Sensor-Based Fast Thermal Evaluation Model For Energy Efficient High-Performance Datacenters

Qinghui Tang
Dept. of Electrical Eng.
Arizona State University
Tempe, AZ 85287

Tridib Mukherjee, Sandeep K. S. Gupta
Dept. of Computer Science and Eng.
Arizona State University
Tempe, AZ 85287

Phil Cayton
Intel Corporation
Hillsboro, Oregon

Abstract

In this work, we propose an abstract heat flow model which uses temperature information from onboard and ambient sensors, characterizes hot air recirculation based on these information, and accelerates the thermal evaluation process for high performance datacenters. This is critical to minimize energy costs, optimize computing resources, and maximize computation capability of the datacenters. Given a workload and thermal profile, obtained from various distributed sensors, we predict the resulting temperature distribution in a fast and accurate manner taking into account the recirculation characterization of a datacenter topology. Simulation results confirm our hypothesis that heat recirculation can be characterized as cross interference in our abstract heat flow model. Moreover, fast thermal evaluation based on cross interference can be used in online thermal management to predict temperature distribution in real-time.

1. Introduction

High performance datacenters are being increasingly deployed with high density computing clusters and server farms to improve the services for typical high performance computing applications. Blade severs are being widely installed in modern datacenters due to their high performance/cost ratio and compact size. These large scale datacenters, limited by power and cooling capacity, can cost millions of dollars, mainly due to the high energy costs required by the computing devices and cooling systems. Therefore, it is essential to improve the energy efficiency of datacenter operation, and consequently maximize the utilization rate and computation capability of datacenters.

We are working on a framework for dynamic thermal management of blade server based datacenter to reduce its energy costs and maximize computation capability. The framework is a unique merger between the physical infrastructure and resource management functions of the cluster operating system to take a holistic view of datacenter management. We contend that to achieve these functionalities, an efficient online thermal-aware scheduling algorithm, based on temperature information provided by onboard sensors, is required to make global, power aware decisions, and to dynamically assign tasks to distributed server nodes. One of the principle challenges of the framework during this decision process is to effectively evaluate the thermal distribution and energy efficiency. In this paper, we propose an abstract heat flow model which characterizes the hot air recirculation inside the datacenter, and accelerates the thermal performance evaluation. The results can be used for online real time thermal-aware scheduling, which can dynamically allocate computing task to servers to achieve optimal energy efficiency.

Traditionally, from infrastructure design and planning perspective, CFD (Computational Fluids Dynamics) simulation has been used to evaluate thermal performance of datacenter given a specific configuration of datacenters (layout, supplied air temperature, and power consumption of servers). But the time spent for the CFD simulation are not appropriate for a dynamic datacenter environment whose utilization rate may vary at a smaller granularity of time compared to that of the CFD simulation. In our study, for a small-scale datacenter, it takes an hour to obtain converged CFD simulation results and will take much longer for simulating a larger datacenter. However, the datacenter's utilization rate and load status might have already changed during this time period. Further, when it is required to compare thermal distributions of several different configurations, a CFD simulation will take a relatively long time to finish these evaluations [1] and is not appropriate for online thermal management and decision making.

Our proposed abstract heat flow model, on the other hand, intends to accelerate and simplify this cumbersome process of traditional thermal performance evaluation. As shown in Figure 1, a fast simulation can be used as short-cut to accelerate the thermal performance evaluation process. Our current study shows that it takes less than 1 minute to evaluate and predict the thermal distribution of a given configuration. The abstract heat flow model can be used for 1) online prediction and fast decision making; 2) integrating with thermal-aware scheduler models to evaluate thermal performance of different policies; and 3) filtering out some potential configurations and verify them with CFD simulation.

The rest of the paper is organized as follows. In Section 2 we review background knowledge of datacenter. We present the abstract heat flow model in Section 3. Simulation results are presented in Section 4. Related work is discussed in Section 5. Finally, we conclude in Section 4.

2. Background

A typical datacenter is laid out with a hot-aisle / cold-aisle arrangement by installing the racks and perforated floor tiles in the raised floor. The air conditioners, normally referred

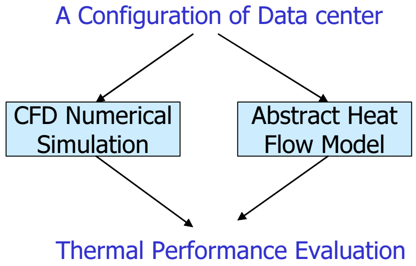
A Configuration of Data center
CFD Numerical Simulation
Abstract Heat Flow Model
Thermal Performance Evaluation

Fig. 1: Abstract heat model can be utilized to obtain fast thermal evaluation instead of using CFD numerical simulation

to as Computer Room Air Conditioner (CRAC) or HVAC (Heating Ventilation Air Conditioner), deliver cold air under the elevated floor. In the sequel, this is referred to as *cooling air*. The cooling air enters the racks from their front side, pick up heat while flowing through these racks, and exits from the rear of the racks. The heated exit air forms hot aisles behind the racks, and is extracted back to the air conditioner intakes, which, in most cases, are positioned above the hot aisles. Each rack consists of several chassis, and each chassis accommodates several computational devices (servers or networking equipment).

Typical cost of operating a datacenter includes energy cost of cooling systems and computing devices, hardware cost for replacing old and dysfunctional computing devices, cost for infrastructure investment and labors. The energy cost of computing devices and air conditioners influences the heat dissipation in the datacenters. The assignment of computation task, the power consumption and the thermo-mechanic properties of different devices, and the cooling capacity of air conditioner will directly or indirectly affect the thermal distribution inside a datacenter. Therefore, *the thermal distribution implicitly correlates with the energy costs of the datacenter. Thus, it is very important to improve thermal distribution performance to enhance the energy efficiency or maximize computation capability of datacenters***.** Improperly designed or operated datacenters may suffer either from overheated servers and potential system failures, or from overcooled system and extra utilities cost.

Due to the complex nature of airflow inside the data center, some of the hot exhaust air from outlets of servers will recirculate into the inlet of other servers. One of the reasons of extra energy cost of cooling system is mainly due to the mixture of cold air with the recirculated hot air in the inlet. To provide acceptable inlet temperature for all servers, the supplied air temperature has to be reduced, increasing the cooling energy cost. Thus, understanding and reducing the hot air recirculation can be explored to reduce the energy costs of datacenters.

Most state-of-art servers used in datacenter provide rich environment information by using onboard sensors, such as humidity sensor or temperature sensor. For example, Dell PowerEdge 1855 blade server used in our study provides inlet temperature, outlet temperature and in-house temperature. The problem for us is how to intelligently process that information to improve thermal management of datacenter.

In our previous work [2], we formalized the total energy cost of a typical datacenter and showed that thermal-aware task scheduling can be utilized to achieve better energy efficiency. Given total incoming tasks C_{Total} of a datacenter, a task assignment $\overrightarrow{\mathbf{C}} = \{C_1, C_2, ...C_N\}$ ($\sum_{j=1}^{n} C_j = C_{Total}$) will result in power consumption distribution vector $\overrightarrow{\mathbf{P}} = \{P_1, P_2, ...P_N\}$, which will lead to inlet/outlet temperature distribution $\overrightarrow{\mathbf{T_{in}}} = \{T^1_{in}, T^2_{in}, ...T^N_{in}\}$ or $\overrightarrow{\mathbf{T_{out}}} = \{T^1_{out}, T^2_{out}, ...T^N_{out}\}$, respectively. The problem was formalized as how to divide the input task set C, into a task vector $\overrightarrow{\mathbf{C}} = \{C_1, C_2, ...C_N\}$ to achieve the minimal total operation energy costs.

In this paper, we are trying to answer the question: ”Given a task scheduling assignment and the thermal profile as obtained from various distributed sensors, can we predict the resulting temperature distribution in a fast and accurate manner taking into account the recirculation characterization of a datacenter topology?”. If we can, then the task scheduling algorithm can be used in online thermal management to achieve better temperature distribution and consequently achieve better energy-efficiency of datacenter.

3. Abstract Heat Flow Model

A datacenter is composed of a set of computing nodes from 1 to n. Those physically separated nodes work individually or cooperatively to accomplish assigned tasks. These nodes are either heterogeneous or homogeneous. There is a scheduler to dispatch incoming tasks C_{Total} (a group of tasks) to individual distributed nodes depending on various scheduling policies, criteria and strategies. Each distributed node i consumes energy at the rate P_i during the execution of task set C_i (a subset of C_{Total}), and the power consumption rate depends on the hardware characteristics of distributed nodes and the task profiles (compute intensive, memory intensive or IO intensive):

$$P_i = G_i(C_i), \tag{1}$$

where G_i is a thermo-mechanic function which depends on the hardware specifications of distributed nodes. G_i can be obtained through experimental measurement or CFD simulation. We characterized the correlation between task and power consumption in our previous work [2].

According to the law of energy conservation and the fact that almost all power drawn by a computing device is dissipated as heat, the relationship between power consumption of a node and its inlet/outlet temperature can be presented as

$$P_i = \rho f_i C_p \left(T^i_{out} - T^i_{in}\right), \tag{2}$$

where C_p is the specific heat of air and ρ is the air density. In other words, the power consumption of node i will change the air temperature from the inlet air temperature T^i_{in} to the outlet temperature T^i_{out}. The generated hot air will also spread to other nodes. The temperature rise can be identified as **self-interference** (heating up air flowing through itself) and **cross interference** (heating up other nodes through recirculation).

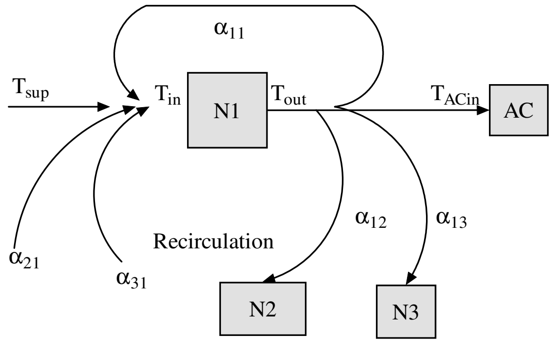

Fig. 2: This figure demonstrates the cross interference among distributed server nodes.

A. Hypothesis

Due to the complex nature of airflow inside the data center, some of the hot exhaust air from outlets of servers will recirculate into the inlet of other servers. Within a room environment, if the dimensions and locations of all major physical objects are fixed, and there are no moving objects inside the room, the air flow pattern should be relatively stable and predictable. Our hypothesis is that the amount of air and heat recirculated from the outlet of one server to the inlet of another server is relatively stable. If we can characterize such heat flow then we can use it to predict temperature distributions based on different power consumption distributions.

B. Definitions

The air recirculation inside a datacenter can be characterized as cross interference among server nodes. Figure 2 demonstrates our abstract heat flow model and the correlation between distributed nodes. Node N1 has inlet temperature T^i_{in}, which is a mix of supplied cold air with temperature T_{sup} and recirculated exhaust warm air from other nodes. The outlet warm air of node N1 will partially return to the air conditioner, and partially recirculate into other nodes with constant rate. It also draws in exhaust hot air from the outlet of other nodes due to recirculation. The rate, or the percentage of recirculated heat a_{ij} is defined as cross interference coefficients.

We assume the amount of recirculated heat from node i to node j is $a_{ij}Q^i_{out}$, where Q^i_{out} is the energy of exhaust air from node i. The coefficient a_{ij} is the percentage of heat flow from node i to node j. Assume cross-interference coefficients are constant, then the matrix $\mathbf{A} = [a_{ij}]_{nxn}$ defines the cross-interference among all server nodes.

For a given node i, the total amount of heat existed in the out air flow, Q^i_{out}, is given by

$$Q^i_{out} = Q^i_{in} + P_i = \rho f_i C_p T^i_{out}, \tag{3}$$

where P_i is the power consumption of node i, f_i is the flow rate (the speed of fan drawing air) of node i, T^i_{out} is the outlet air temperature. Q^i_{in} is the input heat given by

$$Q^i_{in} = \rho f_i C_p T^i_{in}. \tag{4}$$

where T^i_{in} is the inlet air temperature. We assume the air density does not change (where in practice it changes from $1.205 kg/m^3$ at $20°C$ to $1.067 kg/m^3$ at $60°C$).

C. System Function

The heat Q^i_{in} carried by inlet air flow is actually a mixture of supplied cold air and recirculated hot air, and we decompose it into the sum of multiple sub-”air flows” $\sum_{j=1}^{n} a_{ji}Q^j_{out} + Q^i_{\text{sup}}$, so we have

$$\begin{aligned} Q^i_{out} &= \sum_{j=1}^{n} a_{ji}Q^j_{out} + Q^i_{\text{sup}} + P_i & (5)\\ &= \sum_{j=1}^{n} a_{ji}\rho f_j C_p T^j_{out} + Q^i_{\text{sup}} + P_i. & (6)\end{aligned}$$

The total recirculated air flow rate to node i is $\sum_{j=1}^{n} a_{ji}f_j$, then $f_i - \sum_{j=1}^{n} a_{ji}f_j$ is the flow rate of supplied cold air flow drawn by node i, and consequently $Q^i_{\text{sup}} = \rho\left(f_i - \sum_{j=1}^{n} a_{ji}f_j\right) C_p T_{\text{sup}}$.

Then we have

$$Q^i_{in} = \sum_{j=1}^{n} a_{ji}\rho f_j C_p T^j_{out} + \rho\left(f_i - \sum_{j=1}^{n} a_{ji}f_j\right) C_p T_{\text{sup}}. \tag{7}$$

Considering Eq. 3 and substituting $\rho f_i C_p$ with K_i, for node i we have

$$K_i T^i_{out} = \sum_{j=1}^{n} a_{ji}K_j T^j_{out} + \left(K_i - \sum_{j=1}^{n} a_{ji}K_j\right) T_{\text{sup}} + P_i. \tag{8}$$

Then for all the nodes from 1 to n, we can have a group of linear equations presented as a matrix equation

$$\mathbf{K}\cdot\overrightarrow{\mathbf{T}}_{out} = \mathbf{A}'\cdot\mathbf{K}\cdot\overrightarrow{\mathbf{T}}_{out} + \mathbf{K}\cdot\overrightarrow{\mathbf{T}}_{\text{sup}} - \mathbf{A}'\cdot\mathbf{K}\cdot\overrightarrow{\mathbf{T}}_{\text{sup}} + \overrightarrow{\mathbf{P}}, \tag{9}$$

where $\mathbf{A}'$ is the transpose of $\mathbf{A}$, and the diagonal matrix $\mathbf{K}$ is defined as

$$\mathbf{K} = \begin{vmatrix} K_1 & 0 & .. \\ 0 & K_2 & \\ .. & .. & .. \end{vmatrix}, \tag{10}$$

and column vector $\overrightarrow{\mathbf{T}}_{out}$, $\overrightarrow{\mathbf{T}}_{\text{sup}}$, and $\overrightarrow{\mathbf{P}}$ are defined as

$$\begin{pmatrix} T^1_{out} \\ T^2_{out} \\ ... \end{pmatrix}, \begin{pmatrix} T_{\text{sup}} \\ T_{\text{sup}} \\ ... \end{pmatrix}, \text{ and } \begin{pmatrix} P_1 \\ P_2 \\ ... \end{pmatrix} \tag{11}$$

respectively.

Eq. 9 can also be written as

$$\mathbf{K}\cdot\left(\overrightarrow{\mathbf{T}}_{out} - \overrightarrow{\mathbf{T}}_{\text{sup}}\right) = \mathbf{A}'\cdot\mathbf{K}\cdot\left(\overrightarrow{\mathbf{T}}_{out} - \overrightarrow{\mathbf{T}}_{\text{sup}}\right) + \overrightarrow{\mathbf{P}}, \tag{12}$$

which we define as the **system function** of the datacenter. It gives the mathematical representation of the correlations among the outlet temperature, the supplied cold temperature, the power consumption of all nodes, and the cross interference coefficients.

D. Characterizing Cross Interference Coefficients

Suppose we have an initial reference power distribution $\overrightarrow{\mathbf{P}}_{ref}$ for all n nodes, and with this distribution we measure the reference outlet temperature as $\overrightarrow{\mathbf{T}}^{ref}_{out}$, so we have reference system function as:

$$\mathbf{K}(\overrightarrow{\mathbf{T}}^{ref}_{out} - \overrightarrow{\mathbf{T}}_{\text{sup}}) = \mathbf{A}'\mathbf{K}(\overrightarrow{\mathbf{T}}^{ref}_{out} - \overrightarrow{\mathbf{T}}_{\text{sup}}) + \overrightarrow{\mathbf{P}}_{ref}, \tag{13}$$

which has only one unknown matrix variable $\mathbf{A}'$.

Suppose for another new power consumption distribution $\overrightarrow{\mathbf{P}}_k$, we have another outlet temperature distribution $\overrightarrow{\mathbf{T}}^k_{out}$, then we have

$$\mathbf{K}(\overrightarrow{\mathbf{T}}^k_{out} - \overrightarrow{\mathbf{T}}_{\text{sup}}) = \mathbf{A}'\mathbf{K}(\overrightarrow{\mathbf{T}}^k_{out} - \overrightarrow{\mathbf{T}}_{\text{sup}}) + \overrightarrow{\mathbf{P}}_k. \tag{14}$$

Subtracting Eq. 13 from Eq. 14, we have

$$\mathbf{K}(\overrightarrow{\mathbf{T}}^k_{out} - \overrightarrow{\mathbf{T}}^{ref}_{out}) = \mathbf{A}'\mathbf{K}(\overrightarrow{\mathbf{T}}^k_{out} - \overrightarrow{\mathbf{T}}_{\text{sup}}) + \overrightarrow{\mathbf{P}}_k - \overrightarrow{\mathbf{P}}_{ref}. \tag{15}$$

In the above equation, there are n linear equations and n^2 variables (a_{ij}). To solve it we need to establish n different scenarios, e.g., n different power distribution vectors $\overrightarrow{\mathbf{P}}_k$, $k = 1..n$, and consequently n different temperature distributions $\overrightarrow{\mathbf{T}}^k_{out}$. Then we have a group of matrix equations

$$\begin{array}{c}
\mathbf{K}(\overrightarrow{\mathbf{T}}^1_{out} - \overrightarrow{\mathbf{T}}^{ref}_{out}) = \mathbf{A}'\mathbf{K}(\overrightarrow{\mathbf{T}}^1_{out} - \overrightarrow{\mathbf{T}}^{ref}_{out}) + \overrightarrow{\mathbf{P}}_1 - \overrightarrow{\mathbf{P}}_{ref} \\
\mathbf{K}(\overrightarrow{\mathbf{T}}^2_{out} - \overrightarrow{\mathbf{T}}^{ref}_{out}) = \mathbf{A}'\mathbf{K}(\overrightarrow{\mathbf{T}}^2_{out} - \overrightarrow{\mathbf{T}}^{ref}_{out}) + \overrightarrow{\mathbf{P}}_2 - \overrightarrow{\mathbf{P}}_{ref} \\
\cdots \\
\mathbf{K}(\overrightarrow{\mathbf{T}}^n_{out} - \overrightarrow{\mathbf{T}}^{ref}_{out}) = \mathbf{A}'\mathbf{K}(\overrightarrow{\mathbf{T}}^n_{out}\overrightarrow{\mathbf{T}}^{ref}_{out}) + \overrightarrow{\mathbf{P}}_n - \overrightarrow{\mathbf{P}}_{ref}
\end{array}. \tag{16}$$

We define a new outlet temperature distribution matrix $\mathbf{T}^{new}_{out} = \{\overrightarrow{\mathbf{T}}^1_{out}, \overrightarrow{\mathbf{T}}^2_{out}, ...\overrightarrow{\mathbf{T}}^n_{out}\}$, and a new power consumption matrix $\mathbf{P}_{new} = \{\overrightarrow{\mathbf{P}}_1, \overrightarrow{\mathbf{P}}_2, ...\overrightarrow{\mathbf{P}}_n\}$. In addition, we define a constant matrix $\mathbf{T}^{ref}_{out} = \left\{\overrightarrow{\mathbf{T}}^{ref}_{out}, \overrightarrow{\mathbf{T}}^{ref}_{out}, ...\overrightarrow{\mathbf{T}}^{ref}_{out}\right\}$, which is just a n column vector of $\overrightarrow{\mathbf{T}}^{ref}_{out}$, and a constant reference power matrix $\mathbf{P}_{ref} = \left\{\overrightarrow{\mathbf{P}}_{ref}, \overrightarrow{\mathbf{P}}_{ref}, ...\overrightarrow{\mathbf{P}}_{ref}\right\}$ for n scenarios. Now we put n equations of Eq.(16) into one matrix equation as

$$\mathbf{K}\left(\mathbf{T}^{new}_{out}\text{-}\mathbf{T}^{ref}_{out}\right) = \mathbf{A}'\mathbf{K}\left(\mathbf{T}^{new}_{out}\text{-}\mathbf{T}^{ref}_{out}\right) + \mathbf{P}_{new}\text{-}\mathbf{P}_{ref}, \tag{17}$$

which can be rewritten as

$$\mathbf{A}' = \mathbf{I} - (\mathbf{P}_{new} - \mathbf{P}_{ref})\left(\mathbf{T}^{new}_{out} - \mathbf{T}^{ref}_{out}\right)^{-1}\mathbf{K}^{-1}, \tag{18}$$

which can be used to calculate cross interference matrix.

The physical and practical indication of Eq. 18 is that we can characterize cross interference with n different power distribution vectors and record the corresponding outlet temperature distribution, then calculate the cross-interference coefficients matrix by Eq. 18. We call this **cross interference profiling**.

4. Simulation Results

The simulation environment we used is the same as that of our previous work [2]. At the beginning, we set 2500W as the reference power consumption of all nodes and obtain a reference temperature distribution of this power configuration. Next, we keep all node except node i's power consumption unchanged, set the power consumption of node i to new value P_{new}, which is defined as **profiling value**. Then we run CFD simulation and measure the resulting outlet temperature distribution $\overrightarrow{\mathbf{T}}^k_{out}$. We repeat such process totally n times for all n nodes. For CFD simulation, it takes about 18 hours to finish the profiling process.

Figures 3 and 4 draws the cross interference coefficients matrix $\mathbf{A}'$ we calculated for profiling value 3500 Watts and 4500 Watts. Note that most of the high values are distributed along the diagonals, as a node will have much strong interference to its neighboring nodes but a relatively weak interference to nodes far from it. The observation of cross interference coefficients matches the physical phenomenon of datacenters. Obviously, the distribution of the cross coefficients are almost the same in the two cases. The cross interference coefficients we obtained for different profiling values verified our hypothesis that air flow pattern and consequently the amount of heat recirculated are relatively stable and can be characterized as stable cross interference coefficients.

Comparing the numerical values of cross interference coefficients shown in Figures 3 and 4, we observe there are some small numerical differences. These differences can be contributed to the simulation truncation error, the simplified CFD model, and the small change of air properties under different temperature (our previous analysis assumed the air property such as density and specific heat is constant regardless the change of temperature).

A. Exit Coefficients and Recirculation Coefficients

We define the recipient of recirculated heat as **victim**, and the source of recirculated heat as **contributor**. Previously, the coefficient a_{ij} is defined as the percentage of energy flow from node i to node j. Then $\left(1 - \sum_{j=1}^{n} a_{ij}\right) Q^i_{out}$ would be the amount of heat at node i 's outlet which returns to the cooling system without recirculating into other nodes. So we define $EC_i = 1 - \sum_{j=1}^{n} a_{ij}$ as the Exit Coefficient (EC) of node i, it is a measure of whether the node i has a good ventilation or not. Similarly, we define $RC_i = \sum_{i=1}^{n} a_{ij}$ as the Recirculation Coefficient (RC) of node i, the amount of heat at node j 's inlet obtained through recirculated warm air instead of supplied cold air.

Please notice that $\sum_{i=1}^{n} a_{ij}$ or $\sum_{j=1}^{n} a_{ij}$ may not equal to 1, since at the outlets, some exhaust air will return to cooling system instead of recirculate to the other nodes, and at the inlets some air flow comes from supplied cold air coming out from floor tiles. A small EC indicates that the node is a source of hot air recirculation, and most of the exhaust heat will be recirculated; whereas a large RC indicates that the node is a victim of hot air recirculation.

Figures 5 and 6 show the EC and RC of the simulated datacenter when typical power consumptions of servers are 4000 W. In Figure 6, we observe that the nodes located at the upper part of the rack (row D or E) have relatively larger RCs because these nodes could not obtain enough cold air supply, so they absorb significant amount of recirculated hot air compared to the nodes of the lower part.

The recirculated hot air comes from the nodes located at the low part of the rack (row A or B). In Figure 5 we observe that these lower part nodes have lower ECs since a significant amount of exhaust hot air recirculates into other nodes, where nodes at the upper part have larger ECs because their outlets are close to the ceiling vents and most of the hot air returns to AC directly.

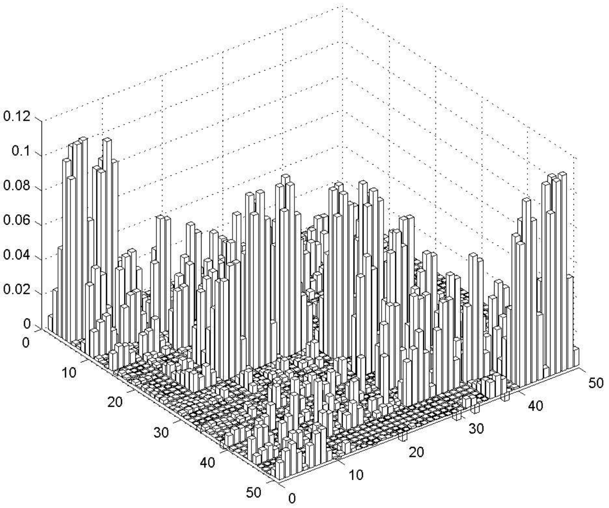

Fig. 3: Cross Interference Coefficient with Profiling Value 3500W

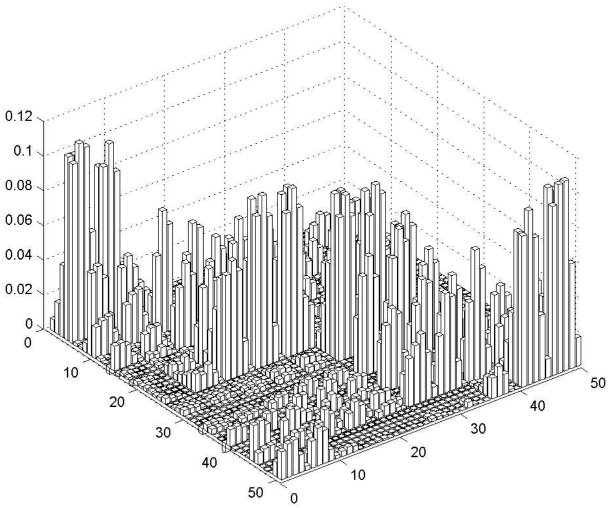

Fig. 4: Cross Interference Coefficient with Profiling Value 4500W

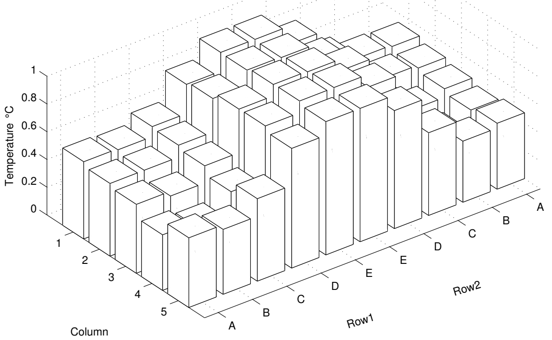

Fig. 5: Exit Coefficient

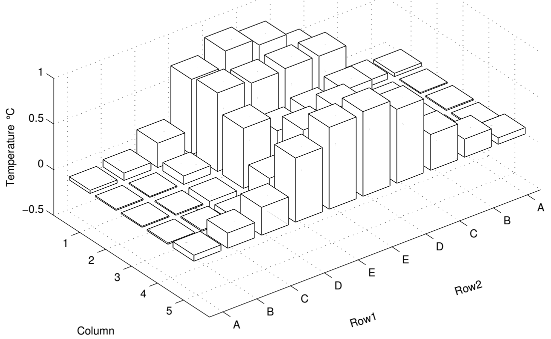

Fig. 6: Recirculation Coefficient

B. Verifying Temperature Prediction

As we mentioned before, the motivation of using abstract heat flow model is to accelerate the thermal evaluation and prediction process. We conducted simulation to verify the accuracy of our cross interference based fast thermal evaluation.

We use fast evaluation to evaluate the inlet temperature distribution of the reference power vector and compare with CFD simulation. Figure 7 shows very similar temperature change trend between predicted temperature and measured temperature. Some numerical difference is due to the simplified heat flow model, the truncation error of CFD simulation and the nonlinear properties of air. We record the difference and use it as **offset parameters** for future temperature evaluations. Then we can greatly improve the prediction accuracy. We randomly generate power consumption vectors, compare the resulting temperatures obtained from CFD to temperatures obtained from fast thermal evaluation. The average temperature prediction error is about $0.38°C$, which is acceptable since this value is even smaller than the natural temperature fluctuation (normally at least 1-2$°C$). One randomly generated power distribution and temperature prediction result are shown in Figure 8. The left figure of Figure 8 shows the power consumption distribution based on a task scheduling result.

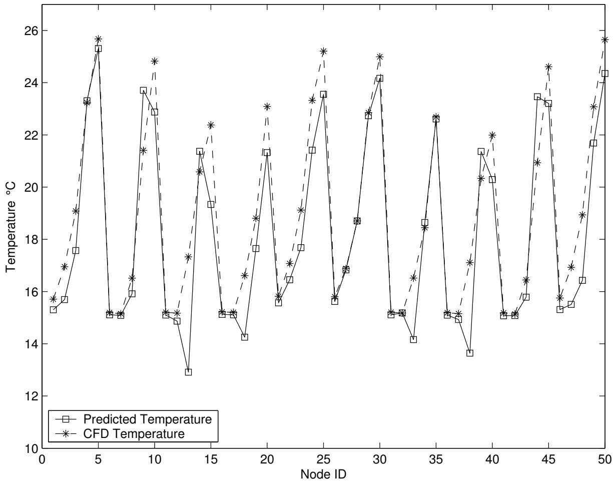

Fig. 7: Similar temperature change trend between predicted temperature and measured temperature can be observed.

The right figure of Figure 8 shows an almost perfect match between predicted temperature and measured temperature. The simulation results confirmed that fast thermal evaluation can be used to calculate temperature distribution which is much faster than CFD simulation, with a very high accuracy.

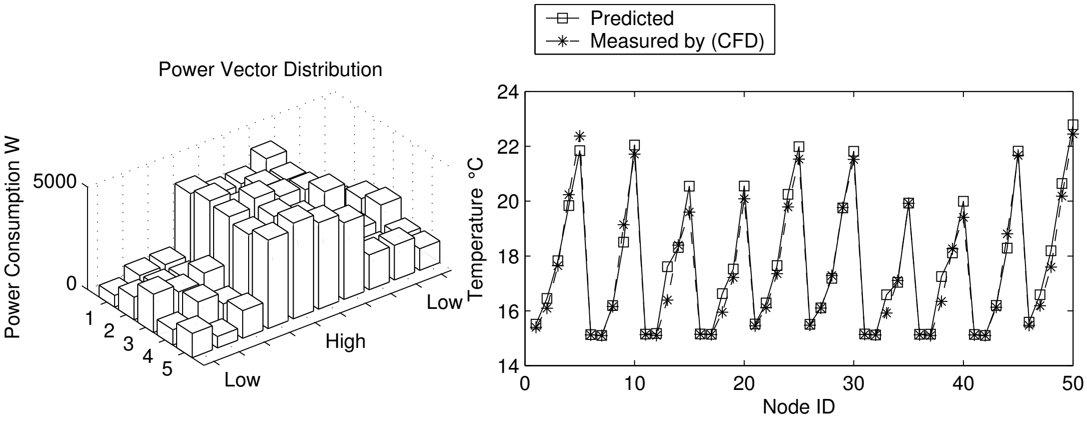

Fig. 8: Given a power consumption distribution of all server nodes, the cross interference based fast evaluation gives quite accurate temperature prediction.

C. Summary

Our observations obtained through above analysis of cross interference coefficients are consistent with the physical phenomena inside datacenter. Thus it confirms that our abstract model and cross interference can characterize the recirculation of the datacenter not only accurately, but also quantitatively. Once we obtain cross interference coefficients with CFD profiling, we no longer need to use CFD simulation in any future temperature evaluations, and achieve fast thermal evaluation with accurate and real-time temperature predictions.

5. Related Work

Researchers at HP Labs and Duke University have published work [3] [4] on smart cooling techniques for datacenters. They have developed online measurement and control techniques to improve energy-efficiency of datacenters, and have defined **Supply Heat Index (SHI)** and **Return Heat Index (RHI)** to characterize the energy efficiency of the cooling systems.

Our cross interference coefficients, as well as EC and RC, provide a much detailed presentation of heat recirculation: at a higher, server node level granularity, instead of the room level description provided by SHI and RHI. The coefficients can be used to identify the main sources of recirculation as well as the victims of recirculation. Accordingly, remedies can be taken to counteract or reduce recirculation effect. Next, we plan to integrate cross coefficients with our previous work of thermal-aware scheduling [2], to further improve the energy efficiency of datacenters.

The work Consil [5] deducted inlet temperatures, and consequently get thermal map, from internal motherboard sensor readings by using neural net algorithms. Their work is based on the assumption that inlet temperature information is not available. Our work, instead, is based on the assumption that temperature information is available from onboard sensors. Based on the information from these sensors prediction and evaluation of temperature can be performed without running the slow CFD simulation.

Thermal-aware scheduling, proposed in [1], [6], and [7], used CFD simulation to conduct thermal evaluation, which can not be used for online real-time datacenter thermal management. To the best of our knowledge, we are the first to propose an abstract heat flow model to predict and evaluate temperature distribution in a more efficient manner, suitable for online, real-time thermal management of datacenter.

6. Conclusions and Future Work

In our previous work [2] we showed the potential of thermal-aware scheduling to improve datacenter energy efficiency. In this paper, we presented an abstract heat flow model which characterizes the air recirculation of datacenters as cross interference among server nodes, and the cross interference coefficients can be used to accelerate the thermal evaluation process. Next, we will combine the abstract heat flow model and fast thermal evaluation with thermal-aware scheduling, to implement online thermal-aware scheduling and to achieve real-time datacenter thermal management.

Acknowledgements

This research is supported in part by MediServe Information Systems, Consortium for Embedded Systems and National Science Foundation grant ANI-0196156.

References

[1] J. Moore, J. Chase, P. Ranganathan, and R. Sharma, “Making scheduling ”cool”: Temperature-aware resource assignment in data centers,” in *2005 Usenix Annual Technical Conference*, April 2005.

[2] Q. Tang, S. K. S. Gupta, D. Stanzione, and P. Cayton, “Thermal-aware task scheduling to minimize enery usage of blade server based datacenters,” in *Procedings IEEE International Symposium on Dependabble, Autonomic and Secure Computing*, Oct 2006.

[3] C. D. Patel, R. Sharma, C. E. Bash, and A. Beitelmal, “Thermal considerations in cooling large scale high compute density data centers,” in *Proceedings of the Eight Inter-Society Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, San Diego, CA, June 2002, pp. 767–776.

[4] M. H. Beitelmal and C. D. Patel, “Thermo-fluids provisioning of a high performance high density data center,” Hewlett Packard Laboratories, Tech. Rep. HPL-2004-146, September 2004. [Online]. Available: http://www.hpl.hp.com/techreports/2004/HPL-2004-146.html

[5] J. Moore, J. Chase, and P. Ranganathan, “Low-cost thermal mapping of data centers,” in *Procedings 1st Workshop on Tacking Computer Systems Problems with Machine Learning Techniques (SysML)*, June 2006.

[6] J. Moore, J. Chase, K. Farkas, and P. Ranganathan, “Data center workload monitoring, analysis, and emulation,” in *Eighth Workshop on Computer Architecture Evaluation using Commercial Workloads*, February 2005.

[7] J. Moore, J. Chase, and P. Ranganathan, “Weatherman: Automated online, and predictive thermal mapping and management of data centers,” in *Procedings 3rd IEEE Int’l Conf. Autonomic Computing*, June 2006.