2013 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing

Operating Cost Reduction for Distributed Internet Data Centers

Yangyang Li, Hongbo Wang, Jiankang Dong, Junbo Li, Shiduan Cheng
State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications
Beijing, 100876, China
Email: {yyli, hbwang, dongjk, lijunbo, chsd}@bupt.edu.cn

***Abstract*—In order to provide low latency, pay-as-you-go infrastructure, Cloud service provider has built a large number of geo-distributed data centers. Nowadays, energy cost has become increasingly important fraction of data center operating cost. Researchers' attentions are attracted to propose many methods to reduce the electricity cost for data centers. However, another significant operating cost, Internet bandwidth cost, has been neglected. Because both the electricity prices and bandwidth prices are varied with time and regions, it is not easy to minimize the operating cost while response to workload timely. As a solution, based on Lyapunov optimization framework, we propose an online control policy to achieve close-to-optimal performance with tradeoff between cost and delay. The advantage of our solution is that we do not need any future knowledge on the stochastic model (e.g. , bandwidth prices and electricity prices). The results from our simulations on real workload trace and price data demonstrated that our solution is effective.**

***Keywords*-Cloud Computing, Data center, Operating cost, Lyapunov optimization**

I. Introduction

With the proliferation of Cloud Computing and Internet online services, large scale, geographical distributed data centers have been built to meet the skyrocketing demand. Considering the typical distributed data centers architecture which is shown in Fig.1. The architecture consists of multiple workload dispatchers handling a diverse mix of workload. These workload dispatchers could be frontend HTTP proxies that route workload requests from a given local to the approximate data centers, the model is adopted by Google and Yahoo!. Or the workload dispatchers could be global load balancers, e.g. DNS lookup servers, which resolves the queries for the names of Websites, this model is adopted by content deliver providers such as Akamai and ChinaCache. The backend data centers response to workload requests which are distributed by workload dispatchers, and administrators can have flexible policies to choose how many servers should be activated.

Data centers typically comprise tens of thousands of servers. To operate such large scale data centers, the energy related cost is estimated to amount to 40% of operating cost [1]. And the total data center power consumption was already 1% of the total US power consumption in 2005 [2]. In this situation, many studies have been carried out to save the electricity cost for data centers. As shown in Fig.1, researchers explored the benefit of electricity price variations across time and locations

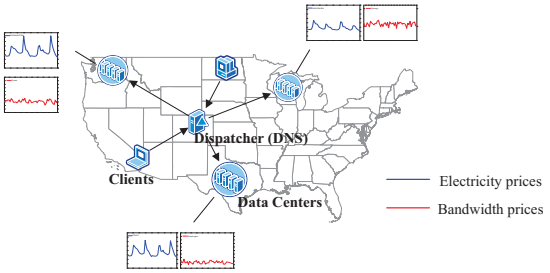


Fig. 1. The architecture of a typical distributed Internet data centers

TABLE I
AVERAGE DAY-AHEAD ELECTRICITY PRICE ($/MWH) ON NOVEMBER 8, 2012 AND MEDIAN BANDWIDTH PRICE ($/MBPS MONTHLY) BETWEEN NOV 13, 2011 AND NOV 13, 2012 IN DIFFERENT REGIONAL MARKETS. SOURCE: PJM [7], NETINDEX [8]

Region	Electricity price	Bandwidth price
Illinois	$34.64	$5.31
Texas	$46.55	$4.60
Washington	$47.29	$3.82
Wisconsin	$38.06	$3.80

and proposed several solutions to cut electricity cost generated by server subsystem in data centers [3], [4], [5], [6]. However, little attention has been paid to Internet bandwidth cost of data centers which also amounts to a significant fraction of data center operating cost.

Actually, hundreds of millions of dollars have been spent on the Internet bandwidth in order to meet the rapid growing of user demand and rich media applications. In some companies, the expenses of Internet bandwidth even account to more than 50% of their revenues. We use the daily day-ahead electricity price and the annual mean bandwidth price published online by PJM [7] and Net Index [8] respectively as shown in Table I to estimate the electricity cost and bandwidth cost of one server. We assume that a 1000-watt sever run no-stop in Texas, it would cost $1000 * 24 * 30/10^{-6} * 46.55 = \$33.516$ a month. Meanwhile, we assume the server has a capacity of 100Mbps (conservatively) connecting to Internet. The bandwidth cost of such server in Texas would be $100 * 4.6 = \$460$ monthly, which is about one order of magnitude higher than electricity cost of one server. Intuitively, reducing the Internet bandwidth cost

978-0-7695-4996-5/13 $26.00 © 2013 IEEE
DOI 10.1109/CCGrid.2013.106

seems to be more significant on reducing the operating cost of data centers. Nevertheless, we argue that the electricity cost of server subsystem should not be neglected. For one hand, currently, the power still mostly sources from nonrenewable resource such as coal and the power price should not be reduced quickly in the near future. For another hand, the price of Internet bandwidth continues to fall. The bandwidth cost is estimated to has a decline of 18% per year. Hence, to reduce the operating cost of data centers, we must consider both the electricity cost and bandwidth cost.

Because both the electricity prices and bandwidth prices are varied with time and regions, it is not easy to minimize the operating cost while response to workload timely. As a solution, based on Lyapunov optimization framework, we propose an online control policy to help the Cloud service providers make the following decision: (1) Decide the quantity of the workload that should be sent to each data center. (2) Determine the number of servers that should be activated in data centers. The advantage of our solution is that we do not need any future knowledge on the stochastic model (e.g. , bandwidth price and electricity price).

In summary, we have made following contributions in this paper:

- Based on real bandwidth price data in U.S, the Internet bandwidth cost of data centers is analyzed. We discover that the Internet bandwidth cost accounts to a significant fraction of operating cost for data centers.
- We present algorithm to approximately achieve the minimum time average expected electricity cost and bandwidth cost for distributed Internet data centers without the knowledge of the statistics of related stochastic models.
- Through extensive simulations using real-world traffic traces from a major Internet television company in China, as well as real-world electricity prices and bandwidth prices. We show that our algorithm can approach the optimal solution within $O(1/V)$ deviation and a tradeoff in average delay that is $O(V)$.

The remainder of this paper is organized as follows. In Section II we introduce related work. The system model is presented in Section III. In Section IV, we propose an algorithm to approximately solve the optimization problem. We evaluate our proposed solution in Section V and conclude in Section VI.

II. Related Work

During past years, researchers have paid much attention on reducing the electricity cost for large scale Internet systems. Asfandyar et al. [9] investigated the wholesale electricity markets in U.S. They found the electricity price vary on an hourly basis and not well correlated at different locations. Rao et al. [3] [4] extend the problem to multi-region electricity markets which better capture the present electricity price in data center locations. They utilized spatial variations of electricity price to minimal electricity cost for distributed internet data centers. Luo et al. [10] studied energy cost minimization for IDCs by exploiting the temporal diversity of electricity price, they designed a two-stage method and eco-IDC algorithm to trade delay for energy cost. The authors in [5] proposed a two time scale approach to reduce the power cost for delay tolerant workload which take both temporal and spatial volatility of power price into consideration. In [6], the author considered the transition cost when workload relocate between data centers, the transition cost in their paper is the extra power cost when activate and deactivate servers in data centers. However, none of the proposals considered the cost of bandwidth.

In fact, the growing number of data center output traffic attracted great interests in researchers. Chen et al. [11] investigated inter-data center traffic characteristics via five Yahoo! datacenters. Laoutaris et al. [12] utilized the distributed data centers to construct a storage node network which improve the utilization of bandwidth. The authors in [13] minimize the bandwidth cost on inter- data center traffic. Li et al. [14] proposed a scheme to reduce the inter-domain traffic between data centers which could be ISP friendly. And then, the authors extended the problem to multiple attribute aware scenarios [15] which consider more network attributes. All these research care about inter-data center traffic, but our problem is quite different and we need to consider the data center to client traffic and the prices of Internet bandwidth charged by ISPs.

The studies most relevant to ours are [9] [16]. In [9], the authors took bandwidth cost as a constraint. They did not consider the time variation of bandwidth cost, so their objective is different from ours. In [16], the authors assumed the workload and electricity prices can be precisely predicted. However, we consider the long term time average expected total cost (electricity and bandwidth cost) from the perspective of Cloud service providers and assume that the future knowledge of related stochastic models are unknown.

III. System Model

In this section, we describe the mathematical models for workload arrival, job distribution, data center operation, and cost model we use in this paper. We also present the control objective, which is to minimize the long term time average expected electricity cost and bandwidth cost.

As shown in Fig.2, we consider I workload dispatchers, denoted by $G = \{G_1, ..., G_I\}$. The system consists of J geographical distributed data centers, denoted by $DC = \{DC_1, ..., DC_J\}$, each data center has N_j^{max} homogeneous servers. We slot the time dimension into multiple time intervals with the same duration, denoted by $t = 0, 1, 2...$.

A. Service Model

In every time slot t, we denote the workload requests at each workload dispatcher as $A_i(t)$, where $A(t) = (A_1(t), ..., A_I(t))$ denotes the request vector. We assume that $A(t)$ are $i.i.d$ every time slot with $E\{A(t)\} = \lambda \triangleq (\lambda_1, ..., \lambda_I)$. We also assume that during one time slot, the maximum value of requests is no more than A_{max}, that is,

$$0 \leq A_i(t) \leq A_{max} \tag{1}$$

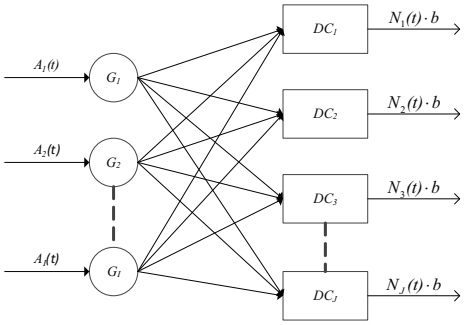


Fig. 2. System model

At the beginning of each time slot, the workload dispatcher need to decide how much requests should be distributed to each data center. We use the vector $\mu_i(t) = (\mu_{i1}(t), ..., \mu_{iJ}(t))$ to represent the dispatch rates from dispatcher i to each data center. There is a maximum distribute rates μ_{max} which the dispatcher i can distribute to data center j for one time slot, i.e.,

$$0 \leq \mu_{ij}(t) \leq \mu_{max} \tag{2}$$

Each request demand resource from data centers, both the requests transiting and data centers' responses can generate the traffic. The traffic volume of users' requests is usually far smaller than that of responses. Hence, to simplify our problem, we only consider the traffic generated by response and assume the traffic volume generated by response for per request is θ.

For each data center, the number of activated servers should be determined at the start of each time slot t, we denote it as $N_j(t)$, and it is bounded by

$$N_j^{min} \leq N_j(t) \leq N_j^{max} \tag{3}$$

Where N_j^{min} and N_j^{max} is the minimum and maximum number of servers be activated at all time slots for data center j respectively. To simplify our model, we assume the servers in each data center are homogeneous and process users' demand with a fixed rate b.

To prevent from increasing the bandwidth cost, the total traffic of each data center transiting during one time slot is no more than $C_j(t)$. In next subsection we will describe the bandwidth pricing model. As the Internet bandwidth capacity of each data center is constrained by ISPs and links of local regions, the traffic volume is bounded by C_j^{max}, that is,

$$N_j(t) \times b \times \theta \leq C_j(t) \leq C_j^{max} \tag{4}$$

B. Cost Model

The total cost we concerned consists of electricity cost and bandwidth cost. We first describe the electricity cost model we use and then give the model of bandwidth cost.

In a real-life electricity market, the power price is fluctuant with varied time and regions. To capture such dynamic pattern, we use vector $p^e(t) = (p_1^e(t), ..., p_J^e(t))$ to represent the electricity price of each data center at time slot t. The power consumption of data centers can be modeled as a function $P_j(\cdot) \times N_j(t)$, where $P_j(\cdot)$ is the power consumption of one server in data center j. We don't specify the form of this function, it could be any continuous and concave function. Then the electricity cost of each data center can be expressed as:

$$f_j^e(t) = P_j(\cdot) \times N_j(t) \times p_j^e(t) \tag{5}$$

Bandwidth is usually charged by ISPs using the basic 95/5 billing model. The ISP records the traffic volume which data center generated during 5-minute intervals and 95^{th} percentile of bandwidth is used for billing. So the bandwidth price is varied with time which depends on the traffic volume by data center. The reason we constrain the traffic in expression (4) is to avoid be charged in a higher bandwidth price. What's more, different ISPs locate in different regions and have different network resources, the bandwidth is charged with different price even at the same time slot. To reflect such a variety, similar to the electricity price, we use a vector $p^b(t) = (p_1^b(t), ..., p_J^b(t))$ to represent. The bandwidth cost of each data center can be calculated as:

$$f_j^b(t) = N_j(t) \times b \times \theta \times p_j^b(t) \tag{6}$$

TABLE II
KEY NOTATIONS IN THE SYSTEM MODEL

Symbol	Description
I	Number of workload dispatchers
J	Number of data centers
$A_i(t)$	Workload arrivals to dispatcher i at time slot t
A_{max}	Maximum of workload demand at each time slot
λ	Expectation of workload demand
$\mu_{ij}(t)$	Number of workload be distributed by dispatcher i to data center j at time slot t
μ_{max}	Maximum distribute rate from dispatcher to each data center
θ	Traffic volume generated for per request
$C_j(t)$	Bandwidth constraint in data center j at time slot t
C_j^{max}	Maximum bandwidth constraint in data center j
$N_j(t)$	Number of activated servers in data center j at time slot t
N_j^{min}	Minimum of activated servers in data center j
N_j^{max}	Number of servers in data center j
b	Servers service rate
$p^e(t)$	Electricity prices of data centers at time slot t
$P_j(\cdot)$	Power consumption function of per server in data center j
$f_j^e(t)$	Electricity cost function
$p^b(t)$	Bandwidth prices of data centers at time slot t
$f_j^b(t)$	Bandwidth cost function at time slot t
$G_i(t)$	Backlog of dispatcher queue i at time slot t
$DC_j(t)$	Backlog of data center queue j at time slot t
$Z_j(t)$	Backlog of virtual queue j at time slot t

C. Design Objective

In this paper, we are interested in long-term electricity cost and bandwidth cost. Hence, our objective is to minimize the long-term time average expected total cost as described below:

$$minimize \quad F = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \sum_{j=1}^{J} \mathbb{E}\{f_j^e(t) + f_j^b(t)\} \tag{7}$$

Where the expectation is with respect to possibly randomized control actions $\mu_{ij}(t)$, $N_j(t)$ as well as the distribution of electricity prices $p^e(t)$ and bandwidth prices $p^b(t)$.

IV. Algorithm and Analysis

A. Lyapunov Optimization Framework

Based on Lyapunov optimization framework, we construct two groups of actual queues and one group of virtual queues to formulate our problem. Firstly, we assume that when workload arrives, the workload are stored in queue $G_i(t)$ to await be distributed. The queueing dynamics are then:

$$G_i(t+1) = \max[G_i(t) - \sum_{j=1}^{J} \mu_{ij}(t), 0] + A_i(t) \quad (8)$$

We call $G_i(t)$ the backlog at time slot t for workload dispatcher i, as it can represent an amount of workload that need to be dispatched.

Similarly, we construct another group of actual queues $DC_j(t)$ for backend data centers. We assume that the workload distributed to data center j is stored in the queues with update equation:

$$DC_j(t+1) = \max[DC_j(t) - N_j(t)b, 0] + \sum_{i=1}^{I} \mu_{ij}(t) \quad (9)$$

Besides, to capture the traffic volume constraint, we define the group of virtual queues with the update equation as following:

$$Z_j(t+1) = \max[Z_j(t) + N_j(t)b - \frac{C_j(t)}{\theta}, 0] \quad (10)$$

Here, we proof that the equation (10) can ensure the time average expected traffic transited by each data center is bounded by C_j^{max}.

Proof: By (10), we have for any slot $t \geq 0$:

$$Z_j(t+1) \geq Z_j(t) + N_j(t)b - \frac{C_j^{max}}{\theta} \quad (11)$$

Here, we using the fact that $C_j(t) \leq C_j^{max}$. Rearranging terms in the above inequality:

$$Z_j(t+1) - Z_j(t) \geq N_j(t)b - \frac{C_j^{max}}{\theta} \quad (12)$$

Summing the above over $t \in 0, ..., T-1$ and using the law of telescoping sums yields:

$$Z_j(T) - Z_j(0) \geq \sum_{t=0}^{T-1} N_j(t)b - \frac{C_j^{max}}{\theta}T \quad (13)$$

Rearranging terms, dividing by T, using the fact that $Z_j(0) \geq 0$, and taking a limit as $T \to \infty$ yields:

$$\lim_{T\to\infty} \frac{Z_j(T)}{T} \geq \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} N_j(t)b - \frac{C_j^{max}}{\theta} \quad (14)$$

Taking expectations of the above, and using the fact that, if $Z_j(t)$ is mean stable then $\lim_{T\to\infty} \frac{\mathbb{E}\{Z_j(t)\}}{T} = 0$ yields:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\{N_j(t)b\} \leq \frac{C_j^{max}}{\theta} \quad (15)$$

This means our desired time average constraint is satisfied. ■

Let $\Theta(t) = [G_i(t), DC_j(t), Z_j(t)]$ be a concatenated vector of all actual and virtual queues, with update equations(8), (9), (10). We define the Lyapunov function:

$$L(\Theta(t)) \triangleq \frac{1}{2}\sum_{i=1}^{I} G_i(t)^2 + \frac{1}{2}\sum_{j=1}^{J} DC_j(t)^2 + \frac{1}{2}\sum_{j=1}^{J} Z_j(t)^2 \quad (16)$$

Define the *one-slot conditional Lyapunov drift* $\Delta(\Theta(t))$ as follows:

$$\Delta(\Theta(t)) \triangleq \mathbb{E}\{L(\Theta(t+1)) - L(\Theta(t))|\Theta(t)\} \quad (17)$$

Following the Lyapunov optimization framework, we minimize the drift-plus-penalty at each time slot t as expression:

$$\Delta(\Theta(t)) + V\{\sum_{j=1}^{J} \mathbb{E}\{f_j^e(t) + f_j^b(t)|\Theta(t)\}\} \quad (18)$$

Where $V \geq 0$ is a parameter that represents an *importance weight* on how much we emphasize electricity cost and bandwidth cost minimization.

The proposed algorithm for minimize time average expected electricity cost and bandwidth cost is shown in Algorithm 1.

Algorithm 1: Online control policy for minimizing electricity cost and bandwidth cost

Input: values of the system states: $A_i(t)$, $p_j^e(t)$, $p_j^b(t)$), $G_i(t)$, $DC_j(t)$, $Z_j(t)$
Output: control actions: $\mu_{ij}(t)$, $N_j(t)$

for *each time slot* t **do**
- Update values of $A_i(t)$, $p_j^e(t)$, $p_j^b(t)$, $G_i(t)$, $DC_j(t)$, $Z_j(t)$
- Find the solution for the following minimization problem:
- **minimize** $\sum_{i=1}^{I}\sum_{j=1}^{J} \mu_{ij}(t)[DC_j(t) - G_i(t)] + \sum_{j=1}^{J} N_j(t)[Z_j(t)b - DC_j(t)b + VP_j(\cdot)p_j^e(t) + Vp_j^b(t) \cdot b \cdot \theta]$
- **subject to** (1)(2)(3)(4)(8)(9)(10)

return $\mu_{ij}(t), N_j(t)$

B. Properties of Algorithm

Lemma 1. Suppose each system state is $i.i.d$ over time slots. Under any control algorithms, the drift-plus-penalty expression has the following upper bound for all t, all possible values of

$\Theta(t)$,and all parameters $V > 0$, we have:

$$\Delta(\Theta(t)) + V\{\sum_{j=1}^{J}\mathbb{E}\{f_j^e(t)+f_j^b(t)\}|\Theta(t)\} \leq$$

$$B+\mathbb{E}\{\sum_{i=1}^{I}G_i(t)A_i(t)|\Theta(t)\} \quad +\mathbb{E}\{\sum_{j=1}^{J}[Z_j(t)-\frac{C_j(t)}{\theta}]|\Theta(t)\}$$

$$+\mathbb{E}\{\sum_{i=1}^{I}\sum_{j=1}^{J}\mu_{ij}(t)[DC_j(t)-G_i(t)]|\Theta(t)\}$$

$$+\mathbb{E}\{\sum_{j=1}^{J}N_j(t)[Z_j(t)b-DC_j(t)b+Vp_j(\cdot)p_j^e(t)+VP_j^b(t)\cdot b\cdot\theta] \tag{19}$$

Here, $B \triangleq \frac{IJ^2\mu_{max}^2+IA_{max}^2+2J(N_j^{max}b)^2+JI^2\mu_{max}^2+J[\frac{C_j^{max}}{\theta}]^2}{2}$.

Proof: See [17]. ■

Comparing with the objective of Algorithm 1, it is obvious that our algorithm always attempts to greedily minimize the right hand side of (19) for each time slot t over all possible feasible control policies.

C. Performance bound

In this section, we analyze the performance bound of our proposed minimize power cost and bandwidth cost algorithm. Before presenting the bound, we characterize the optimal time average expected electricity cost and bandwidth cost $f^{opt} = [f_j^e(t)+f_j^b(t)]^{opt}$ that can be achieved by any other algorithms which stabilize the queues. And we denote Λ as the capacity region of the system.

Theorem 1 (Optimality over stationary randomized policies)**.** For any rate vector $\lambda \in \Lambda$, there exists a stationary randomized control policy *opt* that choose $\mu_{ij}(t)$,$N_j(t)$ every time slot t, and achieves the following

$$\sum_{j=1}^{J}\mathbb{E}\{f^{opt}(t)\} = f_{av}^{opt}(\lambda) \tag{20}$$

$$\mathbb{E}\{A_i(t)\} = \mathbb{E}\{\sum_{j=1}^{J}\mu_{ij}^{opt}(t)\} \tag{21}$$

$$\mathbb{E}\{\sum_{i=1}^{I}\mu_{ij}^{opt}(t)\} = \mathbb{E}\{N_j^{opt}(t)b\} \tag{22}$$

$$\mathbb{E}\{N_j^{opt}(t)\} = \frac{C_j(t)}{\theta} \tag{23}$$

Here, $f_{av}^{opt}(\lambda)$ is the optimal time average expected of electricity cost and bandwidth cost.

Proof: It can be proven using Caratheodory's theorn in [18] and is omitted here for brevity. ■

Theorem 2 (Performance of minimize electricity cost and bandwidth cost algorithm)**.** Suppose each system state is $i.i.d$ over time slots and there exists an $\varepsilon > 0$ such that $\lambda + 2\varepsilon\mathbf{1} \in \Lambda$. The problem (7) is feasible, and that , then:

1) Time averaged expected cost satisfies:

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\sum_{j=1}^{J}\mathbb{E}\{f_j^e(t)+f_j^b(t)\} \leq \frac{B}{V}+f_{av}^{opt}(\lambda) \tag{24}$$

2) All queues $G_i(t)$, $DC_j(t)$, $Z_j(t)$ are mean rate stable.
3) Under our algorithm, we have

$$\lim_{T\to\infty}\sum_{t=0}^{T-1}\{\sum_{i=1}^{I}\mathbb{E}\{G_i(t)\}+\sum_{j=1}^{J}\mathbb{E}\{DC_j(t)\}+\sum_{j=1}^{J}\mathbb{E}\{Z_j(t)\}\} \leq \frac{B+Vf_{av}^{opt}(\lambda+2\varepsilon 1)}{\varepsilon} \tag{25}$$

Here **1** denotes the vector of all 1's.

Proof: See [17]. ■

The above Theorem guarantees that our algorithm with worst case delay of $O(V)$ that comes within $O(1/V)$ of minimizing the operating cost.

V. Performance Evaluation

In this section, we conduct several simulations to evaluate the proposed algorithm using real-world trace data sets and electricity, bandwidth prices.

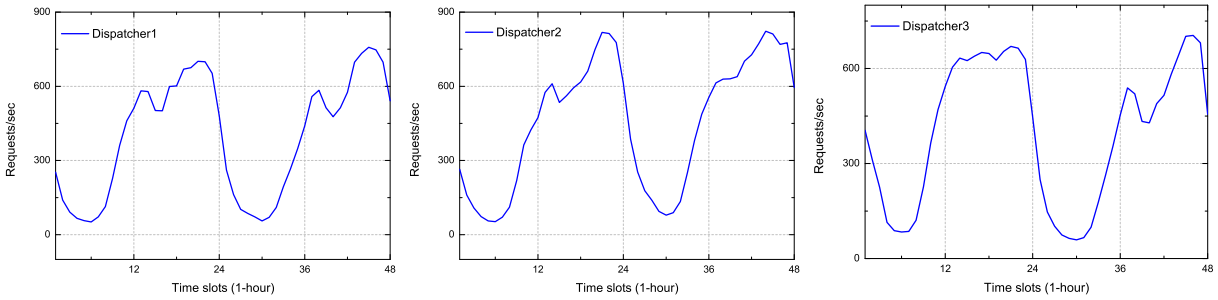
A. Simulation Setup

In our simulations, We model a small architecture of distributed data centers with 3 workload dispatchers and 4 data centers.

1) Workload arrival rates: To capture the workload requests arrival at each dispatcher, we use real-world traces collected from an major Internet television company in China. We acquired the workload request rates of 3 dispatchers distributed in different locations from a subset of the company's dispatcher. The data is sampled every hour in a 48 hours period on November 8, 2012 to November 9, 2012 as shown in Fig. 3. The reason why we use traces from the company is because the traces can reflect a faithful workload. Hence, it is suitable to use the traces for evaluating the proposed algorithm.

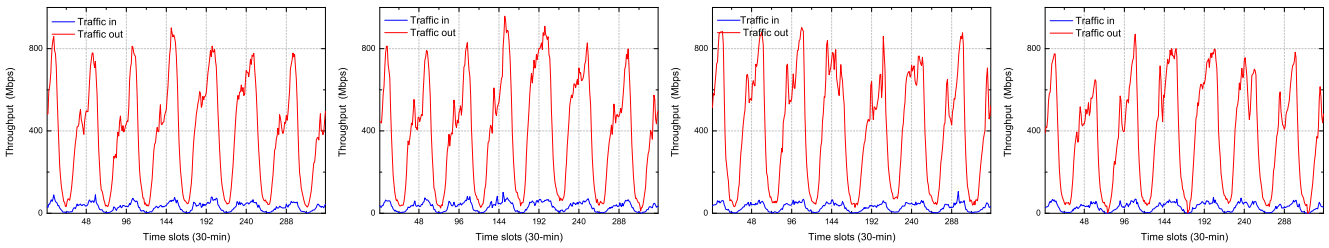
In the trace, we can see that the requests are varied with time. Specifically, at the beginning of the first 10 time slots, the number of user request is quite small. Since then, the number is increasing rapidly. It was considered that the demand of these regions follows strong diurnal patterns [12].

2) Bandwidth capping: A large online service provider usually deploy servers in clusters where each cluster consists of tens of hundreds of servers in a particular data center in a specific locations. In order to estimate the output traffic $C_j(t)$ of each data canter at time slot t, we collected traffic data from 4 servers located in different regions on November 6 to November 13, 2012, as shown in Fig.4. The duration of each time slot is 30 minutes. To simplify our problem, we assume the servers in each data center are all homogenous. And we use a back-of-the-envelop approach to calculate the total output traffic of each data center by summing the traffic of all servers. In this paper, we assume each data center has



(a) Dispatcher 1 (b) Dispatcher 2 (c) Dispatcher 3

Fig. 3. Dispatcher request rates from an Internet television company data set.



(a) Server in city A (b) Server in city B (c) Server in city C (d) Server in city D

Fig. 4. Throughput of servers from an Internet television company data set.

1000 servers and with minimum 800 servers to be activated . And we let C_j^{max} equal to the largest output traffic during 48 hours. Based on the statistics of dispatchers and data centers, we estimated the average response traffic θ for per request. In our simulations, $\theta = 50$ Mb. The trace also shows that the demand follows diurnal pattern. Compared to traffic which is sent out from the data center, the traffic generated from users' demand is neglectable. That's why we only consider output traffic generated by data center in our system model.

3) Electricity prices: To capture the electricity price of the region where each data center located in, we collected day-ahead hourly electricity price on November 8 to November 9, 2012, as shown in Fig.5 [7]. We assume that our 4 data centers are deployed across the continental U.S, and in each region we randomly choose the hourly electricity price of one hub to reflect regional electricity market. The electricity price in Chicago and Wisconsin are far less than that in Texas and Washington. Intuitively, distributing the workload to Chicago and Wisconsin can save the electricity cost. And we can see that the peak electricity price appear during 7:00-9:00 and 17:00-19:00, electricity cost can be reduced by delaying the workload to avoid the peak electricity price.

4) Bandwidth prices: Internet bandwidth price is varied from ISPs and the bandwidth market of the local region. It's difficult to survey the bandwidth price of each ISP charged. However, we can use the mean bandwidth price given by NetIndex [8] which is based on hundreds of thousands of survey and results of test to estimate the bandwidth prices of the regions where our 4 data centers are deployed in. The value is shown in Table I. In order to reflect the temporality of bandwidth price, we generate the hourly bandwidth price by using Poisson distribution with mean values got from Table I.

5) Power consumption model: We use a linear electricity consumption function which is respect to the average CPU utilization as the authors used in [6] [19].

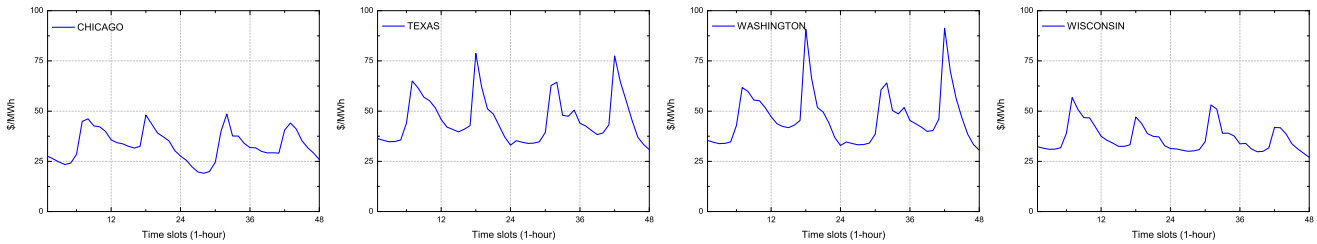
$$P_j(\cdot) = P_{idle} + U_j(P_{peak} - P_{idle}) \tag{26}$$

We assume that each server in data center has the same CPU utilization: 60%, because research on adjusting service rate so as to save the server power consumption is out of the scope of this paper. For simplify, we assume all workload requests are distributed to 4 data centers. So we can estimate the service rate b by calculating how many requests a server can response on average. In this simulation b is equal to 15 requests per time slot. And we set $P_{idle} = 200$ Watts and $P_{peak} = 400$ Watts respectively [20]. We emphasize that the proposed algorithm is suitable for any continuous and concave function rather than the linear one we used in the simulations.

6) Remarks: Limited by the real traces and electricity, bandwidth price we acquired, we conduct our experiment with 1 hour duration of each time slot. Though one hour is too coarse to make a precise control policy, we still get effective operating cost reduction as the results shown in the following evaluation. And we argue that our proposed algorithm is not limited by the duration of the time slots.

B. Algorithms for comparison

In this paper, we compare the following 3 control polices.



(a) Chicago (b) Texas (c) Washington (d) Wisconsin

Fig. 5. Daily averages of day-ahead hourly electricity price at different regions

ECR is short for Energy Cost Reduction, which is the control policy that only consider to minimize the electricity cost as reported in [5]. Here we implement a simplified algorithm just let $P_j^b(t)$ to be zero and do not adjust the server service rate b.

LB is short for Load Balancing. Workload dispatcher distribute the workload to each data center evenly. It means when the workload arrives, the dispatcher distributes the demand by round robin regardless of the states of backend data centers..

OCR is our proposed Operating Cost Reduction algorithm, which optimize the total operating cost consists of both power cost and bandwidth cost.

C. Results and analysis

In this section, we evaluate the impact of long term slotted time T firstly and then consider the impact of parameter V.

1) Impact of long-term slotted time T: The result illustrated in Fig.6 shows that the total cost increases with the long-term slotted time T with a fixed parameter $V = 0.1$ under different algorithms. The result also shows our proposed algorithm can achieve the maximum operating cost reduction, the larger the T is, the more operating cost reduction our proposed algorithm can obtain. Though the result of energy cost reduction algorithm is close to our result, we still can save extra 10000 dollars per 48 hours. It means that our algorithm can save extra 2 million dollars a year! In this paper, we consider a distributed data centers architecture has 4000 servers in total. Organizations such as Google with hundreds of thousands of servers is not uncommon. Our proposed algorithm can save even more operating cost for these large company.

It is shown in Fig.7 that the backlog of data center is varied with workload demand, see Fig.3. Both our proposed method and electricity cost reduction algorithm maintain a higher backlog than load balancing algorithm. It is because that the nature of Lyapunov optimization based method is trade delay for cost reduction. However, backlog increase under our proposed algorithm is not obviously larger than that under electricity cost reduction algorithm. Whereas our solution can save more operating cost. It validates that our algorithm is more effective.

2) Impact of parameter V: As shown in Fig.8. We conduct experiments with different parameter V, and calculate the average hourly total cost with a fixed $T = 48$ hours. It shows

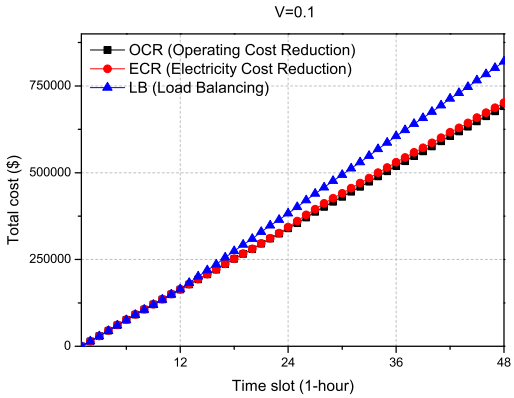


Fig. 6. Comparison of total cost

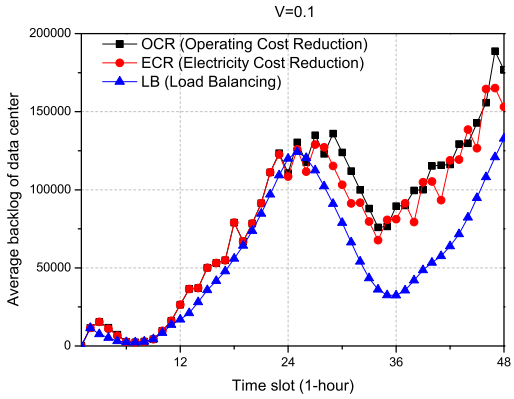


Fig. 7. Comparison of average backlog of data centers.

that both ECR and OCR algorithms can reduce total cost remarkably compared to LB. Whereas our algorithm is more sensitive to parameter V, to mentioned that parameter V is the *importance weight* of cost compared to queueing stability. Hence even the weight is small, our OCR method still can reduce the total operating cost and our algorithm can achieve the optimal value faster with the increasing of V. The result also verifies that OCR confirms *Theorem 2*, which means OCR can approach the optimal solution within diminishing gap of $O(1/V)$.

Fig.9 shows average hourly backlog of data centers versus

parameter V. Similarly, to minimize the total operating cost, our OCR algorithm is more sensitive to delay the workload and await to be done with a cheaper electricity and/or bandwidth prices. The result also confirms *Theorem 2*, which means the average hourly data center backlog is bounded by $O(V)$.

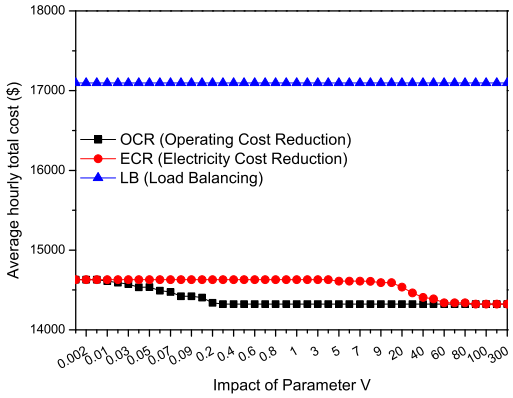


Fig. 8. Average hourly total cost versus V

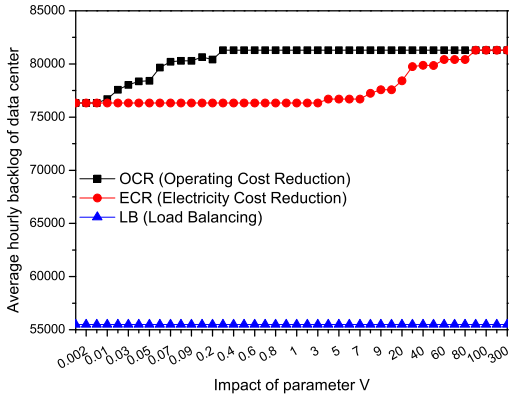


Fig. 9. Average hourly backlog of data centers versus V

VI. Conclusion

In this paper, we have proposed an approach to cut the operating cost for distributed Internet data centers which consider both the power cost and bandwidth cost of the server subsystem. The intuition behind our approach is to trade delay for operating cost. We built a group of virtual queues to guarantee the output traffic of each data center during each charging cycle is constrained so as to avoid rise in bandwidth cost. Both mathematical analysis and real trace driven simulations illustrated that our approach can achieve to $O(1/V)$ of optimal operating cost and within an $O(V)$ tradeoff in time average queue backlogs. Simulations showed that our proposed approach can reduce operating cost effectively.

Acknowledgment

This work is supported by the National Natural Science Foundation of China(No. 61002011); the 863 Program of China(No.s 2013AA013303, 2011AA01A102); the Open Fund of the State Key Laboratory of Software Development Environment(No. SKLSDE-2009KF-2-08), Beijing University of Aeronautics and Astronautics; the 973 Program of China(No. 2009CB320505).

References

[1] A. Greenberg, J. Hamilton, D. A. Maltz, and P. Patel, "The cost of a cloud: research problems in data center networks," *SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 1, pp. 68–73, 2008.

[2] J. G. Koomey, "Worldwide electricity used in data centers," *Environmental Research Letters*, vol. 3, no. 3, 2008.

[3] L. Rao, X. Liu, M. D. Ilic, and J. Liu, "Distributed coordination of internet data centers under multiregional electricity markets," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 269–282, 2012.

[4] L. Rao, X. Liu, L. Xie, and W. Liu, "Minimizing electricity cost: Optimization of distributed internet data centers in a multi-electricity-market environment," in *INFOCOM, 2010 Proceedings IEEE*. IEEE, 2010, pp. 1–9.

[5] Y. Yao, L. Huang, A. Sharma, L. Golubchik, and M. Neely, "Data centers power reduction: A two time scale approach for delay tolerant workloads," in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 1431–1439.

[6] M. S. Ilyas, S. Raza, C. C. Chen, Z. A. Uzmi, and C. N. Chuah, "Red-bl: Energy solution for loading data centers," in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 2866–2870.

[7] "Pjm," http://www.pjm.com/, [Online; accessed 12-Nov-2012].

[8] "Net index by ookla," http://www.netindex.com/, [Online; accessed 12-Nov-2012].

[9] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs, "Cutting the electric bill for internet-scale systems," *SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 4, pp. 123–134, Aug. 2009.

[10] J. Luo, L. Rao, and X. Liu, "eco-idc: Trade delay for energy cost with service delay guarantee for internet data centers," in *Cluster Computing (CLUSTER), 2012 IEEE International Conference on*. IEEE, 2012, pp. 45–53.

[11] C. Yingying, S. Jain, V. K. Adhikari, Z. Zhi-Li, and X. Kuai, "A first look at inter-data center traffic characteristics via yahoo! datasets," in *2011 Proceedings IEEE INFOCOM*, 2011, pp. 1620–1628.

[12] N. Laoutaris, M. Sirivianos, X. Yang, and P. Rodriguez, "Inter-datacenter bulk transfers with netstitcher," in *Proceedings of the ACM SIGCOMM 2011 conference*. Toronto, Ontario, Canada: ACM, 2011, pp. 74–85.

[13] Y. Feng, B. Li, and B. Li, "Postcard: Minimizing costs on inter-datacenter traffic with store-and-forward," in *Proceedings of the 2nd International Workshop on Data Center Performance (DCPerf 2012)*, Macau, China, 2012.

[14] Y. Li, H. Wang, P. Zhang, J. Dong, and S. Cheng, "D4d: Inter-datacenter bulk transfers with isp friendliness," in *Cluster Computing (CLUSTER), 2012 IEEE International Conference on*. IEEE, 2012, pp. 597–600.

[15] ——, "Multi-attribute aware scheduling for inter-datacenter bulk transfers," *Chinese Journal on Communications*, vol. 33, no. Z1, pp. 121–131, 2012.

[16] X. Zheng and Y. Cai, "Energy-aware load dispatching in geographically located internet data centers," *Sustainable Computing: Informatics and Systems*, vol. 1, no. 4, pp. 275 – 285, 2011.

[17] Y. Li, H. Wang, J. Li, and S. Cheng, "Joint electricity cost and bandwidth cost reduction for distributed data centers," Beijing University of Posts and Telecommunications, Tech. Rep., 2012.

[18] L. Georgiadis, M. J. Neely, and L. Tassiulas, "Resource allocation and cross-layer control in wireless networks," *Foundations and Trends? in Networking*, vol. 1, no. 1, pp. 1–144, 2006.

[19] V. Mathew, R. K. Sitaraman, and P. Shenoy, "Energy-aware load balancing in content delivery networks," in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 954–962.

[20] "The problem of power consumption in servers," http://www.infoq.com/articles/power-consumption-servers/, [Online; accessed 12-Nov-2012].