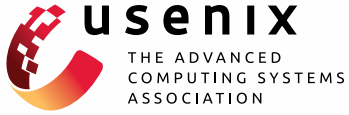


Coordinating Liquid and Free Air Cooling with Workload Allocation for Data Center Power Minimization

Li Li, Wenli Zheng, Xiaodong Wang, and Xiaorui Wang, ***The Ohio State University***

https://www.usenix.org/conference/icac14/technical-sessions/presentation/li_li

This paper is included in the Proceedings of the 11th International Conference on Autonomic Computing (ICAC '14).

June 18–20, 2014 • Philadelphia, PA

ISBN 978-1-931971-11-9

Open access to the Proceedings of the 11th International Conference on Autonomic Computing (ICAC '14) is sponsored by USENIX.

Coordinating Liquid and Free Air Cooling with Workload Allocation for Data Center Power Minimization

Li Li, Wenli Zheng, Xiaodong Wang, and Xiaorui Wang
Power-Aware Computer System (PACS) Laboratory
Department of Electrical and Computer Engineering
The Ohio State University, Columbus, OH 43210
{*li.2251, zheng.691, wang.3570, wang.3596*}*@osu.edu*

Abstract

Data centers are seeking more efficient cooling techniques to reduce their operating expenses, because cooling can account for 30-40% of the power consumption of a data center. Recently, liquid cooling has emerged as a promising alternative to traditional air cooling, because it can help eliminate undesired air recirculation. Another emerging technology is free air cooling, which saves chiller power by utilizing outside cold air for cooling. Some existing data centers have already started to adopt both liquid and free air cooling techniques for significantly improved cooling efficiency and more data centers are expected to follow.

In this paper, we propose SmartCool, a power optimization scheme that effectively coordinates different cooling techniques and dynamically manages workload allocation for jointly optimized cooling and server power. In sharp contrast to the existing work that addresses different cooling techniques in an isolated manner, SmartCool systematically formulates the integration of different cooling systems as a constrained optimization problem. Furthermore, since geo-distributed data centers have different ambient temperatures, SmartCool dynamically dispatches the incoming requests among a network of data centers with heterogeneous cooling systems to best leverage the high efficiency of free cooling. A light-weight heuristic algorithm is proposed to achieve a near-optimal solution with a low run-time overhead. We evaluate SmartCool both in simulation and on a hardware testbed. The results show that SmartCool outperforms two state-of-the-art baselines by having a 38% more power savings.

1 Introduction

In recent years, high power consumption has become a serious concern in operating large-scale data centers. For example, a report from Environmental Protection Agency (EPA) estimated that the total energy consumption from data centers in the US was over 100 billion kWh in 2011. Among the total power consumed by a data center, cooling power can account for 30-40% [14][2]. As new high-density servers (e.g., blade servers) are increasingly being deployed in data centers, it is important for the cooling systems to more effectively remove the heat. However, with the high-density servers being installed, the traditional computer room air conditioner (CRAC) system might not be efficient enough, as its Power Usage Effectiveness (PUE) is around 2.0 or higher. PUE is defined as the ratio of the total energy consumption of a data center over energy consumed by the IT equipment such as servers. With a high PUE, the cooling power consumption of a data center can grow tremendously as high-density servers being deployed, which not only increases the operating cost, but also causes negative environmental impact. Therefore, data centers are in an urgent need to find higher-efficient cooling techniques to reduce PUE.

Two new cooling techniques have recently been proposed to increase the cooling efficiency and lower the PUE of a data center. The first one is liquid cooling, which conducts coolant through pipes to some heat exchange devices that are attached to the IT equipments, such that the generated heat can be directly taken away by the coolant. The second one, which is referred to as free air cooling [7], exploits the relatively cold air outside the data center for cooling and thus saves the power of chilling the hot air returned from the IT equipment. Although both of the two cooling techniques highly increase the cooling efficiency of data centers, each technique has its own limitations. The liquid cooling approach requires additional ancillary facilities (e.g., the valves and pipes) and maintenance, which can increase the capital investment when being deployed in a large scale. The free air cooling technique requires a low outside air temperature, which might not be available all the time in a year. In order to mitigate the problems, hy-

brid cooling system, which is composed with the liquid cooling, the free air cooling and the traditional CRAC air cooling, can be used to lower the cooling cost and ensure the cooling availability. Several hybrid-cooled data centers have been put into production. For example, the CERN data center located in Europe adopts a hybrid cooling system with both liquid-cooling and traditional CRAC cooling systems, in which about 9% of the servers are liquid-cooled [4].

However, efficiently operating such a hybrid cooling system is not a trivial task. Currently, existing data centers that adopt multiple cooling techniques commonly use some preset outside temperature thresholds to switch between different cooling systems, regardless of the time-varying workload. Such a simplistic solution can often lead to unnecessarily low cooling efficiencies. Although some previous studies [11][35] have proposed to intelligently distribute the workload across the servers and manage the cooling system according to the real-time workload to avoid over-cooling, they address only one certain cooling technology and thus the resulted workload distribution might not be optimal for the hybrid cooling system. To the best of our knowledge, no prior research has been done for efficiently coordinating multiple cooling techniques in a hybrid-cooled data center. In addition, for a network of data centers that are geographically distributed, there exist some works focusing on balancing the workload or reducing the server power cost [16][27], but none of them minimize the cooling power consumption, especially when the data centers have heterogeneous cooling systems. In order to minimize the power consumption of a hybrid-cooled data center, we need to face several new challenges. First, the different characteristics of these three cooling systems (liquid cooling, free air cooling and the traditional CRAC air cooling) demand for a systematic approach to coordinate them effectively. Second, workload distribution in such a hybrid-cooled data center needs to be carefully planned in order to jointly minimize the cooling and server power consumption. Third, due to the different local temperatures, a novel workload distribution and cooling management approach is needed for data centers that are geographically distributed at different locations, in order to better utilize free air cooling in the hybrid cooling system more efficiently.

In this paper, we propose *SmartCool*, a power optimization scheme to optimize the total power consumption of a hybrid-cooled data center by intelligently managing the hybrid cooling system and distributing the workload. We first formulate the power optimization problem for a single data center, which can then be solved with a widely adopted optimization technique. We then extend the power optimization scheme to fit a network of geo-distributed data centers. To reduce the computational overhead, we propose a light-weight algorithm to solve the optimization problem for the geo-distributed data centers. Specifically, this paper has the following major contributions:

- More and more data centers are on their way to adopt high-efficient cooling techniques, but many data centers heavily rely on simplistic solutions to separately manage their cooling systems, which often lead to an unnecessarily low cooling efficiency. In this paper, we propose to address an increasingly important problem: Intelligent coordination of cooling systems for jointly minimized cooling and server power in a data center.
- We formulate the cooling management in a hybrid-cooled data center with liquid cooling, free air cooling, and traditional CRAC air cooling, as a constrained power optimization problem to minimize the total power consumption. SmartCool features a novel air recirculation model developed based on computational fluid dynamics (CFD).
- To best leverage the high efficiency of free cooling in geo-distributed data centers that have different ambient temperatures, we extend our optimization formulation to dynamically dispatch the incoming requests among a network of data centers with heterogeneous cooling systems. A light-weight heuristic algorithm is proposed to achieve a near-optimal solution with a low run-time overhead.
- We evaluate SmartCool both in simulation and on a hardware testbed with real-world workload and temperature traces. The results show that SmartCool outperforms two state-of-the-art baselines by saving 38% more power consumption.

The rest of the paper is organized as follows. We review the related work in Section 2, and introduce the background of different cooling technologies in Section 3. Section 4 formulates power optimization problem for a single hybrid-cooled data center, which is extended for geo-distributed data centers in Section 5 with a light-weight algorithm. We present our simulation results in Section 6 and the hardware experiment results in Section 7. Finally, Section 8 concludes the paper.

2 Related Work

Minimizing the power consumption of data centers has recently received much attention, such as [27, 11, 18, 29, 31, 32, 13, 28, 17]. In particular, a lot of work has been done to optimize the traditional CRAC air cooling in data centers. For example, Anto et al. [22] construct

a model of a single CRAC unit which offers flexible selection in different heat exchangers and coolants. Zhou et al. [23] propose a computationally efficient multivariable model to capture the effects of CRAC fan speed and supplied air temperature on the rack inlet temperatures. Tang et al. [20] propose a workload scheduling scheme to make the inlet temperatures of all servers as even as possible. A holistic approach is proposed by Chen et al. [30] that integrates the management of IT, power and cooling infrastructures to improve the data center efficiency. In our work, we adopt the modeling process of traditional CRAC cooling system, but coordinate its work with the liquid cooling and free cooling systems as a hybrid-cooling system for the improvement of data center cooling efficiency.

Free air cooling and liquid cooling have also attracted wide research attentions. Christy et al. [9] study two primary free cooling systems, the air economizer and the water economizer. Gebrehiwot et al. [3] study the thermal performance of an air economizer for a modular data center using computational fluid dynamics. Coskun et al. [1] provide a 3D thermal model for liquid cooling, with variable fluid injection rates. Hwang et al. [8] develop an energy model for liquid-cooled data centers based on the thermo-fluid first principles. Differently, our work focuses on the power optimization of hybrid-cooled data centers, by managing the cooling modes and workload distribution.

For geo-distributed data centers, Adnan et al. [16] save the cost of load balancing by utilizing the flexibility of the Service Level Agreements. Related algorithms are developed in [33][34][12] to minimize the total cost of geo-distributed data centers and the environmental impact. Our global workload dispatching strategy minimizes the total power consumption of all the distributed data center, by leveraging the temperature differences among different locations and maximizing the usage of free air cooling.

3 Different Cooling Technologies

Figure 1 illustrates the cooling system of a hybrid-cooled data center, which includes traditional air cooling, liquid cooling and free air cooling. The liquid cooling system uses chiller and cooling tower to provide coolant. Either the CRAC system or the free cooling system can be selected for air cooling. The CRAC system also relies on chiller and cooling tower to provide the coolant, which is then used to absorb heat from the air in the data center. The free cooling system draws outside air into the data center through the Air Handling Unit (AHU) when the outside temperature can meet the cooling requirement.

Traditional CRAC air cooling is the most widely used cooling technology in existing data centers. This

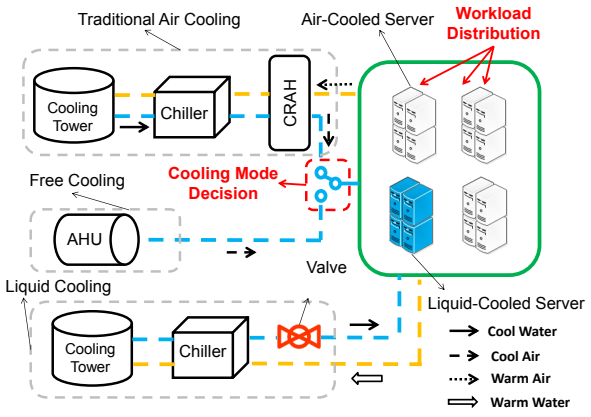


Figure 1: Cooling System of Hybrid-Cooled Data Center. The cold plates used for liquid cooling are installed inside the liquid-cooled servers. The air cooling mode is decided based on the outside temperature.

system deploys several CRAC units in the computer room to supply cold air. The cold air usually goes under the raised floor before joining in the cold aisle through perforated tiles to cool down the servers, as shown in Figure 2. The hot air from the servers is output to the hot aisle and returned to the CRAC system to be cooled and reused. The deployment of cold aisle and hot aisle is used to form isolation between cold and hot airs. However, due to the existence of seams between servers and racks, as well as the space close to the ceiling where there is no isolation, cold air and hot air are often mixed to a certain extent, which decreases the cooling efficiency. The PUE of a data center using CRAC cooling is usually around 2.0 [6].

Liquid cooling technology usually uses coolant (e.g., water) to directly absorb heat from the servers. It can be divided into three categories: *direct liquid cooling* [19], *rack-level liquid cooling* [24] and *submerge cooling* [26]. In *direct liquid cooling*, the microprocessor in a server is directly attached with a cold plate that contains the coolant to absorb heat of the microprocessor, while the other components are still cooled by chilled air flow. *Direct liquid cooling* improves the cooling efficiency by enhancing the heat exchange process. *Rack-level liquid cooling* and *submerge cooling* adopt some other heat exchange devices instead of the cold plate. In this paper, we adopt the direct liquid cooling technology as an example to demonstrate the effectiveness of our solution, due to its low cost.

Free air cooling is a highly efficient cooling approach that uses the cold air outside the data center and saves power by shutting off the chiller system. It is usually utilized within a range of outside temperature and humidity.

Within this range, the outside air can be used for cooling via an air handler fan. The traditional CRAC system is employed by these data centers as the backup cooling system.

4 Power Optimization for a Local Hybrid-cooled Data Center

In this section, we introduce the power models and formulate the total power optimization problem for a local hybrid-cooled data center. We then present how we solve the problem by using computational fluid dynamics (CFD) modeling and optimization techniques.

4.1 Power Models of a Hybrid-cooled Data Center

We use the following models to calculate the server and cooling power consumption in the data center.

- *Server Power Consumption Model*

For a server i, we adopt a widely accepted server power consumption model [11] as:

$$P_i^{server} = W_i \times P_i^{compute} + P_i^{idle} \quad (1)$$

where W_i is the workload handled by server i, in terms of CPU utilization. $P_i^{compute}$ is the maximum computing power when the workload is 100%. P_i^{idle} represents the static idle power consumed by the server. If the workload is 0, the server can be shut down to save power and thus the power consumption is 0. Therefore, the total server power consumption of a data center with N server is

$$P^{server} = \sum_{i=1}^{N} P_i^{server} \quad (2)$$

- *Air Cooling Power Model*

In a hybrid-cooled data center, the components of a liquid-cooled server except for the microprocessor, the rest server components, such as disk and memory, are cooled by the air cooling system, which contributes to the hot air coming out of the server. To characterize this relationship, we assume that in a liquid-cooled server, α percent of the power is consumed by the microprocessor. Therefore, assuming the first M servers among the total N servers are liquid-cooled, we can calculate the total power consumption of all the servers and components that are cooled by the air cooling as:

$$P_{air}^{server} = \sum_{i=M+1}^{N} P_i^{server} + \sum_{i=1}^{M} (1-\alpha) * P_i^{server} \quad (3)$$

The power consumption of the traditional CRAC air cooling system depends on the heat generation (i.e., P_{air}^{server}) and the efficiency of the CRAC system:

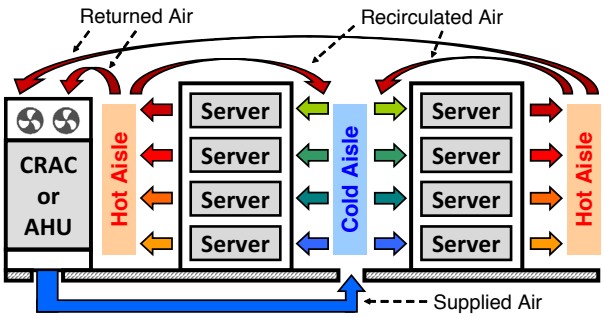


Figure 2: Air circulation in an air-cooled data center. Some hot air can be recirculated to the inlet and mixed with the cold air, degrading the cooling efficiency.

$$P_{CRAC}^{air} = \frac{P_{air}^{server}}{COP^{CRAC}} \quad (4)$$

According to [11], the COP (coefficient of performance, characterizing the cooling efficiency) of a CRAC system can be calculated according to the supplied air temperature T_{sup}:

$$COP_{air} = 0.0068 * T_{sup}^2 + 0.0008 * T_{sup} + 0.458 \quad (5)$$

To avoid overheating the servers, the inlet air temperature of an air-cooled server needs to be bounded by a threshold. Based on a temperature model from [21], we use Equation 6 to first calculate the outlet air temperature, and then get the inlet air temperature with Equation 7:

$$K_i T_{out}^i = \sum_{j=1}^{N} h_{ij} K_j T_{out}^j + (K_i - \sum_{j=1}^{N} h_{ij} K_j) T_{sup} + P_i^{air} \quad (6)$$

$$T_{in}^i = \sum_{j=1}^{N} h_{ji} * (T_{out}^j - T_{sup}) + T_{sup} \quad (7)$$

K_i represents a multiplicative item, $\rho f_i C_p$, where C_p is the specific heat of air. ρ represents the air density, and f_i is the air flow rate to server i. h describes the air recirculation. In Equation 6, the first term characterizes the impact of the air recirculation from server j to server i and the second term models the cooling effect of the supplied air. The third term is the power consumption of server i that heats up the passing cold air. Equation 7 shows that inlet server temperature is determined by the supplied air temperature and the recirculation heat. We explain how to derive h using CFD in Section 4.3.

When the free air cooling method is chosen for the hybrid-cooled data center, the air cooling power is calculated in a different way, according to [7]

$$P_{free}^{air} = (PUE_{free} - 1) * P_{air}^{server} \quad (8)$$

In our experiment, the free cooling PUE is modeled to be proportional to the ambient air temperature according

to [10]. This is because when the outside air temperature is relatively high, more air is needed to take away the heat generated by the servers, and the fan speed of AHU needs to be higher to draw more air.

In our paper, we assume that only one of the two air cooling systems can run at one time in a hybrid-cooled data center. Thus the total air cooling power consumption can be expressed as:

$$P^{air} = \beta P^{air}_{CRAC} + (1-\beta) P^{air}_{free} \quad (9)$$

where β is a binary variable indicating which air cooling system is activated.

- *Liquid Cooling Power Model*

With M liquid-cooled servers and the microprocessor consuming α percent of the server power, we have the liquid cooling power consumption as

$$P^{liquid} = \frac{\sum_{i=1}^{M} \alpha P^{server}_i}{COP^{liquid}} \quad (10)$$

where COP^{liquid} is the COP of the chiller used in the liquid cooling system. Due to the high cooling capacity of liquid medium, the changes of the liquid temperature and the flow rate hardly affect the COP value, and thus COP^{liquid} can be viewed as a constant. To derive a COP that can provide cooling guarantee for all the liquid-cooled servers, we run simple experiments with the worst-case setup by putting all servers to 100% utilization, and adjust the chiller set point and flow rate of the cold plate to ensure the microprocessor temperature is below the threshold. We then use the COP gained in this situation as a constant.

4.2 Power Minimization

We now formulate the power minimization problem of the hybrid-cooled data center. N servers are deployed in the data center and M of them are liquid-cooled. Assuming that the total workload is W_{total}, we minimize the total power consumption as:

$$\min\{P^{server} + P^{air} + P^{liquid}\} \quad (11)$$

Subject to:

$$\sum_{i=1}^{N} W_i = W_{total} \quad (12)$$

$$T^{mp}_i < T^{mp}_{th} \quad 1 \le i \le M \quad (13)$$

$$T^{in}_i < T^{in}_{th} \quad M+1 \le i \le N \quad (14)$$

Equation 12 guarantees that all the workload W_{total} is handled by the servers. Equation 13 enforces that the microprocessors' temperatures of these M liquid-cooled servers are below the required threshold T^{mp}_{th}. Equation 14 enforces that the inlet temperatures of the $(N-M)$ air-cooled servers are below the required threshold T^{in}_{th}.

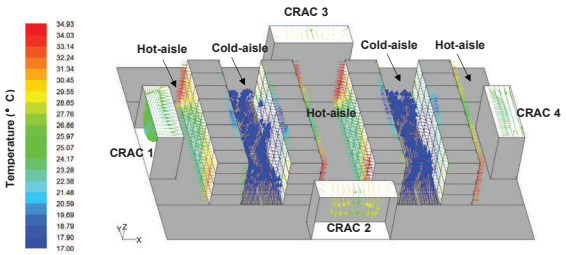


Figure 3: Data center model used in evaluation

4.3 CFD-based Recirculation Matrix and Optimization Solution

We now explain how to get the CFD-based recirculation matrix H. In Equation 6, h_{ij} is an element of the matrix H, indicating the percentage of heat flow recirculated from server i to server j. To simulate the thermal environment of the data center, we use Fluent [5], which is a CFD software package. Figure 3 shows both the layout of the data center model used in this paper and an example of the thermal environment when all the servers are air-cooled. We set the CRAC supply temperature in CFD and use it to get the outlet temperature of each server, in different workload distribution scenarios. After getting the power consumption (P^{air}_i), the outlet temperature of each server (T^i_{out},T^j_{out}) and the CRAC supply temperature (T_{sup}) in all the scenarios, we use them to solve the linear equation shown in Equation 6 and get the recirculation matrix H.

To solve the optimization problems, we use LINGO [15], a comprehensive optimization tool. LINGO employs branch-and-cut methods to break a non-linear programming model down into a list of sub problems to enhance the computation efficiency. It is important to note that our scheme performs offline optimization to determine workload distribution, server on/off and the cooling mode of the data center at different outside temperatures. To dynamically determine those configurations, our scheme can conduct the optimization for different loading levels in an offline fashion and then apply the results online based on the current loading and the current outside temperature.

5 Power Optimization for Geo-Distributed Hybrid-cooled Data Centers

Some big IT companies may have multiple data centers around the world. Although the power minimization for a single hybrid-cooled data center is helpful, it might not be efficient enough for geo-distributed data centers. It is because data centers at different locations have different outside temperatures which lead to different cooling efficiencies. Therefore, it is important to manage the geo-

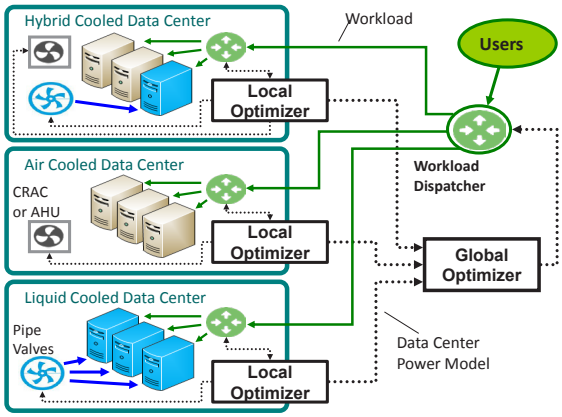


Figure 4: System diagram for geo-distributed data centers. The local optimizer optimizes the cooling configuration and workload distribution locally. The global optimizer optimizes the global workload distribution based on the data center power models.

distributed data centers together. In this section, we extend the total power optimization problem for a single data center to fit geo-distributed data centers. We develop a two-layer light-weight optimization algorithm to lower the computation time.

5.1 Global Optimization

At first, we formulate a global optimization problem for minimizing the total power consumption of geo-distributed data centers, which is very similar to the optimization problem for a single data center. To simplify the notations, we assume that each data center has N servers, which is not a hard requirement for the formulation. We minimize the total power consumption of the data center system that contains K data centers to handle W^{geo}_{total} workload as:

$$\min \sum_{j=1}^{K} P^{DC}_{j} \tag{15}$$

Subject to:

$$\sum_{j=1}^{K}\sum_{i=1}^{N} W_{i,j} = W^{geo}_{total} \tag{16}$$

$$T^{mp}_{i,j} < T^{mp}_{th} \quad 1 \le i \le M \quad 1 \le j \le K \tag{17}$$

$$T^{in}_{i,j} < T^{in}_{th} \quad M+1 \le i \le N \quad 1 \le j \le K \tag{18}$$

P^{DC}_{j} is the total power consumption of data center j. $T^{mp}_{i,j}$ is the microprocessor temperature of liquid-cooled server and $T^{mp}_{i,j}$ represents the inlet temperature of air-cooled server in a data center. $W_{i,j}$ is the workload distributed to server i in data center j. Equation 16 guarantees that all the workload for geo-distributed data centers can be handled. Equation 17 and Equation 18 are the temperature constraints of liquid-cooled and air-cooled servers.

5.2 Two-layer Light-weight Optimization

To solve the global optimization problem for geo-distributed hybrid-cooled data centers, a straightforward solution is to use LINGO directly, as the solution for the single data center power optimization in Section 4. However, as LINGO utilizes the branch-and-bound technique to solve the problem, the computational complexity increases significantly when LINGO solves the problem with geo-distributed data centers. Therefore, we design a two-layer light-weight optimization algorithm to lower the computation complexity.

We first define cPUE for a single data center as

$$cPUE = \frac{P^{server} + P^{cooling}}{P^{compute}(W)} \tag{19}$$

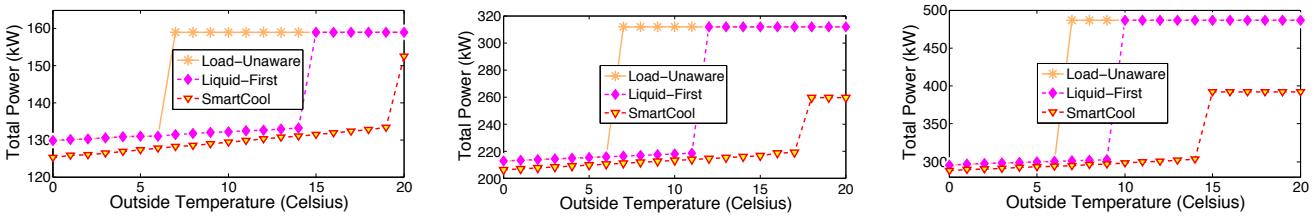
where $P^{compute}$ is the total dynamic computing power consumed by the servers to handle a given workload W. As shown in Figure 4, the local optimizer uses the power optimization process of a single data center (discussed in Section 4) to derive an optimal cPUE for a data center with a given workload W and an outside temperature $T_{outside}$ as $cPUE_{optimal}(T_{outside}, W)$,

$$cPUE_{optimal} = f(T_{outside}, W) \tag{20}$$

In fact, with different amounts of workload and different outside temperatures, $cPUE_{optimal}$ has different values. To get the $cPUE_{optimal}$ model for each data center, we need to obtain a set of sample values of the optimal power consumption with different workloads and outside temperatures. We change the workload from 0% to 100% with a 10% increment step each time, and also change the outside temperature from 0 °C to 20 °C with an increment of 1 °C. We get the power optimization solution of each single data center with different workload and outside temperature combinations. The obtained results are a set of sampled triplets as $(cPUE_{optimal}, W, T_{outside})$. We then use the *Levenberg Marquardt* (LM) algorithm to conduct the linear fitting to find out the $cPUE_{optimal}$ model:

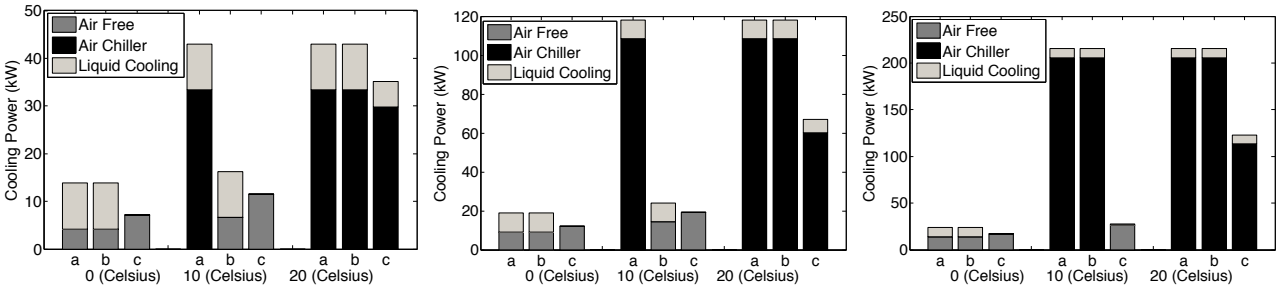
$$cPUE_{optimal} = a * T_{outside} + b * W + c \tag{21}$$

The coefficients *a, b, c* are determined by the data center cooling configuration (e.g., the number of liquid-cooled servers). We choose to do linear model fitting due to the consideration of calculation complexity. Its accuracy is adequate within the acceptable range as we will discuss in Section 6.4. With the $cPUE_{optimal}$ model, we can model the optimal power consumption of a single data center by combining Equations 19 and 21 as:



(a) Total Power at 30% Loading (b) Total Power at 50% Loading (c) Total Power at 70% Loading

Figure 5: Total power consumption with *Load-Unaware*, *Liquid-First* and *SmartCool* at different loadings



(a) Power break at 30% Loading (b) Power break at 50% Loading (c) Power break at 70% Loading

Figure 6: Cooling power breakdown for different schemes (x-axis: a is *Load-Unaware*, b is *Liquid-First*, c is *SmartCool*)

$$P^{DC} = P^{compute}(W) * cPUE_{optimal}(T_{outside}, W) \quad (22)$$

Given a specific moment, the air temperature outside a data center is a constant, and thus P^{DC} only depends on W. As shown in Figure 4, the local optimizer sends the $P^{DC}(W)$ model to the global optimizer, which optimizes the total power consumption by manipulating the workload assigned to each data center. This optimization problem is also solved using LINGO.

Our algorithm successfully decouples the global power optimization problem to a global workload distribution problem and a power optimization problem of each local data center, and thus reduces the optimization overhead significantly.

5.3 Maintaining Response Time

When the workload dispatching is managed by the global optimizer in Figure 4, the request response time needs to be maintained below a threshold. We consider two components of response time: the queuing delay within the data center, and the network delay outside the data center.

A data center can be modeled as a GI/G/m queue [11]. Using the Allen-Cullen approximation for the GI/G/m model, the queuing delay and the number of servers needed to satisfy a given workload demand are related as follows:

$$R = \frac{1}{\mu} + \frac{P_m}{\mu(1-\rho)}\left(\frac{C_A^2 + C_B^2}{2m}\right) \quad (23)$$

where R is the average queuing delay. $\frac{1}{\mu}$ is the average processing time of a request. ρ is the average server utilization. m is the number of servers. $P_m = \rho^{\frac{m+1}{2}}$ for $\rho < 0.7$. $P_m = \frac{\rho^m + \rho}{2}$ for $\rho > 0.7$ and C_A^2 and C_B^2 represent the squared coefficients of the variation of request inter-arrival times and request sizes, respectively. The network delay d_{ij} between the source i and the data center j is taken to be proportional to the geographical distances between them.

When dispatching workload among data centers, we have the constraint that

$$W + d_{ij} < T_{ij} \quad (24)$$

where T_{ij} is the response time threshold of the requests dispatched from source i to data center j.

6 Simulation Results

In this section, we present our evaluation results from the simulation.

6.1 Evaluation Setup

To evaluate different power optimization schemes in a single hybrid-cooled data center, we use a data center model that employs the standard configuration of alternating hot and cold aisles, which is consistent with those used in the previous studies [11]. Figure 3 shows both the data center layout and a thermal environment example when all the servers are air-cooled. The data center consists of four rows of servers, where the first row is composed of liquid-cooled servers. Each row has eight racks, where each rack has 40 servers, adding up to 1,280

servers in the entire data center. The server in our data center model has a 100W idle power consumption and a 300W maximum power consumption when fully utilized. The volumetric flow rate of the intake air of each server is $0.0068 m^3/s$. Each of the four CRAC units in the data center pushes chilled air into a raised floor plenum at a rate of $9000 ft^3/min$ [11]. There also exists a free cooling economizer system that uses outside air when suitable to meet the cooling requirements.

For the liquid-cooled servers we use the rack CDU (coolant distribution unit), in which the CPU of every server is cooled by cold plate and the other components are cooled by the chilled air. We have two chiller systems, of which one is to supply cold water for the cold plates and the other one is to supply the coolant to the CRAC units.

To evaluate different power optimization schemes among different data centers, we consider about three data centers with different cooling configuration, a air-cooled data center (all the four rows of servers are air-cooled), a hybrid-cooled data center (one row of servers are liquid-cooled as discussed in the previous paragraphes) and a liquid-cooled data center (all the servers are liquid-cooled). The only difference between these three data centers are the number of liquid-cooled servers. Other settings of the data centers are the same.

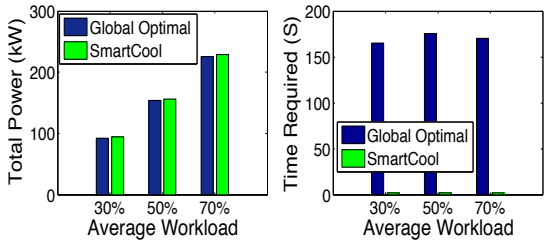
6.2 Comparison of Cooling Schemes

In this section, we compare our *SmartCool* scheme with two baselines: *Load-Unaware* and *Liquid-First*.

Load-Unaware determines the cooling mode by comparing the outside temperature to a fixed temperature threshold, which is equal to the highest CRAC supply temperature that can safely cool the servers when they are all fully utilized. When the outside temperature is below the threshold free air cooling is used, otherwise the traditional air cooling system with chillers and pumps is selected. *Load-Unaware* prefers to distribute the workload to the liquid-cooled servers. If they are fully utilized, the remaining workload is then distributed to the air-cooled servers. The servers in the middle of each row and at the bottom of each rack are prior, as servers located at those places have less recirculation impact and lower inlet temperature [11].

In contrast, *Liquid-First* dynamically adjusts the temperature threshold for free air cooling, based on the real-time workload. It first distributes workload to the servers in the same way as *Load-Unaware*, and then uses the highest CRAC supply temperature that can safely cool the servers as the temperature threshold.

Figure 5 shows the total power consumption of the three different schemes at different loadings with different outside temperature. We can see from the results that



(a) Power Comparison (b) Time Comparison

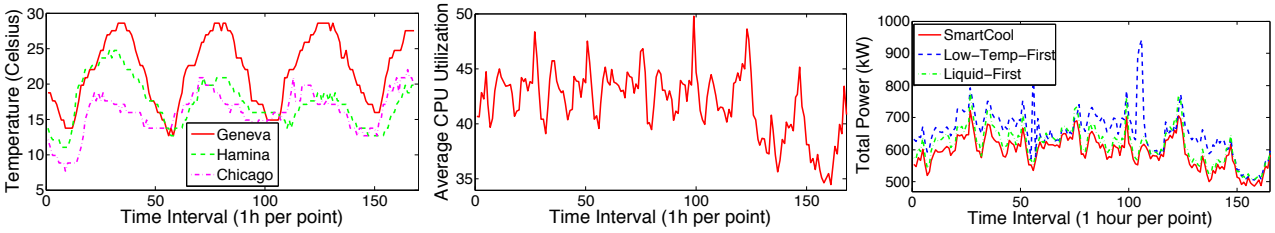
Figure 7: Power consumption and required time comparison between *Global Optimal* and *SmartCool*

all the three cooling schemes achieve a low power consumption when the outside temperature is low, because all of them can use free air cooling. Compared with *Load-Unaware* and *Liquid-First*, *SmartCool* shows the lowest power consumption because it considers the heat recirculation among air-cooled servers when distributing workload.

When the outside temperature increases, we can see that *Load-Unaware* is the first to have a jump in the power consumption curve among the three schemes. This is because *Load-Unaware* uses a fixed temperature threshold to decide whether to use free air cooling or not. The temperature threshold of *Load-Unaware* is determined with the data center running a 100% percent workload. Therefore, it is unnecessarily low for less workload such as 30%. *Liquid-First* is the second one to have a power consumption jump due to switching from free air cooling to CRAC cooling. It can use free air cooling more than *Load-Unaware* when the outside temperature is higher, because its temperature threshold is determined based on the real-time workload (e.g., 30%, 50% or 70%) rather than the maximum workload (100%). Hence *Liquid-First* saves power compared with *Load-Unaware*. *SmartCool* scheme is the last one to have the power consumption jump, because *SmartCool* optimizes the workload distribution among liquid-cooled and air-cooled servers, while the two baselines concentrate the workload on a small number of air-cooled servers and result in some hot spots when air cooling is necessary. Those hot spots require a lower temperature of the supplied air for cooling and thus increase the power consumption. Therefore, *SmartCool* is the most power efficient scheme.

6.3 Power Breakdown of Cooling Schemes

In Figure 6, we break the cooling power consumption of the three schemes (including *Load-Unaware*,*Liquid-First* and *SmartCool*) into *Air Free* (free cooling power consumption), *Air Chiller* (traditional air cooling power consumption) and *Liquid Cooling* (liquid cooling power



(a) Temperature Trace (b) IBM trace file (c) Total Power Consumption Comparison

Figure 8: Total Power Consumption comparison of different workload dispatching schemes

consumption). We choose three outside temperature points, 0 °C, 10 °C and 20 °C to discuss the impact of outside temperature.

Figure 6 (a) shows the cooling power under different outside temperatures at 30% loading. We can see that when the outside temperature is 0 (Celsius), *Load-Unaware* and *Liquid-First* consume the same amount of cooling power. *SmartCool* consumes less cooling power than the two baselines because it distributes all the workload to the air-cooled servers, which are cooled by the more efficient free air, while the baselines prefer to use liquid-cooled servers.

When the outside temperature raises to 10 °C, *Load-Unaware* switches to the traditional air cooling mode, which starts to use the chiller system and leads to the increase of the cooling power consumption. Differently, the other two schemes show similar results as those at 0 °C. When the outside temperature is 20 °C, all the three schemes switch to traditional air cooling mode. However, *SmartCool* still consumes less cooling power because it considers the impact of air recirculation and optimizes the workload distribution.

Figures 6 (b) and 6 (c) show the breakdown of cooling power consumption when the data center is at 50% and 70% loading. They show the same trends as 6 (a) though the total cooling power increases due to the increase of the workload.

6.4 Comparison between SmartCool and Global Optimal Solution

In this section, we evaluate our two-layer power optimization algorithm in the geo-distributed data center settings and compare it with the global optimization scheme in terms of optimization performance, including the optimized total power consumption and the time overhead. The global optimal scheme solves the geo-distributed power optimization problem as a whole including deciding the global and local workload distribution as well as the cooling mode management of each data center. *SmartCool* uses a two-layer optimization algorithm as discussed in Section 5. Due to the long computation time of the global optimization scheme, we use three smaller scale data centers in this set of experiments. Each data center has two rows of racks. Each row contains 4 racks and each rack contains four blocks. There exists 10 servers in each block.

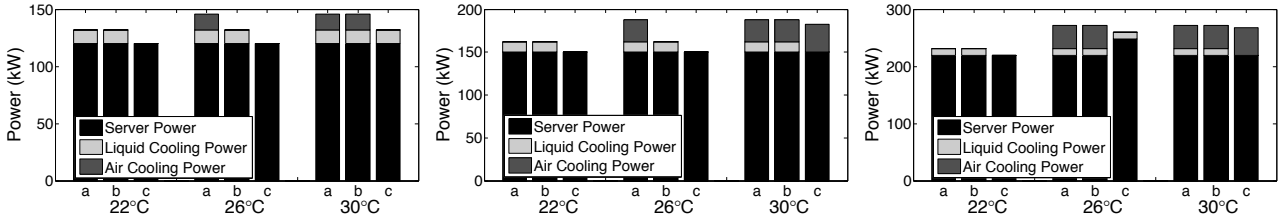
Figure 7(a) shows the total power consumption of the two schemes at different loadings. We can see that *SmartCool* has very close optimization result to the global optimal solution. The performance difference is due to the model fitting error introduced from the cPUE modeling process as discussed in Section 5. However, our *SmartCool* consumes much less time than the *Global Optimal* solution according to Figure 7(b).

6.5 Results with Real Workload and Temperature Traces

We now evaluate different power management schemes on three geo-distributed data centers with real workload and temperature traces. Each data center contains 1,280 servers. To show the diversity of different data centers, we configure them with different cooling system, which are air cooling, liquid cooling and hybrid cooling, respectively. The outside temperatures are shown in Figure 8 (a) which are one week temperature traces of three different locations, Geneva, Hamina and Chicago [25].

We compare the total power consumption under three workload dispatching schemes: *Liquid-First*, *Low-Temp-First* and *SmartCool*. *Liquid-First* dispatches workload to the three data centers according to their cooling efficiencies, which are ranked from high to low as the liquid-cooled data center, the hybrid-cooled data center and the air- cooled data center. Thus workload is first dispatched to the liquid-cooled data center and if it can not handle all the workload, the rest part is dispatched to the hybrid-cooled data center and then the air-cooled data center. For *Low-Temp-First*, workload is first distributed to the data center with the lowest outside temperature since it has the highest possibility to use free cooling system which will consume the least cooling power. For *SmartCool*, workload is distributed according to the approach discussed in Section 5.

Figure 8 (b) shows a one week trace of the average CPU utilization from an IBM production data center [29]. We use this trace to generate the total workload in our experiment. Figure 8 (c) shows the power consumption of the three different schemes. We can see that our *SmartCool* consumes the least power because it considers the impacts of both the outside tempera-



(a) Total power at 30% Loading (b) Total Power at 50% Loading (c) Total Power at 70% Loading

Figure 9: Hardware results under different ambient temperatures at different loadings (x-axis: a is *Load-Unaware*, b is *Liquid-First*, c is *SmartCool*)

ture and the workload on the data center PUE. *Liquid First* consumes more power than *SmartCool* but less than the *Low-Temp-First* solution, because it first concentrates workload on liquid-cooled data center, which has relatively high cooling efficiency. *Low-Temp-First* consumes the most power among the three schemes because when the outside temperature is not low enough or the workload is relatively high, *Low-Temp-First* first dispatches all the workload to the data center with the lowest outside temperature and then traditional air cooling system with chiller and pump must be used to cool down the servers, which will cause high cooling power.

7 Hardware Experiment

In addition to the simulations, we also conduct experiments on our hardware testbed to evaluate *SmartCool* by comparing it with the baselines, i.e., *Load-Unaware* and *Liquid-First*. The testbed includes one liquid-cooled server and three air-cooled servers. A heater is used to set the ambient temperature to be 22°C, 26°C and 30°C.

To compare the total power consumptions of the three schemes, we use power meters to measure the power consumed by the servers and the cold plate used for liquid cooling. For the reason that we do not have an air handler and just use the ambient air to take away heat generated by the server, we assume that the air cooling power is zero under free cooling mode and use Equation 4 and 5 to estimate the power consumption of traditional air cooling.

Figure 9 shows the power consumption of different schemes with different ambient temperatures and loadings. We can see that when the ambient temperature is 22°C, the air cooling power of all the three schemes are zero at different loadings, because they can all adopt free air cooling. *SmartCool* consumes less cooling power than *Load-Unaware* and *Liquid-First*, for it does not consume liquid cooling power when the ambient temperature is at 22°C as all the workload is distributed to the air-cooled servers. In contrast, the two baselines prefer to distribute workload to the liquid-cooled servers, no matter how cold the ambient air is. When the ambient temperature is 22°C and the workload is at 30% or 50%, *Load-Unaware* consumes the most cooling power, because it begins to use traditional air cooling since the ambient temperature exceeds its fixed temperature threshold for cooling mode decision, and thus cause more cooling power. *Liquid-First* still uses free cooling at 30% and 50% loadings, and begins to use traditional air cooling when the workload is 70%. At 26°C *SmartCool* still consumes less cooling power than the other two schemes. The results show the same trend when the outside temperature is 30°C.

8 Conclusion

In this paper, we have presented SmartCool, a power optimization scheme that effectively coordinates different cooling techniques and dynamically manages workload allocation for jointly optimized cooling and server power. In sharp contrast to the existing work that addresses different cooling techniques in an isolated manner, SmartCool systematically formulates the integration of different cooling systems as a constrained optimization problem. Furthermore, since geo-distributed data centers have different ambient temperatures, SmartCool dynamically dispatches the incoming requests among a network of data centers with heterogeneous cooling systems to best leverage the high efficiency of free cooling. A light-weight heuristic algorithm is proposed to achieve a near-optimal solution with a low time overhead. SmartCool has been evaluated both in simulation and on a hardware testbed with real-world workload and temperature traces. The results show that SmartCool outperforms two state-of-the-art baselines by having a 38% more power savings.

References

[1] A.COSKUN, J.L.AYALA, D.ATIENZA, AND T.S.ROSING. Modeling and dynamic management of 3d multicore systems with liquid cooling. In *Proceedings of International Conference on Very Large Scale Integration (VLSI-SoC)* (2009).

[2] A.GREENBERG, J.HAMILTON, A.MALTZ, AND P.PATEL. The cost of a cloud: research problems in data center networks. In *ACM SIGCOMM CCR* (2009).

[3] B.GEBREHIWOT, K.AURANGABADKAR, N.KANNAN, AND D.AGONAFER. CFD analysis of free cooling of modular data centers. In *Proceedings of Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)* (2012).

[4] CERN . Data Centre. http://home.web.cern.ch/about/computing.

[5] COMPUTATIONAL FLUID DYNAMICS (CFD) SOFTWARE BY ANSYS INC. . Fluent. http://www.caeai.com/cfd-software.php.

[6] D.CHEMICOFF. The uptime institute 2012 data center survey. http://symposium.uptimeinstitute.com.

[7] D.C.SUJATHA, AND S.ABIMANNAN. Energy efficient free cooling system for data centers. In *Proceedings of IEEE Third International Conference on Could Computing Technology and Science (CloudCom)* (2011).

[8] D.HWANG, V.P.MANNO, M.HODES, AND G.J.CHAN. Energy savings achievable through liquid cooling: A rack level case study. In *Proceedings of Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)* (2010).

[9] D.SUJATHA, AND S.ABIMANNAN. Energy efficient free cooling system for data centers. In *Proceedings of International Conference on Cloud Computing Technology and Science (CloudCom)* (2011).

[10] EMERSON. Liebert DSE precision cooling system sales brochure . http://www.emersonnetworkpower.com.

[11] F.AHMAD, AND T.N.VIJAYKUMAR. Joint optimization of idle and cooling power in data centers while maintaining response time. In *Proceedings of International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)* (2010).

[12] I.GOIRI, W.KATSAK, K.LE, T.NGUYEN, AND R.BIANCHINI. Parasol and greenswitch: Managing datacenters powered by renewable energy. In *Proceedings of International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)* (2013).

[13] K.ZHENG, X.WANG, L.LI, AND X.WANG. Joint power optimization of data center network and servers with correlation analysis. In *Proceedings of the 33rd IEEE International Conference on Computer Communications (INFOCOM)* (2014).

[14] L.A.BARROSO, J.CLIDARAS, AND U.HOLZLE. The datacenter as a computer: An introduction to the design of warehouse-scale machines. In *Morgan and Claypool Publishers* (2009).

[15] LINDO SYSTEMS INC. LINDO software products. http://www.lindo.com.

[16] M.A.ADNAN, R.SUGIHARA, AND R.GUPTA. Energy efficient geographical load balancing via dynamic deferral of workload. In *Proceedings of 5th IEEE conference on cloud computing (CLOUD)* (2012).

[17] M.BROCANELLI, S.LI, X.WANG, AND W.ZHANG. Joint management of data centers and electric vehicles for maximized regulation profits. In *Proceedings of the fourth IEEE International Greeen Computing Conference (IGCC)* (2013).

[18] M.IYENGAR, M.DAVID, P.PARIDA, AND V.KAMATH. Server liquid cooling with chiller-less data center design to enable significant energy savings. In *Proceedings of Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)* (2012).

[19] M.IYENGAR, M.DAVID, P.PARIDA, AND V.KAMATH. Server liquid cooling with chiller-less data center design to enable significant energy savings. In *Proceedings of Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)* (2012).

[20] Q.TANG, S.K.S.GUPTA, AND G.VARSAMOPOULOS. Thermal-aware task scheduling for data centers through minimizing heat recirculation. In *Proceedings of IEEE International Conference on Cluster Computing* (2007).

[21] Q.TANG, T.MUKHERJEE, S.GUPTA, AND P.CAYTON. Sensor-based fast thermal evaluation model for energy efficient high-performance data centers. In *Proceedings of International Conference on Intelligent Systems and Image Processing (ICISIP)* (2006).

[22] R.ANTON, H.JONSSON, AND B.PALM. Modeling of air conditioning systems for cooling of data centers. In *Proceedings of Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM)* (2002).

[23] R.ZHOU, Z.WANG, C.E.BASH, AND A.MCREYNOLDS. Data center cooling management and analysis-a model based approach. In *Proceedings of Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)* (2012).

[24] S.MOVOTNY. Data center rack level cooling utilizing water-cooled, passive rear door heat exchangers as a cost effective alternative to crash air cooling.

[25] TIMEANDDATE. Weather around the World. www.timeanddate.com.

[26] TREEHUGGER INC. . Data Center Cooling Energy Reduction Thanks to Fluid Submerged Servers. http://www.treehugger.com.

[27] W.HUANG, M.ALLEN-WARE, J.B.CARTER, AND E.ELNOZAHY. TAPO: Thermal-aware power optimization techniques for servers and data centers. In *Proceedings of IEEE International Greeen Computing Conference (IGCC)* (2011).

[28] W.ZHENG, K.MA, AND X.WANG. Exploiting thermal energy storage to reduce data center capital and operating expenses. In *Proceedings of the 20th International Symposium on High Performance Computer Architecture (HPCA)* (2014).

[29] X.WANG, M.CHEN, C.LEFURGY, AND T.W.KELLER. SHIP: Scalable hierarchical power control for large-scale data centers. In *Proceedings of International Conference on Parallel and Distributed Systems (PACT)* (2009).

[30] Y.CHEN, D.GMACH, C.HYSER, Z.WANG, C.BASH, C.HOOVER, AND S.SINGHAL. Integrated management of application performance, power and cooling in data centers. In *Proceedings of Network Operations and Management Symposium (NOMS)* (2010).

[31] Y.ZHANG, Y.WANG, AND X.WANG. Electricity bill capping for cloud-scale data centers that impact the power markets. In *Proceedings of the International Conference on Parallel Processing* (2012).

[32] Y.ZHANG, Y.WANG, AND X.WANG. TEStore: Exploiting thermal and energy storage to cut the electricity bill for datacenter cooling. In *Proceedings of the 8th International Conference on Network and Service Management (CNSM)* (2012).

[33] Z.LIU, M.LIN, AND L.ANDREW. Greening geographical load balancing. In *Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems* (2011).

[34] Z.LIU, Y.CHEN, C.BASH, A.WIERMAN, D.GMACH, Z.WANG, M.MARWAH, AND C.HYSER. Renewable and cooling aware workload management for sustainable data centers. In *Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems* (2012).

[35] Z.WANG, C.BASH, C.HOOVER, AND A.MCREYNOLDS. Integrated management of cooling resources in air-cooled data centers. In *Proceedings of IEEE Conference on Automation Science and Engineering (CASE)* (2010).