Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2022) 30: 1050 – 1066

doi:10.3906/elk-2108-163

TARA: temperature aware online dynamic resource allocation scheme for energy optimization in cloud data centres

Narayanamoorthi THILAGAVATHI[1,*], **Arockiasamy JOHN PRAKASH**[1], **Sridhar SRIDEVI**[1], **Vaidyanathan RHYMEND UTHARIARAJ**[2]

[1]Ramanujan Computing Centre, Information and Communication, Anna University, Chennai, India

[2]Department of Computer Science, Crescent University, Chennai, India

Received: 25.08.2021 • **Accepted/Published Online:** 25.11.2021 • **Final Version:** 21.03.2022

Abstract: Cloud data centres, which are characteristic of dynamic workloads, if not optimized for energy consumption, may lead to increased heat dissipation and eventually impact the environment adversely. Consequently, optimizing the usage of energy has become a hard requirement in today's cloud data centres wherein the major part of energy consumption is mostly attributed to computing and cooling systems. Motivated by which this paper proposes an online algorithm for dynamic resource allocation, namely, temperature aware online dynamic resource allocation algorithm (TARA). TARA demonstrates a novel algorithm design to adapt dynamic resource allocation based on the temperature of a data centre using computational fluid dynamics (CFD). Also, TARA demonstrates a new dynamic resource reclaim strategy for making efficient resource allocations leading to efficient energy consumptions in dynamic environments. The proposed algorithm provides optimal resource allocation considering energy efficiency without being overwhelmed by online dynamic workloads. The optimal energy-efficient dynamic resource allocation for online workloads eventually optimizes the computing and cooling energy consumption. We show through theoretical analysis the correctness, efficiency and optimality bounds given as $TARA(P) \leq 2OPT(P)$, relative to the optimal solution provided by offline dynamic resource allocation algorithm $(OPT(P))$. We show through empirical analysis that the proposed method is efficient and significantly saves energy by 26% when the data centre utilization is 100% compared to batched reclaim. The performance analysis shows significant improvement in optimizing computing and cooling efficiency. TARA can be used in multiple areas of on-demand dynamic resource allocation in cloud computing like resource allocation for virtual machine creation, resource allocation for virtual machine migrations, and virtual resources assignment for elastic cloud applications.

Key words: Thermal aware scheduling, energy efficiency, server consolidation, workload placement, green computing, data centres, bin packing

1. Introduction

Cloud data centres are characterized to face the challenges of handling unpredictable dynamic workloads which often overwhelm the resource allocation strategies. The very nature of the cloud demands online dynamic provisioning of resources for smooth rendering of services. Often, handling peak workloads result in over-provisioning and high energy consumption. Therefore to meet the above challenges, cloud data centres are built as high-density infrastructures to meet the dynamic workloads but such designs consume a huge amount of energy, incur a high cooling cost, suffer from high-temperature dissipation, and eventually produce high

*Correspondence: nthilagavathi2013@gmail.com

carbon emission. Moreover, when the electrical energy consumed by the servers dissipates heat [1], it leads to equipment reliability issues and increased maintenance. A recent study shows that the data centres consume about $1-3\%$ of total the United States energy usage [2]. The unique green features that characterize each data centre are energy consumption, heat dissipation, and carbon emission. It is therefore pertinent that the energy consumption is optimized for environmental sustainability [3]. Workload distribution, physical arrangement of racks, computer room air condition (CRAC) unit placement and functioning [1] are some of the key factors that affect energy consumption optimization in a cloud data centre. Heat recirculation leads to hotspots and consequently increases the cooling energy requirement in cases such as mixing of cold air from the underground plenum and hot air from server outlet, failure to pull the hot air back during air conditioner failures, and presence of obstructions in the airflow (hot air) from the server outlet mixing up with the incoming cold air from CRAC supply [1, 4]. A realistic estimate of the Inlet temperature distribution in a data centre can be obtained by computational fluid dynamics (CFD) methods. CFD models [1] can truly predict the airflow and temperature distribution of the elements inside a data centre and these models can be used to analyse the heat recirculation. Therefore, the online dynamic resource allocation algorithm should allocate resources being aware of the energy usage and heat recirculation characteristics of the cloud data centre.

There have been efforts on energy optimization to improve the cooling performance such as minimizing cooling power consumption and temperature hotspots [5], minimizing hot air recirculation [1], study the effect of server workload on airflow pattern [6, 7], minimizing peak inlet temperature of the servers, a study of factors affecting energy efficiency and thermal management [8]. Even after such efforts of adopting more sophisticated cooling methods in data centres, such efforts have only increased the cooling system cost without a proportionate increase in the amount of energy saved. Hence, it can be inferred that it is necessary to handle issues of computing and cooling energy optimization holistically. Motivated by the above discussion, this paper proposes a holistic design approach that combines server consolidation through a reclaim strategy and temperature-aware workload placement to achieve energy optimization based on the thermal conditions of the nodes. The proposed approach obtains thermal information of the data centre servers using CFD models. Then it groups servers according to the thermal conditions and power states of the nodes for energy optimization. The empirical analysis shows the efficient performance of the proposed design approach. Thus, the main contributions of this paper are:

1. We propose a temperature aware online dynamic resource allocation (TARA) scheme for energy optimization in cloud data centres with dynamic workloads.

2. The proposed TARA introduces a novel technique based on the temperature distribution characteristics in a typical cloud data centre modelled using CFD tools.

3. The proposed TARA introduces an effective grouping mechanism based on the temperature characteristics of the servers. A reclaim strategy is adopted to consolidate the servers which helps in the optimization of energy consumption.

The rest of the paper is organised as follows. Section 2 reviews various thermal management algorithms and approaches for the data centre. The system model, system architecture, design, and algorithms are described in Section 3. The simulation setup and results are described in Section 4, followed by discussions in Section 5 and a conclusion in Section 6.

2. Related work

The research work on thermal management of data centre can be categorized into minimizing heat recirculation, optimizing heat dissipation and energy consumption. Thermal management of data centre constitutes two steps: one is obtaining thermal characteristics of the data centre and second step is optimizing energy consumption. The thermal information was obtained through CFD simulations, using prediction tools and sensor-based thermal mapping. This information is used for temperature-oblivious resource allocation strategies. The data centre thermal profile is based on the factors such as heat circulation, placement of jobs, and cooling systems. Moore et al. [1], proposed a zone-based discretisation (ZBD) algorithm and minimize-heat-circulation (MinHR) algorithm for thermal management to reduce the cooling cost of the data centre. The ZBD algorithm borrows excess power required from the neighbouring nodes in the zones to run the workload. Similar to the power allocation algorithm, this ZBD is designed such that equal temperature throughout the data centre is maintained by thermal state management. Whereas, MinHR algorithm provisions power to minimize the heat air recirculation to reduce the cooling cost.

Chan et al. [9] proposed a convex optimization model for energy optimization with thermal constraints in the data centre systems. The authors designed an optimal fan speed policy for the given workload condition to manage energy and temperature. Chen et al. [10] proposed a dynamic thermal management (DTM) framework for accurate temperature estimation based on the temperature behavior of the server chips. In contrast to the estimation of full chip temperature distribution, this work aims to design an algorithm that identifies the optimal placement of thermal sensors. This framework enables efficient thermal management with limited thermal sensing capability. Tang et al. [11] have addressed minimizing peak inlet temperature problem through task assignment (MPIT-TA) problem through XInt-GA and XInt-SQP algorithms which eventually reduce the cooling cost of a data centre. The authors designed an assignment strategy in which the tasks are assigned to the server such that the supply temperature of the CRAC unit is maximized thereby reducing the cooling cost. They demonstrate that assigning tasks to the servers affects the inlet temperature of the nodes. Wang et al. [12] proposed an energy-saving technique through virtual server consolidation and provision scheme.

An online dynamic scheduling algorithm is proposed by Abbasi et al. [13] for resource provision and workload distribution to minimize inlet temperature but their mechanisms do not take into account real-time temperature changes caused by workload variations. Ilager et al. [5] proposed a dynamic energy and thermal aware scheduling (ETAS) algorithm for VM consolidation and hotspot reduction in cloud data centres. The ETAS algorithm performs host overload, under-load detection, selecting the VM for migration, and identifying the target host for VM placement for energy optimization. CFD tools like FloVENT, OpenFoam, Fluent accurately predicts air flow pattern and temperature dynamics of the elements in the data centre environment. The research work by Patel et al. [14] and Moore et al. [1] used FloVENT for thermal analysis. The FloVENT tool is a widely used tool that has been utilized by many other researchers such as Tang et al. [4], Mukherjee et al. [15] and Banerjee et al. [16]. Zhao et al. [17] used OpenFoam and ANSYS Fluent was used by Nada et al. [18]. Tools like Weatherman (Moore et al. [19]) and C-Oracle (Yang et al. [20], Ramos & Bianchini [21]) were proposed to predict the future thermal information. Moreover, the real time sensors (Bash & Forman [22], Coskun et al. [23]) were placed in the appropriate locations to obtain the temperature information needed for thermal management.

Temperature aware resource allocation schemes fall into static, dynamic or online approaches. Static methods schedule the jobs with prior information about the jobs. Moore et al. [1] proposed power-based task

scheduling whereas, Tang et al. [4] proposed task-oriented scheduling for temperature aware resource allocation. Dynamic allocation schemes were presented by Ramos & Bianchini [21], Yang et al. [20] and Zhao et al. [17]. Recent work by Moulik et al. [24] gives a bilevel proportional dynamic and fair resource allocation strategy to achieve higher resource usage efficiency within a reasonable thermal threshold. The latest strategy by Akbar et al. [25] uses cooperative game theory to improve the thermal balance and avoid the data centre hotspots. Work by Banerjee et al. [16], Abbasi et al. [13] and Ilager et al. [5] proposed temperature aware online schemes. The proposed work in this paper uses a hybrid approach combining the best of thermal profile based approaches and resource allocation approaches. In the proposed work, the thermal profile of the data centre under study is captured using the CFD models accurately and temperature aware online dynamic resource allocation is performed to optimize energy consumption.

3. TARA design

The problem of reducing the energy consumption of servers in a data centre by optimal resource allocation is equivalent to the problem of the online bin packing problem. The online bin packing problem aims at optimizing the assignment of arriving items to a minimum number of bins. Similarly, minimizing the number of servers used for a given workload minimizes the energy expended primarily by a data centre. Secondarily, selecting the right servers being aware of energy characteristics for workload assignment further improves the amount of energy saved. Therefore, minimization of the number of servers used and choosing such servers intelligently are the two primary strategies upon which the proposed mechanism is designed. The description of the overall proposed design is organised into three parts, namely: (1) first, we present the mathematical model of the energy minimization problem. The model formulates the problem of minimizing energy consumption in a data centre and reduces it to the problem of minimizing the number of bins in the online setting, (2) second, we present the TARA algorithm in which the CFD model is created and some chosen algorithm parameters are measured. The measured parameters are used to design an efficient and energy-aware workload assignment algorithm. The TARA algorithm uses the mathematical model to compute various parameters, (3) thirdly, we present the description of the online dynamic workload assignment algorithm, namely online dynamic resource allocation (ODRA), a subroutine of TARA which dynamically assigns workload being aware of the energy consumed by servers. The workload assignment algorithm greedily minimizes the number of servers used and chooses each server intelligently being aware of the energy expended in the assignment.

3.1. Mathematical model

Let b_i denote each item in the finite set of items or workloads/jobs with size $size(b_i) \in \mathbb{Z}^+$ for each i, let a positive integer B denote the capacity of a node and let a positive integer N denote the number of nodes or servers used. Let I be a positive integer that denotes the total number of items in the system. The problem is to minimize the number of servers by identifying k-partitions of N such that the sum of the sizes of the items

or jobs assigned to each N_j is almost B. The problem can then be modelled as:

$$\text{Minimize } N = \sum_{j=1}^{m} N_j$$

$$\text{subject to } N \geq 1,$$

$$\sum_{i=1}^{n} size(b_i) A_{ij} \leq B N_j, \forall j \in \{1, \ldots, m\}$$

$$\sum_{j=1}^{m} A_{ij} = 1, \forall i \in I \tag{1}$$

Where $N_j \in \{0,1\}, \forall j \in \{1, \ldots, m\}$. Here, $N_j = 1$ if server j is used and $A_{ij} \in \{0,1\}, \forall i \in I, \forall j \in \{1, \ldots, m\}$ where $A_{ij} = 1$ if item b_i is assigned to run in server N_j.

3.2. System setting

The system architecture for energy or thermal aware workload assignment approach is modeled into various components namely, system model, data centre model, power consumption model, and cooling cost model representing the functional components of the system setting. The definitions of the symbols used in the model are listed in Table 1.

Table 1. Definition of symbols.

Symbol	Definition
N	Number of nodes used
G	Number of node groups
P_j	Power consumption of node j
TP	Total power usage of all nodes
P_{base}	Base power consumption of node
$P_{computing}$	Power usage of computing nodes
$P_{cooling}$	Power consumption of CRAC unit

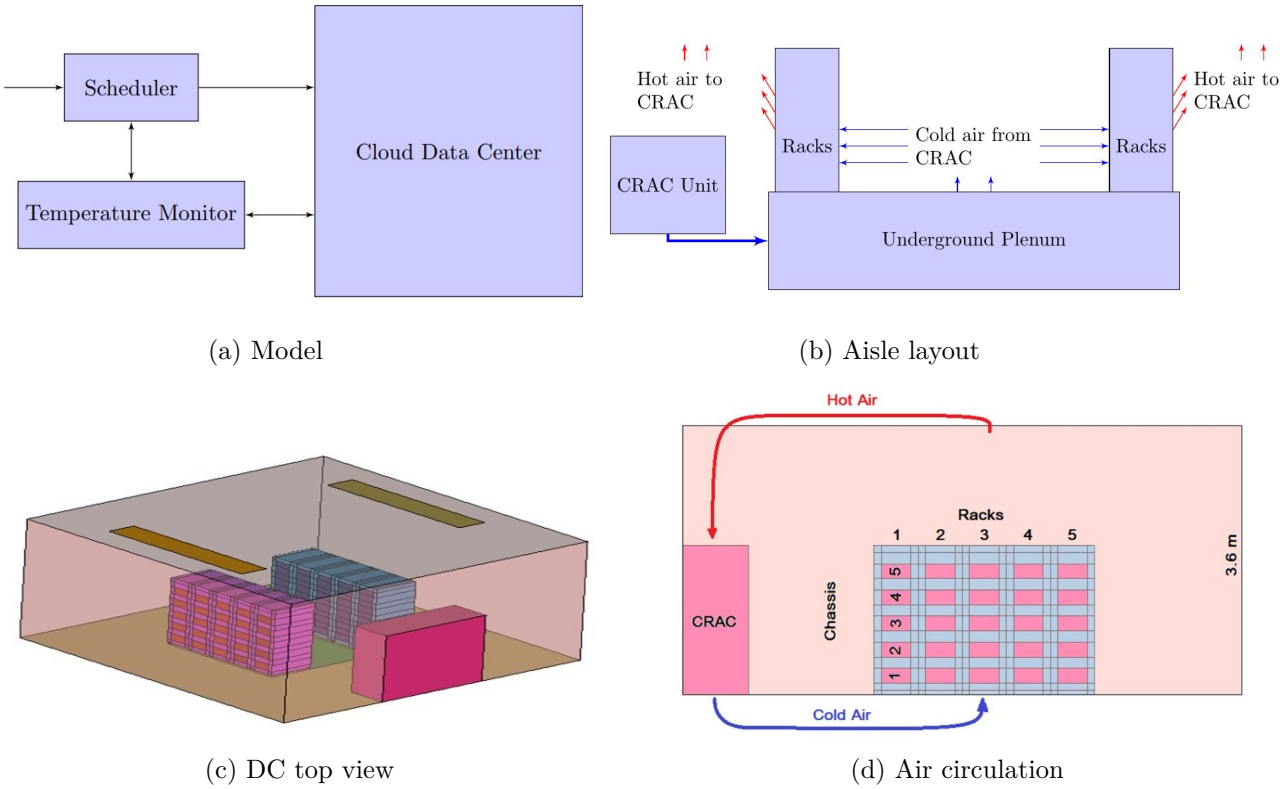
Symbol	Definition
T_{supply}	Supply temperature of the CRAC unit
T^j_{inlet}	Inlet temperature of node j
$T_{redline}$	Redline temperature of the nodes
P_{IDLE}	Power usage in IDLE state of the node
P_{ON}	Power usage in ON state of the node
P_{OFF}	Power usage in OFF state of the node
P_{SLEEP}	Power usage in SLEEP state

3.2.1. System model

This model consists of three components namely, temperature monitor, scheduler and data centre. The temperature monitor collects the thermal state of each server at the data centre using CFD methods to be used by the proposed algorithm to group the servers of similar thermal characteristics. A system model in the chosen data centre setting is shown in Figure 1a with the hot aisle and cold aisle shown in Figure 1b.

3.2.2. Data centre model

The top view of the data centre model and arrangement of racks with homogenous virtualized servers is shown in Figure 1c. Figure 1d shows the arrangement of racks in rows, chassis, and the air circulation convention. Each rack is assumed to have five chassis hosted with blade servers and the CRAC unit supplies cold air through perforated tiles from the underground plenum. The cold air from CRAC is drawn to the inlet of servers by the chassis fans and gets heated up by the server load. Hot air from the servers exits through the outlet into the hot aisle which gets removed by the hot air ducts provided in the ceiling of the data centre.



(a) Model

(b) Aisle layout

(c) DC top view

(d) Air circulation

Figure 1. System model and data centre layout.

3.2.3. Power model

The power consumption P_j of each node N_j with η blade servers is $P_j = b + \eta a$ [4, 11]. The total power consumption of the data centre denoted as TP is given in Equation (2). Here $P_{computing}$ is the power consumed by computing and $P_{cooling}$ denotes the power consumed by cooling. Other sources of energy consumption such as lighting are assumed to be negligible. The power usage of all the nodes at the data centre denoted by $P_{computing}$ is given in Equation (3).

$$TP = P_{computing} + P_{cooling} \tag{2}$$

$$P_{computing} = P_{base} + \sum_{j=1}^{N} Load\%(j) * P_j \tag{3}$$

3.2.4. Cooling model

The cooling cost of the data centre is based on the supply temperature of the CRAC unit. The efficiency of CRAC is measured by the coefficient of performance (CoP), defined as the ratio between the amount of heat extracted to the energy consumed by the CRAC unit. The CoP model used to compute cooling cost is adapted from [1], where T_{supply} is the temperature of cold air supplied by the CRAC unit. The CoP of CRAC unit is calculated using Equation (4). CoP varies mainly based on physical characteristics such as the layout of the

server racks, thermodynamic nature of the walls and ceiling.

$$CoP(T_{supply}) = 0.0068T_{supply}^2 + 0.0008T_{supply} + 0.458 \tag{4}$$

Using Equation (4) the cooling power denoted $P_{cooling}$ is calculated as given in Equation (5).

$$P_{cooling} = \frac{P_{computing}}{CoP(T_{supply})} \tag{5}$$

The overall TARA algorithm is depicted in Algorithm 1. It uses the models described in the previous section to minimize energy usage of servers. Workload placement to the suitable nodes is designed as a bin packing problem where the servers are bins whose size is the utilization level of the node as given by Equation (1). The overall working of the algorithm is:

3.3. TARA algorithm

Algorithm 1: Temperature aware online dynamic resource allocation (TARA)

Input: $N := \{N_1, \ldots, N_m\}$ servers
Output: $G = \{G_{low}, G_{mid}, G_{high}\}$
Simulate a data centre in CFD with $|N|$ servers
for $row \in \{1, 2\}$ **do**
  for $rack \in \{1, 2, 3, 4, 5\}$ **do**
    Arrange the servers
  end for
end for
for $load \in \{25, 50, 75, 100\}$ **do**
  Measure inlet-temperature, computing cost, and cooling cost
  Compute total power cost $TP[N_j]$ for each server as given in Equation (6)
end for
Sort the servers in ascending order of total power cost $\{TP[N_j] : 1 < j < m\}$
Divide the sorted servers into three groups namely $G_{low}, G_{mid}, G_{high}$ such that G_{low} contains the first $1/3$ of servers from the sorted list, G_{mid} contains the next $1/3$ and the rest in G_{high}
Execute ODRA algorithm for each workload assignment

1. Create a CFD model for a data centre as described in Section 3.2.2.
2. Arrange all the $|N|$ chassis hosted with blade servers in the data centre into rows and racks.
3. Divide the servers into three groups namely $G = \{G_{low}, G_{mid}, G_{high}\}$ such that,

$$TP[N_j] : \forall N_j \in G_{low} \leq TP[N_j] : \forall N_j \in G_{mid} \leq TP[N_j] : \forall N_j \in G_{high}$$

where $1 < j < m$, N_j is the j^{th} server among m servers and $TP[N_j]$ denotes the total power consumed by node N_j. The computing nodes are classified into three groups based on the energy consumption computed by the temperature monitor unit, namely low, mid, and high. The set of servers thus grouped into G_{low} then would relatively consume the least power and G_{high} consumes the most in the data centre.

4. The servers will be switched to ON state starting from G_{low} to G_{high} as workload arrives with just one spare server kept always in IDLE state while keeping all other servers powered OFF. Total power consumption of the data centre with 3 groups denoted as TP_3 is given in Equation (6)

$$TP_3 = P_{base} + \alpha * P_{IDLE} + \sum_{j=1}^{N} Load\%(j) * P_{ON} + P_{cooling} \tag{6}$$

Similarly, total power consumption of the data centre with 4 groups with additional group representing the set of servers which are set to SLEEP state denoted TP_4 as given in Equation (7).

$$TP_4 = P_{base} + \alpha * P_{IDLE} + \beta * P_{SLEEP} + \sum_{j=1}^{N} Load\%(j) * P_{ON} + P_{cooling} \tag{7}$$

where α is number of IDLE servers running, β is number of SLEEP servers running, P_{IDLE} and P_{SLEEP} are power consumed in IDLE and SLEEP states respectively.

5. Vary the load in the data centre and measure the inlet temperature (temperature of the nodes). The computing cost is calculated using Equation (3), the cooling cost using Equation (5) and the total power consumption by each server as given in Equation (6).

3.4. ODRA algorithm

The online dynamic resource allocation (ODRA) algorithm finds a feasible assignment for dynamically arriving workload with optimal number of servers. The ODRA being energy aware selects the least power consuming server available at that instance. This online greedy assignment approximately reduces the energy consumption for each arrival and thereby effectively reduces the energy expended asymptotically. The ODRA algorithm for 3-group setup is shown in Algorithm 2 and the execution steps are:

1. The servers are grouped into three groups $G = \{G_{low}, G_{mid}, G_{high}\}$ based on system, power, and cooling models as described in Section 3.3. As the set of servers in the G_{low} consumes least power while G_{high} the most, the algorithm initially starts with switching *spare* servers from G_{low} and remaining servers are set to power status OFF. It is assumed that setting of power status by ODRA leads to power being switched ON, IDLE and OFF in the respective servers.

2. Let each group G has a capacity $size(G) = \sum_{N \in G} size(N)$ where $size(N)$ is the capacity of server N. Let $GAP(G)$ denote the current status of the capacity in group G which is the availability of space for workloads. Similarly, let $GAP(N)$ denote the available space for workloads at server N. Hence, $GAP(G) = \sum_{N \in G} GAP(N)$. These initializations are performed by the subroutine $Preprocess()$. The state information $GAP(G)$ and $GAP(N)$ before and after each assignment are maintained by the algorithm.

3. When a job b arrives at time t, the algorithm invokes the subroutine $Packing(b)$ to determine a feasible server assignment. $Packing()$ invokes $Reclaim(G)$ on all groups to update the state information and power status of servers.

Algorithm 2: Online Dynamic Resource Allocation (ODRA)

Input: Groups $G = \{G_{low}, G_{mid}, G_{high}\}$, $n := |G|$ number of groups, $N := \{N_1, \ldots, N_{|N|}\}$ servers, $k_i := |\forall N_j \in G_i|$ number of servers in group G_i, $size(N)$ size of server for all N and Q wait queue

Preprocess()
if b *arrives or* $b \leftarrow Q$ **then**
  Packing(b)
end if

Procedure *Preprocess()*
  for $i \leftarrow \{1, \ldots, n\}$ **do**
    $\forall N_j \in G_i, GAP(N_j) = size(N_j)$
    $size(G_i) = \sum_{\forall N_j \in G_i} size(N_j)$
    $GAP(G_i) = size(G_i)$
  end for
  $POWER(N_j) \leftarrow \text{OFF} : \forall N_j \in \{G_{low}, G_{mid}, G_{high}\}$
  for $j \leftarrow \{1, \ldots, lastOpen\}$ **do**
    $POWER(N_j) \leftarrow \text{ON} : \forall N_j \in G_{low}$
  end for
  $POWER(N_{j+1}) \leftarrow \text{IDLE}$
  $lastOpen \leftarrow lastOpen + 1$

Procedure *Packing(b)*
  if $\nexists N_j : POWER(N_j) \leftarrow ON \wedge GAP(N_j) \geq Spare$ **then**
    $POWER(N_{lastOpen+1}) \leftarrow \text{ON} : \exists N_j \in G$
    $POWER(N_{lastOpen+2}) \leftarrow \text{IDLE} : \exists N_j \in G$
    $lastOpen \leftarrow lastOpen + 1$
  end if
  $y = \min\{i : Reclaim(G_i), GAP(G_i) \geq size(b), 1 \leq i \leq n\}$
  $z = \min\{j : N_j \in G_y \wedge \text{GAP}(N_j) \geq size(b)\}$
  if $(y, z) \neq NULL$ **then**
    $GAP(G_y) = GAP(G_y) - size(b)$
    $GAP(N_z) = GAP(N_z) - size(b)$
  end if
  else
    if $t_b > 0$ **then**
      $Q \leftarrow Q \cup (b, t_b)$
    end if
    else
      Drop b
    end if
  end if

Procedure *Reclaim(G_y)*
  $GAP(N_j) \leftarrow GAP(N_j) + size(x) : \forall N_j \in G_y \wedge x \in N_j \wedge \text{STATUS}(x) = \text{FINISHED}$
  $POWER(N_j) \leftarrow \text{IDLE} : \forall N_j \in G_y \wedge GAP(N_j) = size(N_j) \wedge G_y \in \text{Mid}$
  $POWER(N_j) \leftarrow \text{OFF} : \forall N_j \in G_y \wedge GAP(N_j) = size(N_j) \wedge G_y \in \text{High}$

4. $Reclaim(G)$ searches for all finished workloads x whose status $STATUS(x)$ are set to FINISHED upon completion in servers $N \in G$. When such x in $N \in G$ were found, the respective state information $GAP(N)$ is updated. Also, after updating the state information, sets the power status to IDLE, for all $N \in G_{mid}$ if $GAP(N)$ equals $size(N)$ and power status to OFF, for all $N \in G_{high}$ if $GAP(N)$ equals $size(N)$.

5. $Packing(b)$ then finds a group G_y having sufficient $GAP(G)$ for $size(b)$. If more than one such candidate groups were found then the group with minimum index value y is preferred.

   (a) In other words, G_{low} is given preference over G_{mid} and G_{mid} is given preference over G_{high}. It can be observed that the preferential order of assignment rule $G_{low} \gg G_{mid} \gg G_{high}$ being imposed implicitly resulting in tight packing eventually leading to containment of workload to low power consuming servers.

   (b) Similarly, a $N_z \in G_y$ is found with a preference to lower value of z is chosen. A lower value for z leads to filling of the left bins or servers before another one is opened. This implicitly reduces the servers being unnecessarily switched ON leading to energy saving at groups especially in G_{mid} and G_{high}.

   (c) $Packing(b)$ then updates the state variables

6. If $Packing(b)$ fails to find such a group $G_y \geq size(b)$ then b is added to queue Q with its tolerant time t_b until which the assignment could be postponed. Before expiration of t_b or upon next arrival, whichever is earlier, the queued job is tried repeatedly for feasible assignment. If no assignment could be found the queued b will be dropped after expiration of t_b.

The ODRA shown in Algorithm 2 can be extended to a 4-group setup with the groups temperature or power consumption arranged in increasing order like $G = \{G_{low}, G_{mid-low}, G_{mid-high}, G_{high}\}$. The intuition behind the working of ODRA is that reclaim procedure when invoked reclaims and frees up space for subsequent assignments and greedily tries to keep the set of servers switched ON to G_{low} if not G_{mid} and so forth. The reclaim also frees up space in servers in which workload have been completed, implicitly facilitates assignment to the left most bins which eventually results servers being switched to IDLE and OFF state except *spares*. This along with group based first fit behaves like best fit with repacking.

4. Performance analysis

This section describes the experimental results to prove the efficiency of TARA. The algorithm was validated by conducting simulation experiments using CFD and CloudSim whose setup and experimental results are presented in this section in three steps as follows: (1) Simulation of the CFD model measurements for a chosen data centre with servers and MIPS/Servers in TARA algorithm. (2) Simulation of ODRA algorithm in TARA using CloudSim and CFD data centre thermal profile measurements for same MIPS/Servers or VMs and groups configuration obtained from the previous step. (3) Using the mathematical model, determine the computing, cooling, and energy saved for the proposed method.

4.1. CFD model setup

A typical data centre with two rows (2×5) of racks with standard 42U sized racks, each row with 5 racks with 7.6m x 7.6m x 3.6m dimension is set up for CFD based simulation. Each rack is set up with five chassis and each having ten blade servers. Each chassis starting from the bottom of the rack is numbered as 1-5. The CRAC unit supplies cold air with the flow rate of $8.5m^3/s$ and the cold air enters the racks from the underground plenum through the perforated tiles. The supply temperature is set to $13°C$ and drawn towards the nodes by the fans. The maximum inlet temperature is set to $35°C$ (red line temperature).

The CFD simulation parameters used for data centre simulation are shown in Table 2. The experiments are performed initially for 3 groups with ON, IDLE, OFF power states and then repeated for 4 groups with ON, IDLE, SLEEP, and OFF power states. The total power consumed by the nodes is 200 kW with a base power consumption of 50 kW. Each server utilizes 350W when in ON state, uses 50W in IDLE state, only 20W in SLEEP state, and 0W in OFF state. The power consumption parameters were chosen similar to Dell PowerEdge 1855 servers. The servers were then divided into groups and states were set based on CFD measurement on parameters like air velocity and temperature distribution (shown in Figure 2).

Table 2. Simulation parameters.

Parameter	**Value**
Cooling medium	Air
Cooling air flow rate	9000 cfm or 8.5 $\frac{m^3}{s}$
Cooling air temperature at inlet vent	$13°C$%
Chassis air flow rate	120 cfm
Room pressure	Atmospheric
Chassis heat rejection	2 kW
Total heat rejection from all racks	100 kW
Processor load for Uniform	33%
Wall heat transfer	0 kW (adiabatic)

4.2. CFD thermal profile measurements

The chassis inlet air temperature of racks in rows was measured and is shown in Figure 3a and 3b. These parameters determine the amount of computing cost, cooling cost, and power consumed by the load. This categorizing or grouping of servers based upon the above parameters into low, mid, and high groups of servers along with the ODRA algorithm minimizes the load assignments to the mid and high group of servers eventually leading to energy savings. Based upon these experimental observations, the servers were grouped and the order of workload assignment was determined for rows.

In the baseline uniform scheme, all nodes were assigned with an equal amount of workload such that each node has a total workload/number of nodes [1]. In the baseline uniform scheme, the power consumption and temperature increase linearly with the system utilization. Simulation experiments were performed to compare the efficiency of the proposed algorithm. It can be observed from Figure 2 and Figure 3 that the temperature distribution inside the data centre is not uniform even though all the nodes in the data centre were assigned equal workload. The experiment was performed with a 33% load assigned to all nodes in the data centre in which changes in temperature distribution for equal workload could be observed because of recirculation of hot

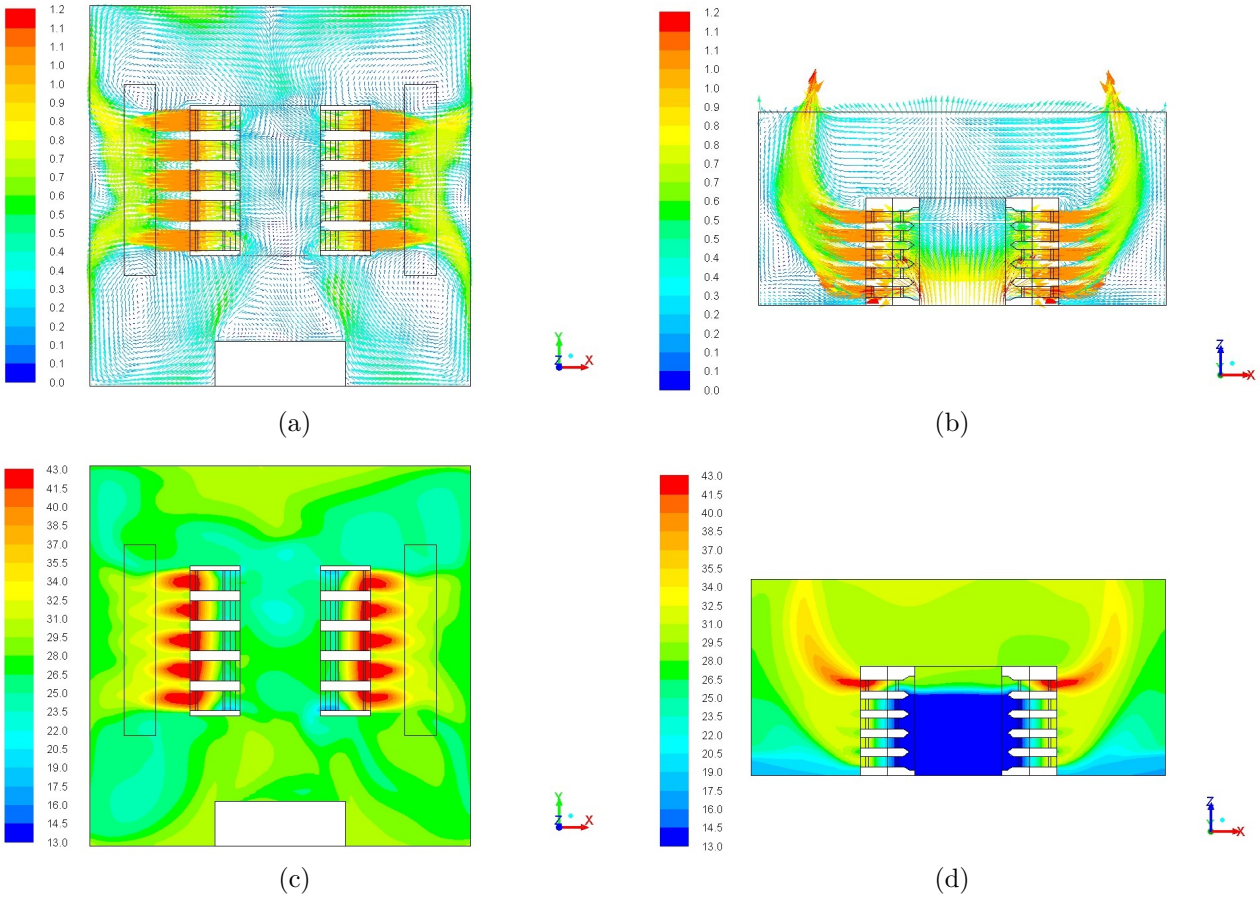


Figure 2. Air velocity and temperature distribution readings from ANSYS Fluent 15.0 (3d, dp, pbns, rke) measured on March 06, 2019. Velocity vectors colored by velocity magnitude (m/s) - (a) aerial view, (b) side view. Contours of air temperature - (c) aerial view, (d) side view.

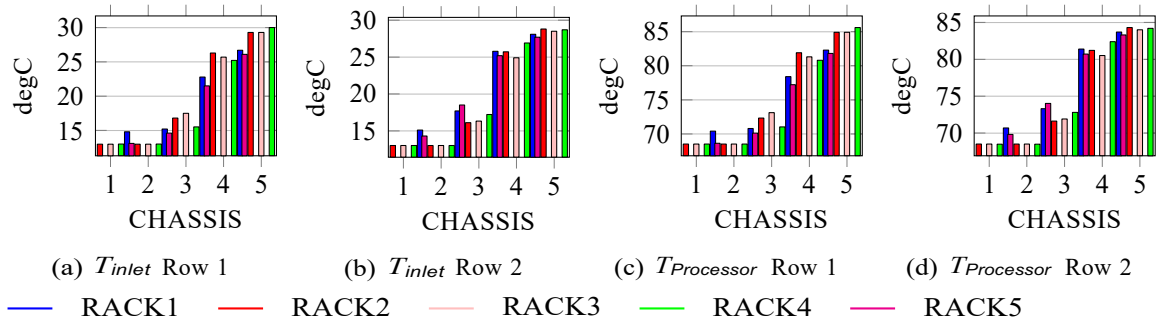
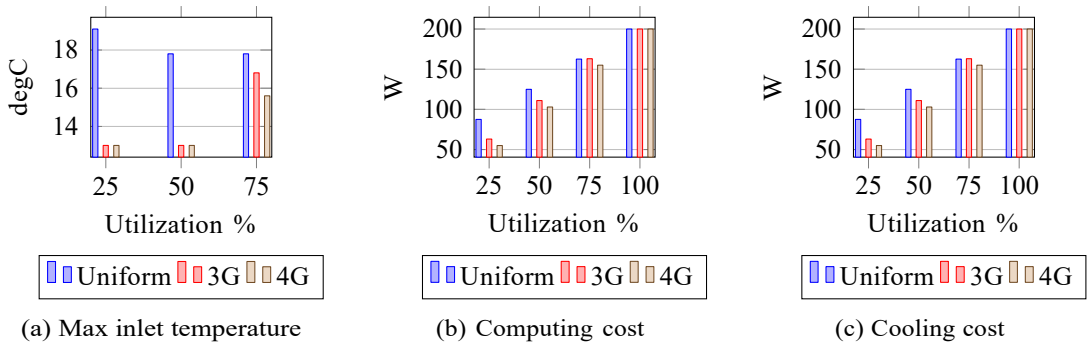


Figure 3. Chassis inlet air temperature of racks in Row-1 and Row-2.

air. Generally, hot air circulation inside the data centre shows an uneven pattern and depends on the data centre configuration and arrangement of nodes. The computing and cooling power consumption were then computed using Equation (2) and Equation (7). The CFD setup is evaluated for three groups (ON, IDLE, OFF) denoted

as 3-Groups and were then repeated for four groups (ON, IDLE, SLEEP, OFF) denoted as 4-Groups. The performance of the algorithm for 25%, 50%, 75% of data centre utilization are compared with the uniform workload. It can be observed from Figure 4a that the inlet temperature of uniform workload is relatively greater than 3-Groups and 4-Groups because the optimized approach uses thermal condition-based grouping and workload assignment to reduce heat recirculation. From Figure 4b and 4c it can be observed that 3-Groups and 4-Groups save energy better than uniform workload. 4-Groups approach can be observed to improve energy saving in computing and cooling costs. This can be attributed to two main factors: minimization of inlet temperature and heat recirculation that affect the cooling cost due to the additional power state (SLEEP). When recirculation is less the cooling supply temperature can be increased to save cooling energy.



(a) Max inlet temperature (b) Computing cost (c) Cooling cost

Figure 4. CFD measurements (a,b,c).

4.3. TARA parameters setup and results

The TARA algorithm, particularly the ODRA subroutine was implemented using the CloudSim simulator and the setup parameters are shown in Table 3. When jobs or cloudlets arrive online, TARA dynamically selects a VM and assigns the workload being temperature aware such that energy consumption is optimized. The parameters like the number of VMs used for varying numbers of jobs were observed during the simulation and the results obtained are shown in Figure 5a - 5d. It can be observed that the number of nodes or VMs used by TARA is less compared to other schemes. A practical batched reclaim scheme is assumed for comparison in which the space allotted for completed workloads is reclaimed on a scheduled basis in batches. A batch reclaim scheme is better compared to uniform or round-robin-based workload assignment. It can be observed that the number of servers used, cooling cost, computing cost and energy saved are better relative to the batch reclaim scheme.

5. Discussion

In this study, the temperature-aware workload assignment schemes are analysed to optimize the computing and cooling power consumption of the cloud data centre. CFD simulation is used to obtain the thermal state of the data centre and grouping while the CloudSim tool was used to analyse the total workload assignment schemes in the cloud data centre environment for energy optimization. The cooling cost, computing cost, and energy saved are observed from the above setup.

A small-scale data centre was modelled using the CFD tool to monitor and obtain the thermal state of the data centre accurately. The variation of power consumption and temperature distribution for uniform

Table 3. CloudSim simulation parameters.

Parameter	Value
Cloudlet length	300-2000 MI
Execution mode	Space shared
Type of servers	Homogenous
Number of VMs	50
Number of cores per VM	1
MIPS/VM	3000 MIPS
Total capacity of data centre	150,000 MIPS

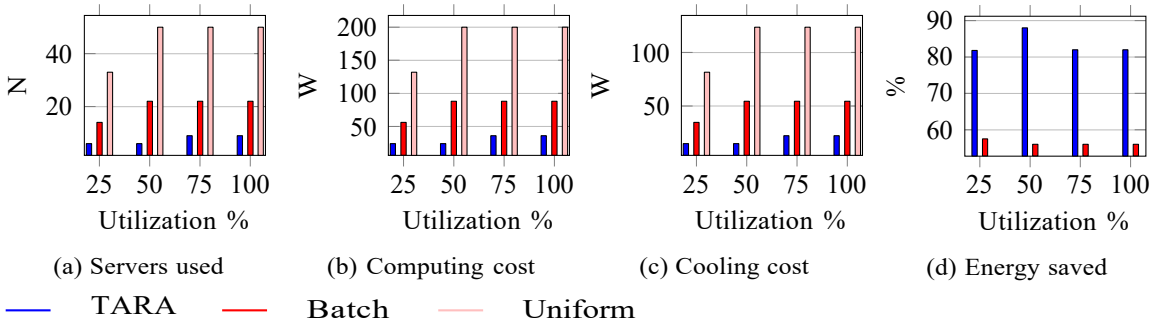


Figure 5. CloudSim results.

workload for computing nodes simulated were measured for different utilization levels varying for every 10% increase up to 100% for power consumption and temperature of the nodes was obtained for uniform workload for the nodes. The air velocity and thermal map were measured and are shown in Figure 2 at the vertical and horizontal section that passes through the racks as a top view to show the distribution pattern of temperature dissipated from the outlet of the racks. This shows the generation of hot air and its movement inside the data centre which is because of hot air recirculation. Figures 3a to 3d show the inlet temperature matrix for 33% utilization (uniform workload) to all the computing nodes. The observed temperature variation for uniform workload was due to the heat recirculation. Figures 3a to 3d show energy saving when load assignments are performed according to the formed 3 groups and 4 groups with the servers ordered based on energy consumption. Group-based resource allocation reduces the number of switch-on nodes thereby reducing the computing power consumption. The percentage of reduction in inlet temperature in the proposed schemes compared with the uniform algorithm is shown in Table 4. This is due to the temperature-aware grouping of nodes and placing jobs in the nodes which have less inlet temperature reduce the hot air circulation which consequently reduces the inlet temperature of the nodes.

The TARA approach with dynamic reclaim performs better than the batched reclaim approach and uniform algorithm which can be observed from Figures 5a to 5d. The dynamic algorithm reclaims the used space dynamically on completion and reuses the reclaimed space for further assignment of the jobs. Whereas, the batched algorithm is set to reclaim after every 50 jobs. The computing cost, cooling cost, and energy saved by TARA relative batched reclaim are summarized in Table 5. The proposed scheme shows better energy-saving

even for increased data centre utilization due to grouping and dynamic reclaim. New nodes were not switched on further, the assignment was made within some nodes in the low energy consumption group itself and thereby saving computing energy which eventually, saves cooling power consumption as well.

Table 4. % Cost saved by 3 groups[a] and 4 groups[b] (relative to uniform).

Parameter	**25%**	**50%**	**75%**
Computing cost saved[a]	28	11	0.3
Computing cost saved[b]	37	17	4.6
Cooling cost saved[a]	27	10	1
Cooling cost saved[b]	37	16	4
Reduction in T_{inlet}[a]	31	30	5
Reduction in T_{inlet}[b]	31	30	12

Table 5. % Energy saved by TARA (relative to batched reclaim).

Parameter	**25%**	**50%**	**75%**	**100%**
Computing cost saved	24	32	26	26
Cooling cost saved	24	33	27	27
Energy saved	24.3	32	26	26

6. Conclusion

Energy efficiency and environmental sustainability of data centres can be improved significantly through the design of efficient holistic algorithms. Particularly, the proposed algorithm for dynamic resource allocation will result in the establishment of energy-efficient cloud data centres. The online dynamic resource allocation problem can be reduced to the bin packing problem which is proved to be computationally intractable. This paper provides a solution to this problem, by proposing the TARA algorithm. TARA utilizes thermal information of the servers obtained from the CFD tool. It groups the servers considering their thermal condition, heat recirculation and energy-saving states of the servers. Based on server groups identified using TARA, the subroutine ODRA assigns workload to the appropriate server using the concept of server consolidation. Due to this, the maximum inlet temperature at the data centre is reduced which eventually reduces thermal hotspots. This helps to efficiently reduce the computing and cooling energy consumption of the data centre. The empirical analysis shows that TARA saves about 24% energy when the data centre is utilized at 25%, 32% when the data centre utilization is 50%, 26% when the data centre utilization is 75%, and 26% when the data centre utilization is 100% compared to batched reclaim. Therefore, the efficient temperature-aware dynamic workload placement strategy as demonstrated through TARA is essential in creating green cloud data centres and a sustainable green environment.

Acknowledgment & Conflict of interest

The authors wish to acknowledge the support of Dr. S. Jayaprakash for his directions and constant support in this work. This work was supported by DST PURSE II fellowship, Government of India and Visvesvaraya PhD Scheme for Electronics and IT, Government of India. The authors have no conflict of interest to disclose.

References

[1] Moore J, Chase J, Ranganathan P, Sharma R. Making scheduling cool: Temperature-aware workload placement in data centers. In: Proceedings of USENIX Annual Technical Conference; Berkeley, CA, USA; 2005. pp. 61–74.

[2] Ayanoglu E. Energy efficiency in data centers. In IEEE Communications Society Newsletter 2019. <https://www.comsoc.org/publications/tcn/2019-nov/>.

[3] Buyya R, Srirama SN, Casale G, Calheiros R, Simmhan Y et al. A manifesto for future generation cloud computing: Research directions for the next decade. ACM Computing Surveys 2019; 51 (5): 1-38. doi: 10.1145/3241737

[4] Tang Q, Gupta SKS, Varsamopoulos G. Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: A cyber-physical approach. IEEE Transactions on Parallel and Distributed Systems 2008; 19 (11): 1458-1472. doi: 10.1109/TPDS.2008.111

[5] Ilager S, Ramamohanarao K, Buyya R. ETAS: Energy and thermal-aware dynamic virtual machine consolidation in cloud data center with proactive hotspot mitigation. Concurrency and Computation: Practice and Experience 2019; 31 (17): 1-15. doi: 10.1002/cpe.5221

[6] Kumar P, Sundaralingam V, Joshi Y. Effect of server load variation on rack air flow distribution in a raised floor data center. In: 27th Annual IEEE Semiconductor Thermal Measurement and Management Symposium; San Jose, CA, USA; 2011. pp. 90-96.

[7] He Z, He Z, Zhang X, Li Z. Study of hot air recirculation and thermal management in data centers by using temperature rise distribution. Building Simulation, Springer 2016; 9: 541-550. doi: 10.1007/s12273-016-0282-7

[8] Thilagavathi N, Uthariaraj VR. A study on energy and thermal management factors of green computing. In: Proceedings of International conference on big data and cloud computing; Coimbatore, India; 2017, pp. 41-50.

[9] Chan CS, Akyurek AS, Aksanli B, Rosing TS. Optimal performance aware cooling on enterprise servers. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 2019; 38 (9): 1689-1702. doi: 10.1109/TCAD.2018.2855122

[10] Chen KC, Tang HW, Liao YH, Yang YC. Temperature tracking and management with number-limited thermal sensors for thermal-aware NoC systems. IEEE Sensors Journal 2020; 20 (21): 13018-13028. doi: 10.1109/JSEN.2020.3003657

[11] Tang Q, Mukherjee T, Gupta SKS, Cayton P. Sensor-based fast thermal evaluation model for energy efficient high-performance datacenters. In: Fourth International Conference on Intelligent Sensing and Information Processing; Bangalore, India; 2006. pp. 203-208.

[12] Wang L, Laszewski GV, Dayal J, He X, Younge AJ et al. Towards thermal aware workload scheduling in a data center. In: 10th IEEE International Symposium on Pervasive Systems, Algorithms, and Networks; Kaoshiung, Taiwan; 2009. pp. 116-122.

[13] Abbasi Z, Varsamopoulos G, Gupta SKS. Thermal aware server provisioning and workload distribution for internet data centers. In: Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing; Chicago, Illinois, USA; 2010. pp. 130-141.

[14] Patel C, Bash C, Stahl L, Sullivan D. Computational Fluid dynamics modeling of high compute density data centers to assure system inlet air specifications. In: Proceedings of ASME IPACK'01 The Pacific Rim/ASME International Electronic Packaging Technical Conference and Exhibition; Kauai, Hawaii, USA; 2001. pp. 1-9.

[15] Mukherjee T, Banerjee A, Varsamopoulos G, Gupta SKS, Rungta S. Spatio-temporal thermal-aware job scheduling to minimize energy consumption in virtualized heterogeneous data centers. Computer Networks, Elsevier 2009; 53 (17): 2888-2904. doi: 10.1016/j.comnet.2009.06.008

[16] Banerjee A, Mukherjee T, Varsamopoulos G, Gupta SKS. Cooling-aware and thermal aware workload placement for green hpc data centers. In: IEEE International conference on green computing; Chicago, IL, USA; 2010. pp. 245-256.

[17] Zhao X, Peng T, Qin X, Hu Q, Ding L et al. Feedback control scheduling in energy-efficient and thermal-aware data centers. IEEE Transactions on Systems, Man, and Cybernetics: Systems 2015; 46 (1): 48-60. doi: 10.1109/TSMC.2015.2434797

[18] Nada S, Said M, Rady M. CFD investigations of data centers thermal performance for different configurations of CRAC units and aisles separation. Alexandria engineering journal 2016; 55: 959-971. doi: 10.1016/j.aej.2016.02.025

[19] Moore J, Chase JS, Ranganathan P. Weatherman: Automated, online and predictive thermal mapping and management for data centers. In: IEEE international conference on Autonomic Computing; Dublin, Ireland; 2006. pp. 155-164.

[20] Yang J, Zhou X, Chrobak M, Zhang Y, Jin L. Dynamic thermal management through task scheduling. In: ISPASS IEEE International Symposium on Performance Analysis of Systems and software; Austin, TX, USA; 2008. pp. 191-201.

[21] Ramos L, Bianchini R. C-oracle: Predictive thermal management for data centers. In: IEEE 14th International Symposium on High Performance Computer Architecture; Salt Lake City, UT, USA; 2008. pp. 111-122.

[22] Bash C, Forman G. Cool job allocation: Measuring the power savings of placing jobs at cooling-efficient locations in the data center. In: USENIX Annual Technical Conference; Santa Clara, CA; 2007. pp. 1-6.

[23] Coskun A, Simunic T, Whisnant K, Gross K. Static and dynamic temperature-aware scheduling for multiprocessor SoCs. IEEE Transactions on Very Large Scale Integration (VLSI) Systems 2008; 16: 1127-1140. doi: 10.1109/TVLSI.2008.2000726

[24] Moulik S, Sarkar A, Kapoor HK. TARTS: A temperature-aware real-time deadline partitioned fair scheduler. Journal of Systems Architecture 2021; 112 (2): 1-14. doi: 10.1016/j.sysarc.2020.101847

[25] Akbar S, Malik SR, Choo KKR, Khan SU, Ahmad N et al. A game-based thermal-aware resource allocation strategy for data centers. IEEE Transactions on Cloud Computing 2021; 9 (3): 845-853. doi: 10.1109/TCC.2019.2899310