Future Generation Computer Systems 71 (2017) 157–170

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Spatio-temporal thermal-aware scheduling for homogeneous high-performance computing datacenters

Hongyang Sun [a,*], Patricia Stolf [b], Jean-Marc Pierson [b]

[a] Ecole Normale Superieure de Lyon & INRIA, France
[b] IRIT, University of Toulouse, France

HIGHLIGHTS

- Thermal model capturing both spatial and temporal temperature correlations in datacenters.
- Formulation of a spatio-temporal thermal-aware scheduling problem for HPC applications.
- Scheduling heuristic using thermal-aware load for job assignment and thermal management.
- Simulations to show the effectiveness of heuristic under a wide range of parameters.

ARTICLE INFO

Article history:
Received 15 March 2016
Received in revised form 7 January 2017
Accepted 5 February 2017
Available online 7 February 2017

Keywords:
HPC datacenters
Thermal model
Spatio-temporal correlation
Thermal-aware scheduling
Makespan
Energy consumption
DVFS

ABSTRACT

Datacenters have become an important part of today's computing infrastructure. Recent studies have shown the increasing importance of thermal considerations to achieve effective resource management. In this paper, we study thermal-aware scheduling for homogeneous high-performance computing (HPC) datacenters under a thermal model that captures both spatial and temporal correlations of the temperature evolution. We propose an online scheduling heuristic to minimize the makespan for a set of HPC applications subject to a thermal constraint. The heuristic leverages the novel notion of thermal-aware load to perform both job assignment and thermal management. To respect the temperature constraint, which is governed by a complex spatio-temporal thermal correlation, dynamic voltage and frequency scaling (DVFS) is used to regulate the job executions during runtime while dynamically balancing the loads of the servers to improve makespan. Extensive simulations are conducted based on an experimentally validated datacenter configuration and realistic parameter settings. The results show improved performance of the proposed heuristic compared to existing solutions in the literature, and demonstrate the importance of both spatial and temporal considerations. In contrast to some scheduling problems, where DVFS introduces performance–energy tradeoffs, our findings reveal the benefit of applying DVFS with both performance and energy gains in the context of spatio-temporal thermal-aware scheduling.

1. Introduction

Datacenters have become an important part of today's computing infrastructure. With the ever increasing power consumption and high packing density of servers, both the heat dissipated in datacenters and the temperature have increased dramatically. High temperature is undesirable in the operation of a datacenter for several reasons: (1) It reduces the reliability of the servers. In particular, some studies have shown that the failure rate of computing nodes will double for every 10 °C increase in temperature [1,2]. (2) Increased temperature induces a larger cooling cost, which has been shown to increase nonlinearly with the temperature [3,4]. (3) A higher temperature also leads to more leakage current, which in turn increases the static power consumption of the servers [5]. As a result, thermal management has been widely recognized as an important technique for optimizing the application performance and reducing the energy consumption in modern datacenters.

Modeling the thermal behavior of datacenters is an essential first step to the design of effective thermal management techniques. The literature contains three main approaches to characterize the thermal map of a datacenter. The first approach approximates the temperatures of the servers using simple *analytical models*, which are usually based on classical heat transfer laws and cyber–physical properties of the datacenters [4,6,7]. The second approach applies complex *predictive models*, which use more sophisticated techniques, such as machine learning and neural

* Corresponding author.
E-mail addresses: hongyang.sun@ens-lyon.fr (H. Sun), patricia.stolf@irit.fr (P. Stolf), jean-marc.pierson@irit.fr (J.-M. Pierson).

http://dx.doi.org/10.1016/j.future.2017.02.005
0167-739X/

networks, to predict the temperatures at various locations of a datacenter [8,9]. The last approach employs elaborate *computational fluid dynamics (CFD)* simulations to model the temperature evolution in a datacenter [10–12]. Although the CFD-based approach offers the best accuracy, it is too slow to facilitate fast and real-time decision making. In contrast, the first two approaches incur much lower overheads while offering reasonable temperature estimates, which can be validated offline by CFD simulations or calibrated online by ambient and onboard temperature sensors (if available). Hence, many researchers have relied on the first two approaches for modeling the temperature in datacenters and for designing scheduling solutions.

In this paper, we present a *spatio-temporal* analytical model to characterize the thermal behavior of datacenters, thus allowing the resource management system to make fast and online scheduling decisions in real time. Indeed, recent studies [4,7] have shown that the server temperature in a datacenter exhibits both spatial and temporal correlations. Spatially, the inlet temperatures of the servers are related to each other via a heat-distribution matrix, which is determined by the complex airflow and heat recirculation in the datacenter. Temporally, the temperature of each server at any time is related to both its current power consumption and its historical temperature due to the physical law of cooling. Although the literature has considered either model separately (see Section 2 for details), to the best of our knowledge, no previous work has studied a holistic spatio-temporal thermal model, which is capable of providing more accurate approximations to the datacenter thermal map. Moreover, the considered spatio-temporal model does not require any knowledge or characteristic of the benchmarks, and conforms only to physical laws and datacenter configurations.

Based on this spatio-temporal model, we study a thermal-aware scheduling problem for homogeneous high-performance computing (HPC) datacenters. The objective is to minimize the makespan for a set of computation-intensive applications subject to a temperature threshold, which cannot be violated at any time during the execution. Indeed, such a threshold is imposed in many resource management systems for either energy reduction considerations or reliability concerns [2,3]. To tackle this problem, we introduce a novel notion, called *thermal-aware load*, to capture more precisely the loads of the servers under the thermal constraint. We propose an online scheduling heuristic that applies this notion to both job assignment and thermal management aspects of the scheduling decision. For job assignment, we strategically choose the server to assign each arriving job in such a way that leads to well-balanced loads (in the thermal-aware sense) among all servers, which helps to minimize the makespan. For thermal management, we rely on dynamic voltage and frequency scaling (DVFS) to regulate the job executions during runtime for respecting the temperature threshold. To further improve the makespan, thermal-aware load is again used for prioritizing the servers while applying DVFS to cope with the complex space–time correlation of the temperature evolution. The proposed scheme guarantees to respect the thermal threshold while reducing the makespan with low computational overhead. The use of DVFS also allows the heuristic to significantly reduce the energy consumption.

We conduct extensive simulations to evaluate the effectiveness of our approach based on an experimentally validated datacenter configuration and realistic parameter settings. The results confirm that our algorithm outperforms several existing solutions in the literature, and hence demonstrate the importance of both spatial and temporal considerations in the context of thermal-aware scheduling. Finally, in contrast to some other scheduling problems, where DVFS introduces performance–energy tradeoffs, our findings reveal the benefit of applying DVFS with both performance and energy gains in the context of spatio-temporal thermal-aware scheduling.

The main contributions of this paper are summarized as follows:

- An analytical thermal model that captures both spatial and temporal behaviors of the temperature evolution in datacenter environments.
- The formulation of a spatio-temporal thermal-aware scheduling problem for high-performance computing (HPC) applications in homogeneous datacenters.
- An online scheduling heuristic that applies the notion of thermal-aware load for both job assignment and thermal management.
- A comprehensive set of simulations to demonstrate the effectiveness of the proposed heuristic under a wide range of parameter settings in a realistic datacenter configuration.

Finally, we stress once again that our proposed solution works for computation-intensive jobs and we leave the consideration of I/O-intensive and communication-intensive jobs for future work. The rest of this paper is organized as follows. Section 2 reviews some related work on thermal modeling and scheduling. Section 3 presents the spatio-temporal thermal model, based on which we formulate a thermal-aware scheduling problem in Section 4. Section 5 describes our thermal-aware scheduling heuristic. The simulation results are presented in Section 6. Finally, Section 7 concludes the paper with future directions.

2. Related work

Many papers have studied thermal-aware scheduling in datacenters with either a spatial model or a temporal model. In this section, we review some related work on thermal modeling and scheduling. Interested readers can refer to [13] for a recent survey of the field.

2.1. Work on spatial correlation

To characterize the spatial correlation of the temperatures in a datacenter, Moore et al. [3] first introduced the notion of heat recirculation. They also proposed "Weatherman", a software tool to predict the thermal profile of a datacenter by taking the topology and heat flow into account [8]. Tang et al. [4] formally defined a heat-distribution matrix via an abstract heat flow model, and applied it in the optimization of the cooling cost of a datacenter. This abstract spatial model has been subsequently adopted by a series of research, and it was also successfully validated by computational fluid dynamics (CFD) simulations in [14,15]. Pakbaznia and Pedram [16] considered the problem of minimizing the total energy consumption of a datacenter by performing server consolidation while accounting for heat recirculation. Mukherjee et al. [17] considered a similar problem while taking the temporal job placements into account (but without a temporal thermal model). Sun et al. [18] studied performance–energy tradeoff in heterogeneous datacenters with heat recirculation effect. They also proposed effective server placement strategies in order to minimize the cooling cost. The latter problem was independently studied by Pahlavan et al. [19], who utilized integer linear programming (ILP)-based methods to find the optimal location of each server in the datacenter. By assuming specific heat recirculation patterns, Mukherjee et al. [20] designed approximation algorithms for a couple of related thermal-aware scheduling problems.

2.2. Work on temporal correlation

The temporal temperature correlation has also attracted much attention. Ramos and Bianchini [6] presented "C-Oracle", a software infrastructure to predict the servers' temperatures in a datacenter based on simple temporal models governed by heat transfer laws. Skadron et al. [21] was the first to apply the lumped-RC model to capture the transient behavior of temperatures in