Energy Efficient Resource Scheduling through VM Consolidation in Cloud Computing

Ahmad Fayyaz[1], Muhammad U. S. Khan[1],
[1]COMSATS Institute of Information Technology,
Abbottabad, Pakistan
e-mail: {afayyaz, ushahid}@ciit.net.pk

Samee U. Khan[2]
[2]North Dakota State University, Fargo, ND, USA
e-mail: samee.khan@ndsu.edu

Abstract— **In modern computing paradigms, energy consumption is of foremost importance because it constitutes a major percentage of the operational expenses. Virtual Machines (VMs) are consolidated on a fewer number of servers to significantly reduce the overall power consumption. The idle servers are then turned off by using Dynamic threshold Voltage Scaling (DTVS). The VM monitor checks for under-utilized, partially filled, over-utilized, and empty servers. The VM monitor then transfers the tasks to suitable servers for execution if subject to a set of constraints. As a result, many servers those were under-utilized get idle and are turned off by using DTVS to save power. Simulation results affirm our study and a substantial reduction is observed in the overall power consumption of the cloud data centers.**

Keywords- Virtual Machine, servers, cloud computing, consolidation

I. INTRODUCTION

The dual influence of increasing cloud computing data center energy consumption and increasing energy costs has raised the significance of cloud computing data center efficiency as a policy to decrease costs, accomplish size and indorse environmental responsibility. The data centers are the most integral part for most of Information Technology (IT) organizations. Many renowned organizations, such as Google, Microsoft, and Amazon have big data centers that contain thousands of computing servers around the world to provide fast and efficient cloud computing services to the customers [1]. The past decade has witnessed a phenomenal increase not only in the size of the existing data centers but also in the number of new data centers. The aforementioned situation has increased the word-wide power consumption that drive many research communities to carry out research on the data center energy consumption, energy efficient techniques for computing units, and power consumption prediction of the data centers [2-3]. In 2006, the Environmental Protection Agency (EPA) conducted a study that revealed that the data centers are consuming more than 61 Tera Watt hour (TWh) of electricity per year that was 1.5% of the total power consumption of the whole US for the same year. Figure 1 shows the energy consumption for a typical data center and evaluated how energy is used within the data center. The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) [4], has published a trend that by 2014, the energy and infrastructure costs of the data center will contribute about 75% in the total data center cost, while IT will contribute the remaining 25% in the overall operating cost of the data center [5].

The quantity of computing resources and hardware power inefficiency are not the only factors that result in enormous energy consumption in the data centers. The inefficient use of resources, such as CPUs and memory also play a major role in the increase of data center energy consumption.

In [6], the authors collected a data from more than 5000 computing servers in a data center over a period of six-months and reported that the data center server utilization is seldom 100% even when the servers are not idle. More than 90% of the servers were running at 10-50% utilization of their total 100% capacity. This phenomenon results in extra expenses on over provisioning that directly increase the total power consumption cost of the data center [6]. Moreover, handling and preserving over-provisioned data center's resources result in increased Cost of Ownership (TCO). In another study [7], authors reported that if the data center servers are completely idle even then the power consumption is 70% of their total peak power consumption. Therefore, it is a known conclusion that the underutilization of the servers in data centers is extremely inefficient in terms of energy consumption.

In [8], a comprehensive study is conducted that monitors energy consumption of Grid'5000 infrastructure. The authors reported significant opportunities for energy saving in the data center via techniques, such as switching servers on and off with respect to utilization or run the servers on low power mode. The idle power consumption of the data centers can be reduced by switching idle servers to low-power modes.

Moreover, higher the energy consumption by the data center's infrastructure, higher is the carbon dioxide (CO2) emissions that contribute to the greenhouse effect [9]. One simple solution for the energy inefficiency in the data centers is to involve virtualization technology [10]. The virtualization technology provides opportunities to the Cloud providers to create several Virtual Machines (VMs) on a single physical computing server.

Moreover, the use of live migration [11], we can dynamically consolidate the VMs to the minimal number of physical servers according to their current resource requirements. The key contributions of the paper are as follows:

- Reducing the energy consumption by defining different DVS levels while the targeted deadlines are met.

- Virtual Machine/task consolidation for energy minimization while taking into account the CPU power and storage capacity of the individual server and the bandwidth of the links between servers is incorporated.

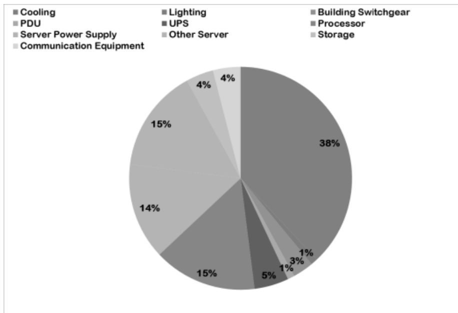

Fig. 1. Data Center Energy Consumption Partition

The remainder of the paper is arranged as follows. In Section II, the related work is elaborated. The implementation details along with the system model, problem formulation, and of the VM Consolidation model are presented in Section III. Section IV presents the simulation results and their comparison, while the concluding remarks are presented in Section V.

II. RELATED WORK

A number of literature works have been published that proposed solutions to reduce the energy utilization of the Cloud data centers. The research carried out in [14] focused on reducing the energy consumption either by only considering the hardware aspects of the data centers or in a single data center. The data centers are benefited from some renowned technologies, such as virtualization. Among the virtualization techniques, there are VMs migration [15] and consolidation [16]. However, the main issue in the VM migration or consolidation is its complexity. Moreover, the VMs resumption and suspension causes system overloading [11]. Furthermore, these methodologies are more of a reactive methods rather than proactive and preventive. Therefore, preventive methods are more important and effective. As stated in the introduction that an idle server consumes almost half of the power compared to the power it consumes at peak load [13]. The authors of [18] introduce a dynamic right-sizing on-line algorithm that predicts how many servers will be required to execute the arriving workload of the data center. The experimental results of [16] stated that dynamic right-sizing achieves significant energy savings, but the technique requires different power levels for the servers with the servers being able to switch between different power level states. In a similar work [15], Green Open Cloud (GOC) architecture is proposed that increases the prediction of the arrived requests by advance resource reservation for the users. The abovementioned technologies do not specifically consider the carbon emission and are implemented within a data center with the aim to decrease the energy consumption. The reduction in the energy consumption of the data center will not inevitably decrease the carbon footprint. The work on the availability of both non-polluting and polluting energy sources in a single data center is presented in [12]. The prediction-based scheduling algorithms are used to increase usage of green energy sources.

The Green Scheduler considers the servers to be in an order [21]. It then starts scheduling the tasks to first server from the pool until that server can execute no more tasks and is overloaded. Consolidation of tasks is achieved at the time of allocation of tasks. Our work is different from Green scheduler as it has variable sized workload as compared to the fixed sized workload of Green scheduler. In addition our technique consolidates the tasks even after the allocation phase is over i.e. we migrate the tasks from one server to the other to minimize the energy consumption even if the task is in execution phase.

The DENS [22] methodology selects the best-fit computing resources for the execution of tasks by considering the communication potential and load level of data center components. Its aim is to achieve balance between traffic demands, job performances, energy consumed by the data center, and the job QoS requirements.

Round Robin [23] scheduler equally distributes the communicational and computing loads among the switches and servers. As a result no server is overloaded and the network traffic is balanced. This scheduler is least efficient in terms of energy consumption because all the switches and servers are busy most of the time. In our work the workload is exponentially distributed to mimic the real time arrival of workload. We are also incorporating consolidation of tasks whereas Round Robin is not using any type of consolidation technique.

III. PROPOSED METHODOLOGY

One of the key causes of high power consumption is the underutilization of servers in a data center. The power consumption will obviously be higher with a higher number of running servers results. Therefore, high power efficiency can be achieved by optimal utilization of servers that will result in smaller number of turned on servers. Cloud data centers can host multiple virtual machines (VMs) on a single physical server by the use virtualization technology.

Important characteristics of the data centers that consolidate the servers resulting in cutback in the amount of hardware usage are Virtual Machine (VM) placement and scheduling. VM placement is a significant research domain in data centers where provisioning is performed manually and the cloud providers are able to enhance energy efficiency by applying proficient VM placement algorithms.

Most of the methodologies proposed in the literature only considered CPU utilization (defined in MIPS) as a decision metric for the above stated approaches. Approaching the problem with a single varying parameter and keeping other parameters static solves the problem in the controlled environment. However, the assumption of non-varying parameters reduces the effectiveness of the proposed methodologies in real environment. Therefore, there is a need of research endeavors that are not based on a single parameter. In this regard we aim to carve a strategy that considers the following constraints, all at once.

- CPU constraints
- Memory constraints
- Bandwidth constraints

A. System Model

Power consumption P(u) is defined as a function of CPU utilization [17].

$$P(u) = f.P_{max} + (1 - f).P_{max}.u, \qquad (1)$$

where P_{max} is the power consumed by computing servers, the fraction of power used by idle server is given by f and CPU utilization is denoted by u.Total energy consumption by a server is given by (Eq. 2). The CPU utilization may vary with respect to time due to variability in workloads

$$E = \int_t P(u(t))\, dt. \qquad (2)$$

B. Problem Formulation

Consider a cloud consisting of S servers, each having its own memory. Let the *i-th* server be denoted by S_i and let D_i denote the total memory capacity of S_i. Suppose that the *i-th* server has a set of VMs, V = {v_1, v_2, v_3, …, v_M}, where M represents the total number of VMs at a particular server at a given time. VM_j^i denotes *j-th* VM on the *i-th* server. Suppose d_j^i is the memory consumed by the VM_j^i and AM_i is the available memory at S_i, such that

$$AM_i = D_i - \sum_{j=1}^{M} d_j^i \qquad (3)$$

$$\sum_{j=1}^{M} d_j^i = d_i \qquad (4)$$

where d_i is the total memory being used by all the VMs at Si at any time. Each VM_j^i has some CPU requirement, i.e., the CPU utilization of VM_j^i is given by CPU_j^i, while the overall consumed CPU utilization of the server Si is represented by cpu_i.

$$\sum_{j=1}^{M} CPU_j^i = cpu_i, \qquad (5)$$

$$CPU_{avail} = CPU_i - cpu_i, \quad (6)$$

where CPU_i is total CPU power and CPU_{avail} is the available CPU power of the server S_i.

The Data Center Network (DCN) is the communicational backbone of the cloud computing [19]. We consider Fat-tree DCN architecture in our study because of its better performance in terms of throughput and average network delay. The Fat-tree DCN architecture is switch-centric network topology consisting of k pods. There are k servers and k switches within each pod. The switches are ordered in two successive layers of k/2 switches. The lower layer (edge layer) switches are linked to k/2 servers and k/2 upper layer (aggregation layer) switches in each pod. Each aggregation layer switch in the pod is linked to k/2 core level switches, out of the total of $(k/2)^2$ core level switches. The Fat-tree DCN architecture is shown in Figure 2. Let the power consumed at core level switches be denoted by P_{core}, power consumed by the aggregation level switches by P_{agg}, and at the edge level switches by P_{edge}, respectively. In addition the total power consumed by all the servers is given by P_{ser} where P_j^i is the power consumed by VM_j^i.

$$P_{ser} = \sum_{i=1}^{S} . \sum_{j=1}^{M} P_j^i$$

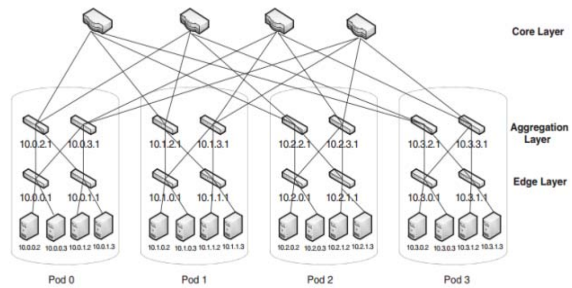

Fig. 2. Fat-tree DCN architecture

The bandwidth required by VM_j^i on server *Si* is represented by BW_j^i and the bandwidth of the slowest link, *l*, between two servers is denoted by BW_l. Suppose that the utilized bandwidth of the link *l* at any given time is denoted by $BW_l^{utilized}$. The free bandwidth on the slowest link, BW_{free} is then given as

$$BW_{free} = BW_l - BW_l^{utilized} \qquad (7)$$

Our goal is to solve the multi-objective optimization problem with multiple constraints. The problem can be formulated as follows:

The overall power consumption on both the servers and the switches in the DCN is minimized.

Mathematically

$$min\,(P_{ser}) \quad + \quad min \sum (P_{core}, P_{agg}, P_{edge})$$

and

$$min \sum_{i=1}^{n} mt_j^i$$

Subject to the following constraints

$$BW_j^i \leq BW_{free} \qquad (8)$$

$$CPU_j^i \leq CPU_{avail} \qquad (9)$$

$$\sum_{j=1}^{M} d_j^i \leq D_i \quad (10)$$

$$size(VM_j^i) \leq AM_i \quad (11)$$

$$\sum_{j=1}^{M} CPU_j^i \leq CPU_i \quad (12)$$

Constraint (8) makes sure that the free BW is more than the BW required by the VM for allocation / migration. Similarly, constraint (9) is there to ensure that there is enough CPU power available for the incoming VM. The memory consumed by all the running VMs must be less than the total memory of the server and is depicted as constraint (10). In constraint (11), the size of the VM must be less than the available memory of server where the VM is being allocated and/or migrated. Constraint (12) guarantees that the CPU power consumed by all the running VMs must be less than the total CPU power of the server.

C. Methodology

The task/VM scheduling is a two-step process. First the tasks arriving at the VM monitor (one server acts as VM monitor) are allocated to the servers in the cloud computing data centers. The tasks, as they arrive, are allocated to the servers starting from one side of the fat-tree network. The first server in the network is allocated tasks until anymore allocation of tasks overload the server. Further, the tasks are allocated to the next server and so on until all the tasks are exhausted. Virtual machines are created for the execution of each task allocated to the servers based on the CPU power and memory requirement of each task. The task continues to be executed on this VM until it is completed or migrated. Next, when the pool of tasks to be allocated is exhausted, VM/task migration is initiated. In this step the VM monitor periodically checks all the servers for the under-utilized servers. The process for migration is depicted in Algorithm 1. A server that is loaded with tasks such that its 90% of its total capacity is being utilized in the execution of tasks is termed as overloaded. We consider 90% utilization of server by tasks as overloaded because we leave 10% of CPU power for the server's own operations. On the other hand we consider a server being under-utilized if 30% of the CPU capacity is used by the tasks execution. The servers having no assigned tasks are turned off using DPM and DTVS to reduce the energy consumption in the idle mode.

When the VM monitor is done calculating the utilization if servers, it then migrates tasks from under-utilized servers to those running servers that can complete the tasks within the deadlines of the tasks. Here a delay is incurred due to the migration process and is dependent on the slowest link between the servers. Therefore, before migrating the task the estimated time to migrate and the time of execution at the target server is calculated so as to check whether the task's deadline can be met at the targeted server. If this condition is satisfied then the task is migrated and the server from which the task is migrated is turned off by using DTVS. The set of conditions that need to be satisfied for a migration to take place are listed and explained in the problem formulation.

Algorithm 1: VM Consolidation

Definitions: *S= set of server machines, tag = status of server, Maxvm = maximum number of Virtual machines that can be hosted a machine, £ = list of over utilized servers, ¥= list of filled servers, µ = list of under-utilized servers, £ = list of not filled servers, Þ = list of OFF servers*

```
1.  for each s ∈ S do
2.      e ← getEnergyCosumption(s)
3.      vmc ← getHostedVMCount(s)
4.      if e > 0.9 then
5.          £ ← £ . append(s)
6.      else if e > 0.5 and vmc = Maxvm
7.          ¥← ¥ . append(s)
8.      else if e < 0.5 and vmc = Maxvm
9.          µ ← µ . append(s)
10.     else if  e > 0.1 and vmc < Maxvm
11.         Ω ← Ω . append(s)
12.     else
13.         Þ ← Þ . append(s)
14.     end if
15. end for
16. for each s ∈ £ do
17.     vm ←getVM(s)
18.     oh ← getServerfromNotFilledOrOffLists(Ω, Þ)
19.     if oh != NULL then
20.         if  migrateVM(vm,oh) then
21.             £ ←£.remove(s)
22.             reemoveServerFromNotFilledOrOffLists(Ω, Þ,oh)
23.         else
24.             goto Line 18
25.         end if
26. end for
27. for each s ∈ Ω do
28.     while vm ←getVM(s)
29.         oh ← getServerfromNotFilledLists (Ω,s)
30.         if oh != NULL then
31.             if !migrateVM(vm,oh) then
32.                 goto Line 29
33.             end if
34.         else
35.             Break
36.         end if
37.     end while
38.     if vm == NULL then
39.         Ω ← Ω.remove(s)
40.     end if
41. end for
```

If the migrations are increased then the links and switches will experience congestion as all the traffic will be routed through the switches. Therefore the power consumed by the communication links and the switches is increased. On the contrary, the power of servers is reduced by migrating tasks from under-utilized servers and turning them off. Migration of tasks even when they are in execution is termed as task consolidation.

IV. PERFORMANCE EVALUATION

A. *GreenCloud Simulator*

We used the GreenCloud simulator to implement our proposed methodology for the purpose of performance evaluation. GreenCloud simulator was developed as an extension of the Network Simulator NS2 [20]. It captures the communication processes of the data center at the packet level. GreenCloud simulator provides users with a tool that monitors the energy consumed by servers, switches, and communication links within a cloud computing data center. Dynamic Voltage and Frequency Scaling (DVFS) and Dynamic Power Management (DPM) are the two energy efficient optimization techniques that are used by the simulator but in our proposed work we incorporate Dynamic voltage Scaling (DVS) and Dynamic Threshold Voltage Scaling (DTVS) in the GreenCloud simulator to further achieve reduction in energy consumption of the cloud data center.

B. *Simulation Results*

The simulation results are evaluated in terms of the power consumption of the cloud data center. Three scheduling techniques namely Green Scheduler, Round Robin, and Random scheduling are implemented. The results are then compared with the same scheduling techniques when VM/task consolidation is incorporated in them. It is important to note here that in the Green scheduler the consolidation of VMs/tasks is achieved at the time of allocation of tasks. It then starts scheduling the tasks to the first server from the pool until it becomes overloaded. In addition, the Green scheduler technique has fixed size workload but we considered a variable sized workload when we used task consolidation. The second technique, Round Robin scheduler, equally distributes the communicational and computing loads among the switches and servers. As a result no server is overloaded and the network traffic is balanced. Therefore, there is more room for performing VM consolidation in this technique as compared to the Green scheduler. Tasks are also scheduled randomly on different servers just for the purpose of comparison with our proposed work. The Random scheduler is included as a reference so that we can then calculate how much improvement in power consumption is achieved after VM migration is performed.

The results for the power consumption of the switches and servers are plotted individually against the changing load of the data center from 30% to 90% in figure 3 and figure 4. Figure 3 depicts power consumption of all the servers in the data center with the changing load of the cloud data center. It can be observed that the power consumption of the servers has decreased significantly after performing VM migration.

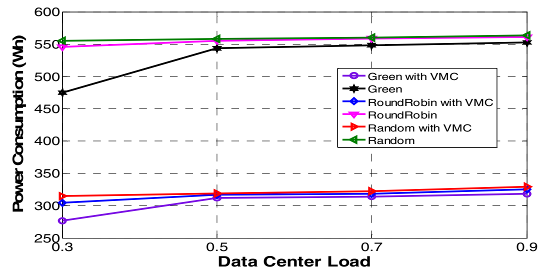

Fig. 3. Power consumption of all servers

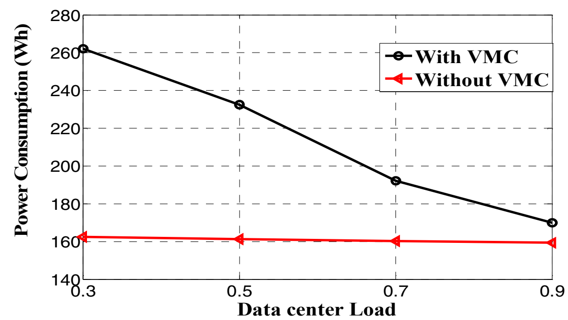

Fig. 4. Power consumption of all switches and communication links

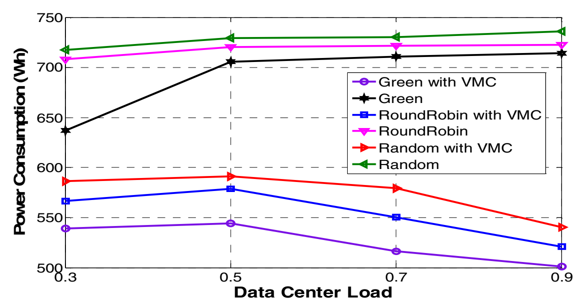

Fig. 5. Net sum of power consumption of switches and servers in the data center

As the load in the data center increases, the number of task migrations is decreased because all of the servers then work near to their full capacity, leaving less or no space for VM migrations and as a result power consumption is slightly increased. In addition, it can also be noted that the scheduling techniques have higher power consumption when VM migration is not performed.

In Figure 4, the results are plotted for the power consumption of all the switches in the cloud data center. When we perform VM migration in the data center, the network traffic increases due to the fact that many tasks are being transferred from one server to the other through the communication links and the switches. This causes congestion at the links and the switches. Due to the increased load, the communication links and switches consume more power than they would consume in the absence of VM migration. The results in Figure 4 affirm our intuition as it can be observed that the power consumption has increased when task migration is performed. It can also be noted that as the load of the data center increases the power consumption in case of VM migration is decreased because with the increase in the load of the data center there is less room for VM migration. The power consumption at the switches and the communication links is almost uniform when VM migration is not performed in the aforementioned schedulers.

Finally we plot the total power consumption of all the switches and servers (the net sum of power consumption of servers, switches and the communication links) in the cloud data center. Figure 5 depicts the results of the net power consumption of the data center both with and without task migration. By combining the effect of the power consumption of the servers and the switches we are still achieving power savings although the switches and the communication links were using more power when VM migrations increased. When the load of the data center is increased the net power consumption decreases because at this point the power consumed by the switches and communication links is significantly decreased as depicted in fig 4.

V. CONCLUSIONS

We proposed a framework that reduces the energy consumption of the cloud data center while maintaining the targeted performance. The framework utilizes the cloud infrastructure to study the effect of VM/task consolidation. By consolidating the tasks on a fewer number of servers the overall power consumed can be significantly reduced. The tasks are first allocated to suitable servers until all the tasks are exhausted. The idle servers are then turned off by using DTVS. The Virtual Machine (VM) monitor checks for under-utilized, partially filled, over-utilized, and empty servers. The VM monitor then migrates the tasks to suitable servers for execution if a set of conditions is met. By this way, many servers those were under-utilized get free and are turned off by using DTVS to save power. Simulations results confirm our study and a substantial reduction in the overall power consumption of the cloud data center is observed.

REFERENCES

[1] E. R. Masanet, R. E. Brown, A. Shehabi, J. G. Koomey, and B. Nordman. "Estimating the energy use and efficiency potential of US data centers." Proceedings of the IEEE 99, no. 8 (2011): 1440-1453.

[2] J. G. Koomey, "Worldwide electricity used in data centers. "Environmental Research Letters 3, no. 3 (2008): 034008.

[3] L. Wang, and S. U. Khan. "Review of performance metrics for green data centers: a taxonomy study." The Journal of Supercomputing 63, no. 3 (2013): 639-656.

[4] Trends, Datacom Equipment Power. "Cooling Applications." ASHRAE, http://www.ashrae.org (2005).

[5] C. L. Belady, "In the data center, power and cooling costs more than the it equipment it supports." Electronics cooling 13, no. 1 (2007): 24.

[6] L. A. Barroso, and U. Hölzle. "The case for energy-proportional computing." Computer 12 (2007): 33-37.

[7] X. Fan, W. D. Weber, and L. A. Barroso. "Power provisioning for a warehouse-sized computer." In ACM SIGARCH Computer Architecture News, vol. 35, no. 2, pp. 13-23. ACM, 2007.

[8] M. D. De Assuncao, J. P. Gelas, L. Lefevre, and A. C. Orgerie. "The Green Grid'5000: Instrumenting and using a Grid with energy sensors." In Remote Instrumentation for eScience and Related Aspects, pp. 25-42. Springer New York, 2012.

[9] The green grid consortium 2011. URL http://www.thegreengrid.org. (Accessed June 2015)

[10] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, Al. Ho, R. Neugebauer, I. Pratt, and A. Warfield. "Xen and the art of virtualization." ACM SIGOPS Operating Systems Review 37, no. 5 (2003): 164-177.

[11] C. Clark, K. Fraser, S. Hand, Jacob G. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield. "Live migration of virtual machines." In Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation-Volume 2, pp. 273-286. USENIX Association, 2005.

[12] C. D. Alfonso, M. Caballer, F. Avarruiz, and V. Hernandez, "An energy management system for cluster infrastructures," Journal of Computer and Electrical Engineering, vol. 39, no. 8, June 2013.

[13] L. Barroso, U. Holzle: "The case for energy-proportional computing." IEEE Computer 40(12), 33–37 (2007).

[14] A. Beloglazov, R. Buyya, Y. Lee, and A. Zomaya: A taxonomy and survey of energy efficient data centers and cloud computing systems. Advances in Computers 82(2), 47–111 (2011).

[15] E. Harney, S. Goasguen, J. Martin, M. Murphy, M. Westall: The e•cacy of live virtual machine migrations over the internet. In: Proc. of the 2nd International Workshop on Virtualization Technology in Distributed Computing. ACM (2007) 16. Lef evre, L., Orgerie, A.: Designing and evaluating an energy e•cient cloud. The Journal of Supercomputing 51(3), 352–373 (2010).

[16] M. Lin, A. Wierman, L. Andrew, E. Thereska: Dynamic right-sizing for powerproportional data centers. In: Proc. of the IEEE INFOCOM, pp. 1098–1106. IEEE (2011).

[17] S. Srikantaiah, A. Kansal, F. Zhao: Energy aware consolidation for cloud computing. In: Proc. of the 2008 Conference on Power Aware Computing and Systems, p. 10. USENIX Association (2008).

[18] A. Beloglazov, J. Abawajy, and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing." Future generation computer systems 28, no. 5 (2012): 755-768.

[19] K. Bilal, M. Manzano, S. U. Khan, E. Calle, K. Li, and A. Zomaya, "On the characterization of the structural robustness of data center networks," IEEE Transactions on Cloud Computing, Vol. 1, No. 1, 2013, pp. 64-77.

[20] The Network Simulator NS2: available at http://www.isi.edu/nsnam/ns (accessed may 2015)

[21] D. Kliazovich, P. Bouvry, and S. U. Khan. "GreenCloud: a packet-level simulator of energy-aware cloud computing data centers." The Journal of Supercomputing 62, no. 3 (2012): 1263-1283.

[22] D. Kliazovich, P. Bouvry, and S. U. Khan. "DENS: data center energy-efficient network-aware scheduling." Cluster computing 16, no. 1 (2013): 65-75.

[23] B. P. Rimal, E. Choi, and I. Lumb. "A taxonomy and survey of cloud computing systems." In INC, IMS and IDC, 2009. NCM'09. Fifth International Joint Conference on, pp. 44-51. Ieee, 2009.