International Journal of Heat and Mass Transfer 71 (2014) 69–78

Contents lists available at ScienceDirect

International Journal of Heat and Mass Transfer

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

Long-term transient thermal analysis using compact models for data center applications

Zhihang Song *, Bruce T. Murray, Bahgat Sammakia

Department of Mechanical Engineering, Binghamton University, Binghamton, NY, USA

A R T I C L E I N F O

Article history:
Received 30 August 2013
Received in revised form 27 November 2013
Accepted 2 December 2013

Keywords:
POD
NLPCA
Compact model
Thermal modeling
Server room

A B S T R A C T

Reduced-order thermal models are necessary to enable real-time assessment of the optimum operating and control conditions to improve data center energy efficiency. A 3D computational fluid dynamics (CFD) and heat transfer model of a basic raised floor, air cooled, hot aisle/cold aisle data center configuration was developed and simulation results were used to generate snapshots for initializing compact models based on the Proper Orthogonal Decomposition (POD) and the Nonlinear Principle Component Analysis (NLPCA) methods. The specific focus of the study was to numerically investigate how the thermal trends can be affected by the long-term transient flow patterns associated with leakage and how the compact models (e.g. POD and NLPCA) can be utilized for much faster implementations and characterizations of the transient flows. Using both the POD and the NLPCA method, good agreement was achieved between the full simulation results and the compact models for predicting the dynamically developed local flow structure over a range of transient cooling air supply operating conditions. In addition, the NLPCA method was implemented to better characterize the nonlinear aspects of the CFD results. The benefits of using both the POD and NLPCA methods are discussed in relation to constructing compact models as real-time predictive tools. A systematic use of the compact models has also been proposed, to enable more robust thermal management and control of data centers.

1. Introduction

Due to the need to improve energy efficiency in the thermal management of data centers, the number of experimental and computational studies in this area is increasing rapidly. A significant portion of this research involves full-scale computational fluid dynamics simulations of the turbulent flow and temperature distributions in air-cooled, raised floor data centers. In order for the full-scale CFD simulations to adequately capture the high-dimensional, complex-flow physics and energy transport, substantial computational effort is required. An extremely large number of discrete equations and degrees of freedom result from approximation methods such as the finite volume or finite element approaches. The solution of the discrete systems, especially for transient behavior, generally requires long computation times, which make it difficult to achieve real-time assessment. Fast response in modeling the thermal environment is necessary in order to develop a feedback control system to minimize energy usage. Thus, reduced-order or compact models are required as a predictive tool for generating the response of the thermal environment to various operating conditions.

The POD method is one approach that can be used to develop a reduced order model based on numerical or experimental data. However, there are some disadvantages in the traditional POD method, because the selection of higher order modes and the number of modes impacts the accuracy and reliability of the predictions. Difficulties occur because a nonlinear coupling relationship exists between the high-order modes and the low-order modes in the nonlinear momentum and energy transport models that are sensitive to the initial and boundary conditions. A long-term perturbation (such as unreasonably truncated higher-order POD modes) may cause a change in the nature of the topology of the system. It was found that the long-term vortex behavior described by the developed reduced-order 12 POD modal low-dimensional systems did not match the numerical simulation results when time increases beyond 25 s [1]. POD reduced-order modeling was applied to the problem of flow around a cylinder in [2], which concluded that the traditional POD cannot accurately describe the transient behavior of the original system and long-term behavior with a varying Reynolds number. Ref. [3] shows that the included shift-mode significantly improves the resolution of the transient dynamics from the onset of the vortex shedding to the periodic von Karman vortex street. The inclusion of the shift-mode can precisely describe the dependence of the flow on Reynolds number. However, an unstable steady-state solution of the original system

* Corresponding author. Tel.: +1 315 380 8882.
E-mail address: zsong1@binghamton.edu (Z. Song).

0017-9310/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijheatmasstransfer.2013.12.007

is required for the approach, which increases both the difficulty and time of the computation.

The Galerkin projection-based method is an alternative procedure to solve for the POD coefficients using a set of coupled nonlinear time-based ordinary differential equations for the transient stage. Earlier investigations, such as [4], utilize this method to create reduced-order models of transient flow and temperature fields in terms of a parametric study, for channel flow with easily homogenized boundary conditions. The POD-Galerkin approach was employed in conjunction with the finite volume method (FVM) to generate a reduced-order model for a transient 1D heat transfer problem [5], and both POD based interpolation and projection methods were applied to multivariate heat conduction for fast thermal predictions [6]. Nie and Joshi [7] combined the flux-matching POD and Flow Network Method (FNM) to achieve a multiple scale reduced model strategy for chip-level conductive through rack-level convective transport.

The POD method has been widely used in the field of data center reduced-order modeling [8–11]. More recently, Ghosh and Joshi [12] developed a dynamic POD model to predict the transient temperature response for a transient start-up scenario (server heat load and cooling air flow rate). When time was used as the only variable, the interpolation technique was used to efficiently determine the POD coefficients instead of the more traditional Galerkin based method. However, the methodology was found to be unsuitable to solve a multivariate problem (e.g., transient flow and temperature field with variable cooling load). In addition, in the study the temperature field was expressed by a set of linear basis functions alone. The current study focuses on the development and implementation of an improved method to characterize the nonlinear behavior such that a reduced-order, dynamic predictive capability can be achieved for real time control and optimization of the energy used to provide sufficient cooling of a data center.

2. Data center computational model

The model of the basic raised floor data center configuration is shown in Fig. 1. For the representative study presented here, the server room is taken to be 8.75 m long and 6.4 m wide. The cooling air is supplied from a single CRAC unit located on one end through the raised floor plenum (0.914 m height) and returned through an overhead plenum mounted on the ceiling (1.524 m height), as seen in Fig. 2. The height between the raised floor and the drop ceiling plenum is 2.7 m. Under steady conditions, the CRAC flow rate (downward to the under-floor plenum) is specified as 5.3 m^3/s (at 80% capacity) and the air temperature is set at 14 °C. The same amount of air returns to the CRAC via the ceiling plenum. In the data center model, two rows of five server racks (2.13 m high) are located on either side of the center cold aisle. This single cold

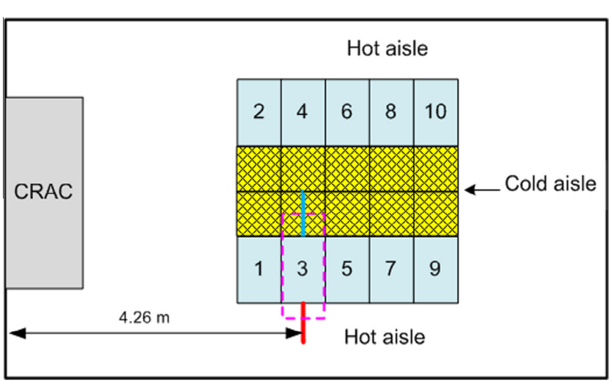

Fig. 1. Data center layout and specifications (top view).

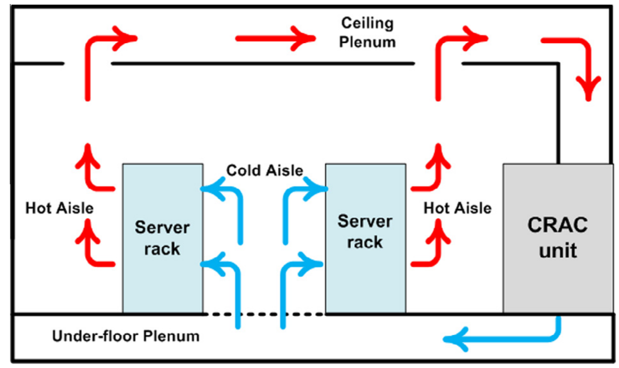

Fig. 2. Basic hot aisle/cold aisle raised-floor data center.

aisle room configuration is similar to the one that was characterized experimentally and modeled using the POD approach by Ghosh and Joshi [13]. The racks are modeled as enclosures with a heat source and internal fans. The total airflow required by the server fans is equal to the CRAC supply. The cold aisle consists of two rows of perforated tiles (0.61 × 0.61 m^2; 56% porosity). Fig. 3 provides more details about the model specifications; air resistance modeling was used to characterize the flow distribution through the perforated tiles [14]. In this approach, a local pressure drop is instantaneously produced at the foot of the server racks due to the air injected from the perforated tiles. In this study, the large-scale flow and temperature simulations were performed using the commercial CFD software package FloTHERM [15].

For the simulations performed using the FloTHERM software package, a finite-volume approximation of the Reynolds-averaged Navier–Stokes (RANS) and energy equations with a standard k–ε turbulence model is used. The continuity, momentum and energy equations of the 3D constant-density mean flow are expressed as follows.

$$\nabla \cdot \vec{u} = 0$$

$$\frac{\partial \vec{u}}{\partial t} + \vec{u} \cdot \nabla \vec{u} = \nabla \cdot (\nu_{eff} \nabla \vec{u}) - \frac{1}{\rho} \nabla p + \vec{g}$$

$$\rho c_p \left[\frac{\partial T}{\partial t} + (\vec{u} \cdot \nabla) T \right] = \nabla \cdot (k_{eff} \nabla T) + S \tag{1}$$

where, ν_{eff} and k_{eff} are effective fluid viscosity and thermal conductivity, respectively. The term $\vec{g}$ is the gravitational acceleration vector, and the term S represents the heat source.

The convective terms were discretized using a second-order upwinding numerical scheme and the SIMPLEST algorithm [16] used to evaluate the coupled velocity and pressure field. A thermal/airflow solver similar to Patankar's formulation in [17] was

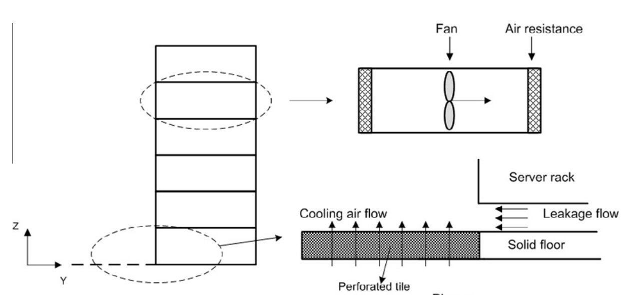

Fig. 3. Model specifications for server racks and perforations.