International Journal of Heat and Mass Transfer 127 (2018) 890–900

Contents lists available at ScienceDirect

International Journal of Heat and Mass Transfer

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

Real-time temperature predictions in IT server enclosures

Hosein Moazamigoodarzi [a,b], Souvik Pal [a], Suvojit Ghosh [a], Ishwar K. Puri [a,b,*]

[a] Computing Infrastructure Research Centre, McMaster University, Hamilton, Ontario, Canada
[b] Department of Mechanical Engineering, McMaster University, Hamilton, Ontario, Canada

ARTICLE INFO

Article history:
Received 11 June 2018
Received in revised form 15 August 2018
Accepted 20 August 2018

Keywords:
Rack mountable cooling units
Temperature prediction
Zonal method
Flowrate mismatch
Energy balance

ABSTRACT

Current data center (DC) cooling architectures are inefficient due to (1) inherent airflow efficiencies and (2) their inability to spatiotemporally control cooling airflow and DC temperatures on demand. Rack-based cooling is a promising recent alternative since it provides more effective airflow distribution and is more amenable to rapid real-time control. A control scheme should be able to predict spatiotemporal temperature changes as the system configuration and parameters change, but a suitable method is not yet available. Existing approaches, such as proper orthogonal decomposition or machine learning are unsuitable because they require an inordinately large number of *a priori* simulations or experiments to generate a training dataset. We provide an alternative real-time temperature prediction tool which requires no *a priori* training for DC server enclosures into which a rack mountable cooling unit (RMCU) has been integrated. This new model is validated with experimental measurements and its applicability is demonstrated by separately evaluating the influence of varying IT server configuration, RMCU flowrate, step changes in system conditions, and interactions between multiple RMCUs. The resulting tool will facilitate advanced control techniques and optimize design for any DC rack-based cooling architecture.

1. Introduction

A third of the energy provided to a conventional data center (DC) cools its IT equipment (ITE) [1–3]. Typically, traditional DCs use air cooling systems because of their relatively reliability, and lower capital and maintenance costs [4,5]. Cooling units for delivering cold air are placed either along the server room perimeter or in-row along racks. These architectures contain two significant air distribution inefficiencies, i.e., hot air recirculation and cold air bypass [6–8]. Therefore, to protect the ITE, typically the provided cooling airflow equals two times the required amount [9].

To reduce redundant airflow, a rack-based cooling architecture that is integrated entirely within a single, IT rack is an alternative means for supplying cold air in close proximity to servers [10,11]. A rack mountable cooling unit (RMCU) with dimensions similar to those of a server can be mounted conveniently inside a standard rack [12–14], considerably shortening airflow paths, lowering fan power and facilitating better airflow distribution [7,11]. The shorter airflow path enables more rapid regulation of cooling in response to dynamically changing ITE demand. The faster cooling control eliminates large temperature fluctuations when there are rapid changes in IT load, significantly mitigating equipment failure [4]. However, this is possible only when control actions occur within a duration $t_e < \tilde{t}$, where the latter denotes the timescale over which significant DC thermal events occur. For realistic rack-based DC cooling, $\tilde{t} \sim 1$ s [15–17]. Real-time control in a rack mountable cooling architecture requires a rapid scheme to predict temperature. This is not yet possible since candidate methods, such as (1) proper orthogonal decomposition (POD), (2) machine learning, and (3) heuristic models, have limitations that prevent practical implementation.

Models based on POD and machine learning predict changes in DC temperature distribution faster than full-field simulations [18–21]. Machine learning methods to train model parameters are classified as black box or grey box. Black box approaches relate outputs, e.g., temperatures, to inputs through equations that ignore the flow physics [22–26]. In contrast, a grey box method considers some aspects of the physics while ignoring others. With both POD, and all machine learning approaches, empirical parameters must be trained using sample datasets that are obtained either from CFD simulations or experiments.

This poses two challenges. First, the development of training datasets that are statistically significant is nontrivial. Based on the numbers of input and output parameters, the corresponding number of simulations can easily range from $\sim 10^2$–10^3, requiring computational time of the order of days for typical 3D DC

* Corresponding author at: McMaster University, 1280 Main St. W., Hamilton, ON L8S 4L7, Canada.
E-mail address: ikpuri@mcmaster.ca (I.K. Puri).

https://doi.org/10.1016/j.ijheatmasstransfer.2018.08.091
0017-9310/

simulations, as well as dedicated access to supercomputing clusters. Performing so many specific experiments is also generally impractical. The second limitation of is that test data are similar to the training data. Hence, when the physical configuration differs from one used to obtain training data, the algorithms must extrapolate, which degrades performance and reliability.

In contrast, heuristic approaches express the physical behavior of the system through strategic simplifications of rigorous physical laws to predict air temperatures at discrete locations, such as server inlets and outlets [27–31]. However, simplification still requires empirical parameters. For instance, a rapid CFD and lumped capacitance hybrid model can predict server inlet temperature changes due to transient events, such as server shutdown, chilled water interruption, and failure of the computer room air handlers [29]. However, the model requires CFD simulations for each case to determine unknown parameters and index values. As another example, a three-dimensional pressurized zonal model for room-based cooling with a raised floor can be employed to predict the temperature distribution [28]. Here, the characteristic dimension is typically limited to ~1 m, which is too large to accurately predict temperatures at server inlets. Furthermore, the model requires information about mass flowrates through computationally expensive CFD simulations. In both examples, obtaining real-time temperature distributions is unfeasible.

Instead, we propose an original parameter-free transient zonal model to obtain real-time temperature distributions inside a typical DC IT rack that is contained within an enclosure cooled by an RMCU with separated cold and hot chambers. The model is based on mass and energy conservation relations for each zone within the enclosure. Because of its geometry, the flow field can be resolved using fluid mechanics principles, avoiding the need for CFD simulations, experiments or training of empirical parameters.

Our objectives are to (1) describe the application of the zonal transient model for a specific configuration, (2) validate the model with experimental results, (3) investigate the effect of RMCU operational parameters on the temperature distribution, (4) analyze the influence of IT load distribution on thermal performance, and (5) compare thermal performance when a single RMCU is used versus two RMCUs placed within the enclosure.

2. Methodology

2.1. System configuration

The zonal model is an intermediate method between full CFD simulations and multi-node lumped models to calculate temperatures. This method considers mass and energy transport in a space that is partitioned into a coarse number of zones to which conservation relations are applied. Physical quantities, such as temperature, are assumed uniform within a zone, eliminating spatial dependence. Hence, the partial differential equations for mass, momentum and energy conservation are reduced to a system of ordinary differential equations, significantly reducing solution time compared to CFD simulations. Various zonal models have been developed for HVAC and building energy management [28,32–36].

The geometry and zones for a DC IT rack within an enclosure that is cooled by an RMCU are shown in Fig. 1. Assuming no heat and mass transfer between the enclosure and the ambient, four control volumes are identified, i.e., (1) the cold chamber in front of each server, (2) the hot chamber at the back of each server, (3) each server itself, and (4) the RMCU. The RMCU is a heat removal module in the form of a plate fin heat exchanger that transfers heat to a chilled water loop supplied from an external chilled water system.

If the entering and exiting air flowrates for the zones are known, the energy balance for each zone can be solved. While a CFD simulation is required to determine the flowrates within a conventional DC room, for the architecture of Fig. 1 these airflows are readily determined from mass conservation relations alone. Fig. 2 depicts a schematic of the airflows within the enclosure, where there are two prime movers, the fans inside the RMCU and in the servers. Server fans draw in cold air from the cold chamber, remove heat from the servers and transport the warmer air to the hot chamber. The RMCU draws warm air from the hot chamber, extracts heat from it and releases cooled air into the cold chamber. A leakage airflow between the hot and cold chambers occurs due to the pressure difference between these two chambers $\Delta P = P_h - P_c$

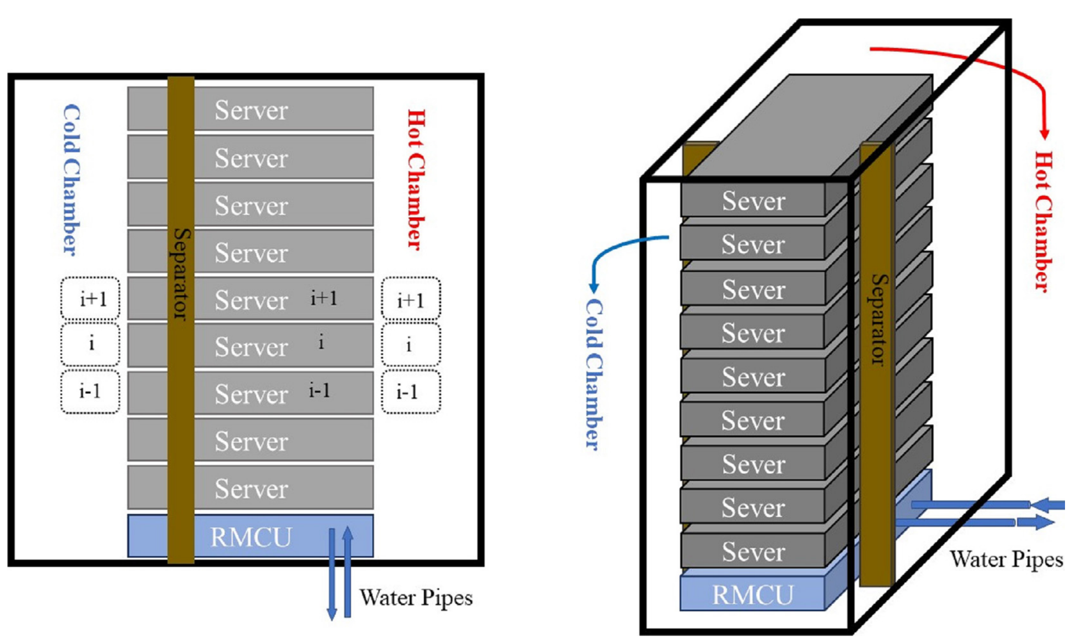

Fig. 1. Schematic of the IT enclosure integrated with a single rack and an RMCU with separated cold and hot chambers. The zones (control volumes) in the front and back chambers are shown.