Cutting the Electricity Cost of Distributed Datacenters through Smart Workload Dispatching

Zehua Guo, Yang Xu, *Member, IEEE*, H. Jonathan Chao, *Fellow, IEEE*, and Zhemin Duan

***Abstract*—This letter proposes a smart inter- and intra-datacenter workload dispatching scheme, Joint Electricity price-aware and Cooling efficiency-aware load balancing (JEC), to cut the electricity cost of distributed datacenters. Evaluation shows JEC outperforms existing schemes and achieves significant reduction on the total electricity cost of distributed datacenters.**

***Index Terms*—Distributed datacenters, workload dispatching, electricity cost, electricity price, cooling efficiency.**

I. Introduction

GEOGRAPHICALLY distributed datacenters have been rapidly expanding in recent years due to the increasing demand on cloud services. In general, a datacenter spends 30%∼50% of its operational expenses toward electricity [1]. To cut the electricity cost of distributed datacenters, many studies have been conducted to seek optimal datacenter workload management mechanisms. Some studies work on the *inter*-datacenter workload dispatching to minimize the electricity cost of active servers [2][3]. Others focus on the *intra*-datacenter workload dispatching to reduce the power consumption of datacenter devices [4][5]. However, each of these existing schemes considers only one part of datacenters. A simple combination of the two aforementioned schemes cannot achieve the global minimization of the total electricity cost of distributed datacenters, as we will detail in this letter.

In this letter, we propose a novel workload dispatching scheme, Joint Electricity price-aware and Cooling efficiency-aware load balancing (JEC), to minimize the total electricity cost of distributed datacenters. JEC jointly considers the variation of electricity prices among datacenters and the impact of workload distribution on the efficiency of the cooling system in each datacenter. Evaluation shows that JEC outperforms existing schemes and achieves significant reduction on the total electricity cost of distributed datacenters.

II. Electricity Cost Models and QoS Constraints of Distributed Datacenters

A. Electricity Cost Models of Distributed Datacenters

1) Electricity Cost Model of Active Servers: Suppose one cloud service provider owns N distributed datacenters. Datacenter i $(1 \leq i \leq N)$ is at location i with hourly electricity price $Pr_i(t)$ $(t > 0)$ at time t. The power consumption of each server in datacenter i is Po_i [6]:

$$Po_i = P_{idle} + (P_{peak} - P_{idle})u_i \tag{1}$$

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.

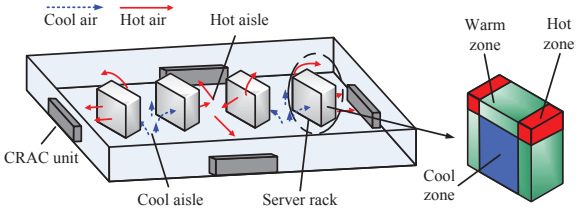

Fig. 1. Water-chilled datacenter cooling system and three nearly temperature isolated zones in a server rack [7].

Assume each datacenter uses homogeneous servers and configurations, and datacenter i contains m_i active servers. The electricity cost of active servers ($ECAS$) of N distributed datacenters is given as:

$$ECAS = \sum_{i=1}^{N} m_i Po_i Pr_i(t) \tag{2}$$

2) Electricity Cost Model of Cooling System: Fig.1 shows the typical datacenter cooling system using the water-chilled Computer Room Air Conditioner (CRAC). CRAC unit takes in hot air produced by active servers and delivers cool air into a datacenter room. The efficiency of cooling system is quantified by Coefficient of Performance (COP) of CRAC units. The cooling power required to remove the heat from active servers is given by the relation: $the\ power\ of\ active\ servers/COP$. As presented in [8], to remove 10 kW of heat for cooling a specific volume of air, the CRAC units taking in hot air at 25°C and pushing cool air at 20°C saves about 40% cooling power, as compared with the CRAC units taking in hot air at 20°C and pushing cool air at 15°C. Therefore, COP is an increasing function of CRAC output temperature, and the efficiency of cooling system can be maximized by raising CRAC output temperature, while preventing the room temperature from crossing the maximum safety temperature [8]. Specifically, assume the outside environment remains unchanged at datacenter i for a certain period of time, the maximum safety temperature of datacenter room is T^{MAX}, and CRAC units push cool air at the temperature of T^{out} into the room. The room temperature is affected by the temperature of servers T_i^{server}, which depends on the number and distribution of processing workload. To prevent the room temperature from exceeding the maximum safety temperature, we have the adjusted new CRAC output temperature and COP function [8]:

$$T_i^{new} = T^{out} + T^{MAX} - T_i^{server} \tag{3}$$

$$COP_i = 0.0068 T_i^{new2} + 0.0008 T_i^{new} + 0.458 \tag{4}$$

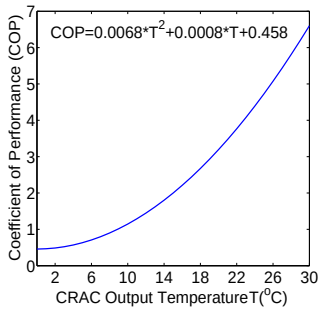

Fig. 2. COP curve of CRAC [8].

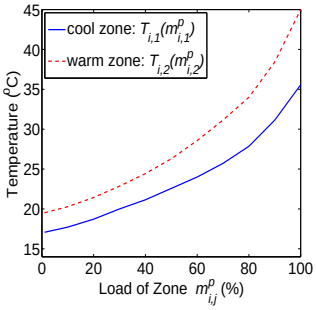

Fig. 3. Temperature curves of warm zone and cool zone [7].

T_i^{server} is affected by the distribution and number of processing workload in datacenter i [7]. Due to physical structure and air flow patterns in a datacenter, a datacenter can be divided into three nearly temperature isolated zones [7]. Specifically, in each datacenter, hot air rises from bottom to top and air does not circulate well at the ends of the aisles. As a result, 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. Hence, m_i active servers in datacenter i can be divided into three parts: $m_i = \sum_{j=1}^{3} m_{i,j}$, $m_{i,1}$ for the cool zone, $m_{i,2}$ for the warm zone, and $m_{i,3}$ for the hot zone [7]. The temperature of zone j in datacenter i is given as $T_{i,j}(m_{i,j}^p)$, which is a linear piece-wise curve associated with $m_{i,j}^p$ ($m_{i,j}^p = m_{i,j}/m_{i,j}^{MAX}$), the load percentage of zone j [7], that is:

$$T_i^{server} = max\left(T_{i,1}(m_{i,1}^p), T_{i,2}(m_{i,2}^p), T_{i,3}(m_{i,3}^p)\right) \quad (5)$$

$$T_{i,j}\left(m_{i,j}^p\right) \leq T^{MAX} \quad (6)$$

For a given workload, we could minimize the temperature difference of the three zones by rationally dispatching service requests to servers of the three zones and, in turn, minimize T_i^{server}. Therefore, the electricity cost of cooling system ($ECCS$) of N distributed datacenters is given as:

$$ECCS = \sum_{i=1}^{N} \frac{\left(\sum_{j=1}^{3} m_{i,j}\right) Po_i Pr_i(t)}{0.0068 T_i^{new\,2} + 0.0008 T_i^{new} + 0.458} \quad (7)$$

3) Total Electricity Cost Model of Distributed Datacenters: The total electricity cost (EC) of N distributed datacenters could be written as:

$$\begin{aligned} EC &= ECAS + ECCS \\ &= \sum_{i=1}^{N} \left(\sum_{j=1}^{3} m_{i,j}\right) Po_i Pr_i(t) \cdot \\ &\quad \left(1 + \frac{1}{0.0068 T_i^{new\,2} + 0.0008 T_i^{new} + 0.458}\right) \end{aligned} \quad (8)$$

B. QoS Constraints of Distributed Datacenters

Assume N distributed datacenters receive λ service requests in a time interval, and datacenter i with server service rate μ_i is assigned with λ_i service requests using a specific workload dispatching scheme. The average delay of datacenter i is given as d_i, which should not exceed a delay constraint D_i. We use the model of average delay similar to a related study [3]. Therefore, we have:

$$\lambda = \sum_{i=1}^{N} \lambda_i \quad (9)$$

$$m_{i,j} \leq m_{i,j}^{MAX} \quad (10)$$

$$d_i = \frac{1}{\left(\sum_{j=1}^{3} m_{i,j}\right)\mu_i - \lambda_i} \leq D_i \quad (11)$$

$$\mu_i = f(u_i, \mu_i^{MAX}) \quad (12)$$

where $(X)^{MAX}$ is the upper limit of variable X, and μ_i is a function associated with u_i and μ_i^{MAX}.

III. Problem Formulation and Solution

A. Transformations and Assumptions

To solve a complicated optimization problem, a common method is to transform the problem into a standard problem (e.g., convex optimization problem) that can be solved using existing optimization techniques or solvers. The complexity of our problem comes from the nonlinear constraint (Eq.(5)) and the nonlinear objective function (Eq.(8)). To solve the problem, we should transform Eq.(5) into a linear constraint, and transform Eq.(8) into a convex function. In Eq.(8), 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 dispatching scheme. Assume any two of the three variables are constant, if the other one decreases, EC decreases, and vice versa. Hence, the three variables have uniform monotonicity with EC. If the minimization of Eq.(8) is set as the core objective, Eq.(5) and Eq.(6) can be transformed into two linear inequalities:

$$T_{i,j}(m_{i,j}^p) \leq T_i^{server} \quad (13)$$

$$T_i^{server} \leq T^{MAX} \quad (14)$$

The complexity of EC comes from T_i^{server} (a nonlinear function related to $m_{i,j}$) in its denominator. Thus, EC can be transformed into a formulation directly related to $m_{i,j}$ by simplifying its fractional component. We combine three linear piece-wise zone temperature curves associated with their zone load percentages (Fig.3) into one datacenter temperature curve associated with the overall datacenter load percentage, which is shown in Fig.4. To accommodate potential workload spikes, we use 10% capacity margin for each datacenter [9]. We bring the datacenter temperature curve into COP function, take reciprocal of COP, and then get $1/COP_i$ curve, shown as the blue diamond curve in Fig.5. $1/COP_i$ is also a nonlinear function associated with the overall datacenter load percentage. We use one dimensional linear regression [10] to substitute the original $1/COP_i$ function with K linear piece-wise functions $g_k(m_i^p)$, $(m_i^p = m_i/m_i^{MAX}, 1 \leq k \leq K)$, which are shown as the red triangle curves in Fig.5, that is:

$$g_k(m_i^p) = a_k + b_k m_i^p, (th_{k-1} < m_i^p \leq th_k) \quad (15)$$

$$m_i^p = m_i/m_i^{MAX} \quad (16)$$

$$g_k(m_i^p) \leq G^{MAX} \quad (17)$$

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

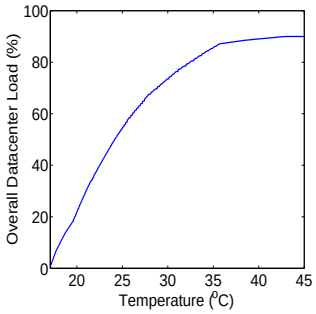

Fig. 4. Temperature curve of overall datacenter load percentage.

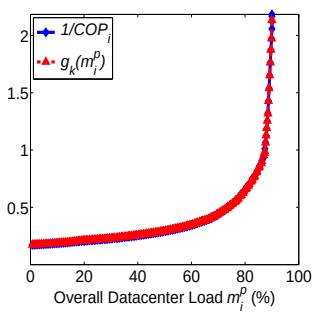

Fig. 5. $1/COP_i$ and $g_k(m_i^p)$ of overall datacenter load percentage.

Therefore, we have a convex function: ($EC^{new\prime\prime}(m_i) = \frac{2Po_iPr_i(t)b_k}{m_i^{MAX}} > 0$, $EC^{new\prime\prime}(u_i) = 0$) :

$$EC^{new} = \sum_{i=1}^{N} m_i Po_i Pr_i(t) \left(1 + g_k(m_i^p)\right) \tag{18}$$

With power-management, a server's idle power accounts for 50%∼65% of its peak power [2]. Thus, we assume $P_{peak} = 2P_{idle}$. Similar to a related study [3], we assume that all the active servers will run close to 100% utilization, because the number of active servers is minimized by the workload dispatching scheme, that is $u_i = 1$. Based on the above assumptions, $Po_i = 2P_{idle}$ and $\mu_i = \mu_i^{MAX}$.

B. Formulation of Electricity Cost Minimization Problem

Given λ service requests in a time interval, the optimization goal is to minimize the total electricity cost of distributed datacenters by a workload dispatching scheme, so that datacenter i activates $m_{i,j}$ servers in zone j to process the distributed λ_i service requests. m_i is first obtained by the solution of Problem One (P1), and $m_{i,j}$ is further obtained by bringing m_i into some equations. P1 is formulated as follows:

$$\min \sum_{i=1}^{N} 2m_i P_{idle} Pr_i(t) \left(1 + a_k + b_k m_i / m_i^{MAX}\right) \tag{19a}$$

subject to

$$1/(m_i\mu_i^{MAX} - \lambda_i) \le D_i \tag{19b}$$

$$m_i \le m_i^{MAX} \tag{19c}$$

$$a_k + b_k m_i / m_i^{MAX} \le G^{MAX} \tag{19d}$$

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

C. Solution of Electricity Cost Minimization Problem

Since variables m_i and λ_i are integers, and EC is a nonlinear function, P1 is a Nonlinear Integer Programming (NIP) problem. However, the objective function of P1 is a convex function and all constraints of P1 are linear. Therefore, the decimal solution of P1 can be first obtained by efficient optimization techniques (e.g., *Interior Point method*) or solvers (e.g., *MINOS 5.5*). In the final solution, the number of active servers in one of the distributed datacenters is at least in the order of thousands. Thus, to reduce the scheme's complexity, the integer solution of m_i can be obtained by rounding up m_i from decimal solution instead of using Branch-and-Cut method. Then, we can get the approximate COP_i by bringing m_i into Eq.(15), and get T_i^{server} by bringing COP_i into Eq.(3) and Eq.(4). Finally, we can obtain $m_{i,j}$ with Eq.(13). The entire solution is named JEC. The evaluation results show that the difference between the optimal solution obtained from Branch-and-Cut method and the round-up solution is less than 0.005%. For a cloud service provider operating 20 datacenters, the processing time of JEC is less than 20ms using MINOS 5.5.

IV. Evaluation

A. Workload Dispatching Schemes for Comparison

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

2) Electricity price-aware InteR datacenter load balancing (EIR)[2][3]: EIR considers only the location and time diversity of electricity prices, and cuts $ECAS$ by dispatching service requests to datacenters with lower electricity prices.

3) Cooling-aware IntrA datacenter load balancing (CIA)[4][5]: CIA takes into account only the physical structure of a datacenter, and reduces cooling power by selecting servers in locations with higher cooling efficiency to process incoming service requests.

4) EIR+CIA: EIR+CIA considers the diversity of electricity prices and the efficiency of cooling system in two separate steps. In the first step, EIR is used to decide the number of active servers for each datacenter to minimize $ECAS$ based on electricity prices. In the second step, CIA is used to decide the placement of active servers in the three zones of each datacenter to improve the efficiency of cooling system. Without jointly considering the two factors, EIR+CIA can cause some undesirable situations. For example, 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 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.

5) JEC: To cut EC, JEC jointly considers the two factors to decide the optimal number and placement of active servers in distributed datacenters. 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 $ECCS$.

Since EIR and CIA are partial workload dispatching schemes as compared with JEC, we complement them with random load balancing.

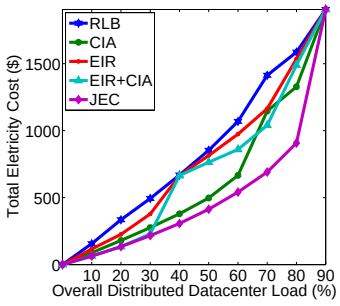

Fig. 6. Total electricity cost of three datacenters for variant overall three datacenters loads under Case 1.

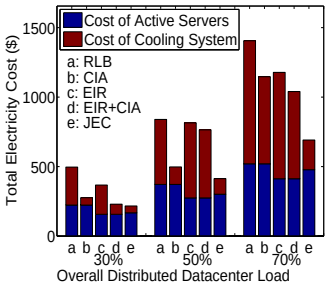

Fig. 7. Electricity cost compositions of three datacenters for overall three datacenters loads under Case 1.

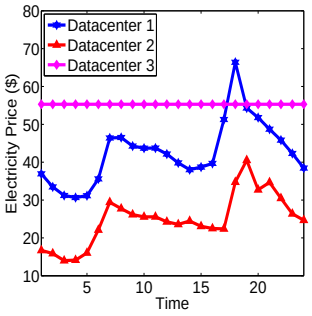

Fig. 8. Hourly electricity prices of three datacenters on Dec. 14, 2012 for Case 2.

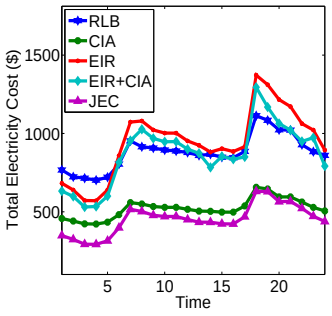

Fig. 9. Total electricity cost of three datacenters for different times under Case 2.

B. Evaluation Setup

In our evaluation, we use a trace containing 10% of Internet traffic arrived at Wikipedia between Oct.1, 2007 and Nov.30, 2007 [11]. We simulate a cloud service provider operating three geographically distributed datacenters located in Long Island, NY; Houston, TX; and Atlanta, GA; respectively. The ones in Long Island and Houston are located in the electricity wholesale market regions, where electricity prices vary based on the condition of grid. The one in Atlanta is in the regulated utility region, where electricity prices are fixed for a certain period of time. The power consumption profile of each server in the three datacenters is assumed to be approximately the same: $P_{idle} = 100$ watts [6]. The maximum numbers of servers are assumed to be 30000, 60000, and 25000 and their processing capacity coefficients are 4.0, 2.5, and 3.5 service requests per second, respectively. The delay constraint is 100 ms. We use the water-chilled cooling system shown in Fig.1. Each datacenter contains four CRAC units which push cool air at 15°C into the room. To prevent the room temperature from exceeding the maximum safety temperature of 30°C, CRAC units adaptively adjust their efficiency.

We use two specific cases to compare JEC with the four baseline schemes. In Case 1, service requests received by the three datacenters vary, but electricity prices are fixed at $Pr_1(t) = 42.92566\$/MWh$, $Pr_2(t) = 20.27\$/MWh$, $Pr_3(t) = 55.3\$/MWh$. In Case 2, the arriving service requests per second are fixed at 50% of the overall load of the three datacenters, but electricity prices of the three datacenters change with their regional electricity price schemes. The maximum overall load of the three datacenters is 90%, since 10% capacity margin for each datacenter is used to accommodate potential workload spikes [9].

C. Results

1) Case 1: Fig.6 shows the total electricity cost of the three datacenters for the five schemes. JEC outperforms RLB, EIR, CIA, and EIR+CIA for all datacenter load spans. Specifically, in the range of 30% to 70% datacenter load where datacenters operate most of the time [7], JEC achieves substantial electricity cost reductions of 48% to 57%, as compared with RLB. Fig.7 shows the composition of electricity cost for JEC and the four baseline schemes at three typical datacenter loads in detail. It verifies the analysis in Section IV-A.

2) Case 2: Due to different regional electricity price schemes, we use the fixed electricity price for Atlanta, and hourly electricity prices for Long Island and Houston on Dec.14, 2012 from NYISO and ERCOT shown in Fig.8. Fig.9 presents the total electricity cost for the five workload dispatching schemes when real electricity prices are applied. Compared with other schemes, the total electricity cost of JEC varies slightly with the variation of electricity prices, and JEC performs better than other schemes all the time.

V. Conclusion

As this letter shows, JEC outperforms the existing schemes and achieves significant reduction of the total electricity cost of distributed datacenters. In our future work, we will study other state-of-the-art cooling solutions (e.g., ambient air cooling, hot aisle/cold aisle containment) to make our JEC scheme more generally applicable.

References

[1] L. Barroso and U. Hölzle, "The datacenter as a computer: An introduction to the design of warehouse-scale machines," *Synthesis Lectures on Computer Architecture*, vol. 4, no. 1, pp. 1–108, 2009.

[2] 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, pp. 123–134, Aug. 2009.

[3] 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*, pp. 1–9, 2010.

[4] R. Sharma, C. Bash, C. Patel, R. Friedrich, and J. Chase, "Balance of power: Dynamic thermal management for internet data centers," *Internet Computing, IEEE*, vol. 9, no. 1, pp. 42–49, 2005.

[5] C. Bash and G. Forman, "Cool job allocation: measuring the power savings of placing jobs at cooling-efficient locations in the data center," in *USENIX '07*, pp. 1–6, USENIX Association, 2007.

[6] M. Ghamkhari and H. Mohsenian-Rad, "Energy and performance management of green data centers: A profit maximization approach," *Smart Grid, IEEE Transactions on*, vol. 4, no. 2, pp. 1017–1025, 2013.

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

[8] J. Moore, J. Chase, P. Ranganathan, and R. Sharma, "Making scheduling" cool": temperature-aware workload placement in data centers," in *USENIX '05*, pp. 5–5, 2005.

[9] J. Li, Z. Li, K. Ren, and X. Liu, "Towards optimal electric demand management for internet data centers," *Smart Grid, IEEE Transactions on*, vol. 3, no. 1, pp. 183–192, 2012.

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

[11] G. Urdaneta, G. Pierre, and M. van Steen, "Wikipedia workload analysis for decentralized hosting," *Computer Networks*, vol. 53, pp. 1830–1845, July 2009.