

## A hybrid algorithm to reduce energy consumption management in cloud data centers

Mehran Tarahomi<sup>1</sup>, Mohammad Izadi<sup>2</sup>

<sup>1</sup>Kish International Campus, Sharif University of Technology, Iran

<sup>2</sup>Department of Computer Engineering, Sharif University of Technology, Iran

---

### Article Info

#### Article history:

Received Aug 3, 2018

Revised Aug 30, 2018

Accepted Sep 15, 2018

---

#### Keywords:

Cloud data center

Cloudsim

Energy consumption

Load balance

Migration

---

### ABSTRACT

There are several physical data centers in cloud environment with hundreds or thousands of computers. Virtualization is the key technology to make cloud computing feasible. It separates virtual machines in a way that each of these so-called virtualized machines can be configured on a number of hosts according to the type of user application. It is also possible to dynamically alter the allocated resources of a virtual machine. Different methods of energy saving in data centers can be divided into three general categories: 1) methods based on load balancing of resources; 2) using hardware facilities for scheduling; 3) considering thermal characteristics of the environment. This paper focuses on load balancing methods as they act dynamically because of their dependence on the current behavior of system. By taking a detailed look on previous methods, we provide a hybrid method which enables us to save energy through finding a suitable configuration for virtual machines placement and considering special features of virtual environments for scheduling and balancing dynamic loads by live migration method.

*Copyright © 2019 Institute of Advanced Engineering and Science.  
All rights reserved.*

---

### Corresponding Author:

Mehran Tarahomi,  
Kish International Campus,  
Sharif University of Technology,  
Tehran, Iran.  
Email: tarahomi@ce.sharif.edu

---

## 1. INTRODUCTION

Providers of cloud services and customers are both motivated towards using cloud computing because of the gained benefits for both of them. Providers make money by providing such services and customers save money related to maintenance of infrastructures of these services. Due to the importance of running applications and user data, the providers guarantee a specific service level agreement (SLA). Virtualization, a key technology in cloud computing, make the separation of virtualized machines possible in a way that each virtualized machines will then be able to be configured on a number of hosts according to user applications [1]. However, still the specific features and attributes of the host are concealed from the user.

Therefore, virtualized machines can be transferred to other hosts without interrupting services in order to increase the efficiency of applications. It is also possible to dynamically alter the amount of resources allocated to a certain customer [2]. Nowadays, modern CPUs can run in different speeds through dynamic frequency scaling (DFS); Dynamic voltage scaling (DVS) or combined dynamic frequency and voltage scaling (DFVS). These techniques lead to formation of a non-linear power-frequency relation. According to [3], at each level of efficiency, the power-frequency relation can be approximated through a second-degree model [4, 5]. By using cloud users no more require high-end computer systems because the entire computing and storing processes are performed by well-equipped and advanced servers [6]. Currently, there are three different methods for providing cloud computing services: 1-Software as a service, 2-platform

as a service, and 3-infrastructure as a service. With respect to their special needs and requirements, organizations and firms can select and make use of a single or a series of service strategies [7-10].

Service level agreement is an important feature of cloud computing and is considered as a contract between the customer and the provider based on provided service. However, nowadays with the emergence of large cloud service providers, most service level agreements are pre-determined and pre-standardized unless the customer is a major consumer of the services. Scheduling for load balancing and energy saving in cloud services depend on response-time of the machine. The load should be balanced on CPU, Memory or the delay in the network. Existence of excessive load on the cloud network will cause several problems for service providers and consumers [11, 12].

Balancing load in servers is highly effective on energy consumption. It will also result in reducing response-time which is the most important factor in service agreement [12-14]. Load balancing is the process of distributing load between several nodes through the network. As a result, the efficiency of resources, amount of consumed energy and response time will optimize. On the other hand, the situations in which some nodes are under heavy loads and some others are simultaneously almost free, are avoided. By balancing the traffic between servers, data packets are sent and received with no major delays [15].

There are several different load balancing algorithms which include water drop exploration algorithm, active clustering algorithm and algorithms of Max-Max; Min-Min and OLB+LBMM. Important factors which should be considered while developing such algorithms are load estimation, load ratio, adaptability of different systems, system efficiency, interaction between nodes, scheduling, nature of transferred task, selection of nodes, etc. [3, 13].

In Figure 1, the following are assumed:

- There are several running applications in a virtual machine with different loads,
- Each host includes n virtual machines,
- All system component is correct at  $T=0$  and will remain that way until occurring an error,
- Failures are permanent and are independent of each other,
- The confidence of entire similar virtual machines is similar.

With respect to the aforementioned hypotheses, when a given job is addressed to a machine, if the accessible resources are used properly, then the load can be executed effectively. There should be a mechanism for selecting those machines with such resources. As you can see in Figure 1, connected users to network use the facilities by sending requests to the network. Since multiple requests from users forces the network to displace virtual machines, increasing number of users will cause serious debate [3].

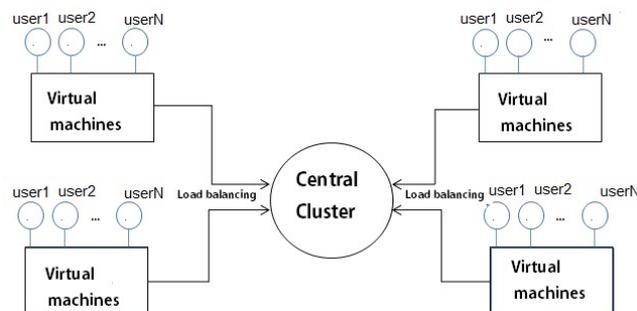


Figure 1. Addressing the applied load to central cluster

Virtualization technologies provide the processes and application data with the ability for being executed in the virtual component (virtual machines and virtual disks) without being dependent on physical resources. In other words, they make applications independent of machines. Virtualization may take place in several layers or levels in order to create a modern computer system. A level of virtualization which is applied to cloud computing data center is a systematic virtualization which enables virtual machines with different operating systems to be executed simultaneously in a real machine. The virtualization software which is called virtual machine monitor (VMM) is responsible for management and controlling resources consumption by each virtual machine. Xen [16] and VMware [17] are examples of virtualization methods which are applied on systems. In the proposed method, we are trying to suggest a method for cloud computing data center which enables it to balance the distribution of loads between hosts as well as reducing energy consumption.

## 2. OBJECTIVES AND INNOVATIONS OF THE RESEARCH

Objectives of this study include:

- a. Reducing energy consumption in virtualized data center infrastructure,
- b. Improving confidence and failure tolerance in execution of applications,
- c. Avoiding hotspots in the data center and balancing load distribution between the existing resources,
- d. Avoiding the situation of thrashing among hosts of the data center.

Innovations of this study includes

- a. Developing a dynamic algorithm based on load balancing and hardware techniques for optimizing energy consumption considering scheduling,
- b. Monitoring the provided level of service by hosts,
- c. Using live migration technology,

Variability of the scale of using processing resources according to demands of exploiters

## 3. RELATED WORKS

Reducing energy consumption is feasible in static and dynamic methods. In terms of the static method and from the perspective of hardware, energy management includes optimization methods [18] which are used at the time of designing different levels including circuits and system architecture. In terms of optimization at circuit level, by designing complex gates and altering the size of transistors, the developers focus on reducing energy consumption in combined and sequential circuits. The purpose of optimization in logic level is to reduce the energy consumption of switching in combined and sequential circuits.

This section divides the virtualization technology into three categories which include scheduling according to efficiency and load balancing, using hardware facilities and concerning thermal features of the environment. This categorization is shown in the following image and the tasks related to each of these three approaches are further explained in the text. Categorization of Related Works Regarding Energy Saving through Virtualization as shown in Figure 2.

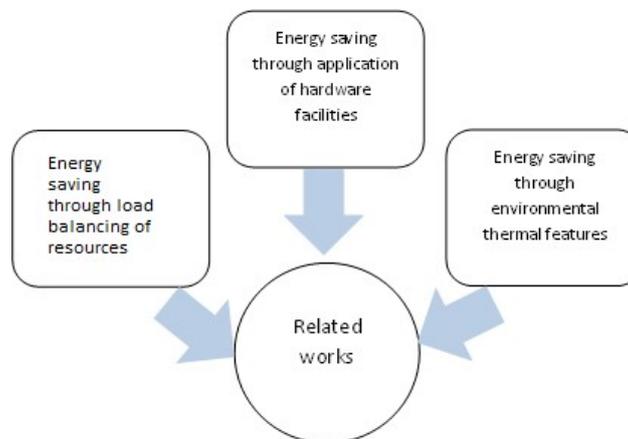


Figure 2. Categorization of related works regarding energy saving through virtualization

### 3.1. Energy saving through load balancing of resources

The RR algorithm balances load between nodes by optimizing energy consumption. This algorithm distributes tasks between the entire auxiliary processors. The entire tasks are allocated to auxiliary processes according to the RR order. This means that a selected processor is operated in an order and if the last processor is reached, the loop will return to the first processor. [19]. Another way of achieving load balancing is using random algorithm. It uses random processor numbers for selection of auxiliary processors. Auxiliary processors are selected according to statistical distributions. [20]. Another algorithm in this domain is central manager model. In this model, the main processor selects an auxiliary processor with minimum load for tasks allocation in each step.

The main processor is able to collect information from all other auxiliary processors; therefore, the choice may be made according to this algorithm. [21]. In dynamic load balance algorithms, tasks are distributed among processors during execution. The main manager allocates the new process to an auxiliary processor according to new collected information. In a distributed system, dynamic load balancing will be

done in two ways: distributed and non-distributed (centralized). In distributed one, load balance algorithms are executed on the entire machines and the responsibility of maintaining the load balance is shared between the entire engaged machines. The relation between the machines for balancing load can be in either ways of collaborated or non-collaborated. For a general balance through the system, processors work together [22, 23].

On the other hand, in non-collaborated algorithms, each machine depends on its own resources' scheduling in which decisions are made independent of the rest of the system. On this basis, the machine may schedule or allocate tasks according to local performance. In collaborated algorithms, the tasks of energy consumption monitoring and maximization of profits are considered as ROI [15].

Dynamic load balance algorithms can create more messages compared to centralized algorithms; because each machine within the system contact with all other machines within the same system. An advantage of this is that if one or some machines of the system are failed, the total procedure of load balancing will not terminate and it can only affect the performance of system. However, in these algorithms the heuristic learning functions are used which is inconsistent with the service level agreement [15, 23, 24]. Currently, several methods and experiences are applied for optimization of efficiency of energy consumption. These methods include: optimization of algorithms of applications [31] and making use of hardware with low energy consumption [25].

By implementing the virtualization technology, we can make several virtual machines on a physical server. As a result, the amount of required hardware is reduces and also efficiency of resources increases. Making use of this technology in cloud computing naturally leads to optimized energy consumption which may be based on the following reasons:

- a. Increased efficiency of resources [26]
- b. Lack of physical and location dependence; virtual machines could be moved and executed in places where energy is cheaper [15]
- c. Alteration of the scale of implementation of resources; implementation of resources can be varied according to recent requirements [10]

In [27], reduced energy consumption is investigated in a virtualized cloud network data center. In this article, the authors suggest an efficient resource management system for cloud network data centers. This system reduces operational costs and improves service quality. This systems' energy consumption saving is done through uninterrupted monitoring of virtual machines according to the current efficiency of resources. The main instrument that is applied here is making use of live migration for virtualized machines of the cloud network. Live migration is a feature for transferring virtual machines between different physical servers which have low overload and provide the cloud provider with the flexibility in which virtual machines can allocate resources dynamically and according to current and recent requirements.

In [28] an architecture is suggested which is aimed at management of energy in virtualized data centers. In this architecture, resources are managed under local and global levels. In terms of the local level, the system makes use of power management strategies embedded in guest operating systems. In other words, the same strategies which would have had been used if the system was running on a physical machine. Also in terms of the global level, for the purpose of allocation of virtualized machines, those general policies are used which make use of live migration technology.

### 3.2. Scheduling through hardware facilities

Dynamic voltage and frequency scaling is a hardware technique for dynamic management of power consumption which can be seen in some hardware pieces such as processors and CPUs. By using this capability, by reducing loads, processors can enter operational statuses with lower power consumption. Therefore, the processing power of processors can be scaled according to current requirements of the system without having a flaw in efficiency of programs [29]. Anyway, in this technique processors are able to choose their operational voltage and frequency from a restricted and limited set of values [30]. In addition, by the use of this approach, energy consumption can only be reduced up to 10 to 20 percent [31]. As a result of existing limitations, many related tasks for reduction of energy consumptions make use of the approach of centralization of loads on a fewer number of brokers and turning off the idle brokers or putting them in sleep modes or hibernate modes or even a combination of the centralization method and dynamic frequency and voltage scaling method [32].

The problem of management of power consumption in brokers with web based applications and under agreements of service level is discussed in [33]. The main goal of the authors of this paper [33] was to reduce operational costs and reduction of downtimes resulting from excessive warming of physical brokers. For this purpose, they have made use of the approach of turning operational nodes on and off in addition to dynamic voltage and frequency scaling method for management of energy consumption. They evaluated their

suggested policies through a number of simulations. Results of their simulations indicated that the policy of dynamic frequency and voltage scaling can reduce energy consumption up to 29 percent.

The mechanism of turning operational nodes on and off can also reduce energy consumption up to 42 percent. This is while combination of these two methods can only save energy up to 18 percent. Anyway, their research has failed to consider for many effective real parameters. For example, the required amount of time for turning brokers back on is not considered. In this approach, it is assumed that only one application is executed on a cluster at a time. They have neglected the changes of tasks' loads and this may lead to making inappropriate decisions. In addition, this research only investigates CPUs and neglects other units.

### 3.3. Reduction of energy consumption through awareness about thermal features of the environment

In another approach of reduction of energy consumption in virtual environments, the thermal features of the environment have been used. One of the most well-known efforts in this field belongs to Beloglazov and Buyya [34] in university of Melbourne. They have deigned a resource manager which dynamically reads thermal features of each node and tries to schedule virtual machines in a way that heavier loads are oriented towards machines which are cooler. Results of implementing this design on a simulated virtual context indicated that making use of this approach can significantly reduce the consumption of energy in virtual environments.

In two other researches in this field which are performed by Tang et al. [35, 36], the emphasis has been put on the point that calculation nodes with high temperatures are not suitable choices for allocation of tasks and activities. On this basis, different methods have been proposed for scheduling activities according to thermal characteristics of working nodes. Results of previous evaluations indicate that scheduling tasks with consideration for the temperature of the physical machine can have significant positive effects on optimization and improvement of accomplishment of tasks.

### 3.4. Comparison of related works and challenges

With daily increase of demands for applications and increased complexity of calculation environments, distributional systems and data centers are beginning to consume more and more amounts of energy. Increased energy consumption in distributional systems such as cloud computing models has led to formation of several challenges. As it can be seen in this section, different methods of reduction of energy consumption in data centers can be divided into three general categories including: 1) methods of reduction of energy consumption through balancing resources' loads; 2) scheduling through hardware facilities and 3) Reduction of energy consumption through awareness about thermal features of the environment. By studying these methods, the challenges of reduction of energy consumption in cloud networks will be as follows: a) Finding a suitable configuration for placement of virtual machines, b) Being aware of special traits of virtual environments for dynamic scheduling of virtual machines.

## 4. THE PROPOSED HYBRID ALGORITHM

By the use of migration technology we are able to transfer a virtual machine from one server to another without interrupting the service. As a result, by transferring a specific amount of load to other servers, resource management is distributed between resources as well as optimization of energy consumption. Considering the ability of modern CPUs to take different speeds, the relation between power and frequency could be well approximated by a second type model. It means that we can write:

$$P(f) = P_{min} + \alpha (f - f_{min})^2 \quad (1)$$

In the upper relation, Power is estimated in watts and Frequency is estimated in GHz and also  $\alpha$  is a coefficient based on  $W/(GH)^2$ . In our model, we have assumed that the CPUs of the hosts obey such a non-linear relation within the data center. The CPUs are run in different and limited values of frequency in a  $[f_{min} - f_{max}]$  range. This leads to a tradeoff between efficiency and power related costs. In our model, all servers and hosts obey this relation. Also  $\alpha$  is based on  $W/(GH)^2$  because the highest amount of energy is consumed when CPU is run with the highest power. With respect to this content,  $\alpha$  can be calculated as (2):

$$\alpha = \frac{P_{max} - P_{min}}{(f_{max} - f_{min})^2} \quad (2)$$

In addition, servers within the data center can go to stand-by if the amount of load on the data center was low. By taking a look at the upper relation, we can see that even if there were no loads on a server, still a minimum amount of power is required to keep the server in active mode. Therefore, putting the host into

sleep or hibernate can have a significant effect on reduction of consumed energy. We have used a relation titled as Imbalance score for the load balance algorithm and calculation of the overload. This relation is as follows:

$$IB\ score(u, T_{upper}) = e^{(u - T_{upper})} \quad (3)$$

U is the utilization level of each of the resources of the server. For example utilization level of the CPU, RAM and etc.  $T_{upper}$  is the upper threshold of server utilization which is used for initialization. The purpose of using this relation is to show the amount of load on each server and also the amount of each overload of server. If the utilization amount of the server is smaller than the threshold, then the value of  $T_{upper}$  will be negative. As a result, as the amount of load on the server is reduced, the amount of the IB score value is reduced as well. In addition, another concept titled as attractiveness is defined as follows. It's aimed at determining the attractiveness of a server for being selected as the destination of a virtual machine.

$$Attractiveness_{host} = IBscore_{host} \times LoadFraction_{vm,host} \quad (4)$$

The proposed algorithm showed in Figure 3 acts dynamically. It means that for a balanced load distribution on different resources, it only considers their current status. The purpose of using this algorithm is to avoid hotspots in servers within the data center because if the host is overloaded, the system is prone to reduction of efficiency and vulnerability. Though, this algorithm tries to keep the amount of load on each server lower than the threshold. Still, in special situations in which the systems in under heavy loads of work, we need to allocate the entire existing infrastructure resources to virtualized machines and makes as many virtual machines as possible on the infrastructure of our cloud computing network. Also in this situation, a large amount of load may be allocated to each virtualized machine. In this case, this algorithm is able to impose an overload on the model.

```

Load-Balancing algorithm
For every power-on host
  If host-utilization > higher-threshold-utilization then
    Add some VMs of host to migrate-set
    Until host-utilization <= higher-Threshold-utilization
  For every VMs in migrate-set
    If (find target host) then
      Migrate VMs.

```

Figure 3. Proposed Hybrid Algorithm

As mentioned, machines are transferred lively. After selection of certain machines, the algorithm selects a number of hosts as destinations of virtualized machines. Among these servers, according to policies, the one is selected which has a smaller attractiveness function value. This algorithm also investigates if the selected overload of overload threshold of server is not exceeded after transferring the virtual machines. If so, another host or server is selected. In fact a host is selected which has a small value of attractiveness function and will not be overloaded after transfer. The former inhibits the situation of thrashing. After selection of the destination host, the process of migration starts and after this process, the execution of the programs continues by the new host and therefore resources of the source host are freed.

## 5. EVALUATION OF THE PROPOSED METHOD

In order to evaluate the proposed algorithm, we use Cloudsim software. In fact, the core of this simulator for modeling the algorithm of load balance is improved as well as modeling the effects of the DVFS technology in energy consumption and calculating added overload. Service level agreement is calculated in each time interval. Our simulated cloud computing data center is consisted of 100 servers which are able to be virtualized. In fact, it is being hypothesized that virtualization software such as Xen are installed on them. Each of the hosts is modeled as a single core host each of which has a power of 1000, 2000 or 3000 MIPS. In addition, each server has 8 GB of RAM and 1 TB of H.D.D. we have created 220 virtualized machines on this center. One application is run in each machine and each application is also consisted of 1.500.000 instructions.

The purpose of proposed method is to balance load between machines and avoiding the formation of hotspots. However, implementing this algorithm violates the service level agreement. Figure 4 shows the simulation output comparison of the DVFS method with the proposed method. This figure shows the amount of violation of service level agreement with a utilization threshold of 80 percent on each server during the execution of programs and applications. Since energy consumption has a large cost for providers of cloud based services, in our simulation we have used the DVFS method in conjunction with putting idle servers to standby.

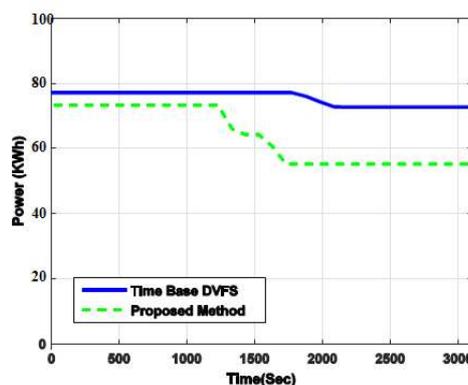


Figure 4. The simulation output comparison of the DVFS method with the proposed method

## 6. CONCLUSIONS

Cloud computing is a novice concept and with respect to its advantages for both the provider and costumers, it is rapidly growing. However, in order to be able to realize the promises of the cloud computing, there are some issues that require further investigation. In this paper, we have proposed a method for reducing consumed energy based on load balancing and hardware techniques in data centers. Using proposed algorithm balance load between the existing resources as well as reduction of consumed energy and avoiding formation of hotspots within data centers. For execution of each requested application, instead of running the request in a single virtualized machine, we run that request in three separate virtualized machines which are created on different servers. Also these servers may be selected from different clusters.

## REFERENCES

- [1] L. Wang, R. Ranjan, J. Chen, and B. Benatallah, "Cloud Computing: Methodology, Systems, and Applications," CRC Press, 2011.
- [2] K. Chandrasekaran, "Essentials of Cloud Computing," Taylor & Francis, 2014.
- [3] D. Schaefer, "Cloud-Based Design and Manufacturing (CBDM): A Service-Oriented Product Development Paradigm for the 21st Century," *Springer International Publishing*, 2014.
- [4] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg and I. Brandic, "Cloud Computing and Emerging IT Platforms: Vision, Hype, And Reality for Delivering Computing as the 5th Utility," *Futur. Gener. Comput. Syst.*, vol. 25, no. 6, p. 17, 2009.
- [5] J. W. Rittinghouse and J. F. Ransome, "Cloud Computing: Implementation, Management, and Security," CRC Press, 2009.
- [6] K. Keahey, M. Tsugawa, A. Matsunaga, and J. Fortes, "Sky computing," *IEEE Internet Comput.*, vol. 13, no. October, pp. 43–51, 2009.
- [7] I. Khalil, A. Khreishah, and M. Azeem, "Cloud Computing Security: A Survey," *Computers*, vol. 3, no. 1, pp. 1–35, Feb. 2014.
- [8] A. Beloglazov and R. Buyya, "Optimal Online Deterministic Algorithms and Adaptive Heuristics for Energy and Performance Efficient Dynamic Consolidation of Virtual Machines in Cloud Data Centers," *Concurr. Comput. Pract. Exp.*, vol. 24, pp. 1397–1420, 2012.
- [9] Y. Ding, X. Qin, L. Liu, and T. Wang, "Energy Efficient Scheduling of Virtual Machines in Cloud with Deadline Constraint," *Futur. Gener. Comput. Syst.*, vol. 50, pp. 62–74, 2015.
- [10] S. Aslazandeh, Z. Chaczko, and C. Chiu, "Cloud Computing #x2014; The effect of Generalized Spring Tensor Algorithm on Load Balancing," in *Computer Aided System Engineering (APCASE), 2014 Asia-Pacific Conference on*, 2014, pp. 5–8.

- [11] L. Chen, H. Shen, and K. Sapra, "RIAL: Resource Intensity Aware Load Balancing in Clouds," in *INFOCOM, 2014 Proceedings IEEE*, 2014, pp. 1294–1302.
- [12] S. G. Domanal and G. R. M. Reddy, "Optimal Load Balancing In Cloud Computing By Efficient Utilization Of Virtual Machines," in *Communication Systems and Networks (COMSNETS), 2014 Sixth International Conference on*, 2014, pp. 1–4.
- [13] M. Mashaly and P. J. Kuhn, "Load Balancing In Cloud-Based Content Delivery Networks Using Adaptive Server Activation/Deactivation," in *Engineering and Technology (ICET), 2012 International Conference on*, 2012, pp. 1–6.
- [14] M. Ajit and G. Vidya, "VM Level Load Balancing In Cloud Environment," in *Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on*, 2013, pp. 1–5.
- [15] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, "Xen and the art of virtualization," *ACM SIGOPS Oper. Syst. Rev.*, vol. 37, p. 164, 2003.
- [16] R. N. Calheiros, "Automated Emulation of Distributed Systems through System Management and Virtualization," PhD Thesis, Pontif. Cathol. Univ. Rio Gd. do Sul Porto Alegre, Brazil, 2009.
- [17] S. B. Shaw and A. K. Singh, "A Survey On Scheduling And Load Balancing Techniques In Cloud Computing Environment," in *Computer and Communication Technology (ICCCCT), 2014 International Conference on*, 2014, pp. 87–95.
- [18] P. Abrahamsson, S. Helmer, N. Phaphoom, L. Nicolodi, N. Preda, L. Miori, M. Angriman, J. Rikkila, X. Wang, K. Hamily, and S. Bugoloni, "Affordable and Energy-Efficient Cloud Computing Clusters: The Bolzano Raspberry Pi Cloud Cluster Experiment," *Cloud Computing Technology and Science (CloudCom), 2013 IEEE 5th International Conference on*, vol. 2, pp. 170–175, 2013.
- [19] S. Devadas and S. Malik, "A Survey Of Optimization Techniques Targeting Low Power VLSI Circuits," *32nd annual ACM/IEEE Design Automation Conference*, pp. 242–247, Spain, 1995.
- [20] V. Petrucci, O. Loques, and D. Mosse, "Dynamic Optimization Of Power And Performance For Virtualized Server Clusters," *Proceedings of the 2010 ACM Symposium on Applied Computing*, pp. 263–264, USA, 2010.
- [21] G. von Laszewski, L. Wang, A. J. Younge, and X. He, "Power-Aware Scheduling Of Virtual Machines In Dvfs-Enabled Clusters," in *Cluster Computing and Workshops, IEEE International Conference on Cluster Computing*, pp. 1–10, 2009.
- [22] E. Rotem, A. Naveh, M. Moffie, and A. Mendelson, "Analysis Of Thermal Monitor Features Of The Intel Pentium Processor," *TACS Workshop*, Hongkong, 2004.
- [23] D. Kusic, J. O. Kephart, J. E. Hanson, N. Kandasamy, and G. Jiang, "Power And Performance Management Of Virtualized Computing Environments Via Lookahead Control," *Cluster computing*, vol. 12, pp. 1–15, 2009.
- [24] E. Elnozahy, M. Kistler, and R. Rajamony, "Energy-Efficient Server Clusters," *Workshop on Power-Aware Computer Systems*, pp. 179–197, Italy, 2003.
- [25] A. Beloglazov and R. Buyya, "Energy Efficient Resource Management In Virtualized Cloud Data Centers," *IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*, pp. 826–831, Sydney, 2010.
- [26] Q. Tang, S. Gupta, and G. Varsamopoulos, "Thermal-Aware Task Scheduling For Data Centers Through Minimizing Heat Recirculation," *IEEE International Conference on Cluster Computing*, pp. 129–138, USA, 2007.
- [27] Q. Tang, S. K. S. Gupta, and G. Varsamopoulos, "Energy-Efficient Thermalaware Task Scheduling For Homogeneous High-Performance Computing Data Centers: A Cyber-Physical Approach," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 19, pp. 1458–1472, 2008.
- [28] S. Banerjee, M. Adhikari, and U. Biswas, "Development of a Smart Job Allocation Model for A Cloud Service Provider," in *Business and Information Management (ICBIM), 2014 2nd International Conference*, pp. 114–119, 2014.
- [29] F. Chang, R. Viswanathan, and T. L. Wood, "Placement in Clouds for Application-Level Latency Requirements," in *Cloud Computing (CLOUD), 2012 IEEE 5th International Conference*, pp. 327–335, 2012.
- [30] A. M. Alakeel, "A Guide to Dynamic Load Balancing in Distributed Computer Systems," *Int. J. Comput. Sci. Inf. Secur.*, vol. 10, no. 6, pp. 153–160, 2010.
- [31] S. Yan, X. Wang, M. Razo, M. Tacca, and A. Fumagalli, "Data Center Selection: A Probability Based Approach," *Transparent Optical Networks (ICTON), 2014 16th International Conference on*, pp. 1–5, 2014.
- [32] H. Wang and K. Bergman, "Optically Interconnected Data Center Architecture For Bandwidth Intensive Energy Efficient Networking," *Transparent Optical Networks (ICTON), 2012 14th International Conference on*, pp. 1–4, 2012.
- [33] A. Beloglazov and R. Buyya, "Energy Efficient Resource Management in Virtualized Cloud Data Centers," *Cluster, Cloud and Grid Computing (CCGrid), 2010 10th IEEE/ACM International Conference on*, pp. 826–831, 2010.
- [34] T. K. Ghosh, R. Goswami, S. Bera, and S. Barman, "Load Balanced Static Grid Scheduling Using Max-Min Heuristic," in *Parallel Distributed and Grid Computing (PDGC), 2012 2nd IEEE International Conference on*, 2012, pp. 419–423.
- [35] R. Uргаonkar, U. C. Kozat, K. Igarashi, and M. J. Neely, "Dynamic Resource Allocation And Power Management In Virtualized Data Centers," *Netw. Oper. Manag. Symp. NOMS 2010 IEEE*, no. Vm, pp. 479–486, 2010.
- [36] J. Adhikari and S. Patil, "Double Threshold Energy Aware Load Balancing In Cloud Computing," in *Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference*, pp. 1–6, 2013.