JET: Electricity Cost-aware Dynamic Workload Management in Geographically Distributed Datacenters

Zehua Guo[a,*], Zhemin Duan[a], Yang Xu[b], H. Jonathan Chao[b]

[a] *School of Electronics and Information, Northwestern Polytechnical University, Xian, China, 710072*
[b] *Department of Electrical and Computer Engineering, Polytechnic Institute of New York University, New York, USA, 11201*

Abstract

The ever-increasing operational cost of geographically distributed datacenters has become a critical issue for cloud service providers. In order to cut the electricity cost of geographically distributed datacenters, several workload management schemes have been proposed, such as Electricity price-aware InteR-datacenter load balancing (EIR), which reduces the electricity cost of active servers by dispatching workload to datacenters with lower electricity prices, and Cooling-aware IntrA-datacenter load balancing (CIA), which decreases the power consumption of a datacenter by consolidating workload on servers with high cooling efficiency. However, these existing schemes could incur some undesired results. For example, EIR may result in high electricity cost of cooling systems due to random workload distribution in datacenters. CIA could lead to high electricity cost of active servers since it does not consider the variation of electricity prices. In this paper, we propose a joint inter- and intra-datacenter workload management scheme, Joint ElectriciTy price-aware and cooling efficiency-aware load balancing (JET), to cut the electricity cost of geographically distributed datacenters. JET uses a short processing time to calculate the optimal workload distribution, which trades off the electricity cost of active servers and cooling systems by alternately selecting the electricity prices or the efficiency of a cooling system as the dominator factor to the electricity cost of geographically distributed datacenters. Extensive evaluations show that JET outperforms the existing schemes and achieves substantial reduction on the electricity cost of geographically distributed datacenters.

Keywords: Geographically distributed datacenters; Dynamic workload management; Electricity cost; Electricity price; Efficiency of a cooling system

1. Introduction

As the demand of resilient and low-latency cloud services increase in recent years, cloud services providers, such as Amazon, Google, and Microsoft, have been rapidly deploying and expanding their geographically distributed datacenters. A recent report shows that the electricity demand of worldwide datacenters increased by about 56% from year 2005 to 2010, and the electricity usage of datacenters accounted for 1.1%$\sim$1.5% of the worldwide electricity usage in 2010 [1]. Generally, a datacenter spends 30%$\sim$50% of its operational expense toward electricity [2]. Therefore, cutting down on the electricity cost of geographically distributed datacenters has become a major effort of cloud service providers.

Many studies have been conducted to seek optimal datacenter workload management schemes on the purpose of reducing the electricity cost of geographically distributed datacenters. Some studies focused on the *intra*-datacenter workload management, which decreases the power consumption of active servers in a datacenter by dynamic server demand management mechanisms [3][4]. Since the power consumption of a cooling system can take up to 50% of the total power consumption of a datacenter [5], other studies considered the impact of workload distribution in a datacenter on the power consumption of a cooling system and proposed the *intra*-datacenter workload management to reduce the power consumption of a cooling system [6][7]. Due to the location and time diversities of electricity prices in the United States, the *inter*-datacenter workload management was proposed to minimize the electricity cost[1] of active servers of geographically distributed datacenters by periodically distributing workload to datacenters with lower electricity prices [8][9].

In recent publications, server facilities and cooling systems are jointly considered to lower the electricity cost of geographically distributed datacenters [10][11]. However, these works only consider a cooling system as a device consuming constant power and neglect variant efficiency of a cooling system resulting from diverse workload distribution in a datacenter. [12] exploited the impact of workload distribution in a datacenter on the efficiency of

*Corresponding author
Email addresses: guolizihao@hotmail.com (Zehua Guo), zhemind@nwpu.edu.cn (Zhemin Duan), yangxu@poly.edu (Yang Xu), chao@poly.edu (H. Jonathan Chao)

[1] The electricity cost of a datacenter equals the power consumption of the datacenter multiplied by the electricity price of the datacenter location.

a cooling system and proposed that a datacenter can be divided into three nearly temperature isolated zones (i.e., cool, warm, and hot zone) based on air flow patterns in the datacenter. The power consumption for cooling the same number of servers in the hot zone is larger than that in the other two zones. Therefore, the workload distribution in different zones of a datacenter significantly affects the power consumption of a cooling system, and it should be taken into consideration as an important factor by cloud service operators to reduce the electricity cost of cooling systems of geographically distributed datacenters.

In this paper, we propose a joint *inter- and intra-*datacenter workload management scheme, Joint ElectriciTy price-aware and cooling efficiency-aware load balancing (JET), to minimize the total electricity cost of geographically distributed datacenters. We first model the electricity cost of active servers as a function of electricity prices and the number of active servers. We also model the electricity cost of a cooling system as a function of Computer Room Air Conditioner (CRAC) output temperature. Based on these models, we propose a electricity cost minimization model of geographically distributed datacenters that effectively integrates both electricity price and cooling system management, and formulate the Electricity Cost Minimization (ECM) problem of geographically distributed datacenters as a constrained nonlinear optimization problem, subject to constraints of Quality of Service (QoS). We then simplify the complicated ECM problem to the Transformed ECM (TECM) problem (a convex optimization problem with linear constraints) by reasonable transformations and assumptions and solve the TECM problem with the proposed JET. Extensive evaluations based on the real-life workload trace and electricity prices of multiple datacenter locations show that JET calculates the optimal workload distribution with a short processing time and substantially reduces the electricity cost of geographically distributed datacenters, as compared with the existing schemes. To the best of our knowledge, our work presents the first study on cutting the electricity cost of geographically distributed datacenters with a joint inter- and intra-datacenter workload management scheme.

The major contributions of this paper are summarized as follows:

1. We build a model for the efficiency of a cooling system in a datacenter with respect to CRAC output temperature and apply this model to the geographically distributed datacenters. Using this model, we further propose the ECM problem through explicitly enforcing the impact of workload distribution in a datacenter on the efficiency of a cooling system in existing works.
2. We make reasonable transformations and assumptions on variables and constraints to transform the ECM problem (a complicated constrained nonlinear optimization problem) to the TECM problem (a convex optimization problem with linear constraints). We solve the TECM problem with the proposed JET, which dynamically dispatches incoming service requests[2] to active servers in three temperature isolated zones of geographically distributed datacenters by jointly considering time-varying locational electricity prices and the impact of workload distribution in a datacenter on the efficiency of a cooling system.
3. We evaluate the performance of JET against the existing schemes based on real-life traces. Extensive evaluations show that JET performs much better than the existing schemes and achieves substantial reduction on the electricity cost of geographically distributed datacenters. Compared with service deadlines in real-life applications, the processing time of JET is really short, and it changes very slightly as the amount of workload changes.

The rest of the paper is organized as follows. Section 2 discusses the motivation of this paper. Section 3 presents system framework. Section 4 describes electricity cost models and QoS constraints used for problem formulation. Section 5 formally proposes the ECM problem and solves the problem by the proposed JET. Section 6 discusses the evaluation strategy and compares JET with the existing schemes using real-life traces. Section 7 reviews the related work and Section 8 concludes the paper.

2. Motivation

2.1. Variation of electricity prices

Electricity is produced by government utilities and independent power producers from different sources, such as coal, natural gas, and renewable energy. Providers and consumers of electricity power are usually connected to an electricity grid, which is a complex electricity transmission and distribution network. Consider the United States as an example, its electricity grid is divided into eight regional grids, each of which is operated and managed by a Regional Transmission Organization (RTO). Each RTO is responsible for setting up and directing electricity flows over the grid, and ensures the short-term reliability of the grid. There are two main parallel electricity markets in the United States: the retail market and the wholesale market. In the retail electricity market, electricity prices are fixed for a certain period of time, while electricity prices may vary on a 5-min or 1-hour basis in the wholesale electricity market. Empirical market analysis from public data archives shows that electricity prices in the wholesale market exhibit a significant amount of day-to-day and hour-to-hour geographically uncorrelated volatility [8]. The electricity prices of locations in different regional markets are never highly correlated, even nearby locations in the same region are not always well correlated [8].

[2]We use service requests and workload interchangeably in this paper.

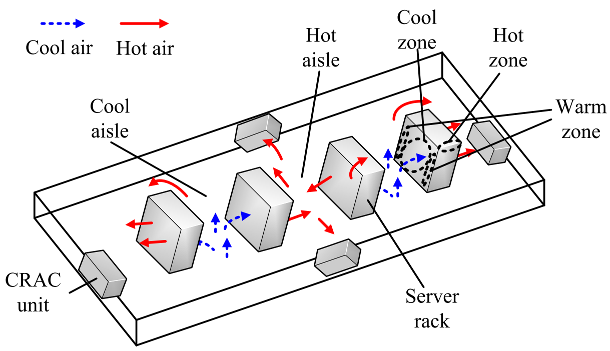


Figure 1: Water-chilled datacenter cooling system [12].

2.2. Dynamic service request routing mechanism

Currently, most cloud service providers have built their datacenters among different regions because of various considerations, such as system reliability, service delay, and electricity cost. To guarantee customer QoS and fault resilience, many cloud service providers have implemented dynamic service request routing mechanism in geographically distributed datacenters [8]. This mechanism enables datacenters to periodically communicate with each other about their system information, which is used to process service requests at geographically distributed sites [8]. Since the electricity cost of geographically distributed datacenters equals the power consumption multiplied by electricity prices, it is possible to reduce the electricity cost by monitoring the fluctuations of electricity prices for distributed datacenter locations and optimally dispatching more workload to those datacenters with low electricity prices, subject to QoS requirements.

2.3. Efficiency of cooling system

As the increase of server density and service amount, the power consumption of a cooling system has been a significant portion of the total power consumption of a datacenter. Typically, a cooling system is used to prevent the temperature of datacenter room exceeding the maximum safety temperature because the performance of a datacenter (e.g., hardware reliability [13]) is affected by the temperature of the datacenter room. Figure 1 describes the typical water-chilled datacenter cooling system using CRAC units [6]. Normally, CRAC units take in hot air produced by active servers and deliver cool air into a datacenter room through floor vent tiles. Thus, in the datacenter, hot air rises from bottom to top of the racks and air flows poorly at the ends of aisles [7][12]. Due to the circulation of air flow in the datacenter room, the top-shelf servers in each rack are hotter than the lower ones, the side racks of each row are hotter than the inside ones, and the servers at the ends of rows are the hottest ones. Therefore, one datacenter can be divided into three nearly temperature isolated zones (i.e., cool, warm, and hot zone) based on air flow patterns in the datacenter [12]. Figure 2 depicts zone temperature curves associated with their zone

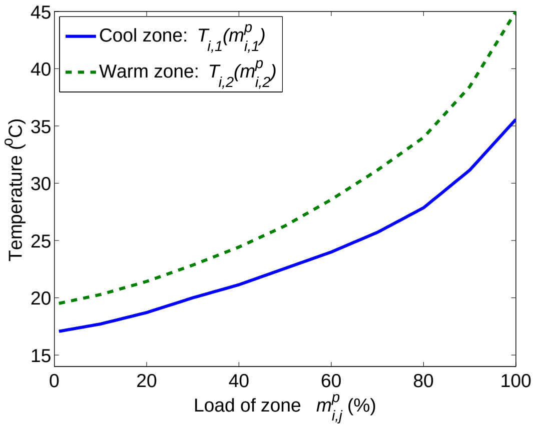


Figure 2: Temperature curves of warm and cool zone [12].

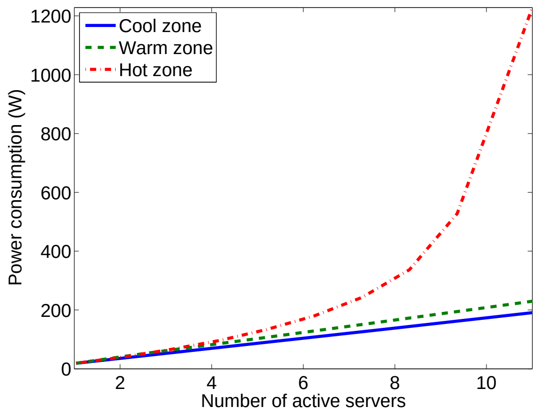


Figure 3: Power consumption of a cooling system in the three nearly temperature isolated zones of a datacenter when the maximum safety temperature of each zone is set at 30°C.

load percentages. In Figure 2, zones 1, 2, and 3 denote the cool, warm, and hot zone, respectively. The temperature of zone j ($j = 1, 2, 3$) in datacenter i ($1 \le i \le N$) is denoted as $T_{i,j}(m^p_{i,j})$, which is a linear piece-wise curve associated with zone j's load percentage $m^p_{i,j}$ [12]. In this figure, when the cool zone is uniformly distributed with 30% of its load, its temperature reaches 20°C. However, the temperature of the warm zone reaches 20°C if the warm zone is uniformly distributed with 7% of its load. A specific case is shown in Figure 3, which describes the power consumption of a cooling system in the three zones of a datacenter when the maximum safety temperature is set at 30°C [12]. Clearly, cooling the same number of active servers in the hot zone consumes more cooling power than that of the other two zones, especially when the number of active servers is over seven.

The power consumption of a cooling system is given by:

$$\frac{the\ power\ consumption\ of\ active\ servers}{Coefficient\ of\ Performance\ (COP)} \qquad (1)$$

COP is an efficiency metric for quantifying CRAC unit

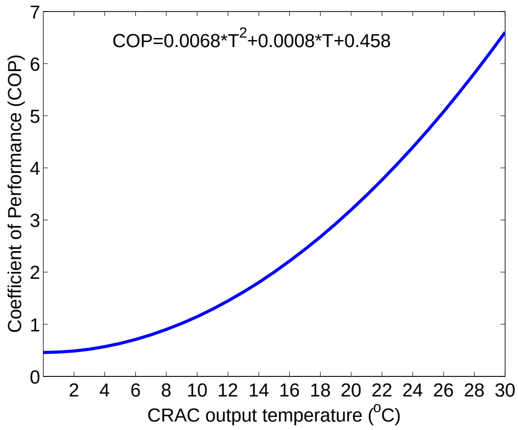


Figure 4: COP curve of CRAC [6].

and is an increasing function associated with CRAC output temperature. Figure 4 shows COP curve of CRAC unit, which is measured from HP labs Utility Datacenter [6]. As we will detailed in Section 4.2, CRAC output temperature depends on the maximum temperature of the three zones [12]. For a given workload, the temperature difference of the three zones can be minimized by rationally dispatching workload to servers of the three zones. Consequently, CRAC units would work in an efficient operating range, and the power consumption of the cooling system could be reduced.

However, inappropriate usage of CRAC could increase the power consumption of the cooling system. For example, there is an intuitive processing of cooling system, which works under two conditions: (1) CRAC units are closed or turned into sleeping mode when the temperature of datacenter room is below the maximum safety temperature; (2) CRAC units are activated only when the temperature of datacenter room exceeds the maximum safety temperature. Under Condition (1), some cooling power can be saved since the heat from active servers is not removed. However, the COP of CRAC under Condition (2) is much lower than the COP of CRAC under Condition (1), because the COP curve is a nonlinear curve associated with CRAC output temperature and the slope of COP curve decreases as the CRAC output temperature decreases, which are shown in Figure 4. Based on Eq.(1), the cumulated heat from past would consume more cooling power under Condition (2) than it does under Condition (1). As a result, the power consumption of a cooling system would be increased. The detail of how to minimize temperature difference of the three zones are detailed in Section 4.2.

2.4. The impact of workload management on the electricity cost of geographically distributed datacenters

Workload management [3][4][6][7][8][9][10][11][12] is a load-balancing application for the datacenter, the goal of which is to enable active servers to have approximately equal processing loads by dispatching incoming service requests based on current datacenter status. The workload management scheme plays a critical role on the electricity cost of geographically distributed datacenters since it determines the number and placement of active servers in datacenters and significantly influences the efficiency of other devices (e.g., cooling system) in each datacenter.

Existing works mainly focus on two aspects: the intra-datacenter workload management and the inter-datacenter workload management. Cooling-aware IntrA-datacenter load balancing (CIA) [6][7] is a representative intra-datacenter workload management scheme. CIA monitors the thermal status of one datacenter and consolidates incoming workload on a subset of servers with high cooling efficiency to reduce the power consumption of this datacenter. Electricity price-aware InteR-datacenter load balancing (EIR) [8][9] is a typical inter-datacenter workload management scheme, which minimizes the electricity cost of active servers in geographically distributed datacenters by periodically distributing workload to different datacenters based on the temporal and spatial variant electricity prices of those datacenter sites.

2.5. Problems of existing schemes

As stated in Section 2.4, each of the existing schemes only considers one factor that impacts the electricity cost of geographically distributed datacenters and overlooks the interaction between the two factors. Thus, some undesirable results could occur. For instance, EIR may incur high electricity cost of cooling systems due to random workload distribution in geographically distributed datacenters. CIA could lead to high electricity cost of active servers without considering the variation of electricity prices among datacenter sites. A simple combination of the two aforementioned schemes cannot achieve the global minimization of the electricity cost of geographically distributed datacenters, as we detail in Section 6.2.

The work presented in this paper differs from the existing works in the way that the proposed workload management scheme JET jointly considers the variation of electricity prices on the electricity cost of active servers and the impact of workload distribution in a datacenter on the electricity cost of a cooling system. For a certain amount of incoming workload, JET alternatively selects either the variation of electricity prices or the workload distribution on the efficiency of cooling systems as the dominator factor for the electricity cost, and directly determines the number and placement of active servers in the three zones of each datacenter to minimize the electricity cost of active servers and cooling systems in the geographically distributed datacenters.

3. System architecture

In this section, we provide a high-level description of our system architecture. In this paper, we assume our sys-

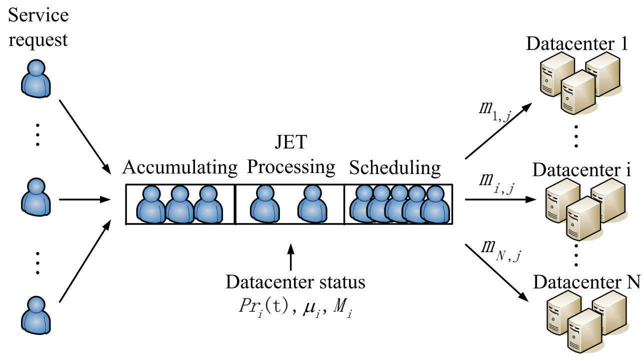


Figure 5: System architecture.

tem is a centralized system that manages a datacenter network for minimizing the electricity cost. While such a centralized architecture is commonly used in the management of geographically distributed datacenters [8][9], our system can be extended to work in a hierarchical/distributed way, which is our future work. Figure 5 shows our system architecture, which includes three components: (1) accumulating, (2) JET processing, and (3) scheduling. In the accumulating component, service requests are accumulated and stored by request aggregators in a fixed period of accumulation time T_{acm} (e.g., 100 ms). For a centralized system, there could be multiple distributed request aggregators working together as a logically centralized request aggregator. In the JET processing component, JET is used to calculate the optimal workload dispatching (the number of active servers in zone j of datacenter i, $m_{i,j}$) for the stored service requests with respect to QoS constraints and the current datacenters status (e.g., electricity prices $Pr_i(t)$, the service rate of servers μ_i, datacenter i's maximum number of servers M_i), so that the total electricity cost of geographically distributed datacenters is minimized. In the scheduling component, dynamic request routing mechanism is used to redirect service requests to geographically distributed datacenters based on the determined request dispatching results in the JET processing component.

4. Electricity cost models and QoS constraints of geographically distributed datacenters

In this section, we present the modeling for electricity cost of active servers and cooling systems of geographically distributed datacenters in detail. We also provide the modeling of different components for QoS constraints of geographically distributed datacenters.

4.1. Notations

The notations used in this paper are summarized in Table I.

Table 1: Notations

N	Total number of datacenter locations
i	Datacenter location i ($1 \leq i \leq N$)
j	Temperature isolated zone j (j =1,2,3)
$Pr_i(t)$	Electricity price of datacenter i at t ($t > 0$)
Po_i	Power consumption of one server at datacenter i
P_{idle}	Average idle power draw of one server at datacenter i
P_{peak}	Average peak power consumption of one server at datacenter i
u_i	CPU utilization of servers at datacenter i
μ_i	Service rate of one server at datacenter i
λ	Total service request rate for the geographically distributed datacenter
λ_i	Service request rate of datacenter i
d_i	Average delay of datacenter i
D_{si}	Delay constraint of datacenter i
m_i	The number of active servers in datacenter i
$m_{i,j}$	The number of active servers in zone j of datacenter i
$m^p_{i,j}$	Load percentage of zone j in datacenter i
M_i	The maximum number of servers in datacenter i
$M_{i,j}$	The maximum number of servers in zone j of datacenter i
$T_{i,j}(m^p_{i,j})$	Temperature of zone j in datacenter i with $m^p_{i,j}$
T^{MAX}	Maximum safety temperature of each datacenter
T^{out}	Temperature of cool air from CRAC unit
T^{server}_i	Temperature of servers in datacenter i
T^{new}_i	Adjusted new output temperature of CRAC unit in datacenter i
COP_i	COP of datacenter i
K	The number of piece-wise functions for approximating $1/COP$
$g_k(m_i)$	k-th piece-wise function for approximating $1/COP$ ($1 \leq k \leq K$)

4.2. Electricity cost model of active servers of geographically distributed datacenters

Suppose one cloud service provider operates N distributed datacenters. Datacenter i ($1 \leq i \leq N$) is at location i with the hourly electricity price $Pr_i(t)$ at time t ($t > 0$). Assume that each datacenter uses homogeneous servers and configurations. The power consumption of each server in datacenter i is Po_i [8]:

$$Po_i = P_{idle} + (P_{peak} - P_{idle})u_i \tag{2}$$

where P_{idle}, P_{peak}, and u_i denote the average idle power draw of a single server, the average peak power, and the

CPU utilization of servers in datacenter i, respectively.

As discussed in Section 2.3, one datacenter can be divided into three temperature isolated zones due to the physical structure of datacenter and air flow patterns [12]. Thus, for a given load in datacenter i, the number of active servers m_i can be divided into three parts, $m_{i,1}$ for the cool zone, $m_{i,2}$ for the warm zone and $m_{i,3}$ for the hot zone. Thus, we have:

$$m_i = \sum_{j=1}^{3} m_{i,j} \tag{3}$$

Therefore, the Electricity Cost of Active Servers (ECAS) of N distributed datacenters is given as:

$$ECAS = \sum_{i=1}^{N} \left(\sum_{j=1}^{3} m_{i,j} \right) [P_{idle} + (P_{peak} - P_{idle})u_i] Pr_i(t) \tag{4}$$

4.3. Electricity cost model of cooling systems of geographically distributed datacenters

In a datacenter, the cooling system is used to maintain the temperature of datacenter room at a reasonable level by delivering cool air into the datacenter room and taking in hot air produced by active servers. As explained in Section 2.3, COP is the metric for quantifying the efficiency of CRAC units in the datacenter. A higher COP indicates a higher efficiency of the cooling system. As presented in [14], if CRAC units take in hot air at 20°C and push cool air at 15°C, CRAC units expend 5.26 kW to remove the 10 kW of heat for cooling a specific volume of air. However, when CRAC units take in hot air at 25°C and push cool air at 20°C, CRAC units only expend 3.23 kW to remove the same 10 kW of heat for cooling the same volume of air, which saves about 40% cooling power. Therefore, the COP is an increasing function of CRAC output temperature, and the efficiency of the cooling system can be maximized by raising the CRAC output temperature, while preventing the room temperature from crossing the maximum safety temperature [14]. Specifically, assume the outside environment remains unchanged at datacenter i for a certain period of time, CRAC units push cool air at the temperature of T^{out} into the datacenter room to keep the temperature of datacenter room under the maximum safety temperature T^{MAX} [14]. However, the temperature of datacenter room is affected by the temperature of active servers T_i^{server}, which depends on the number and placement of current processing workload. To avoid the temperature of datacenter room from exceeding T^{MAX}, the adjusted amount of CRAC output temperature should be $T_i^{adj} = T^{safe} - T_i^{server}$. Thus, the adjusted new CRAC output temperature is $T_i^{new} = T^{out} + T_i^{adj}$. If T_i^{adj} is negative, it means that CRAC units need to provide more cooling, and vice versa. Based on the COP function shown in Figure 4, T_i^{new} determines the actual COP and, consecutively, determines the efficiency of CRAC units. Therefore, we have:

$$COP_i = 0.0068 T_i^{new2} + 0.0008 T_i^{new} + 0.458 \tag{5}$$

$$T_i^{new} = T^{out} + T^{MAX} - T_i^{server} \tag{6}$$

In datacenter i, T_i^{server} denotes the highest temperature of active servers, which is affected by the workload number and placement in datacenter i [12]. As stated previously, one datacenter can be divided into three temperature isolated zones: cool, warm, and hot zone. Therefore, we have:

$$T_i^{server} = max\left(T_{i,1}(m_{i,1}^p), T_{i,2}(m_{i,2}^p), T_{i,3}(m_{i,3}^p) \right) \tag{7}$$

where $m_{i,j}^p$ $(0 \le m_{i,j}^p \le 1)$ denotes the load percentage of zone j in datacenter i, $T_{i,j}(m_{i,j}^p)$ denotes the temperature of zone j in datacenter i with $m_{i,j}^p$ [12].

Moreover, the temperature of each zone in every datacenter should not exceed the maximum safety temperature T^{MAX}:

$$T_{i,j}\left(m_{i,j}^p\right) \le T^{MAX} \tag{8}$$

Therefore, the Electricity Cost of Cooling Systems (ECCS) of geographically distributed datacenters located in N sites is given as:

$$ECCS = \sum_{i=1}^{N} \frac{\left(\sum_{j=1}^{3} m_{i,j}\right) [P_{idle} + (P_{peak} - P_{idle})u_i] Pr_i(t)}{0.0068 T_i^{new2} + 0.0008 T_i^{new} + 0.458} \tag{9}$$

4.4. Total electricity cost model of geographically distributed datacenters

For a cloud service provider operating N geographically distributed datacenters in dispersed N locations, the total Electricity Cost (EC) of N geographically distributed datacenters could be written as:

$$\begin{aligned} EC &= ECAS + ECCS \\ &= \textstyle\sum_{i=1}^{N} \left(\sum_{j=1}^{3} m_{i,j}\right) [P_{idle} + (P_{peak} - P_{idle})u_i] \cdot \\ &\quad Pr_i(t) \left(1 + \frac{1}{0.0068 T_i^{new2} + 0.0008 T_i^{new} + 0.458}\right) \end{aligned} \tag{10}$$

4.5. Workload component of QoS constraints of geographically distributed datacenters

Assume that in a time interval, one cloud service provider operating N geographically distributed datacenters receives λ service requests, and datacenter i is assigned with λ_i service requests using a specific workload management scheme. To ensure that all the service requests are fully processed, we have:

$$\lambda = \sum_{i=1}^{N} \lambda_i \tag{11}$$

To process a large number of service requests, one datacenter usually contains hundreds or thousands of servers. The number of active servers in datacenter i, m_i, should not exceed the maximum number of servers in datacenter i, M_i. The number of active servers in zone j of datacenter i, $m_{i,j}$, should not also exceed the maximum number of servers in zone j of datacenter i, $M_{i,j}$. The load percentage of zone j in datacenter i, $m^p_{i,j}$, equals the number of active servers at zone j of datacenter i divided by the maximum number of servers at zone j of datacenter i. Therefore, we have:

$$m_i \le M_i \tag{12}$$

$$m_{i,j} \le M_{i,j} \tag{13}$$

$$m^p_i = \frac{m_i}{M_i} \tag{14}$$

$$M_i = \sum_{j=1}^{3} M_{i,j} \tag{15}$$

4.6. Delay component of QoS constraints of geographically distributed datacenters

In this paper, we use the M/M/n queueing model to model each server in every datacenter [9]. In the M/M/n queueing model, given the number of active servers n, the service rate μ, the arrival rate λ, and the probability P_Q of service requests waiting in queue, the average delay d can be expressed as $\frac{1}{n\mu - \lambda} P_Q$. In datacenter i, the service rate μ_i is a function associated with the server's CPU utilization u_i and the maximum service rate μ_i^{MAX}. That is:

$$\mu_i = f(u_i, \mu_i^{MAX}) \tag{16}$$

Without loss of generality, we can assume that active servers in a datacenter are always busy to process requests waiting in queue. Hence, we have that P_Q equals 1. Therefore, we have the average delay of datacenter i, d_i:

$$d_i = \frac{1}{\left(\sum_{j=1}^{3} m_{i,j}\right)\mu_i - \lambda_i} \tag{17}$$

Additionally, in datacenter i, d_i should not exceed a delay constraint D_i That is:

$$d_i \le D_i \tag{18}$$

5. Problem formulation and solution

In this section, we first formulate the electricity cost minimization problem of geographically distributed datacenters as a constrained nonlinear optimization problem. We then transform the complicated original problem to a convex problem with linear constraints and finally solve this problem with the proposed JET.

5.1. Problem formulation

Given a workload of λ service requests in a time interval, the optimization goal is to minimize the electricity cost of geographically distributed datacenters by a workload management scheme, so that datacenter i actives $m_{i,1}$ servers in the cool zone, $m_{i,2}$ servers in the warm zone, and $m_{i,3}$ servers in the hot zone to work together to process the assigned λ_i service requests. The optimization problem is formulated as the **Electricity Cost Minimization (ECM) problem** of geographically distributed datacenters, which is shown as follows:

$$\min \sum_{i=1}^{N} \left(\sum_{j=1}^{3} m_{i,j}\right) [P_{idle} + (P_{peak} - P_{idle})u_i] \cdot Pr_i(t) \left(1 + \frac{1}{0.0068 T_i^{new\,2} + 0.0008 T_i^{new} + 0.458}\right) \tag{19a}$$

subject to

$$\frac{1}{\left(\sum_{j=1}^{3} m_{i,j}\right)\mu_i - \lambda_i} \le D_i \tag{19b}$$

$$\mu_i = f(u_i, \mu_i^{MAX}) \tag{19c}$$

$$\lambda = \sum_{i=1}^{N} \lambda_i \tag{19d}$$

$$T_i^{new} = T^{out} + T^{safe} - T_i^{server} \tag{19e}$$

$$T_i^{server} = max\left(T_{i,1}(m^p_{i,1}), T_{i,2}(m^p_{i,2}), T_{i,3}(m^p_{i,3})\right) \tag{19f}$$

$$T_i^{server} \le T^{MAX} \tag{19g}$$

$$m^p_{i,j} = \frac{m_{i,j}}{M_{i,j}} \tag{19h}$$

$$m_{i,j} \le M_{i,j} \tag{19i}$$

$$M_i = \sum_{j=1}^{3} M_{i,j} \tag{19j}$$

In the proposed ECM problem, the constraints T_i^{server} and μ_i are nonlinear, and the objective function EC is nonlinear. Thus, the ECM problem is a constrained nonlinear optimization problem.

5.2. Transformations and assumptions

Generally, to solve a complicated nonlinear optimization problem, a common method is to transform the original problem into a standard problem that can be solved using existing optimization techniques or solvers. In this section, we transform the ECM problem into the TECM problem (a convex problem with linear constraints) by reasonable transformations and assumptions on the nonlinear constraints T_i^{server}, Po_i, u_i, and the objective function EC.

5.2.1. Transformation on the nonlinear constraint T_i^{server}

In the ECM problem, T_i^{server} (Eq.(19f)) is expressed as a maximum function of three nonlinear functions. As discussed in Section 4.2, a low T_i^{server} leads to a high COP_i. Since reducing the electricity cost of cooling systems is one goal of the ECM problem, the efficiency of cooling system in datacenter i can be maximized by minimizing T_i^{server}. Given the maximum number of servers in zone j of datacenter i, $M_{i,j}$, T_i^{server} can be obtained through solving a sub-problem:

$$\min T_i^{server} \tag{20a}$$

subject to

$$T_{i,j}(m_{i,j}^p) \le T_i^{server} \tag{20b}$$

$$T_i^{server} \le T^{MAX} \tag{20c}$$

$$m_{i,j}^p = \frac{m_{i,j}}{M_{i,j}} \tag{20d}$$

$$m_{i,j} \le M_{i,j} \tag{20e}$$

In this sub-problem, $T_{i,j}(m_{i,j}^p)$ is the combination of piece-wise curves associated with $m_{i,j}^p$ [12], which depends on the specific workload management scheme. Thus, it is impossible to solve the sub-problem without the real-time information about the state of each datacenter and electricity prices. Let us go back to the objective function of ECM problem, also known as EC. In EC, there are four types of variables, $m_{i,j}$, T_i^{new}, u_i, and $Pr_i(t)$. Since T^{out} and T^{MAX} do not change for each datacenter and $Pr_i(t)$ only changes once an hour, during each one hour interval, $m_{i,j}$, u_i, and T_i^{server} depend on the applied workload management scheme. Assume any two types of the variables are constant, if the other one decreases, EC decreases, and vice versa. Therefore, $m_{i,j}$, u_i, and T_i^{server} have the uniform monotonicity with EC. When the minimization of EC is set as the core objective, the solution of the above sub-problem is included. So we can transform the nonlinear constraint of T_i^{server} into four linear inequality constraints Eq.(20b), Eq.(20c), Eq.(20d), and Eq.(20e).

5.2.2. Transformation on the nonlinear objective function EC

In the ECM problem, the objective function EC (Eq.(19a)) is a nonlinear function. The complexity of EC comes from T_i^{server}, a nonlinear function related to $m_{i,j}^p$, in its denominator. Thus, EC can be transformed into a formulation directly related to $m_{i,j}^p$ by simplifying its fractional component. We combine three linear piece-wise zone temperature curves associated with their zone load percentages (Figure 2) into one datacenter temperature curve associated with the overall datacenter load percentage, which is shown in Figure 6. To accommodate potential workload spikes, each datacenter should maintain sufficient capacity margins [10]. The capacity margin is based on the volatile characteristics of workload. The high volatile workload

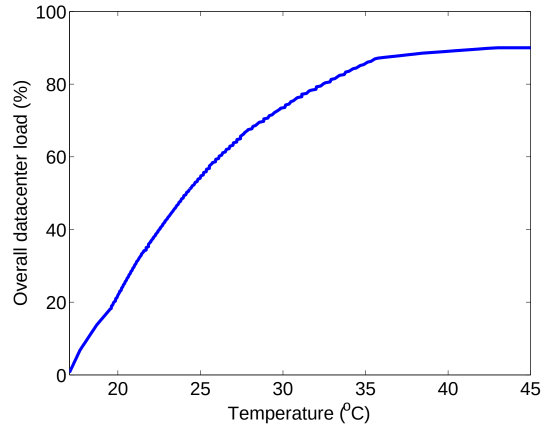


Figure 6: Temperature curve of the overall datacenter load percentage.

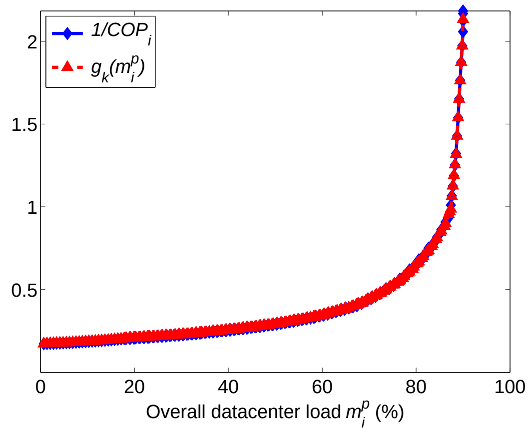


Figure 7: $1/COP_i$ curve of the overall datacenter load percentage.

need a higher capacity margin and vice versa. In this paper, we use 10% capacity margin for each datacenter. We bring the temperature curve of datacenter i into the COP function, take reciprocal of COP, and then get $1/COP_i$ curve, which is shown as the blue diamond line in Figure 7. $1/COP_i$ is also a nonlinear function associated with the overall datacenter load percentage. We use one dimensional linear regression [15] to substitute the original $1/COP_i$ function with K linear piece-wise functions $g_k(m_i^p)$ $(1 \le k \le K)$, which are shown as the triangle dash lines in Figure 7. Therefore, we have:

$$g_k(m_i^p) = a_k + b_k m_i^p, (th_{k-1} < m_i^p \le th_k) \tag{21}$$

$$m_i^p = \frac{m_i}{M_i} \tag{22}$$

$$g_k(m_i^p) \le G \tag{23}$$

where a_k, b_k $(b_k > 0)$, and th_k are, respectively, the intercept, slope, and upper limit for $g_k(\cdot)$, m_i^p is the load percentage of datacenter i, and G is the maximum value of $1/COP_i$.

Therefore, we have a new EC:

$$EC^{new} = \sum_{i=1}^{N} m_i[P_{idle}+(P_{peak}-P_{idle})u_i]Pr_i(t)\,(1+g_k(m_i^p)) \tag{24}$$

The second derivative of EC^{new} are:

$$EC^{new\prime\prime}(m_i) = \frac{2[P_{idle}+(P_{peak}-P_{idle})u_i]Pr_i(t)b_k}{M_i} > 0, \tag{25}$$

$$EC^{new\prime\prime}(u_i) = 0 \tag{26}$$

Thus, EC^{new} is a convex function.

5.2.3. Assumptions on the nonlinear constraints Po_i and μ_i

With power-management, a server's idle power accounts for 50%∼65% of its peak power [8]. Thus, we assume:

$$Po_i = 2P_{idle} \tag{27}$$

Similar to a related literature [9], we assume that all the active servers will run close to 100% utilization because the number of active servers is minimized by the workload management scheme to reduce the power of active servers, that is $u_i = 1$. Based on the above assumptions, we have:

$$\mu_i = \mu_i^{MAX} \tag{28}$$

5.3. Problem solution

Based on the above transformations and assumptions, the proposed ECM problem is simplified to be the **Transformed ECM (TECM) problem**:

$$\min \sum_{i=1}^{N} 2m_iP_{idle}Pr_i(t)\left(1+a_k+b_k\frac{m_i}{M_i}\right) \tag{29a}$$

subject to

$$\frac{1}{m_i\mu_i^{MAX}-\lambda_i} \le D_i \tag{29b}$$

$$\lambda = \sum_{i=1}^{N}\lambda_i \tag{29c}$$

$$a_k + b_k\frac{m_i}{M_i} \le G \tag{29d}$$

$$m_i \le M_i \tag{29e}$$

In the TECM problem, the variables m_i and λ_i are integers, and the objective function EC^{new} is a nonlinear function. Thus, the TECM problem is a Nonlinear Integer Programming (NIP) problem. However, as shown in Section 5.2.2, EC^{new} is a convex function and all constraints of the TECM problem are linear. Therefore, we can solve the TECM problem using the existing optimization techniques and solvers.

The proposed JET includes three steps. In the first step, the decimal result of the TECM problem can be obtained by efficient optimization techniques (e.g., *Interior Point method*) or solvers (e.g., *MINOS 5.5*[16]). In the second step, the integer result of m_i can be obtained by rounding up m_i from the decimal result. In the third step, we can get the approximate COP_i by bringing m_i into Eq.(19) and (20), get T_i^{server} by bringing COP_i into Eq.(4) and Eq.(5), and obtain $m_{i,j}$ with Eq.(7).

JET achieves the trade-off between the performance and computational complexity. Obviously, we can get the optimal result by using Branch-and-Cut method in the second step. However, the optimal result comes with the high complexity since the complexity of Branch-and-Cut method is exponential to the number of introduced variables. In the TECM problem, datacenter i is associated with two variables m_i and λ_i. For a cloud service provider operating 10 datacenters, 20 variables are introduced to the TECM problem. Thus, the TECM problem with the high computational complexity cannot be solved in a reasonable time. However, in the final result, the number of active servers in one of the distributed datacenters is at least in the order of thousands. Thus, without significantly reducing the performance, we can get the integer result by rounding up the decimal result of the first step. Based on the evaluation discussed in Section 6, for a cloud service provider with three distributed datacenters, JET using MINOS 5.5 consumes 15.6 ms on average to determine the optimal allocations of the request-response type of web service received per second, such as real-time Google and Bing search. The electricity cost difference between the optimal result obtained from Branch-and-Cut method and the round-up result is less than 0.001%.

In addition, the processing time of JET changes very slightly as the number of service requests arrived in each accumulation time T_{acm} changes. Nowadays, one cloud service provider only operates a limited number of distributed datacenters. For instance, Google operates just about 13 datacenters around the world [17]. For one cloud provider operating 20 distributed datacenters, if MINOS 5.5 is used as the solver, the processing time of JEC is no more than 20ms, which is less than the typical user delay (e.g., 100ms).

6. Evaluation

In this section, we use the real-life workload and electricity price traces to evaluate the performance of JET against the existing schemes on electricity cost reduction of geographically distributed datacenters at two specific situations. To understand the cost reduction, we analyze the composition of electricity cost, the variation of COPs, and the ratio of the number of active servers and the maximum number of servers on different schemes in detail. These evaluations are primarily targeted towards the request-response type of web service, such as real-time Google and Bing search.

6.1. Evaluation setup

In our evaluation, we simulate three distributed datacenters for a cloud service provider. The datacenters 1, 2, and 3 are assumed to be located in Long Island, NY, Houston, TX, and Atlanta, GA, respectively. As mentioned in Section 2, there are two main parallel electricity markets in the United States. The two locations of Long Island and Houston are located in the electricity wholesale market regions, where the electricity prices vary based on the grids condition, while Atlanta is in the regulated utility region, where the electricity prices are fixed for a certain period of time. For the two locations in the wholesale market, we use the traces of Day-Ahead electricity prices from NYISO [18] and ERCOT [19], respectively. The power consumption profile of each server in the three datacenters is assumed to be approximately the same: $P_{idle} = 60$ watts. The maximum number of servers of the three datacenters are assumed to be 30,000, 60,000 and 25,000, and their processing capacity coefficients are 4.0, 2.5, and 3.5 service requests per second, respectively.

In our evaluation, we use the water-chilled cooling system cooling shown in Figure 2. In each datacenter, there are four CRAC units, each of which pushes cool air at 15°C into the room through floor vent tiles. To prevent the temperature of datacenter room from exceeding the maximum safety temperature 30°C, CRAC units adaptively adjust their efficiency. We validated our cooling system model by matching the temperature profile reported in [6][14] under the uniform distribution of a fixed load in each datacenter.

We use two specific situations in this evaluation to compare JET with the existing schemes. In our evaluation, we use a workload trace containing 10% of Internet traffic arrived at Wikipedia between Oct.1, 2007 and Nov.30, 2007 [20]. Similar to a related study [9], service requests used in this simulation are collected per second. In **Situation 1**, service requests received by the three datacenters vary, but the electricity prices of the three datacenters are fixed at $Pr_1(t) = 42.93\$/MWh$, $Pr_2(t) = 20.27\$/MWh$, $Pr_3(t) = 55.30\$/MWh$. In **Situation 2**, the arriving service requests per second are respectively fixed at 30%, 50%, and 70% of the overall load of the three datacenters, but the electricity prices of the three datacenters change according to the different regional electricity price schemes. In this paper, we refer to the overall load of the three datacenters at 30%, 50%, and 70% as *Light Workload*, *Medium Workload*, and *Heavy Workload*, respectively. The maximum overall load of the three datacenters is 90%, since 10% capacity margin for each datacenter is used to accommodate potential workload spikes [10].

6.2. Schemes for comparison

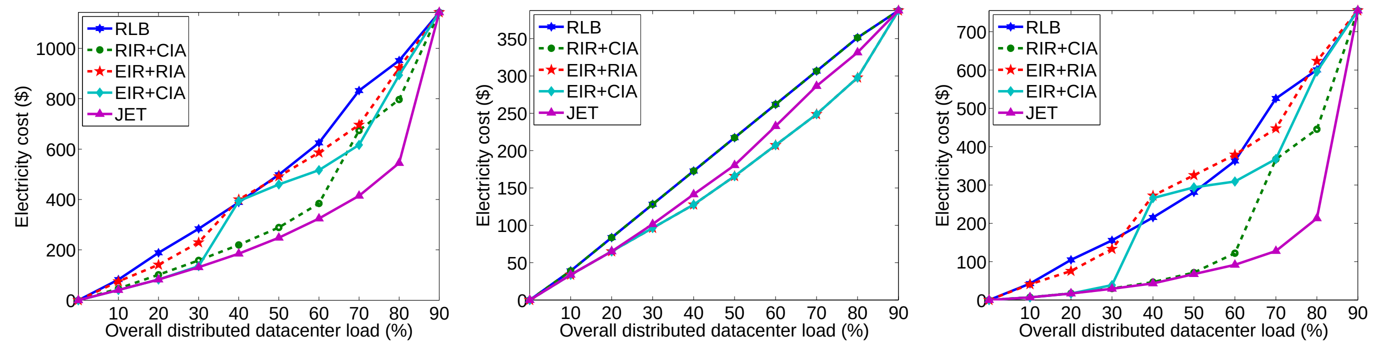
We compare JET with four datacenter workload management schemes, which are explained below. Since CIA [6][7] and EIR [8][9] are partial workload management schemes as compared with JET, we complement them with random load balancing. We ran each of the five schemes 100 times to collect precise statistical characteristics with the intent to improve the accuracy of the simulation.

Random inter-datacenter and intra-datacenter Load Balancing (RLB). Arriving service requests are distributed randomly and uniformly among geographically distributed datacenters; service requests that arrive at a datacenter are further randomly sent to servers in the three zones for execution.

Random InteR-datacenter load balancing plus Cooling efficiency-aware IntrA-datacenter load balancing (RIR+CIA) [6][7]. This schemes includes two steps. RIR is used in the first step, so that incoming service requests are uniformly distributed to geographically distributed datacenters located in different sites. In the second step, CIA works in each datacenter and distributes service requests that arrive at the datacenter to certain servers in order to achieve the minimum temperature differences among the three zones of the datacenter. This scheme cuts down ECCS by selecting servers with the high cooling efficiency to process service requests.

Electricity price-aware InteR-datacenter load balancing plus Random IntrA-datacenter load balancing (EIR+RIA) [8][9]. This scheme also includes two steps. In the first step, EIR is used among geographically distributed datacenters to dispatch service requests to different datacenters based on the current electricity prices of datacenters sites. The datacenter with the lowest electricity price is first fully loaded, and then the datacenter with the second lowest electricity price will be loaded, and so forth. In the second step, RIA is used in each datacenter, and servers are randomly chosen to process the service requests distributed to this datacenter. This scheme reduces ECAS because it explicitly considers the current electricity prices at the locations of geographically distributed datacenters.

Electricity price-aware InteR-datacenter load balancing plus Cooling efficiency-aware IntrA-datacenter load balancing (EIR+CIA). This scheme accounts for the diversity of electricity prices and the efficiency of cooling system in two separate steps. In the first step, among geographically distributed datacenters, EIR is used to decide the number of active servers for each datacenter based on the current electricity prices of datacenters in order to minimize the electricity costs of active servers. In the second step, in each datacenter, CIA is used to decide the placement of active servers in the three zones of each datacenter by considering the efficiency of cooling system in order to reduce the electricity cost of cooling systems. Without jointly considering the two factors, EIR+CIA can cause some undesirable situations. EIR tends to load the datacenter with the lowest electricity price first. Only when that datacenter is full, the datacenter with the second lowest electricity price will be loaded, and so forth. In a fully loaded datacenter, CIA cannot be effectively applied because all servers in the three zones are already activated under the heavy datacenter load, and COP cannot

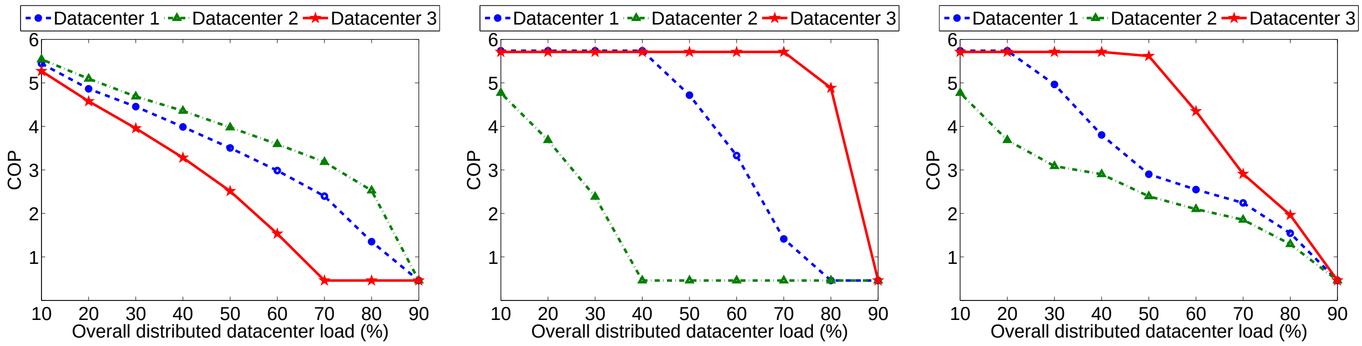


(a) Total electricity cost of three distributed datacenters with the variant overall load of three datacenters.

(b) Electricity cost of active servers of three distributed datacenters with the variant overall load of three datacenters.

(c) Electricity cost of cooling systems of three distributed datacenters with the variant overall load of three datacenters.

Figure 8: Performance of five workload management schemes under Situation 1.



(a) RIR+CIA

(b) EIR+CIA

(c) JET

Figure 9: COP variation for three workload management schemes under Situation 1.

be increased. As a result, for EIR+CIA, EIR may reduce a small portion of ECAS, but more ECCS is incurred by the cooling system. Therefore, EC would be high, as shown in Fig.7.

JET. The proposed JET combines the inter-datacenter workload management and the intra-datacenter workload management into a joint inter- and intra-datacenter workload management. For a given workload, JET jointly accounts for the diversity of electricity prices and the efficiency of cooling system and directly determines both the number and placement of active servers in the three zones of each datacenter with the intent to reduce EC. To cut EC, JEC alternately selects the diversity of electricity prices or the efficiency of cooling system as the dominator factor to EC, and achieves the trade-off between ECAS and ECCS.

6.3. Electricity cost reduction under Situation 1

Figure 8 respectively show the total electricity cost, the electricity cost of active servers, and the electricity cost of cooling systems of the three distributed datacenters for five workload management schemes under Situation 1. In Figure 8(a), RLB costs most since all the service requests are randomly dispatched to servers of the three data centers without any optimization. EIR+RIA and EIR+CIA are aware of the variation of electricity prices among the three datacenters and achieve the lowest electricity cost of active servers of the three distributed datacenters for all overall load spans. To analyze the reduction on the electricity cost in detail, Figure 9 respectively show the variation of COPs for three low-cost schemes (i.e., RIR+CIA, EIR+CIA, and JET) under Situation 1. When the overall load of the three distributed datacenters is lower than 30%, RIR+CIA performs worse than EIR+CIA. This is because the random distribution of the small amount workload does not significantly influence the electricity cost of cooling systems, and the electricity price is the dominated factor for the electricity cost of the three distributed datacenters. However, as the overall load increases from 30% to 60%, the efficiency of cooling systems plays a more important role in the electricity cost of the three datacenters than the electricity price. For EIR+CIA, EIR distributes more service requests to datacenters with lower electricity prices and causes these datacenters heavily or fully loaded. Thus, even if CIA is applied, COP cannot be increased significantly since all servers in the

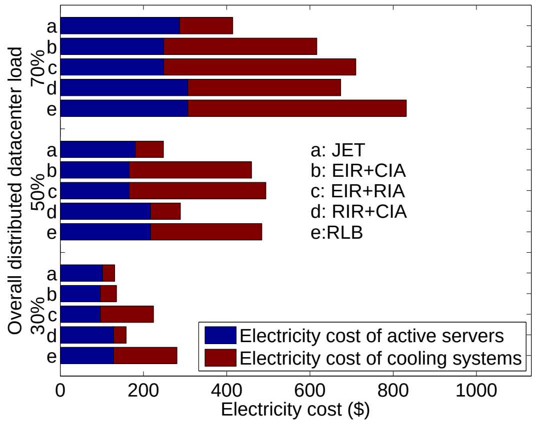


Figure 10: Electricity cost compositions of three distributed datacenters under Situation 1.

Table 2: Percentage reduction of electricity cost of three distributed datacenters under Situation 1

Load	RIR+CIA	EIR+RIA	EIR+CIA	JET
10%	46%	30%	61%	61%
20%	46%	32%	60%	60%
30%	45%	27%	54%	57%
40%	42%	-1%	-1%	53%
50%	42%	3%	10%	52%
60%	36%	6%	17%	48%
70%	18%	17%	25%	51%
80%	16%	3%	6%	43%

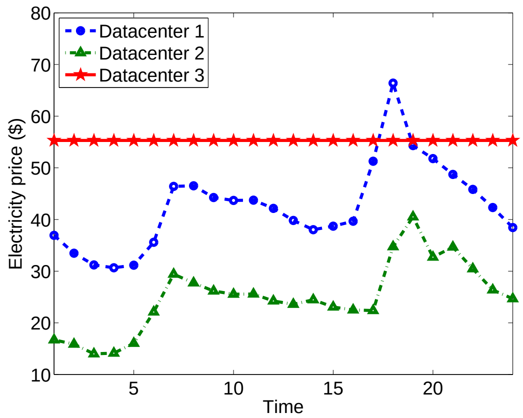
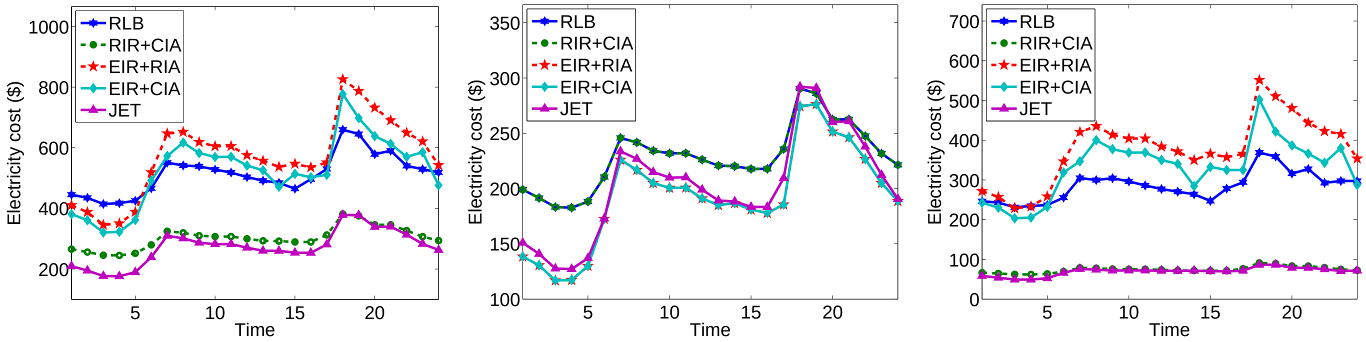


Figure 11: Hourly electricity prices for three distributed datacenters on Dec. 14, 2012 for Situation 2.

three zones of a heavy loaded datarcenter are already activated. In Figure 9(b), when the overall load reaches 40%, the COP of datacenter 2 reaches the efficiency of lower limit because the electricity price of datacenter 2 is the lowest among the three datacenters. Conversely, RIR+CIA achieves randomly and uniformly workload dispatching among the three distributed datacenters, which prevents any one of the three distributed datacenters being heavily loaded and maintains the reasonable efficiency of a cooling system for each datacenter. In Figure 9(a), when the overall load is in the range from 30% to 60%, the COPs of all the three distributed datacenters do not reach the efficiency of lower limit. Thus, in such condition, RIR+CIA costs less than EIR+CIA. When the overall load of the three distributed datacenters exceeds 60%, the performance of RIR+CIA and EIR+CIA varies alternately since either the electricity price or the efficiency of cooling systems alternately becomes the dominated factor for the electricity cost of the three distributed datacenters. In Figure 8(c), RIR+CIA and EIR+CIA have the approximately the same electricity cost of cooling systems at the overall load of 70%. This is because most of servers with the high cooling efficiency in the three distributed datacenters have been already activated when the overall load reaches 60%, and the COPs can not be greatly increased. In Figure 9(a), the COP of datacenter 3 reaches the efficiency of lower limit when the overall load reaches 70%.

JET performs best for all overall load spans, as shown in Figure 8(a). The reason is JET achieves dynamic reduction on the total electricity cost of the three distributed datacenters by adaptively considering either the variation of electricity prices or the efficiency of cooling systems as the major factor of the total electricity cost as the incoming service requests vary. In Figure 9(c), when the overall load of the three distributed datacenters is low, datacenter 2 is assigned with more service requests than others, and its COP is lower than others since it has the lowest electricity price among the three distributed datacenters. As the overall load increases, the efficiency of cooling systems become more important than the variation of electricity prices. Thus, to avoid the fast decrease of the efficiency of cooling systems in datacenter 2, more service requests are dispatched to datacenter 1 and 3 , and the COPs of datacenter 1 and 3 decrease. The COPs of the three datacenters do not reach the lower limit efficiency until the overall load reaches 90%, which is the maximum load of the three datacenters. In Figure 8(b), JET costs a little more on the electricity cost of active servers than EIR+RIA and EIR+CIA. However, as shown in Figure 8(c) JET saves much more on the electricity cost of cooling systems than other schemes. Therefore, JET significantly reduces the total electricity cost of the three distributed datacenters.

To fully understand cost reduction, in Figure 10, we respectively show the electricity cost of active servers and cooling systems for five workload management schemes at three typical datacenter loads in the form of histogram. RLB and RIR+CIA uniformly distribute service requests into the three distributed datacenters, and they have the same electricity cost of active servers for the three particular loads. However, RLB incurs the high electricity cost of

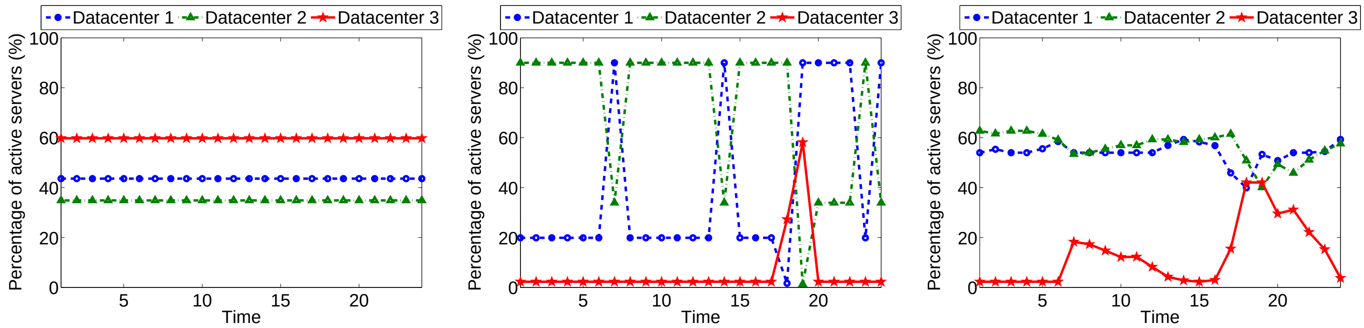


(a) Electricity cost of three distributed datacenters at different times

(b) Electricity cost of active servers of three distributed datacenters at different times

(c) Electricity cost of cooling systems of three distributed datacenters at different times

Figure 12: The performance of five workload management schemes at Medium Workload under Situation 2.



(a) RIR+CIA

(b) EIR+CIA

(c) JET

Figure 13: Percentage of active servers at Medium Workload under Situation 2.

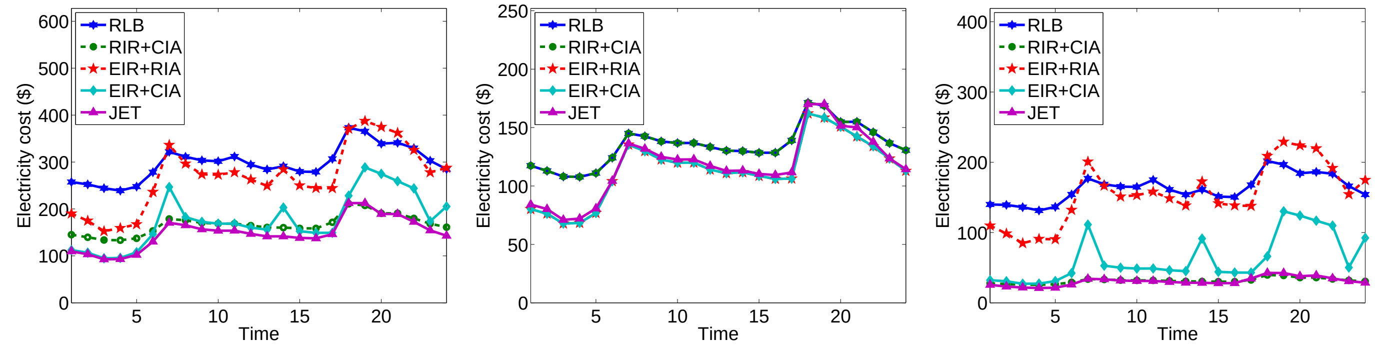
cooling systems due to the uneven workload distribution in the three zones of datacenters. Similarly, EIR+RIA and EIR+CIA also have the same electricity cost of active servers for the three particular loads since both of them allocate service requests based on the electricity prices to reduce the electricity cost of active servers. EIR+CIA further lowers the electricity cost of cooling systems by improving the efficiency of cooling systems. As explained above, JET costs slight more on the electricity cost of active servers than EIR+CIA, but it cuts much more on the electricity cost of cooling systems by rationally dispatching workload in the three zones of each datacenter to improve the COP of each datacenter. Hence, JET achieves substantial reduction on the total electricity cost of the three distributed datacenters.

In Table 2, we show the percentage reduction of the electricity cost of the three distributed datacenters achieved by RIR+CIA, EIR+RIA, EIR+CIA, and JET over RLB at various overall datacenter loads. It shows that for the range of 30∼70% of the overall distributed datacenter load where datacenters operate most of the time [21][22], JET achieves substantial electricity cost reductions of 48∼57%. Recall from Section 1 that since the electricity cost is a great portion in the operational cost of geographically distributed datacenters, such improvements are significant.

6.4. Electricity cost reduction under Situation 2

Due to different regional electricity schemes, we use the fixed electricity price for Atlanta and the hourly electricity prices on Dec. 14, 2012 from NYISO [18] and ERCOT [19] for Long Island and Houston, which are shown in Figure 11. Clearly, the electricity prices in Long Island and Houston vary significantly per hour, which makes it possible to minimize the electricity cost of the three datacenters by dispatching arriving service requests to those datacenters with lower electricity prices.

We respectively simulate the five workload management schemes at Light Workload, Medium Workload, and Heavy Workload when the real-life hourly electricity prices are applied. The simulation results under the three workloads are shown in Figure 12, 14, and 15, respectively. We use the Medium Workload (50% of the overall load of the three distributed datacenters) as an example to analyze the performance difference among the five workload management schemes. Figure 12 respectively show the total electricity cost, the electricity cost of active servers, and the electricity cost of cooling systems of the three distributed datacenters for the five workload management

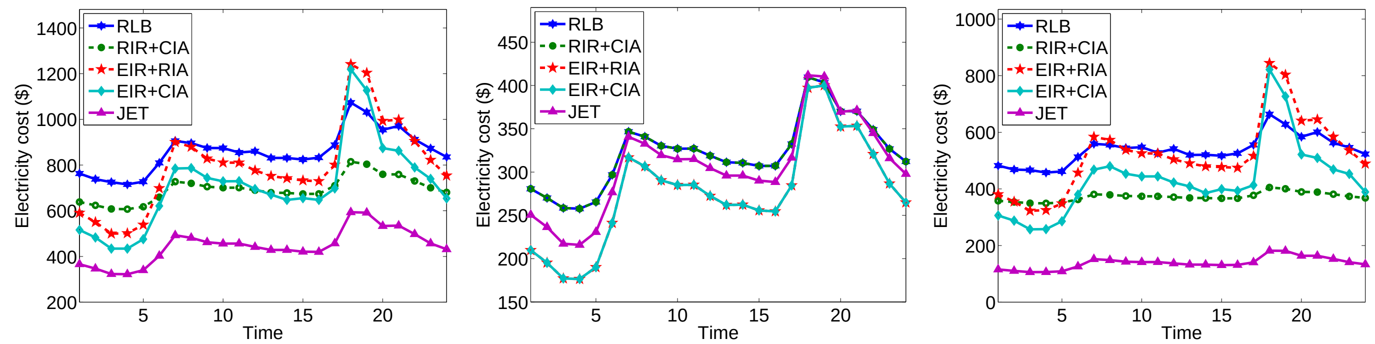


(a) Electricity cost of three distributed datacenters at different times

(b) Electricity cost of active servers of three distributed datacenters at different times

(c) Electricity cost of cooling systems of three distributed datacenters at different times

Figure 14: The performance of five workload management schemes at Light Workload under Situation 2.



(a) Electricity cost of three distributed datacenters at different times

(b) Electricity cost of active servers of three distributed datacenters at different times

(c) Electricity cost of cooling systems of three distributed datacenters at different times

Figure 15: The performance of five workload management schemes at Heavy Workload under Situation 2.

schemes at Medium Workload when the real hourly electricity prices are applied. Figure 13 respectively show the ratio of the number of active servers and the maximum number of servers for three low-cost schemes under the same configuration as Figure 12. In Figure 12(a), both EIR+RIA and EIR+CIA perform worse than RLB for most of time. This is because the EIR-related schemes dispatch service requests among distributed datacenters only considering the variation of electricity prices. Thus, the datacenter with low electricity price is easily got heavily or fully loaded. In Figure 13(b), in each hour, one of the distributed three datacenters reaches 90% load (the maximum load of a datacenter), but the heavily loaded datacenter varies hourly according to the variation of electricity prices. Therefore, the datacenter with low electricity price costs more on the high electricity cost of cooling systems, as shown in Figure 12(c). For RIR+CIA, since the number of active servers for RIR does not change with the variation of electricity prices, RIR costs more on the electricity cost of active servers than EIR. However, CIA reduces the electricity cost of cooling systems as much as possible by rationally assigning the workload in each datacenter. JET achieves the trade-off between EIR and CIA since it is aware of both the variation of electricity prices and the efficiency of cooling systems. In Figure 13(c), the number of active servers of JET slightly changes with the variation of electricity prices, as compared with those of EIR+CIA. In Figure 13 and 14, JET outperforms other four schemes, especially when the three distributed datacenters are under Heavy Workload.

7. Related work

Electricity cost has become the major proportion of the operational cost of geographically distributed datacenters. Many efforts have been spent to reduce the electricity cost of geographically distributed datacenters by the optimal datacenter workload management mechanisms. Some studies focused on the reduction of the power consumption of active servers in a datacenter by the intra-datacenter workload management. Elnozahy et al. [4] reduced the aggregate power consumption of server farms by dynamically adjusting voltage scaling and server node on/off states. Heo et al. [3][23] investigated dynamic cluster server configuration to minimize power consumption through consolidating workload on a subset of servers and turning off the rest during low workload periods. Barrsos et al. [21] did research on how to use Chip Multi-Processor (CMP) to achieve power management. Raghavendra et al. [24] suggested a coordination architecture to regulate different individual approaches in multi-level power management.

With the increasing high density of servers, the power consumption of cooling systems has become a major component of the power consumption of geographically distributed datacenters. Research studies show that the power consumption spent on a cooling system can take up to 50% of the total power consumption of a datacenter [5]. Hence, some studies proposed the intra-datacenter workload management to reduce the power consumption of servers and the cooling system in a datacenter through enforcing the impact of workload distribution on the power consumption of the cooling system. Sharma et al. [6] and Bash et al. [7] analyzed the physical structure of the datacenter and achieved substantial power savings by allocating heavy workload onto servers, which are in more efficient cool places.

The ultimate goal of cloud service providers is to reduce the electricity cost of geographically distributed datacenters, which depends on not only the power consumption of datacenters but also the locational electricity prices of geographically distributed datacenters. Due to various power generation and transmission profiles in different regions, the electricity prices of United States exhibit location and time diversities. Several works proposed the inter-datacenter workload management to reduce the electricity cost of active servers by considering the location and time variation of electricity prices. Qureshi et al. [8] proposed to monitor the time-varying locational electricity prices and periodically to distribute service requests to server clusters with lower electricity prices in a content delivery network to cut the electricity cost of the content delivery network. Rao et al. [9] achieved the reduction on the electricity cost of active servers under a multi-electricity-market environment while guaranteeing QoS.

Both the intra-datacenter workload management and the inter-datacenter workload management have obvious disadvantages since each of them only considers one factor that influences the electricity cost of geographically distributed datacenters and can not globally and minimize the total electricity cost of geographically distributed datacenters. Recently, some publications jointly took into account more devices besides servers to minimize the electricity cost of geographically distributed datacenters. Li et al. [10] have proposed to lower the electricity cost of server facilities as well as cooling systems. Zhang et al. [11] have proposed to minimize the electricity cost of server facilities, cooling systems, and networking devices. However, these studies have not considered the real-life cooling system and made an unrealistic assumption that the cooling system consumes constant power in the datacenter. The assumption is invalid for geographically distributed datacenters since the uneven workload distribution in a datacenter could affect the efficiency of a cooling system. Based on the analysis in [6][7][12], the workload distribution in a datacenter can significantly affect the power consumption of a cooling system, and it should be taken into consideration as an important factor by cloud service operators to reduce the electricity cost of geographically distributed datacenters.

Besides the studies that focus on minimizing the electricity cost of active servers and cooling systems, there are some other researches aiming at minimizing the power consumption of datacenter network. Heller et al. [25] have achieved power saving on datacenter network using a network-wide power manager ElasticTree, which consolidates traffic flows onto as few routes as possible and turns off unused switches and links. Wang et al. [26] have proposed a correlation-aware power optimization scheme CARPO, which consolidates flows with low correlation together to further save power based on a observation that bandwidth demands of different flows do not peak at the same time in real datacenter networks.

To fully utilize the time diversity of electricity prices, some recent proposal present to reduce the electricity cost of datacenters by using the electricity which are stored into energy storage devices when the electricity price is relatively low. Urgaonkar et al. [27] have explored the chances in reducing server power bill by the use of uninterrupted power supply (UPS) units as energy storage devices for datacenters. Zhang et al. [28] have proposed to leverage thermal and energy storage techniques (e.g., ice/water-based thermal tanks, UPS batteries) to store electricity when the electricity price is relatively low and to use those stored energy to cool the datacenter when the electricity price is high. However, the devices for energy storage techniques (e.g., thermal tanks, UPS) need extra cost for cloud service providers.

Due to the global energy crisis and environmental concerns, green datacenter using the renewable energy is becoming an increasingly important topic for clouds service providers. Steward et al. [29] have tried to maximize the use of renewable energy in datacenters. Zhang et al. [30] have proposed to maximally use the renewable energy subject to the cost budget of the cloud service provider as well as being aware of the time-varying electricity prices.

8. Conclusion

One of the key questions faced by many cloud service providers is how to reduce the electricity cost of geographically distributed datacenters. Current studies mainly focus on either the impact of variant electricity prices on the electricity cost of active servers or the impact of datacenter workload distribution in a datacenter on the electricity cost of a cooling system, but neglect the joint optimization of these two aspects. In this paper, we propose a joint inter- and intra-datacenter workload management scheme JET, which jointly takes into account the two factors to reduce the total electricity cost of geographically distributed datacenters. As the incoming service requests vary, JET alternately selects the variation of electricity prices or the efficiency of cooling systems as the major factor for the electricity cost of geographically distributed datacenters, and achieves substantial reduction on the electricity cost of geographically distributed datacenters.

In future, we plan to extend JET in the following aspects: (1) consider other types of workloads (e.g., service requests with different processing durations), (2) design a hierarchical/distributed architecture for managing of geographically distributed datacenters, (3) consider more real-life constraints (e.g., propagation delay, the temperature dynamics of the three zones), (4) study other state-of-the-art cooling solutions (e.g., ambient air cooling, hot aisle/cold aisle containment).

Acknowledgments

We thank Junjie Zhang at the Polytechnic Institute of New York University, for his help on the problem formulation.

References

[1] J. Koomey, Growth in data center electricity use 2005 to 2010, Oakland, CA: Analytics Press, 2011.

[2] L. Barroso, U. Hölzle, The datacenter as a computer: An introduction to the design of warehouse-scale machines, Synthesis Lectures on Computer Architecture 4 (1) (2009) 1–108.

[3] J. Heo, D. Henriksson, X. Liu, T. Abdelzaher, Integrating adaptive components: An emerging challenge in performance-adaptive systems and a server farm case-study, in: RTSS 2007, 2007, pp. 227–238.

[4] E. Elnozahy, M. Kistler, R. Rajamony, Energy-efficient server clusters, Power-Aware Computer Systems (2003) 179–197.

[5] U. E. P. Agency, Report to congress on server and data center energy efficiency: Public law 109-431, 2007.

[6] R. Sharma, C. Bash, C. Patel, R. Friedrich, J. Chase, Balance of power: Dynamic thermal management for internet data centers, Internet Computing, IEEE 9 (1) (2005) 42–49.

[7] C. Bash, G. Forman, Cool job allocation: measuring the power savings of placing jobs at cooling-efficient locations in the data center, in: 2007 USENIX Annual Technical Conference on Proceedings of the USENIX Annual Technical Conference, USENIX Association, 2007, pp. 1–6.

[8] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, B. Maggs, Cutting the electric bill for internet-scale systems, ACM SIGCOMM Computer Communication Review 39 (4) (2009) 123–134.

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

[10] J. Li, Z. Li, K. Ren, X. Liu, Towards optimal electric demand management for internet data centers, Smart Grid, IEEE Transactions on 3 (1) (2012) 183–192.

[11] Y. Zhang, Y. Wang, X. Wang, Electricity bill capping for cloud-scale data centers that impact the power markets, ICPP '12, 2012, pp. 440–449.

[12] F. Ahmad, T. Vijaykumar, Joint optimization of idle and cooling power in data centers while maintaining response time, in: ACM Sigplan Notices, Vol. 45, ACM, 2010, pp. 243–256.

[13] D. Anderson, J. Dykes, E. Riedel, More than an interface – scsi vs. ata, in: FAST03, 2003, pp. 245–257.

[14] J. Moore, J. Chase, P. Ranganathan, R. Sharma, Making scheduling" cool": temperature-aware workload placement in data centers, in: Proceedings of on USENIX Annual Technical Conference, 2005, pp. 5–5.

[15] M. Pióro, D. Medhi, Routing, flow, and capacity design in communication and computer networks, Morgan Kaufmann, 2004.

[16] P. Gill, W. Murray, M. Saunders, M. Wright, Minos 5.5 user's guide. report 83-20r, department of operations research (1998).

[17] Google, `http://www.google.com/about/datacenters/inside/locations/index.html`.

[18] NYISO, `http://www.nyiso.com`.

[19] ERCOT, `http://www.ercot.com`.

[20] G. Urdaneta, G. Pierre, M. van Steen, Wikipedia workload analysis for decentralized hosting, Computer Networks 53 (11) (2009) 1830–1845.

[21] L. Barroso, U. Holzle, The case for energy-proportional computing, Computer 40 (12) (2007) 33–37.

[22] P. Bohrer, E. Elnozahy, T. Keller, M. Kistler, C. Lefurgy, C. McDowell, R. Rajamony, The case for power management in web servers, Power aware computing 78758.

[23] J. Heo, P. Jayachandran, I. Shin, D. Wang, T. Abdelzaher, X. Liu, Optituner: On performance composition and server farm energy minimization application, Parallel and Distributed Systems, IEEE Transactions on 22 (11) (2011) 1871–1878.

[24] R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, X. Zhu, No power struggles: Coordinated multi-level power management for the data center, in: ACM SIGARCH Computer Architecture News, Vol. 36, ACM, 2008, pp. 48–59.

[25] B. Heller, S. Seetharaman, P. Mahadevan, Y. Yiakoumis, P. Sharma, S. Banerjee, N. McKeown, Elastictree: saving energy in data center networks, in: Proceedings of the 7th USENIX conference on Networked systems design and implementation, 2010, pp. 17–17.

[26] X. Wang, Y. Yao, X. Wang, K. Lu, Q. Cao, Carpo: Correlation-aware power optimization in data center networks, in: INFOCOM, 2012 Proceedings IEEE, IEEE, 2012, pp. 1125–1133.

[27] R. Urgaonkar, B. Urgaonkar, M. Neely, A. Sivasubramaniam, Optimal power cost management using stored energy in data centers, in: Proceedings of the ACM SIGMETRICS, ACM, 2011, pp. 221–232.

[28] Y. Zhang, Y. Wang, X. Wang, Testore: Exploiting thermal and energy storage to cut the electricity bill for datacenter cooling, in: Network and Service Management (CNSM), 2012 8th International Conference on, IEEE, 2012, pp. 19–27.

[29] C. Stewart, K. Shen, Some joules are more precious than others: Managing renewable energy in the datacenter, in: Proceedings of the Workshop on Power Aware Computing and Systems, 2009.

[30] Y. Zhang, Y. Wang, X. Wang, Greenware: Greening cloud-scale data centers to maximize the use of renewable energy, in: Middleware 2011, Springer, 2011, pp. 143–164.