs' operators is a high cost due to the large power consumption. IDCs' energy consumption grew by 56% from 2005 to 2010 [4], and data centre dynamics recently revealed that data center power usage increased by 63% in 2012 [5]. Therefore, the need for optimizing power management is urgent to save the energy cost in IDCs all over the world.

Numerous professionals facing IDCs' power cost control have designed a large number of problem models and algorithms. Z. Zhou et al. put forward a diffused algorithm in order to make autonomous decisions on winning bid victory and workload management by combining techniques in the Gribb's sample method and the multipliers' alternating direction method of multipliers [6]. They concluded that incentive mechanism consist of a win-win mechanism for the geo-distributed cloud and the smart grid. L. Rao et al. researched the minimization of the total electricity cost under different electricity markets while guaranteeing quality of service geared to the various and changing electricity price affected by multiple locations and time [7]. They formulated the question as a constrained mixed-integer programming and figured out a valid solution. Evaluating real electricity price data from multiple IDC locations attested the efficiency and efficacy of their algorithm. P. Lama and X. Zhou proposed a new system, PERFUME, which synchronously guaranteed power and performance targets with flexible tradeoffs and service differentiation among co-hosted applications while assuring control accuracy and system stability [8].

Though the above models considered the energy consumption of servers, they overlooked the IDCs' renewable resource utilization. Renewable energy is generally defined as energy collected from resources which are naturally replenished on a human timescale, such as sunlight, wind, rain, tides, waves, and geothermal heat. Renewable energy often provides energy in four important areas: electricity generation, air and water heating/cooling, transportation, and rural (off-grid) energy services. Based on 2014 report [9], the renewable contributed 19% to humans' global energy consumption and 22% to their generation of electricity in 2012 and 2013, respectively.

In recent years, renewable energy has stepped into researchers' fields of vision. GreenSlot, a scheduler proposed by I. Goiri et al. [10], was used to process parallel batch jobs in a data center powered by a photovoltaic solar and the electrical grid (as a backup). Compared to conventional schedulers, GreenSlot can make solar energy consumption ascending and energy cost descending. C. Li et al. modeled and evaluated data centers powered by intermittent renewable energy. Using real-world data center and renewable energy source records, they demonstrated that designing sustainable high performance data centers needed two crucial metrics, renewable power utilization and load tuning frequency [11]. The variable and intermittent characteristics of many renewable energy sources made them and electric grid challenging, which limited the entrance of renewable energy. A. Krioukov et al. proposed a solution: supply-following loads could be used to adjust power consumption matching the available renewable energy sources. A. Krioukov and his mates utilized wind power to schedule the workload and used a simple supply-following job scheduler to produce 40-60% better renewable energy entrance than supply-oblivious schedulers [12].

Although the above papers studied utilization of renewable energy of IDCs, the renewable energy is not the main source of IDCs. Renewable energy alone does not perfectly maintain the servers running. Taking the servers' energy consumption as well as renewable energy into account in some proportion is a hot issue in the control of energy consumption and electricity cost in IDCs.

However, it is difficult to predict most renewable energy resources because the average energy consumption must be optimized. Traditional algorithm must acquire the future renewable energy information and greedily optimize the current cost rather than the average cost during a period of time. Although some models predict the average cost, unfortunately their results are inaccurate so that their applications are very limited. We design an algorithm leveraging the Lyapunov optimization technique to combine server energy with renewable energy.

This paper is organized as follows: Section 2 proposes the system model of data center; Section 3 formulates the cost minimization of IDCs' server and renewable power

(CMISRP) problem; Section 4 applies the Lyapunov optimization theory to approximately handling (CMISRP) problem and analyzes the implementation issues; theoretical performance analysis is presented in Section 5; Section 6 further investigates the performance using simulations; Section 7 concludes the paper.

2. **System Model.** The architecture shown in Figure 1 illustrates IDC's applications, each of which hosts racks of servers. The infrastructures located in IDCs provide electricity for each server to run its service. All requests entering an application share the same workload queue and are scheduled to a server once it is currently idle. Quality-of-service (QoS) requirement associated with each application is also in our consideration.

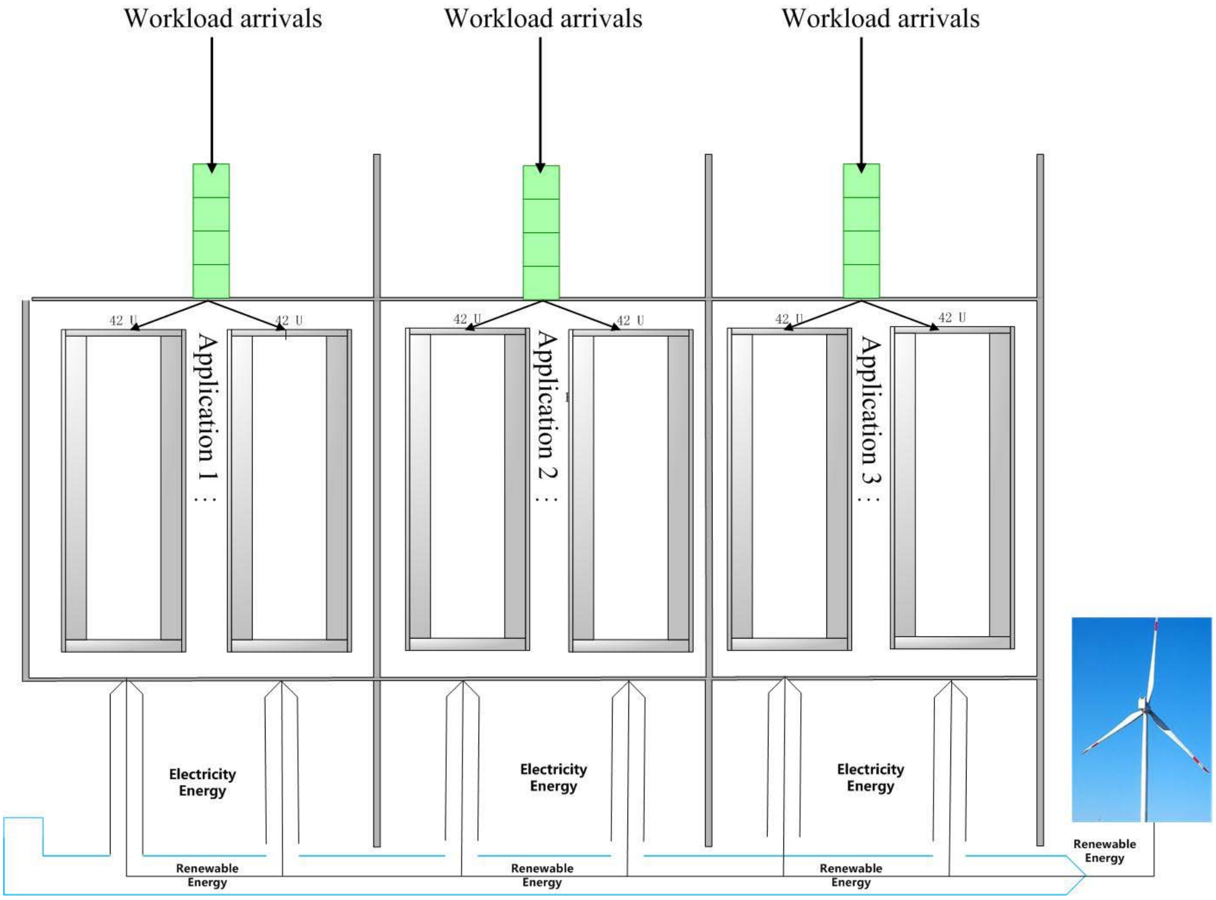

FIGURE 1. System model

Though the huge cost of electricity consumption troubles most IDCs' operators, renewable wind energy will validly cut the total energy cost of a data center. In this paper, we focus on how to minimize the total energy cost subject to QoS.

$$\min \textit{Power cost}$$

subject to:

$$\textit{Queue stability, for each application}$$
$$\textit{The server allocation policy is under control.}$$

The control variable of this problem is the number of servers belonging to each application. The queue stability constraint ensures a limited queue length, which implies a limited queueing delay according to the *little's law*. This paper will discuss the details of this problem in the following sections.

3. **Problem Formulation.**

3.1. Server power consumption model. Let J be the total number of applications hosted in Ordos Uni-Cloud Co.'s IDC. At any time slot t, define $p_j(t)$ and $m_j(t)$ as the power consumption for a single server and the number of servers allocated to the corresponding application $j \in [1, \ldots, J]$ in slot t, respectively. Then the power consumption for application j is

$$p_j(t) = m_j(t)w$$

where w is the energy consumption per server. The total energy consumption for all servers can be expressed as

$$P(t) = \sum_{j=1}^{J} p_j(t) \tag{1}$$

3.2. Queueing dynamic. Assume all requests into a single application share the same queue and all servers in each application also fetch data from the queue. Any request in accordance arrives the system and directly into service if there is a free server, and otherwise joins the queue. Once a server finishes serving a request, it immediately leaves the system, and the next one in line, if there is any, enters service. Suppose that the service rate for application j is μ_j. Let $Q_j(t)$ be the queue length at the time slot t of application j. So update the formula for $Q_j(t)$ as:

$$Q_j(t+1) = \max\left\{Q_j(t) - m_j(t)\mu_j, 0\right\} + \lambda_j(t) \tag{2}$$

where $\lambda_j(t)$ represents the amount of new work that arrives on slot t. $m_j(t)\mu_j$ represents the amount of work the queue in application j can process on slot t. Obviously, both $\lambda_j(t)$ and $m_j(t)\mu_j$ are non-negative.

3.3. Problem formulation. After analyzing the distinguished problem of minimizing the energy cost in IDCs, we can now define the objective function as follows:

$$\min \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathrm{E}\left\{c(t)(P(t) - R(t))\right\} \tag{3}$$

subject to:

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathrm{E}\{Q(t)\} < \infty \tag{4}$$

$$0 \le m_j(t) \le m_j^{\max} \tag{5}$$

$$\sum_{j=1}^{J} m_j(t) \le M \tag{6}$$

where $R(t)$ is wind energy as the main source of renewable energy, $c(t)$ is the instant electricity price on slot t and M is the number of servers in the data center.

4. **Solution.**

4.1. **Lyapunov approach.** The Lyapunov optimization framework provides a method to design a control algorithm, choosing actions for all t to yield a time average expectation of the objective function value close to optimal solution. Meanwhile, make the queue $Q_j(t)$ mean rate stable. In this way, the original problem is transformed into that of minimizing the time average of a cost function subject to queue stability. Not only do we need to minimize IDCs' energy cost, but control the length of queues. Lyapunov drift and optimization theory can hold a balance between the two problems. Define the Lyapunov function:

$$\mathbf{L}(t) = \frac{1}{2}\sum_{j=1}^{J} Q_j(t)^2.$$

Let

$$\Delta(\mathbf{Q}(t)) = \mathrm{E}\left\{\mathbf{L}(t+1) - \mathbf{L}(t)|\mathbf{Q}(t)\right\}.$$

The drift-plus-penalty objective function is

$$\Delta(\mathbf{Q}(t)) + V\mathrm{E}\left\{c(t)(P(t) - R(t))|\mathbf{Q}(t)\right\} \tag{7}$$

Theorem 4.1. *The drift-plus-penalty objective Function* (7) *can be bounded by the following*

$$\Delta(\mathbf{Q}(t)) + V\mathrm{E}\left\{c(t)(P(t) - R(t))|\mathbf{Q}(\mathbf{t})\right\} \tag{8}$$

$$\leq \sum_{j=1}^{J} Q_j(t)\mathrm{E}\left\{(\lambda_j(t) - m_j(t)\mu_j)|\mathbf{Q}(\mathbf{t})\right\}$$

$$+ V\mathrm{E}\left\{c(t)\left(\sum_{j=1}^{J} m_j(t)w - R(t)\right)\middle|\mathbf{Q}(\mathbf{t})\right\} + B \tag{9}$$

where B is a constant defined as $B = \frac{\sum_{j=1}^{J}\left(\lambda_{\max}^2 + (m_{\max}\mu_{\max})^2\right)}{2}$

Proof: Squaring the queue update Equation (2), we have

$$\begin{aligned} Q_j^2(t+1) &\leq \left(Q_j(t) - m_j(t)\mu_j\right)^2 + \lambda_j^2(t) + 2Q_j(t)\lambda_j(t) \\ &= Q_j^2(t) + \lambda_j^2(t) + 2Q_j(t)\left(\lambda_j(t) - m_j(t)\mu_j\right) + \left(m_j(t)\mu_j\right)^2 \end{aligned} \tag{10}$$

Therefore,

$$\frac{Q_j^2(t+1) - Q_t^2(t)}{2} \leq \frac{1}{2}\left(\lambda_j^2(t) + (m_j(t)\mu_j)^2\right) + Q_j(t)\left(\lambda_j(t) - m_j(t)\mu_j\right).$$

Taking conditional expectation of the above equation and summing over $j \in \{1, \ldots, J\}$ give a bound on $\Delta(\mathbf{Q}(t))$. Adding $V\mathrm{E}\left\{c(t)(P(t) - R(t))|\mathbf{Q}(\mathbf{t})\right\}$ to both sides leads to Lyapunov drift-plus-penalty expression. Hence, we have

$$\begin{aligned} &\Delta(\mathbf{Q}(t)) + V\mathrm{E}\left\{c(t)(P(t) - R(t))|\mathbf{Q}(\mathbf{t})\right\} \\ &\leq \sum_{j=1}^{J} Q_j(t)\mathrm{E}\left\{(\lambda_j(t) - m_j(t)\mu_j)|\mathbf{Q}(\mathbf{t})\right\} + V\mathrm{E}\left\{c(t)\left(\sum_{j=1}^{J} m_j(t)w - R(t)\right)\middle|\mathbf{Q}(\mathbf{t})\right\} \\ &+ \frac{\sum_{j=1}^{J}\left(\lambda_j(t)^2 + (m_j(t)\mu_j(t))^2\right)}{2} \end{aligned} \tag{11}$$

Under any control algorithm, the drift-plus-penalty expression satisfies (9) for all t, all possible values of $\mathbf{Q}(t)$, and all parameters $V \geq 0$. B is a positive constant. In the model, assume the maximum of $\mu_j(t)$ is $\mu_{\max}$ and $0 \leq m_j(t) \leq m_{\max}$. For all t: $B =$

$\frac{\sum_{j=1}^{J}(\lambda_{\max}^2+(m_{\max}\mu_{\max})^2)}{2}$, such a constant B exists because $\mathbf{Q}(t)$ is i.i.d. and the boundedness assumptions in (3) hold. *Q.E.D.*

We apply *Lyapunov drift-plus-penalty* to the control of queue length and power cost from IDC (Internet data center) in place of the original objective function. While taking actions to minimize a bound on $\Delta(\mathbf{Q}(t))$ every slot t would stabilize the system, the resulting average power cost might be extremely large, and vice versa. *Lyapunov drift-plus-penalty* intends to reach a balance between queue length and power cost. Regardless of directly minimizing (3), the Lyapunov optimization seeks to minimize a bound of the drift-plus-penalty Function (8).

Our algorithm is designed to minimize (9). This is done via the framework of opportunistically minimizing a conditional expectation and the resulting algorithm is given below.

CMISRP algorithm: observe the current queue states $\mathbf{Q}(t)$ in each time slot t, and make a control decision based on solving the following problem:

$$Minimize: \sum_{j=1}^{J} Q_j(t)\left(\lambda(t)_j - m_j(t)\mu_j\right) + V\left\{c(t)\left(\sum_{j=1}^{J} m_j(t)w - R(t)\right)\right\} \tag{12}$$

subject to constraints (5) and (6).

In the next section, we will discuss the implementation of this algorithm.

4.2. **Implementation.** To easily minimize Lyapunov optimization, the objective function of problem (12) can be rewritten as

$$Minimize: \sum_{j=1}^{J} Q_j(t)\lambda_j(t) - \sum_{j=1}^{J} m_j(t)\left(Q_j(t)\mu_j - Vc(t)w\right) - Vc(t)R(t) \tag{13}$$

We choose Algorithm 1 to enlarge the second term. Once $(Q_j(t)\mu_j - Vc(t)w) \leq 0$, $m_j(t)$ will be set as 0. Confronted with $(Q_j(t)\mu_j - Vc(t)w) > 0$, the algorithm records its value. Algorithm 2 accumulates all results for all applications and sets them in a descending order. According to the order, we will assign the corresponding maximum of servers to the ranked-well application, but the total number of servers is no more than the capacity of IDC, i.e., M. In this way, $m_j(t)$ tends to be determined for the lower energy-cost application. Suppose that $Q_j(1) = 0$, and then update the values of all $Q_j(t)$ by letting $m_j(t)$ into (2). While conserving $Q_j(t)$ for each application, we easily figure out the energy of servers reference to (1) and calculate the mean power cost depending on (3).

Algorithm 1: Determining the coefficient of m_j for the Drift-Plus-Penalty Algorithm

```
if co > 0 then
    return co
else
    return 0
end if
```

Finally, adjust V until we constantly minimize (13) and find the balance between data queue Q and the objective function.

5. **Performance Analysis.** In this section, the minimum time-average IDC power cost can be achieved using a randomized stationary control policy independent of the data queue state $\mathbf{Q}(t)$. The performance bound for objective function can be derived.

Algorithm 2 CMISRP: Choosing the best m_j for the Drift-Plus-Penalty Algorithm

Determine the upper bound of m_j as $m_j^{\max}$, the capacity of IDC M;
Define $co = (Q_j(t)\mu_j - Vc(t)w)$
Study the coefficient of m_j:
for all $j \in J$ **do**
  $co[j] =$ Algorithm 1
end for
for all $j \in J$ **do**
  if $co[\text{The } j\text{th max}] > 0$ and $M > 0$ **then**
    if $M - m_j^{\text{The } j\text{th max}} > 0$ **then**
      $m_j = m_j^{\text{The } j\text{th max}}$
      $M = M - m_j^{\text{The } j\text{th max}}$
    else
      $m_j = M$
    end if
  else
    $m_j = 0$
  end if
end for
return m_j

Define $\mathrm{E}\left\{c^{\pi}(t)\left(P^{\pi}(t) - R^{\pi}(t)\right)\right\}$ as the expected power cost in the interval $[1, \ldots, T]$ under any control policy π, and e^* as the minimum achievable $\mathrm{E}\left\{c^{\pi}(t)\left(P^{\pi}(t) - R^{\pi}(t)\right)\right\}$ over all possible π. We have the following result.

Theorem 5.1. *Suppose that $\lambda_j(t)$ is stationary stochastic process. If problem (3) is feasible, then for any $\delta > 0$, there is an ω-only policy π^*, for all t, and:*

$$\mathrm{E}\left\{c^{\pi^*}(t)(P^{\pi^*}(t) - R^{\pi^*}(t))\right\} \leq e^* + \delta,$$
$$\mathrm{E}\left\{\lambda_j^{\pi^*}(t) - m_j^{\pi^*}(t)\mu_j\right\} \leq \delta.$$

Proof: It is a direct application of Theorem 4.5 in [13] *Q.E.D.*

Theorem 5.1 describes that there is a queue-independent randomized stationary policy yielding a power cost arbitrary close to the optimum as long as the problem is feasible. In the next step, the performance bound is given.

Theorem 5.2. *Suppose that $\mathrm{E}\{L(\boldsymbol{Q}(0))\} < \infty$. We have the following results.*

1) The achievable total energy consumption of the CMISRP algorithm can be bounded by

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathrm{E}\{c(t)(P(t) - R(t))\} \leq \frac{B}{V} + e^* \tag{14}$$

2) The queue $\mathbf{Q}(t)$ is mean rate stable, i.e.,

$$\limsup_{T\to\infty} \frac{\mathrm{E}\{Q_j(t)\}}{T} = 0 \tag{15}$$

Proof: Part 1: In every slot t, CMISRP tries to minimize the right-hand-side of the drift expression (9) over all π, and we have for each slot t:

$$\Delta(\mathbf{Q}(t)) + V\mathrm{E}\left\{c(t)(P(t) - R(t))|\mathbf{Q}(t)\right\}$$

$$\leq \sum_{j=1}^{J} Q_j^{\pi}(t)\mathrm{E}\left\{\left(\lambda^{\pi}(t)_j - m_j^{\pi}(t)\mu_j\right)|\mathbf{Q}(t)\right\} + V\mathrm{E}\left\{c^{\pi}(t)\left(P^{\pi}(t) - R^{\pi}(t)\right)|\mathbf{Q}(t)\right\} + B \quad (16)$$

where $c^{\pi}(t)$, $P^{\pi}(t)$, $R^{\pi}(t)$, and $\lambda^{\pi}(t)_j - m_j^{\pi}(t)\mu_j$ are the instant electricity price, the server power, the wind energy and data queue under any alternative (possibly randomized) decision π.

Now consider the ω-only policy π^* in Theorem 5.1. Because this is an ω-only policy, and the values of $c^{\pi}(t)$, $P^{\pi}(t)$, $R^{\pi}(t)$, and $\lambda^{\pi}(t)_j - m_j^{\pi}(t)\mu_j$ are independent of the current queue $\mathbf{Q}(t)$, we have:

$$\mathrm{E}\left\{c^{\pi^*}(t)(P^{\pi^*}(t) - R^{\pi^*}(t))|\mathbf{Q}(t)\right\} = \mathrm{E}\left\{c^{\pi^*}(t)(P^{\pi^*}(t) - R^{\pi^*}(t))\right\} \leq e^* + \delta,$$
$$\mathrm{E}\left\{\left(\lambda^{\pi^*}(t)_j - m_j^{\pi^*}(t)\mu_j\right)|\mathbf{Q}(t)\right\} = \mathrm{E}\left\{\left(\lambda^{\pi^*}(t)_j - m_j^{\pi^*}(t)\mu_j\right)\right\} \leq \delta.$$

Plugging these into the right-hand-side of (16) and taking $\delta \to 0$ yield:

$$\Delta(\mathbf{Q}(t)) + V\mathrm{E}\left\{c(t)((P(t) - R(t))|\mathbf{Q}(t))\right\} \leq B + Ve^* \quad (17)$$

Further, from the above drift expression, we take iterated expectations and telescope sums for any $t > 0$:

$$\frac{1}{T}\sum_{t=0}^{T}\mathrm{E}\{c(t)(P(t) - R(t))\} \leq \frac{B}{V} + e^* + \frac{\mathrm{E}\{L(\mathbf{Q}(0)\}}{Vt} \quad (18)$$

which proves the theorem by taking a lim sup as $t \to \infty$.

Part 2. As the research refers to the actual power cost, suppose that the expected penalty is lower bounded by a finite value $e^{\min}$ for all t and all possible control actions. We have:

$$\mathrm{E}\{c(t)((P(t) - R(t))\} \geq e^{\min} \quad (19)$$

Plug the above function into (17):

$$\mathrm{E}\{L(\mathbf{Q}(t+1))\} - \mathrm{E}\left\{L(\mathbf{Q}(t))\right\} + Ve^{\min} \leq B + Ve^* \quad (20)$$

Summing the above over $t \in \{1, \ldots, T\}$ for some slot $t > 0$ and using the law of telescoping sums yield:

$$\mathrm{E}\{L(\mathbf{Q}(t))\} \leq TB + TV\left(e^* - e^{\min}\right) + \mathrm{E}\{L(\mathbf{Q}(0))\}.$$

Using the definition of $L(\mathbf{Q}(t))$ yields:

$$\frac{1}{2}\sum_{j=1}^{J}\mathrm{E}\{Q_j(t)^2\} \leq TB + TV\left(e^* - e^{\min}\right) + \mathrm{E}\left\{L(\mathbf{Q}(0))\right\} \quad (21)$$

Therefore, for all $j \in \{1, \ldots, J\}$, we have:

$$\mathrm{E}\{Q_j(t)^2\} \leq 2TB + 2TV(e^* - e^{\min}) + 2\mathrm{E}\{L(\mathbf{Q}(0))\}.$$

However, because the variance of $|Q_j(t)|$ cannot be negative, we have $\mathrm{E}\{Q_j(t)^2\} \geq \mathrm{E}\{|Q_j(t)|\}^2$. Thus, for all slots $t > 0$, we have:

$$\mathrm{E}\{|Q_j(t)|\} \leq \sqrt{2TB + 2TV\left(e^* - e^{\min}\right) + 2\mathrm{E}\left\{L(\mathbf{Q}(0))\right\}}.$$

Dividing by T and taking a limit as $t \to \infty$ prove that:

$$\limsup_{T\to\infty}\frac{\mathrm{E}\{Q_j(t)\}}{T} \leq \sqrt{\frac{2B}{T} + \frac{2V(e^* - e^{\min})}{T} + \frac{2\mathrm{E}\{L(\mathbf{Q}(0))\}}{T^2}} = 0.$$

Thus, all queues $Q_j(t)$ are mean rate stable, proving Part 2. *Q.E.D.*

According to the above detailed analysis, we conclude the $[O(1/V), O(V)]$ tradeoff, i.e., the power expenditure can be pushed arbitrarily close to the optimum as $V \to \infty$.

6. **Numerical Evaluations.** In this section we conduct extensive simulations to evaluate the proposed CMISRP algorithm.

6.1. **System configuration.** Our team members chose Poisson process to simulate workload arrivals and obtained the numbers of servers from applications in the time granularity of one hour, which is shown in Figure 2.

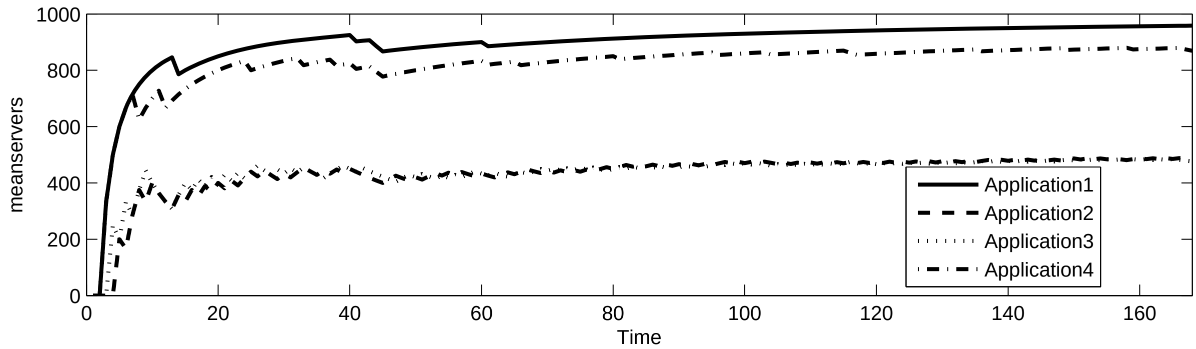

FIGURE 2. Mean number of servers vs. application

This IDC's capacity is 3000 servers in all and four applications can respectively accommodate at most 1000, 500, 500, 900 servers. Without loss of generality, the service rate of a single server μ_j is normalized as 1 across all applications. The power consumption of a single server is 100 Watt. In this study we consider the wind energy as the main source of renewable energy and obtain a one-week wind speed trace from a 70-meter high wind mill of a production wind farm. To translate the wind speed into electricity power, we use a polynomial power curve [16] to approximate the power generation:

$$P(v) = \begin{cases} 0 & v < v_{in} \\ (c_1 + c_2 v + c_3 v^2) \times P_r & v_{in} \leq v < v_r \\ P_r & v_r \leq v < v_{out} \\ 0 & v > v_{out} \end{cases} \tag{22}$$

where v_{in}, v_r, and v_{out} are the cut-in, rated, and cut-out wind speed, respectively. Typically, these values can be set to $v_{in} = 5$, $v_r = 15$ and $v_{out} = 20$. Though currently renewable energy is intermittent and unpredictable, its contribution to global energy generation is constantly increasing. The renewable energy used to power Facebook data centers has grown from 19% in 2012 to 25% in 2015 [14], and is even expected to reach 50% in 2018 [15]. Google reports that over 30% of their services are supported by renewable energy [16]. Based on the above certificate, the rated power Pr, whose value is set as $8.9 * 10^5$, accounts for approximately 30% of the power demand in average.

6.2. **Result analysis.** Based on the parameters above, we investigate the performance of the CMISRP algorithm.

Figure 3 discovers the change of mean cost while V is set to different values from 0.2 to 1. Lyapunov optimization imposes effects on the simulation that the mean cost decreases with the increase of V, especially in $V = 1$. The mean power cost when $V = 1$ is 12% less than that when $V = 0.2$.

Figure 4 depicts that during the whole process, server consumption for all applications basically reaches peak values and occasionally show jitter phenomenons. The above two figures illustrate that PCR algorithm reduces the mean power cost while cutting the energy consumption.

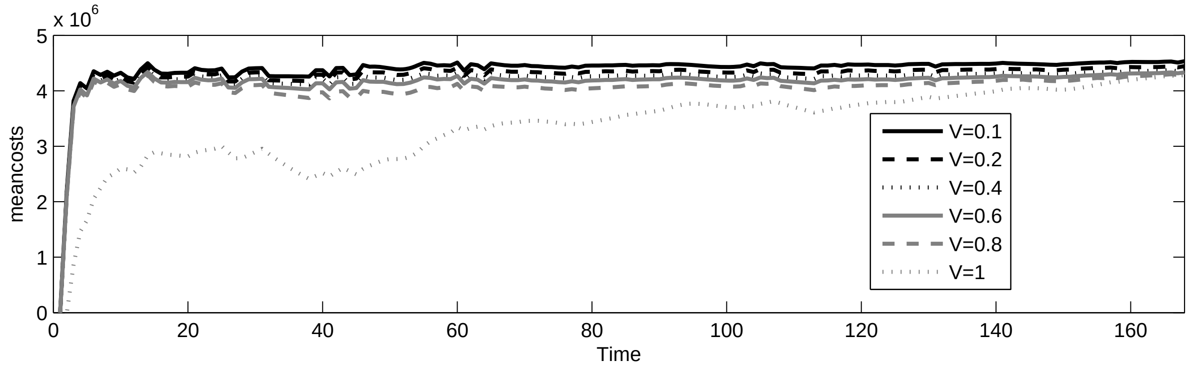

FIGURE 3. Mean prices vs. V

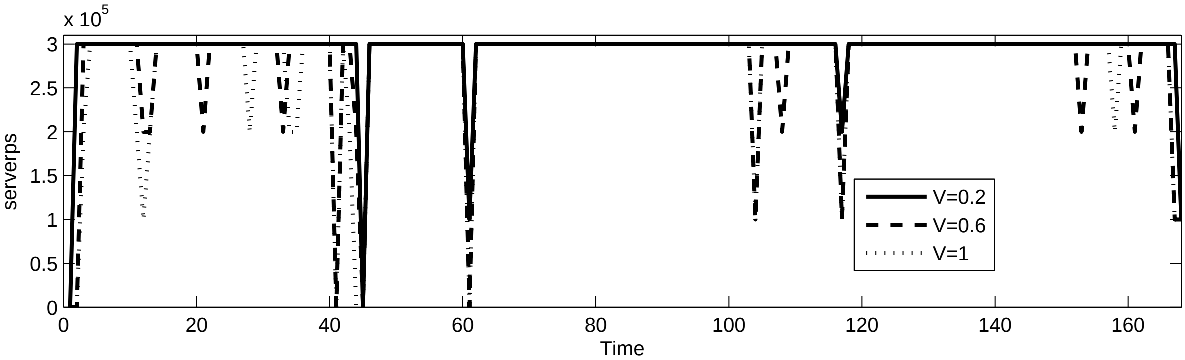

FIGURE 4. Server powers vs. V

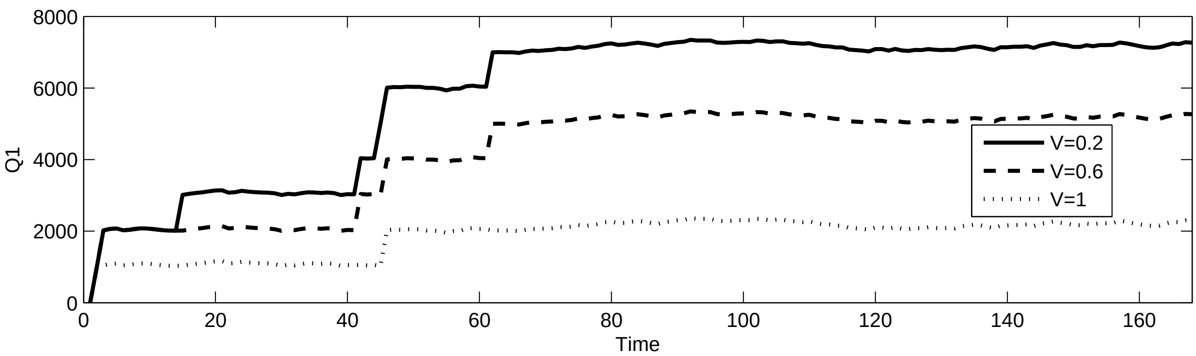

FIGURE 5. Queue length in application 1 vs. V

Figures 5, 6, 7 and 8 have the same ordinates to respectively illustrate data queue lengths from different applications. It is easy to discover that $\mathbf{Q(t)}$ shows a general declining trend with the promotion of V, the weight of Lyapunov drift-plus-penalty objective function. The fluctuation in Figure 6 does not affect the overall decreasing effect.

7. **Conclusion and Future Works.** This paper concentrates on cutting the energy cost of the data center by leveraging the electricity fluctuation and renewable energy. To design a dynamic control algorithm to achieve this goal is challenging since the workload distribution cannot be obtained in advance. To address this issue, we leverage the Lyapunov optimization technique and develop an algorithm to approximately solve energy cost minimization problem. The research shows that our algorithm can be pushed arbitrarily close

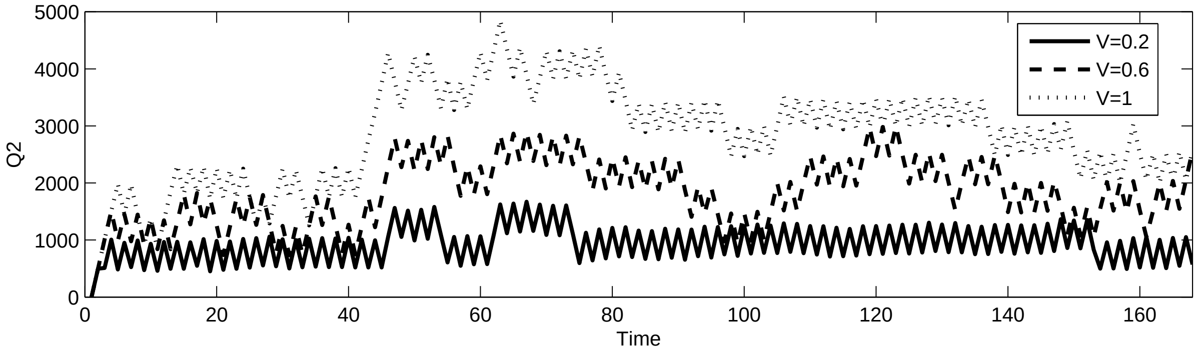

FIGURE 6. Data queue in application 2 vs. V

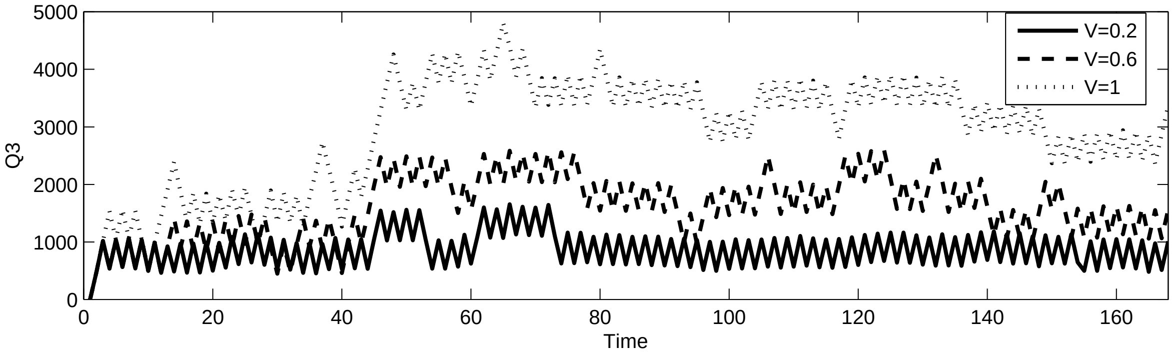

FIGURE 7. Data queue in application 3 vs. V

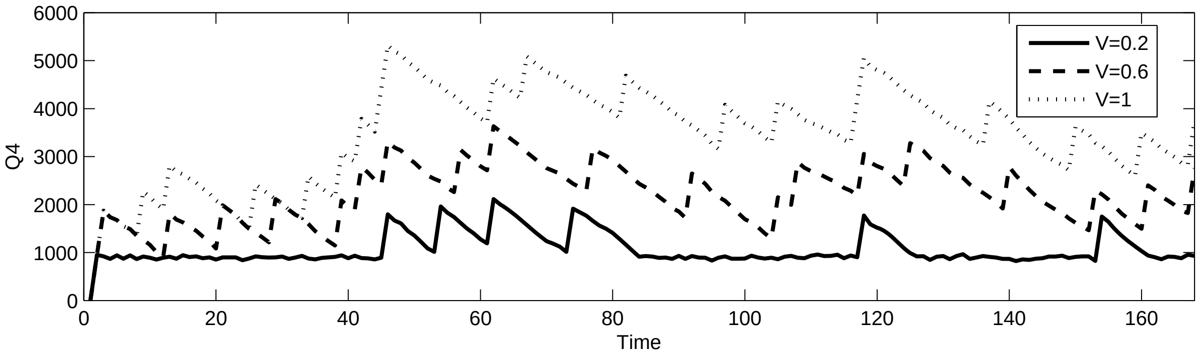

FIGURE 8. Data queue in application 4 vs. V

to optimal solution as the control parameter V increases. Extensive simulations are implemented to explore the performance of the proposed algorithm. Numerical results show that our algorithm can reduce total energy cost while guaranteeing QoS constraints.

In the near future, our research will expand to further consider other factors that influence the energy cost and consumption of data centers, such as the cooling system (refrigerating machine in summer, cooling tower in winter), lighting systems and UPS. Research team will continue to cooperate with Uni-Cloud Technology Co., Ltd and extend the model into a multiple geo-distributed data centers scenario. Once the time is ripe, a more general form of dynamic pricing strategy will be implemented in several typical data centers.

Acknowledgments. This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grant Nos. 61502255, 61540004, the Inner Mongolia Provincial Natural Science Foundation under Grant No. 201-4BS0607, and the Science Research Project for Inner Mongolia College under Grant No. NJZY14064.

REFERENCES

[1] S. Zaman and D. Grosu, A combinatorial auction-based mechanism for dynamic VM provisioning and allocation in clouds, *IEEE Trans. Computers*, vol.1, no.2, pp.129-141, 2013.
[2] P. Bonacquisto, G. D. Modica, G. Petralia and O. Tomarchio, A strategy to optimize resource allocation in auction-based cloud markets, *IEEE International Conference on Services Computing*, Anchorage, AK, pp.339-346, 2014.
[3] J. Zhao, H. Li, C. Wu, Z. Li, Z. Zhang and F. C. Lau, Dynamic pricing and profit maximization for the cloud with geo-distributed data centers, *IEEE Conference on Computer Communications (INFOCOM)*, Toronto, ON, pp.118-126, 2014.
[4] *Growth in Data Center Electricity Use 2005 to 2010*, http://www.analyticspress.com/datacenters.htm.
[5] *Data Centre Power Usage Surges 63 Percent*, http://www.techweekeurope.co.uk/news/data-centre-power-increase-cloud-95359.
[6] Z. Zhou, F. Liu, Z. Li and H. Jin, When smart grid meets GEO-distributed cloud: An auction approach to datacenter demand response, *Proc. of IEEE Conference*, Kowloon, Hong Kong, pp.2650-2658, 2015.
[7] L. Rao, X. Liu, L. Xie and W. Liu, Minimizing electricity cost: Optimization of distributed Internet data centers in a multi-electricity-market environment, *Proc. of the 29th Conference on Information Communications*, San Diego, CA, pp.1-9, 2010.
[8] P. Lama and X. Zhou, Coordinated power and performance guarantee with fuzzy MIMO control in virtualized server clusters, *IEEE Trans. Computers*, vol.64, pp.97-111, 2015.
[9] The free encyclopedia, *Wikipedia*, https://en.wikipedia.org/wiki/REN21.
[10] I. Goiri, M. E. Haque, K. Le, R. Beauchea, T. D. Nguyen, J. Guitart, J. Torres and R. Bianchini, Matching renewable energy supply and demand in green datacenters, *Ad Hoc Networks*, vol.25, pp.520-534, 2015.
[11] C. Li, A. Qouneh and T. Li, Characterizing and analyzing renewable energy driven data centers, *Acm Sigmetrics Performance Evaluation Review*, vol.39, no.1, pp.131-132, 2011.
[12] A. Krioukov, C. Goebel, S. Alspaugh, Y. Chen, D. Culler and Y. Katz, Integrating renewable energy using data analytics systems: Challenges and opportunities, *Bulletin of the IEEE Computer Society Technical Committee*, vol.34, pp.3-11, 2011.
[13] M. J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*, Morgan & Claypool, San Rafael, CA, 2010.
[14] *Facebook's Next Data Center to Rely Only on Renewable Energy*, http://www.computerworld.com/article/2945045/data-center/facebooks-next-data-center-to-rely-only-on-renewable-energy.html.
[15] *New Facebook Data Center Powered by 100% Renewable Energy*, http://apps3.eere.energy.gov/greenpower/news/news_template.shtml?id=1882.
[16] *Renewable Energy*, https://www.google.com/about/datacenters/renewable/.