International Journal of Heat and Mass Transfer 108 (2017) 374–385

Contents lists available at ScienceDirect

International Journal of Heat and Mass Transfer

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

Steady and transient behavior of data centers with variations in thermal load and environmental conditions

Jingru Zhang, Yogesh Jaluria *

Mechanical Engineering Department, Rutgers University, Piscataway, NJ 08854, United States

A R T I C L E I N F O

Article history:
Received 3 October 2016
Received in revised form 8 December 2016
Accepted 9 December 2016

A B S T R A C T

Data centers are extensively used for the storage and retrieval of large amounts of data. Efficient removal of the energy dissipated by the electronic circuitry is a major concern since the temperature must not rise beyond acceptable levels that affect the performance and since the energy required for cooling is generally quite substantial, with typical data centers requiring hundreds for megawatts for heat removal. Though steady operation has been of interest in most studies on data center cooling, transient behavior is also of considerable interest since the load is generally not a constant but varies with time. Also, distribution of load between various data centers may be used effectively to reduce the overall power consumption. The environmental conditions may be used advantageously in the selection of a particular data center in order to reduce the cooling load. All these considerations make it imperative to study both the steady and time-dependent behavior of data centers and to obtain results that may be used to optimize the cooling systems and reduce the overall energy consumption. This paper presents an analytical and numerical study that focuses on load distribution, as well as data center location that determines the relevant environment, in order to obtain characteristic results on the resulting temperature distributions for a range of operating conditions. Such results could form the basis for optimizing the cooling strategy and the system.

1. Introduction

The increased demands on data storage make it necessary to investigate the cooling system for data centers since it is critical to meet the desired temperature constraints on the components for satisfactory performance. Also, the corresponding power required for cooling is generally quite substantial [1–4]. Recent years have seen considerable interest and research activity in this area, particularly for steady state operation of data centers [5–9]. However, it is also important to study the effect of thermal load changes on the response of the system and possible optimization to reduce the energy usage. This could lead to significant savings in costs and in environmental impact. Thus, both steady and transient operations are of interest.

In this paper, temperature and flow distributions for both steady state and transient circumstances are studied. The objective of the study is to control the temperature in the data center at an allowable range, as well as reduce the energy consumed and the cost for the cooling system. Microchips with high heat dissipation result in significant cooling challenges at the data center facility level. In a typical data center, the power consumption on cooling and thermal management is as much as 45% of the total power consumption. For a 10 MW facility, this would lead to a cost of up to $3.9 M annual cost assuming energy price of $0.1 per kW h. Therefore, the thermal management is one of the most importance challenging tasks for data center and computer room managers.

There are two classes of thermal management policies for data centers: those that manage temperature under normal operation and those that manage thermal emergencies. The objective for normal operation thermal management is to reduce the cooling cost. On the other side, a large increase in load that causes temperatures to rise quickly can be considered a thermal emergency. The main objective for managing thermal emergencies is to control temperatures while avoiding unnecessary performance degradation without excessive cooling capacity. In this work, both thermal management policies have been studied.

The air distribution within a data center has a major impact on the temperature distribution of the equipment located in the rooms. In some data centers, the cold air enters the data center from the ceiling through diffusers, and exits the room via vents on the sides of the room. Most of the data centers use hot aisle/cold aisle layout, which is designed to supply cold air through a raised floor. Computer room air conditioning units (CRACs) are used to pump the cooling air into the plenum underneath the data center

* Corresponding author.
E-mail address: jaluria@jove.rutgers.edu (Y. Jaluria).

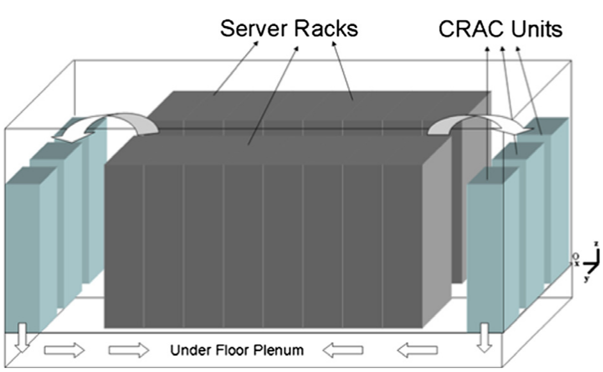
http://dx.doi.org/10.1016/j.ijheatmasstransfer.2016.12.028
0017-9310/

room. There are perforated tiles on the floor that are used to replace the solid tiles and allow air to enter the space above the floor. The aisles with perforated tiles are the cold aisles. The aisles without the cold air delivery are the hot aisles. Raised floor data center is popular because of its flexibility. The floor tiles are designed to be removable. If the layout arrangement of the server racks is changed, the corresponding perforated tile locations can be changed so that the cold air can be delivered to where the hot rack is located. Different configurations and operating conditions have been considered in the literature to obtain the flow and temperature distributions, and heat transfer rates [10–13]. Various numerical models have been developed to simulate data centers, which represent a fairly complex thermal system [14,15].

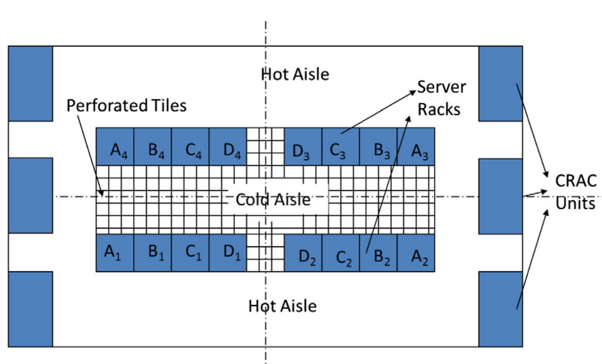
This paper presents the results of a numerical study on the steady and transient operation of data centers under different environmental conditions. Changes in the thermal load are also considered. Different scenarios where the cooling system is started before the data center is subjected to a major load increase are considered and the corresponding results presented. The study could be used for optimization of multiple data centers, with respect to load and location, to achieve considerable savings in energy usage.

2. Mathematical and numerical modeling

A physical model based on the configuration of a simple, but realistic, data center, including the under-floor plenum, server rack, CRACs, and perforated tiles, is shown in Fig. 1. The cooling system is typical of common data centers. This data center model consists of 6 CRAC units and 16 electronic racks (each 1 m × 1 m × 2 m tall). The overall dimension of the data center is 7 m × 8 m × 3 m. The under-floor plenum height is 0.5 m. Each



(a) Sketch of the Data Center



(b) Top View of the CFD model used to simulate a typical data center.

Fig. 1. Schematic of a typical data center and of the computational model. (a) Sketch of the Data Center. (b) Top View of the CFD model used to simulate a typical data center.

rack is assumed to contain 24 servers, and the dimension of each server is 0.43 m × 0.22 m × 0.046 m (17 in × 8.5 in × 1.8 in). Each rack has a power of 10 kW when fully utilized.

The CRAC units discharge cold air in the under floor plenum and the air is delivered to the raised floor through perforated tiles, then the hot air returns to the CRAC units. The entire data center is symmetric so one fourth of the room was selected as the computational domain to reduce the simulation overheads. Different utilization levels are modeled as racks in of the data center being turned on/off.

3. Governing equations

The flow in the data center is taken as turbulent flow, which instantaneously satisfies the Navier-Stokes equations as given below from Kundu and Cohen [16]:

$$\frac{\partial \tilde{u}_i}{\partial x_i} = 0 \tag{1}$$

$$\frac{\partial \tilde{u}_i}{\partial t} + \tilde{u}_j \frac{\partial \tilde{u}_i}{\partial x_j} = -\frac{1}{\rho}\frac{\partial \tilde{p}}{\partial x_i} - g[1-\alpha(\tilde{T}-T_0)]\delta_{i3} + v\frac{\partial^2 \tilde{u}_i}{\partial x_j \partial x_j} \tag{2}$$

$$\rho C_p\left(\frac{\partial \tilde{T}}{\partial t} + \tilde{u}_j \frac{\partial \tilde{T}}{\partial x_j}\right) = k\frac{\partial^2 \tilde{T}}{\partial x_j \partial x_j} \tag{3}$$

where u_i represents the velocity components, T the temperature, x_i the coordinates, t the time, g the magnitude of the gravitational acceleration, α the coefficient of volumetric thermal expansion and k the thermal conductivity. The tilde is used to indicate physical dimensional quantities. However, it is very challenging to predict the flow in detail since there are different scales to be resolved. The averaged equations are used to find to mean velocity and temperature of a turbulent flow. The simplest "complete models" of turbulence are the two equation models in which the solution of two separate transport equations allows the turbulent velocity and length scales to be independently determined.

A two-equation turbulent model $\kappa - \varepsilon$ model in Ansys Fluent 12.0 was applied with standard wall functions to solve the equations listed above. The SIMPLE algorithm is used to fully resolve the linear pressure-velocity coupling. The QUICK scheme is used to solve the convection-diffusion equations.

The turbulence kinetic energy κ and its rate of dissipationε are obtained from the following transport equations:

$$\frac{\partial}{\partial t}(\rho\kappa) + \frac{\partial}{\partial x_i}(\rho\kappa u_i) = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_k}\right)\frac{\partial \kappa}{\partial x_j}\right] + G_k + G_b - \rho\varepsilon - Y_M \tag{4}$$

and

$$\frac{\partial}{\partial t}(\rho\varepsilon) + \frac{\partial}{\partial x_i}(\rho\varepsilon u_i) = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial x_j}\right] + C_{1\varepsilon}\frac{\varepsilon}{k}(G_k + C_{3\varepsilon}G_b) - C_{2\varepsilon}\rho\frac{\varepsilon^2}{\kappa} \tag{5}$$

Here, G_k is the generation of turbulent kinetic energy due to the mean velocity gradient and G_b that due to buoyancy. Y_M represents the contribution of the fluctuating dilatation in compressible turbulence to the overall dissipation rate:

$$Y_M = 2\rho\varepsilon\frac{\kappa}{a^2} \tag{6}$$

In this study, air is treated as incompressible, so this term is zero. Also, $C_{1\varepsilon}$, $C_{2\varepsilon}$, $C_{3\varepsilon}$ are constants, σ_k and μ_t are the turbulent Prandtl numbers for κ and ε, respectively. The corresponding values are 1.44, 1.92, 0.09, 1.0 and 1.3, as given in many references, such as [17,18].