Energy and Buildings 50 (2012) 81–92

Contents lists available at SciVerse ScienceDirect

Energy and Buildings

journal homepage: www.elsevier.com/locate/enbuild

Minimizing the thermal impact of computing equipment upgrades in data centers

Jayantha Siriwardana[a,*], Saman K. Halgamuge[a], Thomas Scherer[b], Wolfgang Schott[b]

[a] Department of Mechanical Engineering, The University of Melbourne, Parkville, Victoria, Australia
[b] IBM Zurich Research Laboratory (ZRL), Rüschlikon, Switzerland

ARTICLE INFO

Article history:
Received 11 November 2011
Received in revised form 6 February 2012
Accepted 12 March 2012

Keywords:
Data center energy efficiency
Data center cooling
Hot air recirculation
Thermal aware equipment upgrading
Load spreading
Computational fluid dynamics (CFD)
Particle swarm optimization (PSO)

ABSTRACT

Upgrading of today's air-cooled data centers (DCs) with high-performance computing, networking, and storage equipment is a challenging task due to typically higher power needs and more adverse cooling requirements of new equipment. To cope with the increase of the power and heat load in DCs, load spreading is commonly applied. This technique distributes the thermal load of new equipment over multiple racks if the power requirement and heat generation of new equipment exceeds the rack's capacity. We present a novel load spreading technique that allows upgrading of the computing equipment with minimal thermal impact on the existing optimized DC cooling environment. Our approach is based on an abstract heat-flow model of the DC, whose parameters are determined by performing a measurement campaign in the DC and with support of computational fluid dynamics simulations. The optimum placement of the new equipment in the racks of the DC is found by applying a particle swarm optimization technique to this model. The effectiveness of our method was assessed based on experiments performed in a production DC. The results show that our holistic approach for optimizing the placement of the upgraded computing equipment in the DC outperforms the conventional load spreading technique.

1. Introduction

The successful operation of an air-cooled data center (DC) requires an efficient cooling environment to ensure that the DC operator can provide its services to customers with maximum availability and reliability at minimal operational cost. An efficient cooling system guarantees that the temperatures at the inlets of all devices in the computer racks of the DC never exceed a given threshold value to prevent device overheating, and achieves this goal with a minimum amount of cooling energy.

In today's DCs, a hot-/cold-aisle cooling concept is usually employed to efficiently cool the computing equipment [1]. For this purpose, the racks are arranged in rows to form hot and cold aisles with alternating airflows. In the cold aisles, chilled air from the computing room air conditioners (CRACs) blown through the raised-floor plenum and perforated floor tiles is directed to the device inlets, while in the hot aisles heated exhaust air from the racks circulates back to the CRACs. To ensure that the inlet temperatures at the devices never exceed a given threshold value and no cooling energy is wasted, a nominal airflow and temperature distribution is provided in the DC by adjusting various parameters of the cooling system (e.g. supply airflow, supply temperature), determining the proper placement of the perforated tiles in the cold aisles, and carefully selecting the best-suited placement of the computing devices in the racks. These steps of setting the DC cooling environment to an optimal operating point have to be carefully performed and successfully verified by using, for example, IBM's Mobile Measurement Technology (MMT) [2] before putting the DC into operation.

Due to the increasing demand for supporting new on-line services such as video on-demand, Internet banking, cloud computing, social networking, etc., DC operators are continuously compelled to upgrade their DC with high-performance computing equipment such as blade servers. These devices are usually smaller and process information at significant higher rates than their predecessors, but typically consume more power and thus dissipate more heat. This additionally generated heat should be efficiently removed from the devices to avoid creating hot spots and other cooling inefficiencies. Hot air exhausted from the rack outlets can recirculate into the cold air stream supplied to the rack inlets, causing the equipment to under-perform, malfunction, or fail all together [3].

To combat the undesired thermal effects of the upgraded equipment on the existing optimized cooling environment, various strategies can be applied: The DC operator can lower the supply air temperature of the cooling system. This approach avoids device overheating, but will significantly increase the annual cost to be spent for cooling energy. Another approach, which is commonly applied in practice, is load spreading [4]. To determine the best placement of the new equipment in the computer racks, the

* Corresponding author. Tel.: +61 425177770.
E-mail addresses: [redacted]@pgrad.unimelb.edu.au, [redacted]@gmail.com (J. Siriwardana).

0378-7788/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.enbuild.2012.03.026

maximum allowed power and heat capacity are specified for each rack. Whenever the load of the devices put to a rack exceeds any of theses design parameters, the equipment is split and distributed over multiple racks. This simple approach more evenly distributes the power and cooling demands of the new equipment over the DC, but does not take into consideration the airflow and temperature distribution in the DC when spreading the heat load.

In this article, we present a novel load spreading technique that allows upgrading of the computing, networking, and storage equipment with minimal thermal impact on the existing optimized cooling environment of a DC. Our approach is based on an abstract heat-flow model that describes the nominal operating point of the DC before upgrading the equipment. The parameters of the model are determined by means of performing an extensive measurement campaign and with support of CFD simulations. The model takes into account the temperature distribution within the DC, various airflows, and the power consumption of the computing devices. The best-suited placement of the upgraded equipment in the racks is found by formulating a constrained non-linear optimization problem based on this model and solving it with a particle swarm optimization (PSO) technique [5]. The effectiveness of our method is assessed based on experiments performed in a production DC. The results show that our holistic approach for minimizing the thermal impact on the DC cooling environment in case of an equipment upgrade outperforms the conventional load spreading technique.

The article is organized as follows: Section 2 describes related work. Section 3 presents the proposed strategy for minimizing the thermal impact of a computing equipment upgrade in a DC. Section 4 gives results of a case study, in which we applied our optimization strategy to a production DC. The article concludes with a brief summary.

2. Related work

Recently considerable attention has been paid to the optimization of the cooling performance of air-cooled DCs with a hot-/cold aisle layout. In this section, we give a brief overview on three previously published approaches that focus on minimizing the hot air recirculation in DCs and are relevant to the proposed equipment upgrading methodology.

2.1. Cluster cooling performance optimization

Shrivastava et al. [6] proposed to combine a genetic algorithm with a neural network (NN) model to obtain the optimum layout of a cluster of DC equipment in terms of thermal performance. They assumed that a cluster typically comprises two equal length rows of racks and coolers bounding a common hot aisle. The NN model predicts the cooling performance by observing the power demand for each rack, and the rack inlet and exhaust airflow rates. Based on the observations, the genetic algorithm assesses all possible cluster configurations to find the best distribution of the heat load among the racks or to identify the best physical cluster layout.

The applicability of this method to a production DC is limited as the approach does not take into account the thermal interference between various clusters covering the DC room. Moreover, when upgrading DC equipment, a complete layout change of the DC is undesirable due to the high installation cost and the required long DC down time.

2.2. Heat recirculation minimization

Moore et al. [7] proposed a system-level solution to control the heat generation in a DC through temperature aware workload placement to servers. A server workload placement policy called minimum heat recirculation (MinHR) has been proposed to assign tasks to servers that have minimum contribution to the hot air recirculation in the DC, and a metric called heat recirculation factor (HRF) to assess the contribution to the hot air recirculation of each server. MinHR maintains a priority list of servers according to their HRF values that is used in allocating tasks.

Even though the scalar dimensionless HRF values can serve as a simple metric for placing tasks on servers, they cannot be used for accurately capturing and modeling thermal effects in the DC room. Moreover, the benefit of using the MinHR server workload placement policy diminishes if the utilization rate of the DC equipment approaches 100%.

2.3. Thermal aware task scheduling

Thermal aware task scheduling as proposed by Tang et al. [8] is a thermodynamic based, but yet simple approach for assigning tasks to servers to minimize the hot air recirculation from the server outlets to the inlets. Tang's method assumes a homogeneous DC with identical equipment. It predicts the temperature at the server inlets with a low-complexity abstract heat-flow model [9] and a power model that linearly maps the CPU utilization to the server's power consumption. The parameters of the heat-flow model can be obtained with CFD simulations. Based on the predicted result, the incoming tasks are distributed among the servers so that the temperature of the chilled air supplied by the CRACs is maximized while keeping the temperature at the server inlets below a given maximum allowed threshold value.

The method has mainly been proposed for temperature aware scheduling of the processing workload among servers to minimize DC cooling requirements. It has not been applied for optimizing the layout of the DC to minimize undesired thermal effects such as hot air recirculation and, in particular, not for spreading the heat load over the DC while upgrading equipment.

3. Strategy for minimizing the thermal impact of server upgrades

The goal of the proposed load spreading strategy is to find the best placement of the upgraded equipment in the DC with minimal thermal impact to the existing cooling environment. This is achieved by formulating a constrained non-linear optimization problem that uses Tang's abstract heat-flow model to quickly predict the outlet and inlet temperatures of the server equipment for various upgrading scenarios, and applies a PSO algorithm to search for the best-suited position for the new servers in the racks of the DC. After having identified the best placement, the servers currently located at these positions are moved to the location of the servers to be replaced before the new servers are installed. In this section, we first review Tang's abstract heat-flow model due to its importance to our work, and then present the proposed PSO-based load spreading technique.

3.1. Abstract heat-flow model

We assume that the DC consists of n nodes. In case of blade servers, a node can be considered as a chassis that houses a number of blade servers as shown in Fig. 1. Each node i thus consist of m servers. Each node i draws air with inlet temperature T_{in}^{i} and dissipates hotter air with average outlet temperature T_{out}^{i}. In a typical DC, the difference between inlet and outlet temperatures of a node is approximately 11 °C [4]. The outlet temperature of a node is determined by the inlet temperature and the heat load of the equipment inside the node, while the inlet temperature is determined by

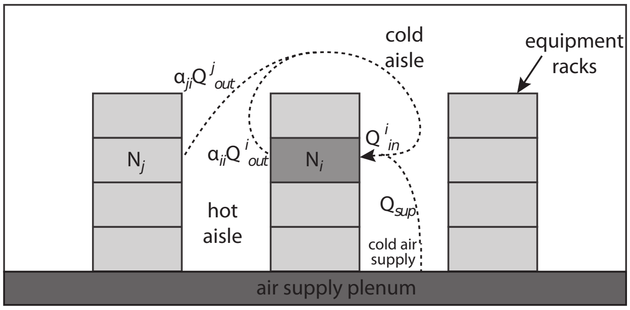


Fig. 1. Schematic of typical thermal cross-interferences among equipment nodes housed in racks in a data center.

the mixture of cold air supplied from the CRAC and hot air recirculated from the node outlets.

3.1.1. Derivation of abstract heat-flow model

Considering that the total power drawn by a node is dissipated as heat, the law of energy conservation can be used to express the steady-state relationship between the power consumption P_i and the in-/outlet temperature of node i as

$$P_i = \rho f_i c_p (T^i_{out} - T^i_{in})$$

or equivalently as

$$T^i_{out} = T^i_{in} + K_i^{-1} P_i, \tag{1}$$

where ρ is the air density, f_i is the airflow rate through the node i, c_p is the constant-pressure specific heat of air and $K_i = \rho f_i c_p$. The heat load P_i thus causes the temperature of the air passing through node i with a constant airflow rate f_i to increase from T^i_{in} at the inlet side to T^i_{out} at the outlet side of the rack.

The air recirculation inside the DC is characterized as cross-interference among the nodes. The inlet temperature T^i_{in} of node i can be considered as a mixture of the supplied cold air with temperature T_{sup} from the CRACs and the recirculated hot exhaust air from other nodes. The total amount of heat exhausted per time unit in the outlet airflow Q^i_{out} is given by

$$Q^i_{out} = Q^i_{in} + P_i = K_i T^i_{out}. \tag{2}$$

Since the inlet heat Q^i_{in} carried by the inlet airflow is a mixture of the heat Q_{sup} of the supplied cold air and the recirculated hot air as illustrated in Fig. 1, the following expression holds:

$$Q^i_{in} = \sum_{j=1}^{n} \alpha_{ji} Q^j_{out} + Q_{sup}, \tag{3}$$

where the amount of recirculated heat from node j to node i is defined as $\alpha_{ji}Q^j_{out}$. Hence, the matrix $\mathbf{A} = [\alpha_{ij}]_{n\times n}$ defines the cross-interference among all the server nodes and is thus called cross-interference matrix.

From Eqs. (2) and (3), we have

$$\begin{aligned} Q^i_{out} &= \sum_{j=1}^{n} \alpha_{ji} Q^j_{out} + Q_{sup} + P_i \\ &= \sum_{j=1}^{n} \alpha_{ji} K_j T^j_{out} + Q_{sup} + P_i. \end{aligned} \tag{4}$$

If the total recirculated airflow rate from all nodes to node i is $\sum_{j=1}^{n}\alpha_{ji}f_j$, then $f_i - \sum_{j=1}^{n}\alpha_{ji}f_j$ is the flow rate of supplied cold airflow drawn by node i and, consequently, we have $Q_{sup} = \rho\left(f_i - \sum_{j=1}^{n}\alpha_{ji}f_j\right) c_p T_{sup}$. Thus, we can rewrite Eq. (4) as

$$K_i T^i_{out} = \sum_{j=1}^{n} \alpha_{ji} K_j T^j_{out} + \left(K_i - \sum_{j=1}^{n} \alpha_{ji} K_j\right) T_{sup} + P_i. \tag{5}$$

For all the nodes from 1 to n, we can derive a group of linear equations from Eq. (5) above as follows:

$$\mathbf{K}\vec{\mathbf{T}}_{out} = \mathbf{A}'\mathbf{K}\vec{\mathbf{T}}_{out} + \mathbf{K}\vec{\mathbf{T}}_{sup} - \mathbf{A}'\mathbf{K}\vec{\mathbf{T}}_{sup} + \vec{\mathbf{P}}, \tag{6}$$

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

$$\mathbf{K} = \begin{pmatrix} K_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & K_n \end{pmatrix},$$

and column vectors $\vec{\mathbf{T}}_{out}$, $\vec{\mathbf{T}}_{sup}$ and $\vec{\mathbf{P}}$ are defined as

$\vec{\mathbf{T}}_{out} = [T^1_{out}, \dots, T^n_{out}]'$,
$\vec{\mathbf{T}}_{sup} = [T_{sup}, \dots, T_{sup}]'$, and
$\vec{\mathbf{P}} = [P_1, \dots, P_n]'$,

respectively. Eq. (6) can also be rewritten as

$$\mathbf{K}(\vec{\mathbf{T}}_{out} - \vec{\mathbf{T}}_{sup}) = \mathbf{A}'\mathbf{K}(\vec{\mathbf{T}}_{out} - \vec{\mathbf{T}}_{sup}) + \vec{\mathbf{P}}, \tag{7}$$

which is referred to as the system function of the DC.

3.1.2. Determining the cross-interference matrix

In Eq. (7), we can consider $\vec{\mathbf{P}}$ as the system input and $\vec{\mathbf{T}}_{out}$ as the system output. We can eliminate the term $\vec{\mathbf{T}}_{sup}$ using a known reference power distribution $\vec{\mathbf{P}}^{ref}$ as follows:

$$\mathbf{K}(\vec{\mathbf{T}}_{out} - \vec{\mathbf{T}}^{ref}_{out}) = \mathbf{A}'\mathbf{K}(\vec{\mathbf{T}}_{out} - \vec{\mathbf{T}}^{ref}_{out}) + \vec{\mathbf{P}} - \vec{\mathbf{P}}^{ref}. \tag{8}$$

Eq. (8) denotes a system of n equations with cross-interference matrix $\mathbf{A}$ comprising n^2 unknown parameters. We can determine the elements of $\mathbf{A}$ by generating n systems of Eq. (8) using n different known power distribution vectors $\vec{\mathbf{P}}$ as follows:

$$\mathbf{K}(\mathbf{T}_{out} - \mathbf{T}^{ref}_{out}) = \mathbf{A}'\mathbf{K}(\mathbf{T}_{out} - \mathbf{T}^{ref}_{out}) + \mathbf{P} - \mathbf{P}^{ref}, \tag{9}$$

where the matrices $\mathbf{T}_{out}$, $\mathbf{P}$, $\mathbf{T}^{ref}_{out}$, and $\mathbf{P}^{ref}$ are defined as

$\mathbf{T}_{out} = [\vec{\mathbf{T}}^1_{out}, \dots, \vec{\mathbf{T}}^n_{out}]$,
$\mathbf{P} = [\vec{\mathbf{P}}_1, \dots, \vec{\mathbf{P}}_n]$,
$\mathbf{T}^{ref}_{out} = [\vec{\mathbf{T}}^{ref}_{out}, \dots, \vec{\mathbf{T}}^{ref}_{out}]$, and
$\mathbf{P}^{ref} = [\vec{\mathbf{P}}^{ref}_1, \dots, \vec{\mathbf{P}}^{ref}_n]$,

respectively. Eq. (9) can now be rewritten as

$$\mathbf{A}' = \mathbf{I} - (\mathbf{P} - \mathbf{P}^{ref})(\mathbf{T}_{out} - \mathbf{T}^{ref}_{out})^{-1}\mathbf{K}^{-1}. \tag{10}$$

The above system of $n \times n$ equations can be used to determine the elements of the cross-interference matrix $\mathbf{A}$ under the assumption that the chosen n different power distributions vectors are mutually independent.

3.2. Optimization problem

In general, the cross-interference matrix $\mathbf{A}$ is unique for a given physical rack layout of the DC and depends on the airflow rates through the rack enclosures and perforated tiles on the floor. In this paper, we assume that the physical rack layout is kept unchanged and the airflow rates through the equipment are constant during the equipment upgrade. In reality, new DC equipment can increasingly be operated at higher temperatures allowing minimal change to pressure head of cooling fans [10]. In the case of blade servers,

servers can be swapped with new ones without changing the fans fixed to their chassis. In this case also the pressure head change of fans is minimal since each fan provides cooling for multiple servers within the chassis. Therefore, our assumption can be considered as reasonable for our purpose. As a consequence, the elements of matrix **A** remain unchanged while performing load spreading.

Moreover, we assume that the equipment consumes a constant amount of power at idle state and its power consumption increases linearly with the utilization rate. The power consumption of node i can thus be expressed as

$$P_i = b_i + a_i F_i(t), \tag{11}$$

where b_i is the power consumption at idle state, a_i is the additional power required to achieve 100% server utilization, and $F_i(t)$ is the utilization rate of equipment i at time t. Note that $0 \leq F_i(t) \leq 1$. In this work we assume that $F_i(t)$ is the average day utilization rate of equipment i at time t. In most cases, new equipment requires more power than their predecessors and dissipates more heat than the equipment to be replaced. We can account for this fact by increasing b_i and a_i as necessary, and introducing a new expected utilization rate function $F_i(t)$ in Eq. (11).

When the cross-interference matrix **A** is known, the following equation can be derived from Eqs. (1) and (7) to predict the inlet temperatures of all equipment nodes:

$$\vec{\mathbf{T}}_{in} = \vec{\mathbf{T}}_{sup} + [(\mathbf{K} - \mathbf{A}'\mathbf{K}^{-1}) - \mathbf{K}^{-1}]\vec{\mathbf{P}}. \tag{12}$$

We then define the optimization function, which is later referred to as fitness function as

$$\min\{avg(T^i_{in})\} \tag{13}$$

after having changed the power profile vector $\vec{\mathbf{P}}$ according to the properties of the upgraded equipment. Our goal is thus to minimize the average inlet temperature of the equipment nodes in the racks by altering the position of the new equipment to be installed.

Since the rack inlet airflow is a mixture of cold supply air and hot air recirculated from other racks, minimizing the average rack inlet temperature means minimizing hot air recirculation. Therefore, we are able to obtain the best possible cooling performance for the given upgrade. In other words, we place new equipment at locations where it is less prone to recirculating hot air back to rack inlets.

3.3. Particle swarm optimization

We use a particle swarm optimization (PSO) technique for minimizing the fitness function Eq. (13) to optimally spread the increased heat load of the upgraded equipment over the racks of the DC. PSO is a population-based self-adaptive optimization technique that was introduced by Kennedy and Eberhart [5] in 1995 and since then went through several modifications [11–13]. Even though this technique cannot always find the optimum solution, we have chosen this biology-inspired approach because of its simple applicability to the load spreading optimization problem, and the easiness to incorporate non-linear constraints into the iterative search algorithm. PSO technique has also been applied for optimization tasks in related topics such as district heating systems [14] and HVAC systems [15].

Similar to other population-based optimization methods, the proposed PSO algorithm starts with randomly initializing a population of candidate solutions, which are commonly called particles, and computes for these particles the corresponding value given by the fitness function. To further improve the initial solution, the PSO algorithm iteratively moves around the particles in a pre-defined search space by imitating the social behavior of animals in a swarm, whereas the movement of each particle is influenced by its own best known position and also the best known position found by other particles. The swarm characterized by the locations and velocities of the particles at a given time instant is often referred to as a generation. The velocity of the particles can be used as a termination condition of the algorithm, as the velocities of the particles approach to zero if they reach the global optimum of the fitness function.

In our PSO algorithm, the trajectory of each particle in the search space is adjusted by dynamically altering the velocity of each particle according to its best known position and the best known positions found by other particles in the search space. Let us denote the position and velocity vector of the ith particle in a d-dimensional space by $\vec{\mathbf{X}}_i = [x_{i1}, \ldots, x_{id}]'$ and $\vec{\mathbf{V}}_i = [v_{i1}, \ldots, v_{id}]'$, respectively. According to the fitness function, the best fitness value obtained by the particle up to time t is $\vec{\mathbf{S}}_i = [s_{i1}, \ldots, s_{id}]'$ and the fittest particle found in the entire swarm up to time t is $\vec{\mathbf{S}}_g = [s_{g1}, \ldots, s_{gd}]'$. Then, the new velocities and the position of the particles of the next generation at time $t+1$ can be calculated as

$$v_{ij}(t+1) = v_{ij}(t) + c_1 R_1 [s_{ij} - x_{ij}(t)] + c_2 R_2 [s_{gj} - x_{ij}(t)], \tag{14}$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1), \tag{15}$$

where c_1 and c_2 are known as acceleration coefficients, R_1 and R_2 are two independent random variables uniformly distributed in the range [0, 1] and j is an integer in the range [1, d].

To optimize the load spreading problem with the PSO technique, we model the change of the location of the new equipment during the heat load optimization as a change of the positions of particles in a PSO search space. The position vector $\vec{\mathbf{X}}_i = [x_{i1}, \ldots, x_{ik}]'$ of particle i contains the physical locations x_{ij}, $j = 1, \ldots, k$, of the new equipment, while the elements v_{ij} of the velocity vector $\vec{\mathbf{V}}_i = [v_{i1}, \ldots, v_{ik}]'$ reflect the rate of the position change of equipment being upgraded represented by the elements of $\vec{\mathbf{X}}_i$. The fitness value for each new equipment configuration $\vec{\mathbf{X}}_i$ can be computed by using Eqs. (12) and (13). Note that the dimension of the optimization is given by the number k of equipment nodes to be upgraded and is always less than the total number n of nodes available in the DC.

4. Case study

We applied the novel load spreading technique for minimizing the thermal impact of computing equipment upgrades that was introduced in the previous section to a data center at the IBM Zurich Research Laboratory (ZRL). Fig. 2 shows the floor plan of the DC, which has a raised floor area of 286 m^2 and houses 27 racks with servers. A summary of the equipment in the DC is given in Table 1.

We investigated the effect of replacing a number of servers in the existing DC with new ones that had similar airflow characteristics but considerably higher heat load. To apply our optimization approach, we first determined the parameters of the heat-flow model. We considered 10 racks in the DC for this case study. The area of interest is highlighted in Fig. 2. Each rack houses a number of servers. This number varies from rack to rack since there are various types of servers of different sizes. To reduce the complexity of the model, we divided each rack into 3 blocks. Each of these blocks corresponds to one node in the heat-flow model with a total of 30 nodes. The calculation of the cross-interference matrix for the heat-flow model required measurement data for 30 different heat load scenarios in the DC. We collected one set of measurement data for one distinct heat load scenario and obtained the rest of the data with support of CFD simulations. We then built a detailed 3-dimensional model of the room and used the measurement data to define the boundary conditions for the CFD simulations. We used the commercial CFD software ANSYS CFX [16] to determine the temperature distribution for the different heat load scenarios. A similar approach is described in [17]. Based on the measurement

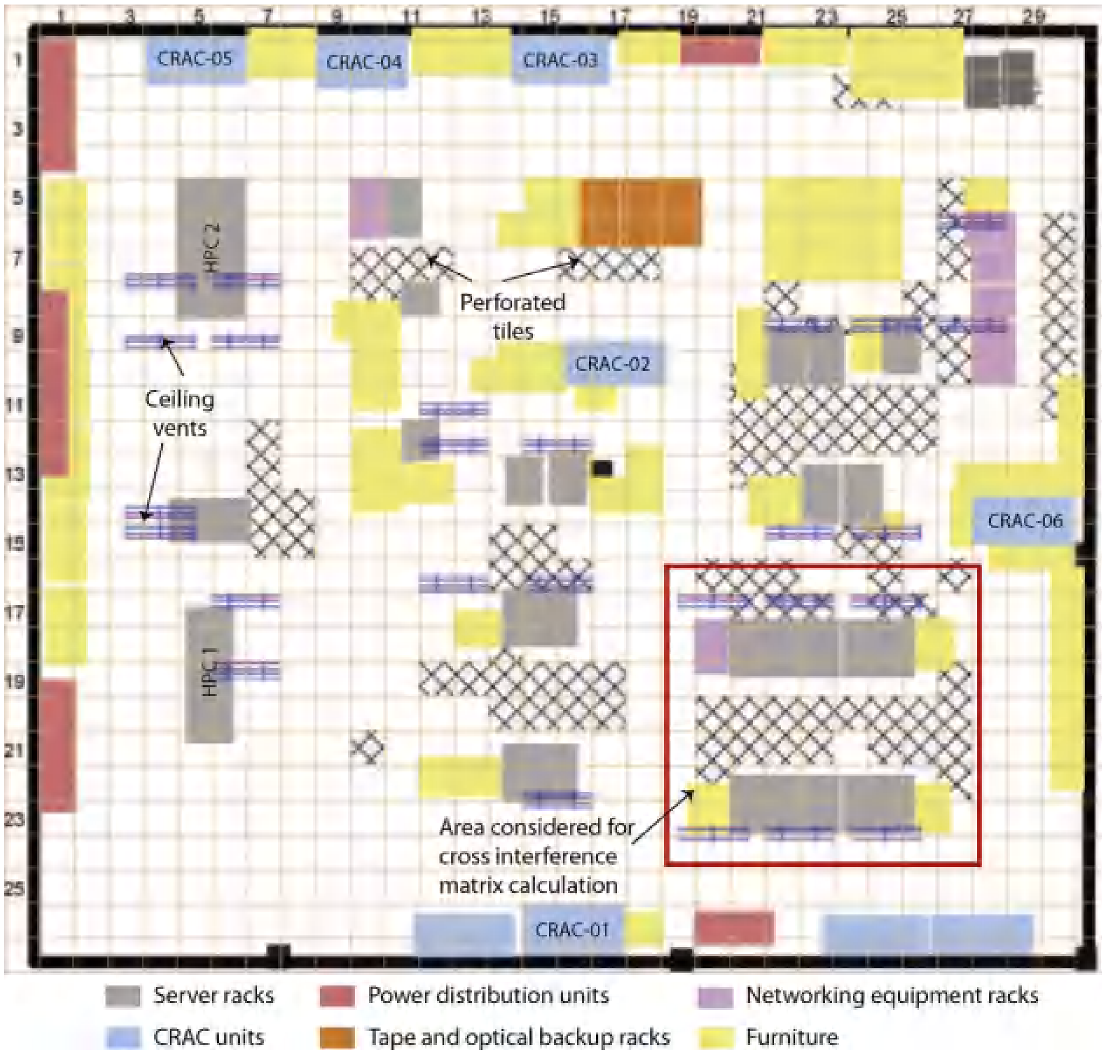


Fig. 2. Floor plan of the data center.

data and CFD simulation results, we calculated the parameters for the heat-flow model. Finally, we applied the novel load spreading technique to determine the optimal location for the new servers and evaluated the results.

Section 4.1 describes the measurement campaign that we conducted in the IBM ZRL DC to determine the temperature distribution within the room, airflow rates through perforated floor tiles, ceiling vents, server racks and CRACs, and power consumption of the servers in the DC. In Section 4.2, we give an overview of the CFD modeling and compare simulation results with actual measurement data for verification. Parameter identification for the heat-flow model is described in detail in Section 4.3. Finally, we illustrate the optimization approach in Section 4.4 and discuss the results in Section 4.5.

Table 1
Overview of data center layout.

Item	Value
DC room raised floor area	18 m × 15.9 m
DC room height	2.7 m
Server racks[a]	27
Racks with networking equipment and storage equipment	9
Computer room air conditioners (CRACs)	6
Power distribution units (PDUs)	4
Perforated floor tiles	92
Ceiling vents	25

[a] Includes two high performance computers (HPCs).

4.1. Measurement campaign

We used IBM's Mobile Measurement Technology (MMT) to obtain the temperature distribution in the data center. This technology has been successfully used to obtain a temperature distribution in DCs [18]. Fig. 3 shows a MMT measurement cart with a total of 72 temperature sensors. Nine sensors are mounted on each of the eight quadratic metal frames to measure the temperature information at the respective height. The cart is on wheels and can thus be readily moved from tile to tile to record the temperature information above every unoccupied floor tile. This allowed us to create a 3-dimensional temperature map of the DC with a spatial resolution of 20 cm in horizontal *x*- and *y*-direction and 30 cm in vertical *z*-direction.

To measure the volumetric airflow rates through perforated floor tiles, ceiling vents, server racks and CRACs, we used a Shortridge Instruments FlowHood CFM-88L [19]. This air capture hood is equipped with a 16-point flow-sensing grid to measure the dynamic pressure in front of an air inlet or outlet. A floor tile of size 60 cm × 60 cm is entirely covered by the hood, while two measurements are required to cover a ceiling vent, and three measurements are required to cover the rear side of a rack. The three

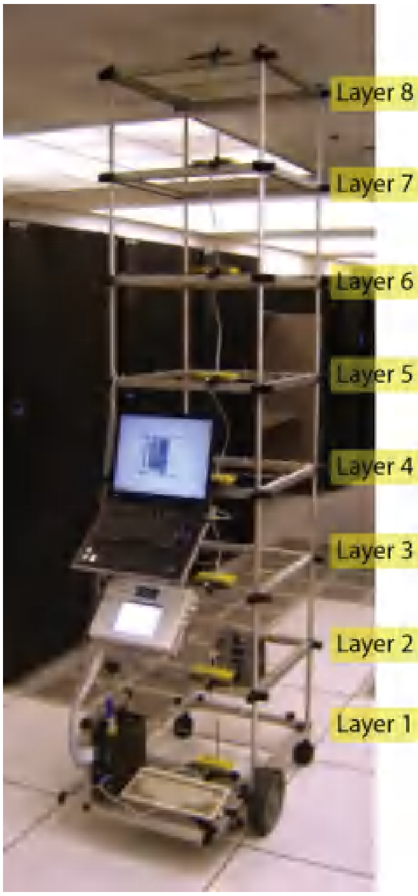


Fig. 3. MMT measurement cart [2] used to obtain the 3-dimensional temperature distribution in the data center.

measurements at the rear of the racks correspond to the airflow through the three blocks as defined at the beginning of this section. When an air capture hood is in place, the airflow is always reduced to some degree. The hood is equipped with flaps that allow us to take two separate measurements with defined air resistance for compensating this effect. Based on the two measurements, the device automatically calculates the volumetric airflow rate. The power consumption per rack was read from the power meters at the power distribution units. For our model, we assumed that the power consumption of the individual servers was proportional to the corresponding measured airflow rates.

4.2. Model for computational fluid dynamics simulations

Computational fluid dynamics based models can be used to simulate the complex airflow patterns and temperature distribution in data centers as evidenced by number of previous studies [20,21]. We used the CFD software ANSYS CFX to determine the temperature distribution for 30 different heat load scenarios in the DC. The model was validated by comparing simulation results with actual measurements.

Fig. 4 shows the geometry of the DC. The fluid domain had to be discretized into a number of control volumes to solve the governing equations numerically; we created a mesh of 627,076 equilateral hexahedral elements. Increasing the number of elements did not change the simulation results significantly, thus grid independence was achieved. The properties of fluid domain are given in Table 2. Table 3 gives an overview of the measurement data used to define

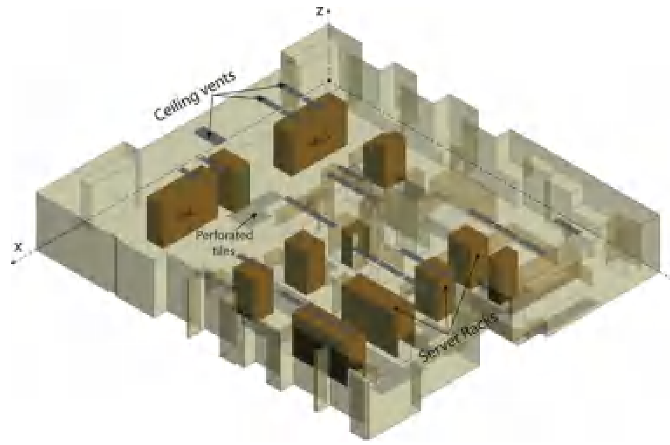


Fig. 4. Model of the data center used for CFD simulations.

Table 2
Material properties of fluid domain in CFD simulations.

Item	Value and/or description
Material	Ideal gas
Reference state	25 °C and 1 atm
Density	1.2 kg m^{-3}
Molar mass	28.96 kg kmol^{-1}
Specific heat at constant pressure	1.0044 kJ kg^{-1} K^{-1}
Dynamic viscosity	1.831×10^{-5} kg m^{-1} s^{-1}
Thermal conductivity	2.61×10^{-2} W m^{-1} K^{-1}

the boundary conditions of the CFD simulations. The measured volumetric airflow out of the DC was about 20% higher than the total measured airflow into the DC. This difference can be explained by air leaking through openings in the floor that were not covered by our measurements. In particular, a significant amount of air streams from the plenum directly into the racks because of openings underneath the racks used for cabling. We therefore extended Eq. (1) to account for this leakage airflow, i.e.

$$T_{out}^i = \frac{f_{in}^i}{f_{out}^i} T_{in}^i + \frac{f_{leak}^i}{f_{out}^i} T_{leak} + \frac{P_i}{\rho c_p f_{out}^i}, \tag{16}$$

where f_{in}^i is the airflow rate through the front of the rack into node i, f_{leak}^i is the airflow rate through the floor into the node, and f_{out}^i is the airflow rate out of the node. The temperature of the air leaking into the racks T_{leak} was assumed to be the same as the supply air temperature T_{sup}. With this assumption, the rest of the heat-flow

Table 3
Data used to define boundary conditions for the CFD simulations.

Parameter	Value
Minimum airflow rate at the server inlets[a]	0 m s^{-1}
Maximum airflow rate at the server inlets[a]	0.73 m s^{-1}
Total volumetric airflow into the DC through perf. floor tiles[a]	4.725 m^3 s^{-1}
Total volumetric airflow into the DC through HPCs[a]	7.670 m^3 s^{-1}
Total volumetric airflow out of the DC through ceiling vents[a]	15.720 m^3 s^{-1}
Total volumetric airflow into the DC through leakage[b]	3.325 m^3 s^{-1}
Flow rate of air leaking into the racks[c]	0.139 m s^{-1}
Supply air temperature	16.5 °C
Total power consumption by IT equipment[d]	219.73 kW

[a] Measured value.
[b] Calculated value: total outflow minus total inflow.
[c] Calculated value: total volumetric flow divided by total rack outlet area.
[d] Including 140 kW consumed by two HPCs.

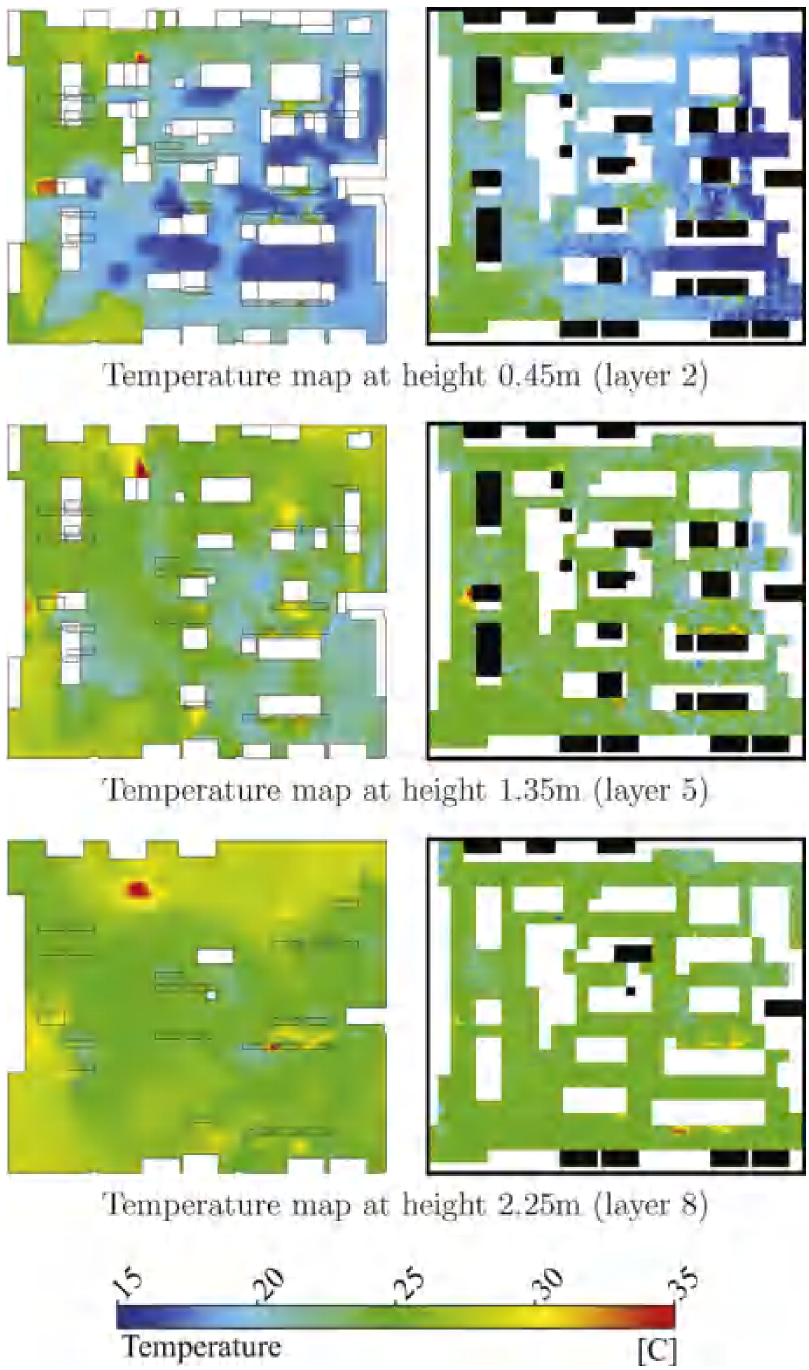


Fig. 5. Comparison of CFD simulation results (left) with measurement data obtained with the MMT cart (right).

model remains unchanged and Eq. (10) for calculating the cross-interference matrix is still valid.

ANSYS CFX was used to numerically solve the Reynolds Averaged Navier–Stokes equations with the Shear Stress Transport (SST) turbulence model to determine the airflow field and temperature distribution. The convergence criterion for the maximum normalized values of the equation residuals was set to 0.0001. The model converged in roughly 20 h on a workstation with a 3 GHz dual core processor.

Fig. 5 shows a comparison of CFD simulation results and measurement data. The temperature distributions match well except for two locations where the simulation indicates significantly higher temperatures around some rack outlets. This is partly because of the simplified model with 3 blocks of servers per rack. As described in the previous section, the power consumption was measured per rack and then distributed among the 3 nodes based on the assumption that the heat load of the individual servers is proportional to the measured airflow.

4.3. Identification of cross-interference matrix

To calculate the cross-interference matrix that describes the heat-flow among the 10 racks in the lower right corner of the data center, we used the CFD model introduced in the previous section. We set the power consumption of each node to $P_{max} = 2000\,\text{W}$ at full load and to $P_{idle} = 500\,\text{W}$ at idling state. The other parameters remained unchanged to match the operating conditions that

Table 4
Overview of heat-flow model.

Item	Value
Number of racks	10
Number of equipment nodes per rack	3
Total number of nodes	30
Power consumption of a node at full load	2000 W
Power consumption of a node at idle state	500 W

were determined during the measurement campaign. Table 4 gives a summary of the parameters for the heat-flow model.

We conducted a CFD simulation to obtain the temperature distribution for the reference state, in which all servers are idling, i.e.

$$\vec{\mathbf{P}}^{ref} = [P_{idle}, \ldots, P_{idle}]'. \tag{17}$$

To calculate the cross-interference matrix with dimension 30×30, the node inlet and outlet temperatures have to be determined for 30 distinct operating points, which are defined by linearly independent power distribution vectors. Furthermore, to minimize the influence of small errors in the CFD simulation on the resulting cross-interference matrix, it is essential that the operating points sufficiently differ from each other. In each case $J = 1, \ldots, 30$, we therefore set the power consumption of exactly one node to maximum load, i.e. $P_J = P_{max}$, while the other 29 nodes were idling, i.e. $P_j = P_{idle}$, where $1 \leq j \leq 30$ and $j \neq J$. The power distribution vector $\vec{\mathbf{P}}_{\mathbf{J}}$ for case J therefore was

$$\vec{\mathbf{P}}_{\mathbf{J}} = [P_{idle}, \ldots, P_{idle}, P_J = P_{max}, P_{idle}, \ldots, P_{idle}]'. \tag{18}$$

By using the CFD simulation results from the reference state to define the initial conditions, we significantly reduced the simulation time, since we only changed the power consumption of one node at a time: The simulation for the reference state converged in roughly 20 h, while the subsequent simulations only required about 6.5 h each.

The cross-interference matrix of the heat-flow model could then be calculated by using Eq. (10). The resulting coefficients are shown in Fig. 6. It is clear that along the diagonal the peaks are higher than in other areas of the bar chart. This implies that the hot air recirculation of a node causes the most impact on inlet temperatures of adjacent nodes. Fig. 7 gives an analysis on the recirculation that occurs at each layer of equipment nodes. This indicates that more hot air recirculation occurs at higher layers of server nodes. These results are consistent with the temperature maps obtained during the measurement campaign.

4.4. Optimization

We applied the novel load spreading technique to determine the optimum placement for upgraded equipment in the data center. As shown in Table 5, we investigated 5 different upgrading scenarios S1, S3, S5, S7, and S10, replacing 1, 3, 5, 7, and 10 out of 30 nodes.

For each upgrading scenario we considered 4 node positioning vectors $U_{n,k}$, $k = 1, 2, 3, 4$, where $n = 1, 3, 5, 7, 10$ is the number of nodes to be upgraded. Each element of $U_{n,k}$ corresponds to one of the 30 available node locations shown in Fig. 8 for the new equipment. Table 5 gives an overview of the upgrading scenarios and includes the number of possible combinations for the node positioning vector for each scenario. For each upgrading scenario, the first three node positioning vectors, $U_{n,1}$, $U_{n,2}$, $U_{n,3}$, were selected randomly for comparison, while $U_{n,4}$ is a result of the particle swarm optimization of the proposed load spreading technique.

The parameters for the PSO are given in Table 6. The power consumption of each new node was set to 500 W at idling state and to 3000 W at full load. We assumed a constant server utilization rate of 60%. In each iteration step of the PSO, the fitness of all particles was assessed by using the fitness function stated in Eq. (13). We defined some penalty functions in order to avoid undesired solutions. For example, the particle with the lowest average inlet temperature could represent an undesired positioning vector, i.e. lead to a situation where some node inlet temperatures heavily

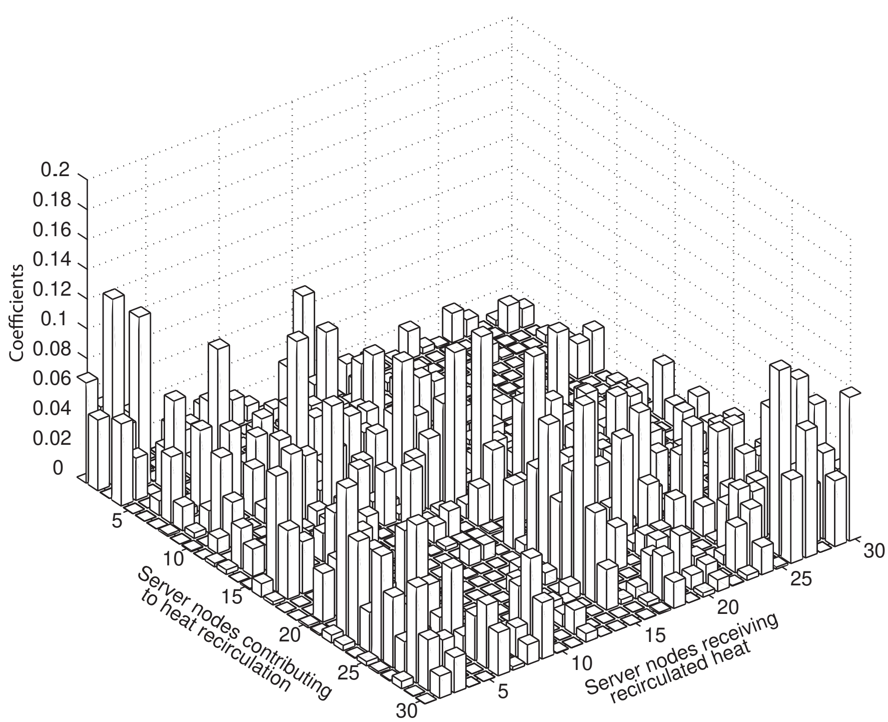
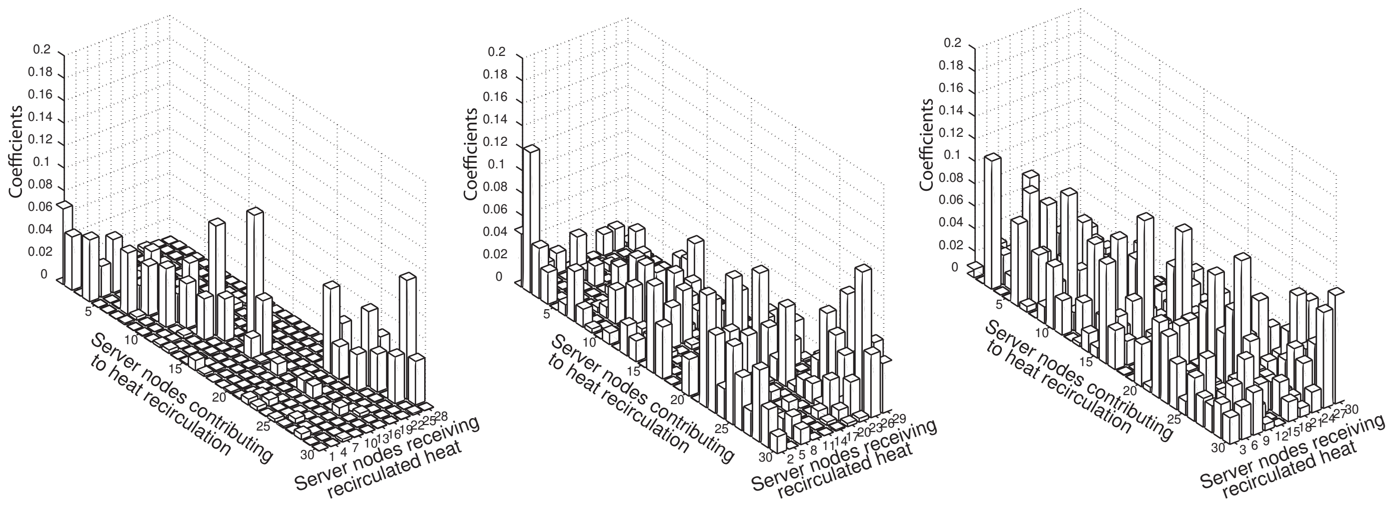


Fig. 6. Cross-interference matrix **A** calculated using CFD simulations.



(a) Bottom layer (b) Middle layer (c) Top layer

Fig. 7. Analysis of hot air recirculation at each layer of equipment nodes by decomposing the cross-interference matrix.

Table 5
Equipment node upgrading scenarios.

Upgrading scenario	Number of combinations	Positioning vectors for new equipment
S1	$\binom{30}{1} = 30$	$U_{1,1} = [8]$
		$U_{1,2} = [16]$
		$U_{1,3} = [23]$
		$U_{1,4} = [3]$
S3	$\binom{30}{3} = 4060$	$U_{3,1} = [7, 11, 16]$
		$U_{3,2} = [25, 28, 29]$
		$U_{3,3} = [16, 17, 20]$
		$U_{3,4} = [3, 6, 9]$
S5	$\binom{30}{5} = 142, 506$	$U_{5,1} = [7, 10, 11, 13, 16]$
		$U_{5,2} = [2, 15, 23, 25, 26]$
		$U_{5,3} = [3, 13, 17, 20, 30]$
		$U_{5,4} = [3, 6, 9, 18, 21]$
S7	$\binom{30}{7} = 2, 035, 800$	$U_{7,1} = [2, 12, 15, 23, 25, 26, 27]$
		$U_{7,2} = [4, 7, 20, 22, 24, 28, 30]$
		$U_{7,3} = [7, 10, 11, 13, 16, 23, 26]$
		$U_{7,4} = [1, 3, 6, 9, 15, 18, 21]$
S10	$\binom{30}{10} = 30, 045, 015$	$U_{10,1} = [17, 20, 23, 24, 25, 26, 27, 28, 29, 30]$
		$U_{10,2} = [8, 11, 13, 14, 16, 17, 18, 20, 22, 23]$
		$U_{10,3} = [1, 7, 10, 11, 13, 16, 17, 23, 26, 28]$
		$U_{10,4} = [1, 2, 3, 4, 6, 9, 12, 15, 18, 21]$

exceed the maximum allowed value. We therefore added a penalty to the average inlet temperature in Eq. (13) in order to make it very unlikely that such particles were selected as the global best particle. Repeating elements in the positioning vector also lead to invalid solutions. This was also countered by adding a penalty to the average inlet temperature.

Table 6
Parameters for particle swarm optimization (PSO).

Parameter	Value
Number of particles	30
Problem dimension	1, 3, 5, 7, 10
Maximum number of iterations	500
Range for each element of position vector	$1 \leq x_i \leq 30$
Range for each element of velocity vector	$-10 \leq v_i \leq 10$
Acceleration coefficient c_1	$0.5 \leq c_1 \leq 2.5$
Acceleration coefficient c_2	$0.5 \leq c_2 \leq 2.5$

4.5. Evaluation

We ran CFD simulations for all upgrading scenarios to determine the inlet temperatures and compared the results for all positioning vectors defined in Table 5. Except for the power consumption of the nodes, all model parameters were kept as described in Section 4.3. A summary of the results is given in Table 7.

At the bottom layer, where abundant cold air supply is available, the average inlet temperature increase is marginal in all trials. It never exceeds 0.5 °C. However, at the middle and even more at the top layer, differences between random upgrading and the proposed load spreading technique become evident. The number of

Table 7
Averaged equipment node inlet temperatures, average temperature increases, maximum temperature increases and the number of occurrences where inlet temperature increase by more than 0.5 °C.

Positioning vector	Bottom layer				Middle layer				Top layer			
	Average inlet temperature [°C]	Average temperature increase [°C]	Maximum temperature increase [°C]	No. of occurrences above 0.5 °C increase	Average inlet temperature [°C]	Average temperature increase [°C]	Maximum temperature increase [°C]	No. of occurrences above 0.5 °C increase	Average inlet temperature [°C]	Average temperature increase [°C]	Maximum temperature increase [°C]	No. of occurrences above 0.5 °C increase
$U_{1,1}$	17.00	0.01	0.1	0	21.74	0.08	0.7	1	26.94	0.11	0.9	1
$U_{1,2}$	17.01	0.02	0.1	0	21.80	0.14	0.9	1	26.97	0.14	1	1
$U_{1,3}$	17.04	0.05	0.1	0	21.80	0.14	0.3	0	26.98	0.15	0.5	1
$U_{1,4}$	17.00	0.01	0.1	0	21.78	0.12	0.4	0	26.98	0.15	0.4	0
$U_{3,1}$	17.02	0.03	0.1	0	21.71	0.05	0.3	0	26.99	0.16	0.8	1
$U_{3,2}$	17.13	0.14	0.3	0	21.94	0.28	0.6	3	27.45	0.62	1.4	6
$U_{3,3}$	17.01	0.02	0.1	0	21.79	0.13	0.7	1	27.01	0.18	0.6	1
$U_{3,4}$	17.03	0.04	0.1	0	21.81	0.15	0.3	0	26.97	0.14	0.3	0
$U_{5,1}$	17.02	0.03	0.1	0	21.90	0.24	0.7	1	27.22	0.39	1.3	3
$U_{5,2}$	17.09	0.1	0.2	0	21.94	0.28	0.6	2	27.18	0.35	0.6	4
$U_{5,3}$	17.03	0.04	0.1	0	21.80	0.14	0.5	1	27.10	0.27	1	2
$U_{5,4}$	17.04	0.05	0.1	0	21.90	0.24	0.4	0	27.00	0.17	0.4	0
$U_{7,1}$	17.13	0.14	0.3	0	22.06	0.4	0.8	3	27.34	0.51	0.9	6
$U_{7,2}$	17.04	0.05	0.1	0	21.79	0.13	0.4	0	27.20	0.37	0.9	3
$U_{7,3}$	17.04	0.05	0.1	0	21.79	0.13	0.4	0	27.15	0.32	0.7	4
$U_{7,4}$	17.07	0.08	0.2	0	21.93	0.27	0.8	1	27.12	0.29	0.6	2
$U_{10,1}$	17.15	0.16	0.4	0	22.54	0.88	1.6	9	27.61	0.78	1.9	7
$U_{10,2}$	17.03	0.04	0.1	0	21.98	0.32	1.5	2	27.37	0.54	1.1	4
$U_{10,3}$	17.08	0.09	0.2	0	21.97	0.31	0.8	1	27.62	0.79	1.3	8
$U_{10,4}$	17.13	0.14	0.3	0	21.93	0.27	0.7	2	27.27	0.44	0.7	7

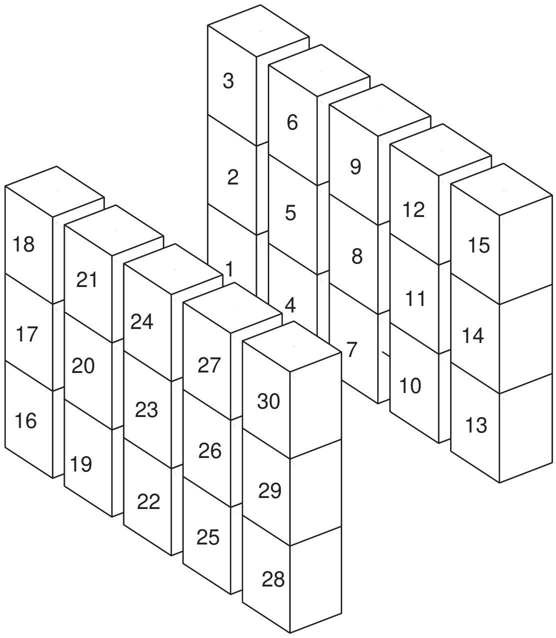


Fig. 8. Equipment nodes in the area of interest in the DC (cf. Fig. 2).

occurrences where the inlet temperature increases by more than 0.5 °C gets higher as the number of replaced nodes increases. For upgrading scenarios S1, S3 and S5, the results for the positioning vectors obtained with proposed load spreading technique show no inlet temperature increase above 0.5 °C. It is also apparent that the proposed method does not lead to inlet temperatures that exceed the maximum allowed inlet temperature. On the other hand, we notice that random equipment upgrades can cause hot spots even if just one server is replaced: For instance, positioning vector $U_{1,2}$ leads to an inlet temperature increase of more than 0.5 °C at the middle and top layers. For the upgrading scenario S7, the result obtained with the proposed load spreading technique is among the best results in terms of average inlet temperature increase. Given that there are more than 2 million possible ways to place the new equipment in that case, PSO achieves a good result after only 500 iterations. For the upgrading scenario S10, large temperature increases at the inlets are inevitable. However, the proposed method spreads the temperature increase evenly across all the nodes. Even though we have a high number of nodes with temperature increase above 0.5 °C in the top layer, the proposed method allows us to significantly reduce the maximum temperature increase compared to random upgrading.

It is important to maintain the inlet temperatures within the manufacturer's specified temperature ranges to ensure proper operation of the equipment. Hot spots caused by unplanned equipment upgrades are often countered by DC operators by reducing the supply air temperature. This leads to an overall lower cooling efficiency of the DC. With the proposed novel load spreading technique, DC operators can maintain a stable cooling environment. Cooling system performance deterioration and risk of equipment malfunction is minimized by keeping a uniform equipment inlet temperature distribution.

5. Conclusions

The provision of an efficient cooling environment is important for the reliable and economically successful operation of an air-cooled data center. Whenever computer equipment has to be upgraded, care has to be paid to avoid significant changes of the airflow and the thermal distribution in the DC room. For this purpose, we developed a novel upgrading strategy that allows replacing the equipment in the computer racks with minimal thermal impact on the existing optimized DC cooling environment. Our approach is based on an abstract heat-flow model of the DC, whose parameters are determined by means of performing a measurement campaign in the DC and with support of CFD simulations. The optimum placement of the new equipment in the racks of the DC is found by applying a particle swarm optimization technique to this model. The effectiveness of our method has been assessed by comparing it to the conventional necessity based equipment upgrading. Based on experiments performed in a production DC and simulations, we demonstrated that it is possible to systematically distribute the inlet temperature increases among equipment nodes in such a way that no hot spots are formed. The proposed method outperforms the commonly applied load spreading technique because our holistic approach takes into account not only the increased heat load of the upgraded equipment, but also the airflow and temperature distribution in the DC room.

Acknowledgments

The authors would like to acknowledge the support of G08/DAAD grant (2010/2011) on optimization in wireless sensor networks for travel support.

References

[1] R. Schmidt, E. Cruz, M. Iyengar, Challenges of data center thermal management, IBM Journal of Research and Development 49 (2005) 709–723.
[2] H.F. Hamann, J.A. Lacey, M. O'Boyle, R.R. Schmidt, M. Iyengar, Rapid three-dimensional thermal characterization of large-scale computing facilities, IEEE Transactions on Components and Packaging Technologies 31 (2008) 444–448.
[3] J. Siriwardana, W. Schott, S. Halgamuge, The power grabbers, IEEE Power and Energy Magazine 8 (2010) 46–53.
[4] N. Rasmussen, Cooling strategies for ultra-high density racks and blade servers, APC White Paper 46 (2009).
[5] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of IEEE International Conference on Neural Networks, vol. 4, 1995, pp. 1942–1948.
[6] S.K. Shrivastava, J.W. VanGilder, B.G. Sammakia, Optimization of cluster cooling performance for data centers, in: Proceedings of 11th IEEE Intersociety Conference on Thermal and Thermochemical Phenomena in Electronic Systems, 2008, pp. 1161–1166.
[7] J. Moore, J. Chase, P. Ranganathan, R. Sharma, Making scheduling "cool": temperature-aware workload placement in data centers, in: Proceedings of the Annual Conference on USENIX Annual Technical Conference, USENIX Association, Berkeley, CA, USA, 2005, pp. 61–75.
[8] Q. Tang, S.K.S. Gupta, G. Varsamopoulos, Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: a cyber-physical approach, IEEE Transactions on Parallel and Distributed Systems 19 (2008) 1458–1472.
[9] Q. Tang, T. Mukherjee, S. Gupta, P. Cayton, Sensor-based fast thermal evaluation model for energy efficient high-performance data centers, in: Proceedings of 4th International Conference on Intelligent Sensing and Information Processing, 2006, pp. 203–208.
[10] M.K. Patterson, The effect of data center temperature on energy efficiency, in: Proceedings of 11th Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems ITHERM 2008, 2008, pp. 1167–1174.
[11] A. Ratnaweera, S. Halgamuge, H. Watson, Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients, IEEE Transactions on Evolutionary Computation 8 (2004) 240–255.
[12] F. van den Bergh, A. Engelbrecht, A cooperative approach to particle swarm optimization, IEEE Transactions on Evolutionary Computation 8 (2004) 225–239.
[13] S. Patchararungruang, S. Halgamuge, N. Shenoy, Optimized rule-based delay proportion adjustment for proportional differentiated services, IEEE Journal on Selected Areas in Communications 23 (2005) 261–276.
[14] K. Steer, A. Wirth, S. Halgamuge, Control period selection for improved operating performance in district heating networks, Energy and Buildings 43 (2011) 605–613.

[15] Y. Bichiou, M. Krarti, Optimization of envelope and HVAC systems selection for residential buildings, Energy and Buildings 43 (2011) 3373–3382.
[16] ANSYS CFX, 2011. http://www.ansys.com/Products/Simulation+Technology/Fluid+Dynamics/ANSYS+CFX.
[17] P. Biller, P. Chevillat, F. de Lorenzi, T. Scherer, W. Schott, R. Ullmann, C. Vmel, Efficient cooling of data centers, in: Proceedings of the 2011 World Engineers' Convention, 2011.
[18] 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, M. O'Boyle, Uncovering energy-efficiency opportunities in data centers, IBM Journal of Research and Development 53 (2009) 10:1–10:12.
[19] FlowHood CFM-88L, Shortridge Instruments, 2011. http://www.shortridge.com/flowhood_specs.htm.
[20] J. Cho, T. Lim, B.S. Kim, Measurements and predictions of the air distribution systems in high compute density (internet) data centers, Energy and Buildings 41 (2009) 1107–1115.
[21] J.F. Karlsson, B. Moshfegh, Investigation of indoor climate and power usage in a data center, Energy and Buildings 37 (2005) 1075–1083.