Considering Thermal-aware Proactive & Reactive Scheduling and Cooling for Green Data-centers

Muhammad Tayyab Chaudhry, T.C. Ling
Faculty of Computer Science and
Information Technology,
University of Malaya,
50603, Kuala Lumpur, Malaysia
email: mtayyabch@yahoo.com
tchaw@um.edu.my

Atif Manzoor
Department of Computer Science,
COMSATS Institute of Information Technology,
54000, Lahore, Pakistan
email: atif.manzoor@ciitlahore.edu.pk

***Abstract*— Data-centers require huge amount of electricity to continue meeting the computing demands of consumers each year. Fossil fuel based electricity is utilized due to lack of abundant renewable energy resources, resulting in the emission of CO_2 in atmosphere and causing global temperature hike. The world is in dire need of efficient utilization of electricity. At the same time, advent of cloud computing has brought the innovation of everything as a service. This has led to proliferation of cloud services in every computing field. It has increased the load on cloud hosting data-centers, resulting in excessive use of electricity. A cloud data-center can manage to save electricity by reducing resource exploitation through either or both of the efficient utilization based and thermal based scheduling and monitoring. In this paper we take a peek into recently proposed thermal aware scheduling and monitoring techniques to maintain a cost effective Green Cloud Computing environment.**

***Keywords*—data center; data-center; cloud computing; green data center; green data-center; thermal-aware cooling; thermal-aware scheduling**

I. INTRODUCTION

There is a huge economic potential in pay-as-you-go type of rental and is one of the main reasons of rapid shift of computing trends to Cloud Computing. A cloud user has to pay nothing for maintenance and care of the computing resources on cloud. It only takes a few mouse clicks for easy online provisioning and management of virtual computing assets on cloud. These benefits make the acquiring of even the High Performance Computing (HPC) assets without investing a penny in infrastructure and maintenance[1-3]. A user can rent out any type and number of cloud services with pay-as-you-go rental [2, 4-6]. For example Oracle Cloud offers the services such as Software as a Service, Platform as a Service, Middleware as a Service, Database as a Service, Infrastructure as a Service and even Cloud as a Service [1, 7, 8] to users that acquire them and use them to fulfill their IT needs. While Google has a whole set of GoogleApps [3] which are much economical as compared to in-house hosting. Amazon EC2 [5] is unique with its elasticity of computing resources according to load and is billed accordingly.

The vast economic application of cloud has lured investors to enter in cloud services rental business and to set up or merge large scale data-centers with thousands of interconnected loosely coupled servers and high network bandwidth.

Of the Total Cost of Ownership (TCO) of a data-center, a dominant part goes to electricity usage. According to past survey [9, 10] it was estimated that the power usage by the data-centers around the globe will be doubled or from year 2005 to year 2010. However in recent report [11], it is observed that the total growth reached to only 36% than hundred percent. Thus as a percentage of total electricity usage in world, it remained between 1.1% and 1.5% till 2010 as compared to 1% in 2005. The IT industry recognized standard for calculating data-center Power Usage Efficiency (PUE) [12] is the ratio of total energy usage of data-center and IT energy consumption. In 2005 the PUE was 2.0 [9, 11] which has dropped to 1.92 in 2010 [11]. But still the data-centers around the globe spent 130 billion KWH for cooling from the total 271 billion KWH electricity usage in 2010[11]. Cooling cost is the biggest contributor in raising the PUE since year 2000, it contributed to half of the electricity consumption in data-center[9]. However, lowering of PUE in year 2010 shows a positive progress in reduction of cooling cost. Koomey et al.[11] regards virtualization technology and the energy efficient equipment and data-center design as the main reasons for global energy efficiency in data-centers. In this paper we classified various algorithms and techniques proposed in recent past to improve the cooling overhead in data-centers.

II. BACKGROUND

A typical data-center scenario[10] in research literature consists of rows of racks containing IT equipment and the Computer Room Air Conditioning (CRAC) units. The air circulation and cooling unit mechanism is under the floor of data-center and cold air is blown from perforated tiles on the floor. The racks are placed on the raised floor in data-center. The racks are arranged in such a way that the cold air is sucked in from the front and hot air is blown out from the back as shown in fig.1. The hot air is sucked by ceiling vents of CRAC unit. The hot air is the result of running of IT equipment such as blade servers and network equipment inside

978-0-7695-4959-0/13 $25.00 © 2013 IEEE
DOI 10.1109/ACSAT.2012.45

the racks. Inside the CRAC unit, the hot air is blown through cooling ducts containing chilled water which cools the air. The cold air is again blown back to the data-center through the floor tiles. The blade servers are arranged in the form of groups called chassis. All the blades inside a chassis share a single power source. Multiple chassis are integrated inside each rack.

Data-centers use a significant amount of electricity for cooling [10, 11] and it is a major reason of increasing the PUE. The cooling cost can become equal or even dominate the IT operations cost as in past [9, 11] and is certainly not desirable for cloud venders and data-center owners. A typical rack fully loaded with blade servers can consume up to 25 kW of power which is sufficient for about 15 houses in USA [10]. These servers on the contrary are stuffed in a very confined space. And there are multiple rows of racks inside data-center hall to get the full benefit of free space available. Thus the servers and other IT equipment such as network switches create enormous amount of heat when turned on. The power consumed by servers is converted into heat. CPU is the most power consuming and most heat emitting component of computer hardware. Each server needs to be operated in vendor specified limits or thresholds such as environment temperature and humidity. Energy efficient load balancing on servers can effectively control temperature across data-center. Our focus in this paper is to analyze different approaches proposed for lowering the cooling cost in recent past and identify the possibilities of energy efficient workload scheduling. We remain to software based controls and management.

Maintenance of temperature through air circulation is the responsibility of Computer Room Air Conditioning (CRAC) unit. The hot air from data-center is sucked in from the top vent of CRAC unit and is passed through water chilled coils to cool it. The water chiller is located outside the data-center[10]. The cool air is then blown inside data-center through perforated tiles. Each server has an air aisle in front called cold aisle to suck in air from outside and blow it over its components to keep them operational. As shown in fig.1, the cold air is supplied to servers in racks from front via cold aisles. The CRAC unit blows the cold air through perforated floor tiles on the raised floor. The hot air is blown out through the rear exhaust fan or hot aisle of the blade servers. The racks are arranged in lines between the hot and cold aisles as shown in fig 1.

As the server load increases in terms of CPU utilization, it consumes more electricity and its temperature rises. The temperature of hot air coming from hot aisle of the server rises up. Therefore it might require more cold air to keep the temperature in desired range. Otherwise some jobs can be migrated from that server to reduce the workload and thus to reduce the power consumption and heat emission.

Thus the workload scheduling for each server can be done by considering the thermal state of that server. In cloud data-center architecture as shown in fig.1, there is a possibility of hot air getting mixed with cold air because both of these are blown in the data-center environment. The temperature of cold air coming from floor tiles and the cold air reaching the server inlets gets different. This is called heat-recirculation. This is a typical challenge for supplying uniform temperature cold air through-out the data-center.

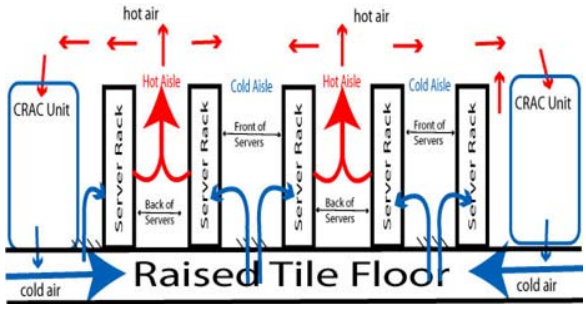

Fig. 1. Typical data-center Hot Aisle / Cold Aisle layout. Source: CIT [13]

Modern day blade servers have thermal sensors inside to monitor cold air inlet temperature and the internal temperature of CPU for example. Thus the thermal status of every server on premises can be monitored. The combined thermal status of all the servers in data-center is called thermal map. For legacy systems, multiple external thermal sensors are mounted on racks at front and back of the racks. These sensors can be monitored via Wireless Sensor Network (WSN). A thermal map can provide excellent information for workload scheduling, migration and management of cooling in a cloud data-center. A scheduling technique that considers the thermal conditions of data-center can be called a "Thermal aware scheduling". It includes the heat emission and recirculation considerations. In this paper we examine various thermal aware work load scheduling techniques proposed in literature and analyze them with respect to cloud computing for the implementation of Green Cloud Computing. These techniques and algorithms can be classified as "reactive" or "proactive". In this paper we classify thermal-aware scheduling techniques. This will be helpful for understanding the trade-offs of implementing either of them in data-center. In our future work, we shall test these techniques in our test bed and propose a new technique.

III. THERMAL-AWARE REACTIVE & PROACTIVE SCHEDULING AND COOLING

A data-center management can be regarded as proactive if it schedules and provisions the resources in anomaly avoidance way. In this paper we remain to thermal anomalies such as maximum inlet temperature violation and overcooling. A proactive approach requires planning which includes prediction and estimation. A prediction can be of thermal map after a batch of jobs is scheduled and may require building of thermal profiles of servers or racks. A thermal profile stores the trend of temperature change within or at outlet air aisle of the server with the change in workload. In this paper we focus on CPU utilization as a measure of workload. Thermal profiles of unique tasks such as benchmarks and software applications can also be created. The thermal profiles help in deciding

about distribution of jobs across data-center. Other techniques can help in predicting a thermal map. Such techniques can be neural networks[14], machine learning[15] and as simple as moving average[15]. The efficiency of proactive approach depends upon the prediction methods which have their limitations[15].

A reactive approach takes steps after an anomaly has occurred. It is the management without planning. A reactive approach does not require prediction and estimation. Neither is it required to maintain thermal profiles. It performs remedy measures after the happening of an anomaly. For example after the inlet temperature of a server has exceeded the maximum threshold. A reactive approach can thus lead to disaster such as equipment failure due to overheating. But a reactive approach is faster than proactive approach. It monitors the real time environment. A proactive approach requires a training phase or a learning phase and is therefore slow to implement in start. A reactive approach can perform well where the jobs show unstable behavior with respect to resource utilization or where the jobs show seldom deviation from a static behavior. In the latter case the reactive approach will manage to stabilize the data-center after a few steps depending upon the static utilization of resources by the jobs being executed. [16].

HP laboratories [17, 18] have done implementation work on improving data-center cooling management and thermal aware workload placement. Workload prediction is done periodically for multi-tiered web application based on the knowledge of resource utilization at each tier of application in [17] while in [18] the execution time of all job is already known. Both [17] and [18] place workload on the basis of how efficiently each server is cooled. The scheduler checks LWPI [17, 18] ranking which is the index of how efficiently a server is cooled, before allocating workload. A low LWPI value means the server is not being cooled efficiently. VMs are migrated away from servers with low LWPI value[17]. Whereas in [18] the longest jobs are scheduled on coolest servers. Idle servers are turned off and cooling is shut down[17, 18] from the nearby CRAC unit. Both [17] and [18] require prediction based or test-run based knowledge of job execution statistics. Without the prior knowledge of arriving work load, these two approaches will just be scheduling workload on the basis of thermal ranking of servers in data-center. Therefore we conclude that the scheduling for workload is proactive and cooling is reactive in both [17] and [18].

In other case the cooling can be increased for such servers through DSC[19] when the scheduler puts maximum workload on minimum servers through a genetic algorithm. Dynamic Smart Cooling (DSC) [19] technique is a reactive cooling technique that dynamically increase or decrease the cooling according to temperature of servers within the thermal zone of each CRAC unit. One of the thermal sensors is chosen as the respective CRAC unit controller from each zone. This avoids over cooling because DSC is implemented zone wise per CRAC unit. But they have not specified how they chose the external sensor.

Another work by M. Marwah et al. [15] of HP labs is testing of various predictive methods based on simple threshold, moving average, weighted moving average and machine learning naïve Bayesian classifier. We comment that this [15] approach can be used to identify a CRAC controller external sensor of [19]. The classifier of [15] should have been tested for multiple sensors for thermal anomaly prediction and the sensor giving best accurate prediction should be used for DSC controller of [19].

A set of proactive thermal aware workload scheduling algorithms were given in [20]. These algorithms are based on the phenomena of power budget distribution among servers on the basis of minimizing resulting heat re-circulation at that power usage. They created thermal profiles of group of servers called *pods.* These profiles are helpful in decision making for power budget allocation. They have quantified the heat generated and heat recirculation for each pod. Each pod is allocated power budget based upon its contribution to total heat recirculation in data-center. However these profiles were made by turning on only one pod at one time throughout an idle data-center rather than in a data-center that is being utilized.

The writers in [21-24] created few thermal aware proactive scheduling techniques to distribute power and workload based upon the thermal states and heat recirculation. Writers have quantified heat from law of energy conservation and captured the heat recirculation in matrix form for the entire data-center. The power consumed by a server is turned into heat which can be calculated through known temperatures of inlet and outlet air of server and the thermodynamic constants. By knowing or keeping constant two of the power, the cold air coming in and the hot air going out, experiments have been done for energy efficiency in [21-24] and proactively workload allocation. Algorithms in [21] are the discrete forms of baseline algorithms of [20] The authors in [23] suggested a genetic algorithm for proactive scheduling with the fitness function to minimize the peak inlet temperature of the servers due to heat recirculation. They have however not specified that the genetic algorithm considers the thermal profiles as in [22] or it requires CFD simulation to evaluate each of the candidate solutions.

Tang et al. in [24] projected the problem of minimizing heat recirculation as a minmax problem. They have proposed a proactive thermal aware scheduling with the consideration of minimizing the maximum inlet temperature of servers. This is the same idea proposed by [25] to alter the thermostat set-point higher inorder to save cooling energy. This is a proactive approach which keeps the supplied temperature within maximum threshold and yet raises the CRAC unit's thermostat set-point. But the thermostat settings are to be updated for each job placement. This requires a programmable thermostat. Another notable thing is that the CRAC unit takes some time to change the mode and therefore job scheduling and

execution will be strained. And frequent mode changing will affect the CRAC unit service life. This is different from DSC [19] in which each CRAC unit is associated with a thermal zone but in [25] the writers have not considered this.

The writers in [14] have suggested to use neural networks and artificial intelligence to predict the thermal map. They have not given any test results. However this approach can be compared to fast thermal evaluation method of [22] and the machine learning prediction method of naïve Bayesian classifier[15].

A proactive approach by Banerjee et. al [25] considers to update thermostat settings of CRAC unit to supply the cold air at maximum temperature. This is done by not violating the redline temperature for the servers given the power distribution vector and heat recirculation matrix. This can be a good power saving algorithm but it requires a programmable thermostat, further in a data-center with number of parallel running jobs with variable ending time, it seems hard to implement such a technique to dynamically update thermostat setting without delaying job ending time and break-up of thermostat. Also the new thermostat settings require the power distribution across data-center. The writers have calculated the power required from the estimated run-time of the jobs considering the jobs to be CPU intensive. Where as in real time, the jobs may not be CPU intensive.

Another proactive approach based on task based thermal profiling was introduced in [26]. In it, the authors have scheduled jobs on the basis of hottest job on coolest server and hot-job-before-cold-job scheduling. Writers have used the RC thermal model and thermal profile of each job to predict the post-job temperature of a node. This scheduling can avoid the maximum temperature threshold proactively and reduce cooling cost. Jobs are held till suitable node with respect to post-job temperature is available. The cooling cost is reduced because executing hot job before cool job results in lower overall system temperature [27]. This has a disadvantage of longer response time due to job holding. A rack level thermal predictor was proposed in [28] to choose a chassis within rack for workload placement. It uses neural network to predict a chance of hotspot. It gives six options from which one is chosen on the basis of lowest air outlet temperature. But writers did not implement it. Writers allocate server on the basis of power usage efficiency instead of thermal efficiency. Writers in [29] have proposed a reactive technique to lower the processor frequency through DVFS in when temperature goes up. The technique however requires constant monitoring and suffers from lower response time. However it is energy efficient and allows raising the supple air temperature of CRAC unit.

A hierarchical proactive thermal monitoring technique was proposed in [30] involving sensor nodes and thermal cameras for monitoring. The writers used neural network unsupervised learning method to predict thermal behavior of VMs and anomaly avoidance. The VM allocation is done through multi-dimensional bin packing, taking heat imbalance as one of the dimensions. But the writers did not mention how they will predict the behavior of VMs that they create dynamically to avoid a hotspot after a VM is provisioned.

A proactive cooling technique was proposed in [16]. The authors tried to manipulate the CRAC unit compressor cycles and fan blow rate to proactively manage cooling system to match the thermal trend of benchmark programs. They compared reactive and proactive cooling and found that proactive cooling with multiple fans will be energy efficient. But the proactive technique requires thermal signatures of the application prior to proactively control cooling.

IV. CONCLUSION

During our review we came across various techniques for data-center cooling and scheduling for cost efficiency. Classification of these approaches is the first step. These techniques can be updated to include virtualization and can be applied to cloud computing. In our future work, we shall provide in-depth review of these techniques and include topics such as thermal map generation techniques, heat quantification approaches and heat recirculation management. We shall present our work with practical implementation on cloud computing.

REFERENCES

1. Han, Y., *Cloud computing: case studies and total costs of ownership*. Information Technology & Libraries, 2011. **30**(4): p. 198-206.
2. D. Kondo, B.J., P. Malecot, F. Cappello and D. P. Anderson, *Cost-benefit analysis of cloud computing versus desktop grids*, in *23rd IEEE International Symposium on Parallel Distributed Processing*2009: Rome, Italy.
3. Google *Cloud Computing—What is its Potential Value for Your Company?* 2009.
4. Hayes, B., *Cloud computing.* Commun. ACM, 2008. **51**(7): p. 9-11.
5. Amazon *Getting Started with Amazon EC2 and Amazon SQS*.
6. Skobel, R.W.a.E. *Hardware vs. Amazon EC2 Cloud Performance In the Cloud*.
7. Oracle *Oracle's Cloud Solutions for Public Sector*. 2012.
8. Oracle *The Most Complete and Integrated Virtualization: From Desktop to Datacenter*. 2010.
9. Koomey, J., *Worldwide electricity used in data centers*, 2008.
10. Program, U.S.E.P.A.E.S., *Report to Congress on Server and Data Center Energy Efficiency Public Law 109-431*, 2007.
11. Koomey, J., *Growth in Data center electricity use 2005 to 2010*, 2011.
12. *Data Center Industry Leaders Reach Agreement on Guiding Principles for Energy Efficiency Metrics*. U.S. Environmental Protection Agency ENERGY STAR Program 2010.
13. *Energy Efficient Equipment in the NIH Data Center*. Available from:

http://datacenter.cit.nih.gov/interface/interface240/energy_efficiency.html.

14. Michael Jonas, G.V., Sandeep K. S. Gupta, *On developing a fast, cost-effective and non-invasive method to derive data center thermal maps*, in *IEEE International Conference on Cluster Computing*2007: Austin, TX, U.S.
15. Manish Marwah, R.S., Cullen Bash, *Thermal Anomaly Prediction in Data Centers*, in *12th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*2010: Las Vegas, Nevada, U.S.
16. Eun Kyung Lee, I.K., Dario Pompili, Manish Parashar, *Proactive thermal management in green datacenters.* Journal of Supercomputing, 2010. **51**: p. 1-31.
17. Yuan Chen, D.G., Chris Hyser, Zhikui Wang, Cullen Bash, Christopher Hoover, Sharad Singhal, *Integrated Management of Application Performance, Power and Cooling in Data Centers*, in *IEEE Network Operations and Management Symposium (NOMS)*2010.
18. Cullen Bash, G.F., *Cool Job Allocation: Measuring the Power Savings of Placing Jobs at Cooling-Efficient Locations in the Data Center*, in *Technical Reports*2007, HP Laboratories Palo Alto.
19. Cullen E. Bash, C.D.P., Ratnesh K. Sharma, *Dynamic Thermal Management of Air Cooled Data Centers*, in *Intersociety Conf. on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM)*2006: San Diego, CA, U.S.
20. J. Moore, J.C., P. Ranganathan, R. Sharma, *Making scheduling "cool": Temperature-aware workload placement in data centers*, in *USENIX Annual Technical Conference*2005: Anaheim, CA, U.S. p. 61 – 75.
21. Qinghui Tang, S.K.S.G., Daniel Stanzione, Phil Cayton, *Thermal-Aware Task Scheduling to Minimize Energy Usage of Blade Server Based Datacenters*, in *2nd IEEE International Symposium on Dependable, Autonomic and Secure Computing*2006: Indianapolis, IN, U.S.
22. Qinghui Tang, T.M., Sandeep K. S. Gupta, Phil Cayton, *Sensor-Based Fast Thermal Evaluation Model For Energy Efficient High-Performance Datacenters*, in *Fourth International Conference on Intelligent Sensing and Information Processing ICISIP*2006: Bangalore, India.
23. Qinghui Tang, S.K.S.G., Georgios Varsamopoulos, *Thermal-Aware Task Scheduling for Data centers through Minimizing Heat Recirculation*, in *IEEE International Conference on Cluster Computing*2007: Austin, TX, U.S.
24. Qinghui Tang, S.K.S.G., Georgios Varsamopoulos, *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): p. 1458 - 1472.
25. Ayan Banerjee, T.M., Georgios Varsamopoulos, Sandeep K. S. Gupta, *Cooling-aware and thermal-aware workload placement for green HPC data centers*, in *International Green Computing Conference, 2010* 2010: Chicago, IL, U.S.
26. Lizhe Wangt , G.v.L., Jai Dayalt, Thomas R. Furlanit, *Thermal aware workload scheduling with backfilling for green data centers*, in *IEEE 28th International Performance Computing and Communications Conference (IPCCC)*2009: Phoenix, Arizona, USA.
27. Jun Yangt, X.Z., Marek ChrobakV, Youtao Zhang, Lingling Jint, *Dynamic Thermal Management through Task Scheduling*, in *IEEE International Symposium on Performance Analysis of Systems and software, ISPASS*2008: Austin, Texas, USA.
28. Hui Chen, M.S., Junde Song, Ada Gavrilovska, Karsten Schwan, *HEaRS: A Hierarchical Energy-aware Resource Scheduler for Virtualized Data Centers*, in *2011 IEEE International Conference on Cluster Computing*2011: Austin, TX, U.S.
29. Shen Li, T.A., Mindi Yuan, *TAPA: Temperature Aware Power Allocation in Data Center with Map-Reduce*, in *International Green Computing Conference and Workshops (IGCC)*2011: Urbana, IL, USA
30. Hariharasudhan Viswanathan, E.K.L., Dario Pompili, *Self-Organizing Sensing Infrastructure for Autonomic Management of Green Datacenters.* Network, IEEE, 2011. **25**(4): p. 34 - 40.