A Unified Approach to Coordinated Energy-Management in Data Centers

Rajarshi Das, Srinivas Yarlanki, Hendrik Hamann, Jeffrey O. Kephart, Vanessa Lopez
IBM T.J. Watson Research Center,
19 Skyline Drive,
Hawthorne, NY 10532, USA

***Abstract*— Energy consumption has become a critical issue for data centers, triggered by the rise in energy costs, volatility in the supply and demand of energy and the widespread proliferation of power-hungry information technology (IT) equipment. In response, researchers are developing energy-efficient data centers by incorporating energy-aware systems both at the IT level (shutting down servers or temperature-aware workload placement) as well as at the facilities level (shutting down air-conditioning units, increasing temperature of air supplied to the plenum). In this work, we explore a novel approach to coordinated-management of IT systems *and* its cooling infrastructure to joint power and temperature objectives. In particular, for a given total IT workload, we present a method to determine the optimal settings of computer room air-conditioning units (CRACs) in a data center so as to minimize overall energy consumption in the data center while satisfying specified temperature constraints. Using potential fluid-flow theory, our approach identifies distinct thermal zones associated with each CRAC and then provisions cooling power to match the heat generated from the racks in that zone. We illustrate the resulting range in behavior and potential for energy savings in a large 10,000 sqft commercial data center with 10 CRACs and 186 racks.**

Keywords-Data center, energy management, self-managing systems, utility functions, thermal model, workload provisioning.

I. Introduction

The sharp rise in energy usage in data centers, fueled by increased IT workload and high server density, and coupled with a concomitant increase in the cost and volatility of the energy supply, have triggered urgent calls to improve data center energy efficiency. Energy management in data centers lies very naturally within the scope of self-managing computing because data centers encompass both the large, complex, difficult-to-manage IT environment and the analogously complex physical infrastructure that supports that IT environment [1]. In data centers, myriad interacting physical components such as power distribution units, power supplies, water cooling units, and air conditioning units interact not just with one another, but also with the software components in IT systems, resulting in a management problem that is both qualitatively similar to and quantitatively harder than that of managing IT alone.

A powerful, principled and practical approach to self-management that has been advocated entails defining high-level objectives in terms of utility functions, and then using a combination of modeling, optimization and learning techniques to set the values of system control parameters so as to maximize the utility. Authors have used utility functions for diverse autonomic computing applications, including negotiation among autonomic performance managers to resolve conflicting resource demands, and managing power-performance tradeoffs in servers [2].

In this paper we demonstrate that utility functions can be applied fruitfully on a larger scale, to the data center facility as a whole. Administrators who operate at this scale tend not to be concerned with application-level issues such as performance, availability, or security. They are more concerned with issues such as energy utilization, temperature, hardware lifetime, and (at the bottom line) cost. Accordingly, we formulate simple, plausible utility functions that express a tradeoff between energy and temperature considerations, and then show how to combine modeling with optimization to find a setting of data center control parameters to maximize that utility. In our case, we aim to manage the spatial provisioning of cooling power in the data center by controlling the fan speeds and on/off state of individual computer room air conditioning units (CRAC) units and the spatial provisioning of heat distribution in the data center by controlling the spatial arrangement of IT workload placement.

Intelligent automated approaches to data center cooling were pioneered by researchers at Hewlett Packard [6], who identified common patterns of data center cooling inefficiencies. Other researchers like Moore et al. [7] have pointed out that workload placement could be used as an additional control parameter for cooling data centers. In previous work Hamann et al. [3-4] have investigated energy-balance and computational fluid dynamics (CFD) models to study the provisioning of cooling power in data centers. Das et al. [5] have employed energy-balance models to investigate the applicability of utility functions to optimize the provisioning of cooling power in a small data center. In this paper, we build on such previous work and attempt for the first time to combine cooling power provisioning along with IT workload provisioning at a more granular level in a large data center by using the concept of CRAC thermal zones.

II. CRAC Thermal Zones

A thermal zone is the space cooled by a CRAC which blows air into the data center plenum. To determine the thermal zone associated with a CRAC, we use potential flow theory to calculate a two-dimensional velocity field in the plenum. The potential, ϕ, satisfies the Poisson equation

$$\nabla^2 \phi = f \qquad \text{Eqn. 1}$$

where f is the source or sink of the potential. The gradient of potential gives the velocity field $\vec{v}$:

$$\nabla \phi = \vec{v} \quad \text{Eqn. 2}$$

In this model, CRACs are treated as sources while perforated tiles in the raised floor of the data center are treated as sinks. The strengths of sources are obtained in real time from flow-sensors mounted on every CRAC. The strengths of the sinks are calculated by correlating the pressure drop across a tile to the flow resistance offered by the tile. The pressure drop across a perforated tile is obtained in real time by pressure sensors placed in the plenum. Once the velocity field is obtained thermal zones are calculated using the velocity field based on the ray tracing approach [3-4]. The thermal zones for an arbitrary setting of CRACs are constructed by exploiting the linearity of the Poisson equation using the superposition principle. This is done by first constructing a set of velocity fields for each CRAC in the data center assuming it to be the only active CRAC. Then the velocity field for any arbitrary combination of CRAC settings can be obtained by superposition of the velocity fields from this set. Once a velocity field for a desired CRAC setting is obtained thermal zones for that setting are obtained by an algorithm which calculates the trajectory of the air flow using:

$$\vec{x} = \int \vec{v} dt \quad \text{Eqn. 3}$$

A trajectory is calculated until it either intersects with a CRAC or with a previously assigned point.

III. Energy Balance In Thermal Zones

Energy balance for a thermal zone under steady state conditions requires that the total power dissipated in the thermal zone above the raised floor, P^d_{RF}, should be equal to the total cooling power, P_{Cool}. The total power dissipated in a thermal zone above raised floor, P^d_{RF}, can be found by adding the power dissipated by summing up the power dissipated by the racks (= P^d_{IT}), CRACs (= P^d_{CRAC}), PDUs (= P^d_{PDU}) and other assets (= P^d_{Misc}) which lie in that zone:

$$P^d_{RF} = P^d_{CRAC} + P^d_{PDU} + P^d_{IT} + P^d_{Misc} \quad \text{Eqn. 4}$$

The cooling power for a thermal zone is calculated from the fluid volumetric-flow rate, φ, air density, specific heat of the air and the difference between the air inlet and exit temperatures to the thermal zone, ΔT.

$$P_{Cool} = \sum P_{i,Cool} = \sum \rho\varphi_i C_p \Delta T \quad \text{Eqn. 5}$$

Here the summation is performed over all perforated tiles in that thermal zone. Energy balance gives:

$$P_{Cool} = P^d_{RF} \quad \text{Eqn. 6}$$

This energy balance is used to determine the temperature difference across the thermal zone, ΔT. The inlet temperature to the thermal zone is obtained using real time temperature sensors mounted on the corresponding CRACs (alternatively we can also use temperature sensors located in the plenum). Hence ΔT can be used to obtain the return temperature to the CRAC for the thermal zone.

To determine the heat dissipated in the raised-floor area in a thermal zone, we mapped each rack to its nearest PDU based on Manhattan distance. Given this mapping, and (for simplicity) assuming that all racks are identical, the power dissipated by a given rack can be estimated by the total power distributed by the PDU and the number of racks mapped to it.

IV. Data Center Management With Utility Functions

In this work, we make the simplifying assumption that the issues of concern for data center management can be reduced to temperature and energy consumption. Such a simplification is sensible, for the following reasons. First, we can eliminate either cost or energy because the two are related by a multiplicative constant: the cost per kWh. Second, a key reason why temperature is of concern is because excessively high temperatures significantly reduce equipment lifetimes and endanger people. Thus temperature constraints can serve as a proxy for ensuring acceptable equipment lifetime.

Here we seek a utility function U(E, T), a scalar function of E, the energy consumed during a specified time interval, and T, a temperature vector that represents a set of temperatures that are either measured directly by sensors or inferred from sensor readings. In general, the dependence of utility upon energy could be complex, especially when the energy supplier uses a nonlinear price schedule that includes either volume discounts or energy savings incentives. In this work we avoid these complexities by assuming that the utility is linear in the energy consumption, which is consistent with a flat price per kWh.

Now suppose that the administrator wishes to minimize energy consumption subject to the constraint that a set of temperatures of interest, T, is kept within acceptable bounds. In practice, data center operators do not take the trouble to look up safe temperature ranges for every piece of equipment in the data center. Instead, the common practice is to adhere to guidelines for safe operating limits published by The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). In the 2008 publication, a maximum temperature threshold of T_{max} of 80.6F (27C) was recommended for all IT equipment.

A utility function which captures the above objectives is a product of the energy utility U(E), and a dimensionless temperature utility $U_T(T)$ defined over a vector of temperatures T :

$$U(E,T) = U_E * U_T(T) \quad \text{Eqn. 7}$$

with

$$U_E(E) = C\,(E_0 - E) \quad \text{Eqn. 8}$$

For purposes of this exposition, we replace the vector of temperatures T with the maximum of the estimated CRAC return temperature among all the CRACs zones and allow the temperature utility to be expressed through a soft-threshold function as follows:

$$U_T(T) = 1/(1 + e^{-(T_{max} - \max(T))}) \qquad \text{Eqn. 9}$$

Here, C is the annual cost of cooling and its units are \$/KW/year, E_0 is the baseline energy consumption defined as the annual energy consumption when no energy-saving measures are taken [5].

V. Methodology For Unified Approach to Coordinated Energy Management in Data Centers

The overall utility in a data center can be maximized by provisioning the right amount of cooling power to match the heat dissipated in any thermal zone. Provisioning cooling power to various sections of the data center is achieved by turning individual CRACs either on or off. In a data center with N CRACs, 2^N combinations of CRAC settings are possible. Note that some of these combinations may be infeasible because if too many CRACs are turned off there may not enough cooling capacity left among the remaining CRACs to maintain the average return temperature over all CRACs below the desired threshold. For example, if energy balance dictates that out of N CRACs only 3 can be turned off, then the number of feasible CRAC settings is $N_{F,3} = C(N,0) + C(N,1) + C(N,2) + C(N,3)$.

Two different scenarios of IT workload provisioning were studied: (1) Real time IT workload distribution before any energy-saving measures are taken; (2) Distributing IT workload across all active thermal zones. Since all CRACs in the test data center have identical capacity distributing the IT workload equally amongst all the active thermal zones (corresponding to active CRACs) ensures maximum CRAC utilization.

For a given scenario of IT workload provisioning the following was repeated for all feasible combinations of CRAC settings: (1) Velocity field and thermal zones were determined using the superposition principle; (2) IT workload in each thermal zone was determined using rack to PDU mapping; (3) Outlet temperature for each thermal zone was determined using energy balance; (4) Utility function for this configuration was calculated. Of all the feasible combinations of CRAC settings the one with the maximum value of the utility function was selected as optimal configuration.

VI. Case Study In A Commercial Data Center

As a detailed case study, we applied our methodology in a 10,360 sq. ft commercial DC. The layout of this data center (Figure 1) with 2x2 ft floor-tiles (in blue-outline) shows that the DC has N = 10 CRACs (blue rectangles), 143 perforated tiles (hashed-squares), 186 racks (grey rectangles), 7 PDUs (red rectangles), and an assortment of network devices (purple rectangles) and other furniture (orange rectangles). The DC is instrumented with pressure, temperature and flow sensors to give real time data.

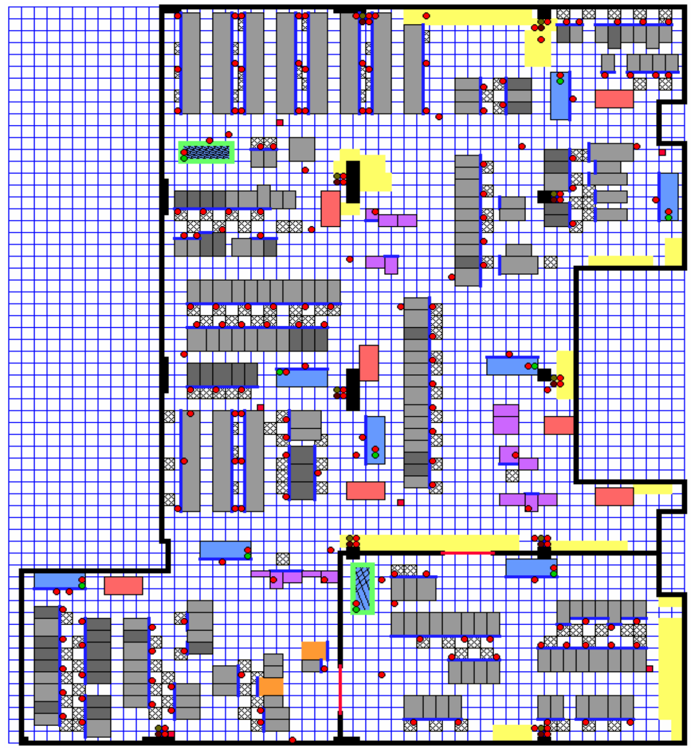

Figure 1. Layout of the data center

All CRACs in this DC are identical and each CRAC is rated for a maximum flow rate φ = 12500 cfm with corresponding energy consumption of P^d_{CRAC} = 8.2 kW (including measured 0.1 CRAC power-dissipation factor). The total IT power P^d_{IT} was measured to be 583 kW and assuming a power-dissipation factor of 0.05 for PDUs ($\sum P^d_{PDU}$ = 29kW), the total power dissipated in the raised floor of the DC summed over all thermal zones, $\sum_{zones} P^d_{RF}$ = 695 kW when all 10 CRACs are turned on.

The actual measured temperature differential between the return and supply air temperatures of the 10 CRACs was 16.5 F. Employing global energy balance over all thermal zones $\sum_{zones} P_{Cool} = \sum_{zones} P^d_{RF}$ under the above assumptions, our estimate of the mean difference between return and supply air temperatures for the CRACs was found to be 17.4 F; an estimation error of less than 6%. Figure 2 shows the thermal zones obtained by using potential flow model by starting with a configuration when 2 CRACs are on (top left) and turning on one CRAC at a time to end at a scenario when all 10 CRACs are on (bottom-right). In each figure, the black rectangles show the location of the active CRACs.

Given the spatial distribution of P_{IT} in the data center, we first explored if it is possible to improve upon the existing data center configuration (i.e., all N = 10 CRACs are turned on). We found that to maintain the average CRAC return temperature below T_{max}, only four CRACs could be turned off. Based on optimization over all feasible configurations of

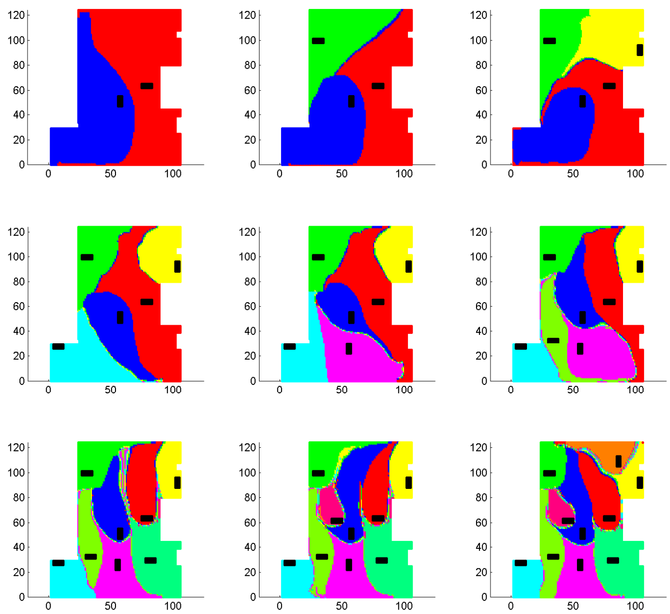

Figure 2. Thermal Zones for 9 CRAC configurations.

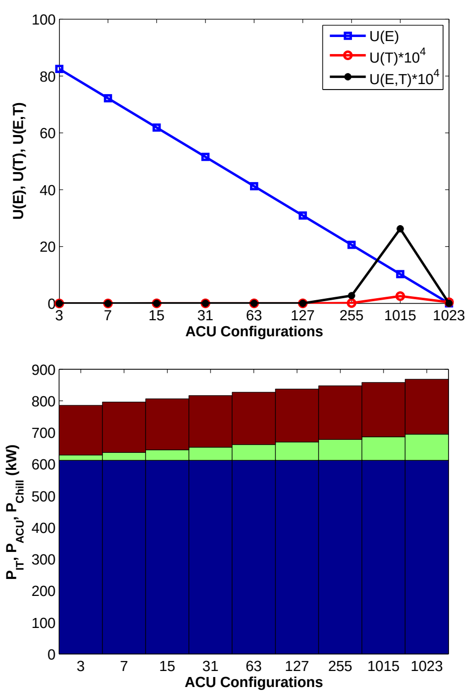

Figure 3(a). Utility analysis for 10 CRAC configurations, (b) Power dissipation vs. CRAC configurations : chiller (brown), CRACs (green) and IT (blue)

thermal zones from $N_{F,4}$ = C(10,0) + C(10,1) + C(10,2)+ C(10,3) + C(10,4) CRAC settings, we found a configuration with 9 CRACs turned on that had the highest utility. Figure 3 shows the power consumed in the data center at the chiller (brown), CRACs (green) and IT (blue) for the same 9 configurations of the CRACs as shown in Figure 2.

Power saved (compared to E_0) in the optimal CRAC configuration is 10.3kW, while the maximum return temperature is 88.9F. Since this is significantly above T_{max} = 80.6F, the temperature utility, and hence the overall U(E,T) is very small. Table 1 shows the ΔT and cooling power at each of the CRACs the under optimal CRAC settings. Here the maximum ΔT = 24.9 F and the maximum P_{Cool} = 98.2 kW.

We applied our approach to maximize utility to three additional scenarios: (2) Same total IT power distributed equally over active zones; (3) Consolidated IT power distributed equally over active zones; and (4) VFD in CRACs for total consolidated IT power distributed equally over all zones.

In scenarios (3) and (4) we assumed that we could reduce the IT power by consolidating the IT workload from an estimated 10% CPU utilization across all servers to a smaller number of servers running at 90% CPU utilization with the remaining servers turned off. Table 1 summarizes the power savings and maximum ΔT for the four different scenarios we considered in our case study.

Scenario	Max ΔT (F)	Power (kW) Consumption	Power (kW) Saved
1. Actual IT power distribution	24.9	98.2	10.3
2. IT power equalized over zones	19.1	76.2	10.3
3. Consolidated IT power	12.1	48.5	554
4. VFD in CRACs with consolidated IT power	10.8	21.6	598

Table 1. Summary of power savings for different scenarios

VII. CONCLUSIONS

We have presented a method for determining the optimal CRAC settings which attempt to provision the right amount of cooling power to match the IT load in a data center. We also considered consolidating IT load and determined the optimal CRAC settings for consolidated IT power.

For the data center used in our case study, our method yields energy saving of about 10%, for unconsolidated IT power while it gives energy savings of about 50% for consolidated IT power. Since we used an enumeration scheme bounded over all feasible combinations of CRAC settings, the approach becomes computationally expensive for large data centers with many CRACs. In future work we explore machine learning techniques to speed up the search process for larger data centers. The thermal zones were based on 2D potential flow in this work which assumes that cross-flow in the data center is not significant. While this seems to be a plausible assumption we aim to derive thermal zones from 3D flow fields in future and determine the error in 2D based thermal zoning.

REFERENCES

[1] J. O. Kephart and D. M. Chess. The vision of autonomic computing. Computer, 36(1):41–52, 2003.

[2] J. O. Kephart and R. Das. Achieving self-management via utility functions. IEEE Internet Computing, 11:40–48, 2007.

[3] H. F. Hamann, V. López, and A. Stepanchuk, *Thermal zones for more efficient data center energy management*, 12-th Intersociety Conference on Thermomechanical Phenomena in Electronic Systems (ITherm), Las Vegas, Nevada, June 2010.

[4] H. F. Hamann, T. G. van Kessel, M. Iyengar, J.-Y. Chung, W. Hirt, M. A. Schappert, A. Claassen, J. M. Cook, W. Min, Y. Amemiya, V. López, J. A. Lacey, and M. O'Boyle, *Uncovering energy efficiency opportunities in data centers*, IBM J.Res. & Dev. Vol.53, No.3, 10:1-10:12 (2009).

[5] R. Das, J. O. Kephart, J. Lenchner, H. Hamann, *Utility-function driven energy efficient cooling in data centers*, International Conference on Autonomic Computing (ICAC), Washington, DC, June 2010.

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

[7] J. Moore, J. Chase, and P. Ranganathan. *Making scheduling "cool": Temperature-aware workload placement in data centers*, In Proc. 2005 USENIX Annual Technical Conference (USENIX '05), 2005.