Available online www.ejaet.com

European Journal of Advances in Engineering and Technology, 2015, 2(2): 80-91

Review Article ISSN: 2394 - 658X

A survey on Energy-Aware Cloud

S Sohrabi and I Moser

Swinburne University of Technology, Melbourne, Australia
ssohrabi@swin.edu.au

ABSTRACT

The Cloud computing has enabled businesses and individuals to utilise the potential infrastructure in the Cloud without dealing with the cost and complexities associated with large computations. It saves businesses the initial setup, updates and maintenance cost. And individuals are provided by the physical resources they might need for a time they need them. They pay as they use the resources. Cloud computing has revolutionized the way processing is carried out. Cloud led to the establishment of large data centers that contribute in the energy consumed worldwide and consequently the carbon emission and environmental drawbacks. Green Cloud computing evolves around the development of algorithms that decreases the energy consumption and became an active research area. Green cloud strategies are proposed and tested via a broad range of assumptions. Surveying these strategies can identify the fitness of them in achieving the common objectives along with the energy consumption. We identified the way energy consumption is observed and what energy saving methods are applied. Based on that we present a taxonomy and analysis of their strength and weakness of the existing methods. Ultimately, regarding the result of the analysis, the challenges are discussed and trends for future research in green Cloud computing are identified.

Key words: Green cloud, Energy-aware cloud, cloud provisioning

INTRODUCTION

Cloud computing has revolutionized the way processing is carried out. Cloud led to the establishment of large data centres that contribute in the energy consumed worldwide and consequently the carbon emission and environmental drawbacks. A study shows that data centres are contributing to a fast growing energy consumption [1] and in 2010 it was approximated to contribute 1.1 to 1.5 percent of the total electricity use [2]. According to the trends the power consumption by data centres grows at 18% rate annually [3]. Energy is reported as the second-highest operating cost for data centres in 2007 [4]. The reason might lie in the low average utilisation of resources [5] although excessive resource utilisation leads to performance degradation. Furthermore, the carbon dioxide emission by the IT industry contributes to 2 percent of the global emissions [6]. It is reported that the data centres emit as much CO2 as the whole Argentine, and this will be four times bigger by 2020 [7]. Energy usage has a direct relation with the temperature of the physical hardware. The more energy used, the more cooling systems are required which then adds to the energy consumption and the cost of the Cloud service provider. The importance of energy saving strategies has been widely addressed in the literature in different contexts including: scheduling and re-scheduling/migration or both, in regard to the objectives such as: QoS/SLA, completion/response time, resource utilisation/wastage, temperature, performance or a combination of these along with the energy consumption. Currently, a variety of green cloud algorithms and strategies are applied.

The focus of our work is to represent the differences of the strategies developed to deal with the energy consumption in the Cloud with respect to the objectives. It would assist the Cloud providers to make an informed decision of suitable algorithms according to the respective circumstances. The energy-aware taxonomy provided in this paper aims to review the literature and identify the strength and weaknesses. The main contributions of this paper can be summarized as follows: (i) it surveys the area of energy-aware resource provisioning in the Cloud to provide insights, (ii) it identifies the assumptions related to how the energy consumption is either measured or calculated and its potential impact on the total energy consumption, (iii) it then reviews how and when the energy saving strategies are deployed, (iv) it analyses the strategies according to the other common objectives to identify the capabilities and possible drawbacks, (v) in conclusion, the trends for future research in Cloud provisioning is discussed.

The flow of this paper will be in a way that first provides a short background on the provisioning software in the Cloud. Next section focuses on the way energy consumption is measured or calculated. It caters the discussions followed in the energy-aware provisioning section that offers a review of when -in the process of provisioning- the energy saving strategies are applied. It also includes the other objectives covered in the literature along with energy consumption. It is then followed by a review of some surveys in the field of Cloud computing. Challenges and Conclusion discuss the identified challenges and the future research trends.

BACK GROUND

In an IaaS Cloud, virtualisation enables the system to control the allocation of the resources to multiple applications. Virtualisation is defined [8] as "a technology that combines or divides computing resources to present one or many operating environments". IT organisations have used virtualisation for decades in mainframes and distributed systems, where it helps provide copies of hardware to support test, development and staging activities of different tasks [9]. Virtualising physical resources and specifically Virtual Machine (VM) was introduced by Intel during 1960s, which was providing interactive access to the mainframes to improve the utilisation of the system resources [10].

In a non-virtualised system the operating system has control over the hardware. In a virtualised system, the hardware supervision and access control is carried out by a software layer known as Virtual Machine Monitor (VMM) or hypervisor. The most common VMMs are VMware [11], Xen [12], Denali [13], Kernel-based Virtual Machine (KVM)[14] and Virtual PC [8]. Virtualisation can be classified according to its usage as workload isolation, workload consolidation and workload migration [15-16]. Workload isolation is the process of virtualising a task and assigning it to a VM. Workload consolidation refers to the process of assigning several heterogeneous workloads to a single physical platform. Workload migration, live migration or application mobility [12] refers to moving a VM from a server to another. The motivation for migration can be load balancing or potential failure prediction. The VMM arbitrates the access to the physical resources so that different operating systems in VMs can share the host infrastructure which can be termed resource provisioning. Energy efficient resource provisioning has become known as the concept of 'Green Cloud', which has attracted much attention among researchers. The allocation of the physical resources to tasks impacts the energy consumption and other related factors.

ENERGY CONSUMPTION

Energy consumption has been an important consideration in the IT industry for some time. It has been investigated in the context of Grid [17-18], Clusters [19-21], High Performance Computing (HPC) [22-23] or simple computer systems [24]. More recently, Cloud computing has led to the emergence of large data centres. Energy consumption contributes to 50% of the operating cost in data centres [25]. Consequently, existing research has been applied to the Cloud and new specialised approaches have been developed. Saving energy in large-scale data centres requires hardware and software optimisation. Hardware improvement is an active area of research. For instance, the peak energy consumption of servers is optimised by modifications of the hardware such as the introduction of a power controller proposed by Lefurgy et al [26]. However, the optimal use of the available hardware depends on the effectiveness of the algorithms that control it. To make data centres more energy-efficient there is a need for developing energy-efficient provisioning algorithms. Newly devised energy-efficient provisioning algorithms are tested either on a data centre or using simulation software. In the case of experiments on a data centre, the energy/power data is collected from the specific hardware. A simulation package applies a mathematical formula to approximate the energy/power consumption.

Energy Measurement

One of the factors in energy related studies is the way the energy/power is measured or calculated. Energy refers to the total flow of current in a period of time and is measured in Watts (W), whereas power refers to the flow of current in a unit of time and is measured in Watts per hour/second. Whether the explicit objective is energy or power, the goal of many studies is to minimise the cost of electricity necessary to execute tasks in the Cloud. In some studies, the actual consumption is measured by connecting a meter to the servers, in others a formula or a set of rules are developed to approximate the consumption. Power meters that are connected to motherboards also consume energy that is included in the measurements, but this consumption is generally regarded as negligible. Real hardware for experimental evaluation is not always available to researchers, who find it easier to use simulation packages instead. Some energy-aware simulation packages are GreenCloud [27-28], MDCSim [30], CloudSim [31-32] and GSSIM [33-34]. A 2013 study by Kaur compares the characteristics of MDCSim, CloudSim and GreenCloud [35]. Among the available packages, CloudSim has been the most widely used by scholars. The reason can lie in the fact that the CloudSim package is constantly being updated by Cloud Laboratory in University of Melbourne. New trends in research are being implemented /tested and different workload from physical hardware is included. A review of CloudSim and its various versions can be found in research by Goyal et al [36]. An energy-aware simulation package has to include formulae for calculating the energy consumption associated with a specific resource utilisation.

CPU utilisation is commonly assumed to be the main contributor to energy consumption, an assumption that has been fortified by experiments in the Green Computing Lab at Swinburne University of Technology, Melbourne, Australia [37]. CPU Energy/power consumption can be divided into constant and dynamic consumption. Constant consumption is hardware-dependent and measured when the system is idle. Dynamic energy/power consumption depends on the frequency of the processor while executing the workload. Dynamic power was defined as $P = C.f^3$ by Kim et al [38], where P is the power consumption, C is a coefficient and f is the frequency of the processor. To obtain total energy consumption, this value should be added to the constant power consumption. The experiments by Lien et al shaped a set of rules that were used [41] and modified [42] by Hsu et al.

Eq.1 [40] provides a set of rules that determines the energy consumption for VM number i, V_i, at the time t using α, the idle energy consumption and β = α.
$$E_t = \begin{cases} \alpha & idle \\ \beta+\alpha & 0\%<CPU\ util\ll 50\% \\ 2\beta+\alpha & 50\%<CPU\ util\ll 70\% \\ 3\beta+\alpha & 70\%<CPU\ util\ll 80\% \\ 4\beta+\alpha & 80\%<CPU\ util\ll 90\% \\ 5\beta+\alpha & 90\%<CPU\ util\ll 100\% \end{cases} \quad (1)$$

The quantities were adjusted later in Hsu et al[42] as follows:

$$E_t = \begin{cases} \alpha & idle \\ \beta+\alpha & 0\%<CPU\ util\ll 20\% \\ 3\beta+\alpha & 20\%<CPU\ util\ll 50\% \\ 5\beta+\alpha & 50\%<CPU\ util\ll 70\% \\ 8\beta+\alpha & 70\%<CPU\ util\ll 80\% \\ 11\beta+\alpha & 80\%<CPU\ util\ll 90\% \\ 12\beta+\alpha & 90\%<CPU\ util\ll 100\% \end{cases} \quad (2)$$

According to the above mentioned energy models, a specific frequency in the system is associated with different values in different models. Although, they might behave relatively the same, the differences in details can lead to different results when it comes to comparison.

Other Goals and Objectives

Energy/power savings have been a major consideration in the complex context of the Cloud that affect many aspects of the system. Many studies have investigated the Cloud from the point of view of these aspects which often present themselves as trade-offs between optimisation goals. These aspects include:

Temperature: High energy consumption leads to the production of more heat that again necessitates more cooling devices, which, in turn, add to the energy consumption.

Resource utilisation: Resource utilisation, CPU utilisation in particular, is the most significant contributor to the energy consumption of the system. Increasing the utilisation of the available resources helps keep the number of active servers at a minimum, which decreases the energy consumption. Completion or response time: Keeping a minimal number of servers active increases the probability an increase in completion/response time.

SLA/QoS: Compressing the workload on the minimum number of servers can also cause violation of QoS/SLA.

Deadline: Deadlines are an issue in the provisioning problem in real-time systems. Deadline violations can be incurred by strategies used for decreasing the energy consumption.

Performance: The definition of performance varies in the literature. It is sometimes reported as the number of finished tasks or the utilisation of a resource per unit of time.

Cost: Cost is usually discussed as a consequence of either energy consumption or resource utilisation or both.

Figure 1 depicts the relationships between these factors. Energy/power consumption is directly related to resource utilisation, Higher the resource utilisation, larger the power consumption and shorter the completion time. Resource utilisation influences the temperature of the resources and entails cooling systems, which leads to higher costs. While a high resource utilisation ideally indicates a well utilised system with high efficiency, it can increase the chances of hardware failure. Shorter completion times improve performance and can be a crucial factor when dealing with real-time systems and application deadlines. SLA/QoS may define requirements regarding completion time, deadline, performance or cost, or a combination of these.

Some studies attempt to achieve a balance between these factors. Torres et al [43] minimised the number of active nodes to decrease the resource wastage and energy consumption while aiming for a minimum performance degradation. Bobroff et al [44] decreased the resource utilisation by dynamically migrating VMs between hosts while satisfying the SLA. Moore et al [45] investigated the trade-off between the resource utilisation and the cost of cooling the system. Several studies investigated the impact of resource provisioning for real-time tasks in an environment sensitive to deadlines [38][46-48].

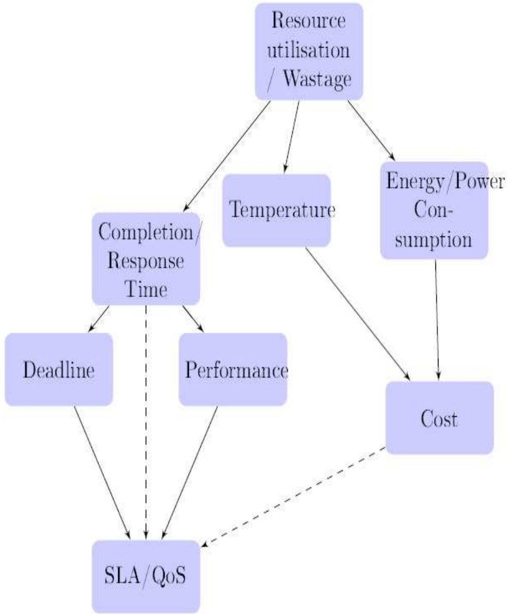

Fig. 1 Other goals and objectives related to energy/power consumption

ENERGY-AWARE PROVISIONING

Energy-aware provisioning requires the deployment of algorithms that lead to lower energy/power consumption. According to the time these strategies come into action it can be categorized as energy-aware scheduling or re-scheduling.

Energy Saving Strategies

The two main strategies applied for energy saving in Cloud-related studies are host switching and Dynamic Voltage and Frequency Scaling (DVFS). Switching off idle hosts reduces the energy consumption and the system responds to the requests with the available hosts. The strategy of switching hosts on and off has been studied by Mao et al [49] and extended in Mao and Humphrey[50], where the effect on deadlines and cost were investigated. The significance of the energy savings achieved by switching off hosts may be due to the fact that an idle host still consumes up to 70% of its peak power [51-52].

Switching hosts on again incurs a short interval of peak energy consumption and possible delays to the system. DVFS saves energy by reducing the frequency of hosts while keeping them active. A strict implementation of DVFS decreases the frequency of the processors to a level where the deadlines are barely met, which can lead to deadline violations in a sensitive system. Nonetheless, the server switching and frequency alteration have proven to be effective [53] and are commonly used.

Scheduling and Re-Scheduling

In the Cloud, the available resources are limited and shared among multiple applications using virtualisation. Each application/request is assigned to a Virtual Machine (VM) then mapped to the physical hardware. Virtual Machine Monitoring (VMM) software is responsible for hardware virtualisation and the access control of the VM. VMM software manages the Cloud system by assigning VMs to the physical hardware. This is known as VM placement or scheduling.

Scheduling algorithms distribute VMs on multiple hosts by allocating multiple VMs to a hardware/host (hardware sharing). If a host is saturated with VMs and cannot provide the required resources, it is overloaded and considered a hotspot. An overloaded host increases the chance of hardware failure[54] which may lead to longer completion times or deadline violations in real-time systems. A potential solution for resolving hotspots is VM (live) migration, where a VM is copied from a source server to the destination without stopping the execution apart from a short time for transferring the VM status [55] with the aim of balancing the load in the system. An unbalanced load leads to longer completion times as the overloaded host cannot provide resources for VMs when they are needed. A longer completion time keeps a host active for longer and increases the energy consumption.

An important aspect of a well-managed data centre is its ability to prevent or alleviate the hotspot problem. Prevention strategies attempt to address the issue at the time of scheduling where a detection strategy monitors the system to find the hotspots. A strategy in load balancing is to schedule applications on the hosts and then monitor the system to detect hotspots or an imbalanced load [22][45][56-64].

A load balancing method based on Honey Bee behaviour (HBB-LB) was devised by Babu and Krishna [56]. The honey bee foraging strategy divides the processors into three groups, overloaded, under loaded and balanced. It finds an overloaded server and removes VMs from it which are then assigned to a server of an under loaded group. The server is then added to the balanced group if it is no longer over- or under loaded. Labelling a server as under-/overloaded is considered a threshold based strategy. The scheduling algorithms used in the experiment include primitive FIFO and WRR (Weighted Round Robin).

In a study by Singh et al the goal is to balance the load between the different entities in a data centre. Entities include memory nodes and network switches as well as processing nodes. The memory units are virtualized as well as processors. A hotspot detection algorithm monitors all entities to provide evidence of an overloaded node to trigger the migration. This study is one of the instances of considering hotspots on switching nodes in network and memory units as well as processing units. However the virtualisation in memory level increases the dependency on network connections and potential delays due to extra load on network switches.

To react to a hotspot a multi-objective Bayesian game based genetic algorithm was proposed by Sallam and Li [58], which selects a VM to migrate from an overloaded host. The approach minimises the load volume on each host as well as the energy consumption in its multi-objective formulation. Minimizing the load on each node does not guarantee a higher saving on energy as all processors might be kept working for a total load that could be processed by a smaller number of hosts while switching others off.

In research by Wood et al [59] a hotspot is defined as a load which exceeds a predefined threshold value at the end of each interval. In their evaluation this threshold is 75%. The authors do not investigate the optimal threshold value. The definition of load volume varies in the literature. It is often defined as CPU utilisation. More general formulae have been described in studies by Wood et al [59] and Tian et al [60]. Eq.3 [59] considers CPU U_{cpu}, memory U_{memory}, and network utilisation $U_{network}$ to calculate the load volume LV.

$$LV = \frac{1}{(1-U_{cpu})(1-U_{memory})(1-U_{network})} \tag{3}$$

Also Eq.4 represents a formula proposed by Hongyuan et al [61] that is used in a study by Tian et al[60], where a reference server m is selected first. Each server i is compared to server m. $N1_i$ is the CPU capability, $N2_i$ is the memory capability and $N3_i$ represents the hard disk. C_i and M_i are the average utilisations of CPU and memory respectively, D_i represents the transfer rate of the hard disk, Net_i signifies network throughput. a, b, c and d are weighting factors for CPU, memory, hard disk and network bandwidth respectively. The strategy of this algorithm is to choose the host with the smallest value B among all physical servers to allocate VMs.

$$B = \frac{a\,N1_iC_i}{N1_mC_m} + \frac{b\,N2_iM_i}{N2_mM_m} + \frac{c\,N3_iD_i}{N3_mD_m} + \frac{d\,N1_iNet_i}{Net_m} \tag{4}$$

Detecting an imbalance solely based on CPU utilisation disregards resources such as memory and network, where their overuse can also incur delays and hardware failures. Ghanbari et al [65] proposed a feedback-based optimisation method for a private Cloud. While the impact of different resources in combination is not the focus of their study, a general formula as shown in Eq.3 and Eq.4 averages the load on all resources. This means that an excessive load on CPU might be overlooked due to negligible memory utilisation. Once the load has been defined, upper and/or lower thresholds are define to identify hosts with imbalanced loads. In a study by Singh et al [57] an imbalance score is defined for each entity in the system as Eq.5, where f is the load usage fraction and T is the corresponding threshold for the resource.

$$IBscore(f,T) = \begin{cases} 0 & f < T \\ e^{(f-T)/T} & otherwise \end{cases} \tag{5}$$

A predefined threshold is short sighted in dealing with an online problem. According to the study by Borodin and El-Yaniv [66] problems that do not have complete knowledge of future events are online problems. Cloud's incoming requests are not completely known in some situations that makes it an online problem. These situations are often disregarded [58][46-47]. A strict threshold based approach for imbalance detection decreases the system's ability in responding to its online and unpredicted requests. A slight difference in the incoming requests that changes the utilisation from below to slightly higher than threshold alters the systems reaction even if the difference was for a short time and wouldn't cause performance degradation. An impractical lower/upper threshold can lead to unnecessary host switching that causes delays and extra energy consumption due to the wake-up energy peak. A summary is provided in table1.

Table -1 Energy-aware scheduling and re-scheduling

Title	Energy Measurement		Energy Saving		Virtualisation	Migration
	Meters	Formulae	Switch	DVFS		
Power-aware scheduling for periodic real-time tasks[48]	×	✓	×	✓	×	×
Virtualpower: coordinated power management in virtualized enterprise systems[68]	✓	×	✓	✓	Xen	✓
No "power" struggles: Coordinated multi-level power management for the data center[69]	✓	×	✓	✓	VMware	✓
Autonomic multi-agent management of power and performance in data centers[70]	✓	×	✓	×	✓	×
Energy aware consolidation for cloud computing[71]	✓	×	✓	×	✓	×
pMapper power and migration cost aware application placement in virtualized systems[62]	✓	×	✓	✓	VMware	✓
Power and performance management of virtualized computing environments via lookahead control[52]	✓	×	✓	×	✓	✓
PADD: Power-Aware Domain Distribution[72]	✓	×	✓	×	Xen	✓
Energy-efficient Scheduling of HPC Applications in Cloud Computing Environments[73]	×	✓	×	✓	×	×
Power-aware Provisioning of Cloud Resources for Real-time Services[46]	×	✓	×	✓	Xen	×
Enacloud: an energy-saving application live placement approach for cloud computing environments[74]	✓	×	✓	×	Xen	✓
GreenCloud A New Architecture for Green Data Center[75]	✓	×	✓	×	Xen	✓
Energy-efficient management of data center resources for cloud computing: A vision, architectural elements, and open challenges[76]	×	✓	✓	✓	✓	✓
Online self-reconfiguration with performance guarantee for energy-efficient large-scale cloud computing data centers[77]	×	✓	✓	×	VMware	✓
Linear Combinations of DVFS-Enabled Processor Frequencies to Modify the Energy-Aware Scheduling Algorithms[78]	×	✓	×	✓	×	×
Mistral: Dynamically Managing Power, Performance, and Adaptation Cost in Cloud Infrastructures[79]	✓	×	✓	×	Xen	✓
Multi-objective Virtual Machine Placement in Vitualized Data Center Environments[80]	×	✓	×	×	✓	✓
An Energy-efficient Scheduling Approach Based on Private Clouds[81]	✓	×	✓	×	✓	×
An Energy-Efficient Scheme for Cloud Resource Provisioning Based on CloudSim[82]	×	✓	×	✓	✓	×
Energy-aware ant colony based workload placement in Clouds[83]	✓	✓	✓	×	✓	×
Energy-aware task consolidation technique for cloud computing[41]	×	✓	×	×	✓	✓
Energy-aware application-centric VM allocation for HPC workloads[23]	✓	×	×	×	Xen	×
Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing[63]	×	✓	✓	✓	✓	✓
Energy efficient utilisation of resources in cloud computing systems[84]	×	✓	×	✓	✓	✓
Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in Cloud data centers[64]	✓	×	✓	✓	✓	✓
GreenCloud: a packet-level simulator of energy-aware cloud computing data centers[27]	×	✓	✓	✓	×	×
A multi-objective ant colony system algorithm for virtual machine placement in cloud computing[85]	✓	×	✓	×	✓	×
A green energy-efficient scheduling algorithm using the DVFS technique for cloud data centers[86]	×	✓	×	✓	✓	×
A Multi-objective Virtual Machine Migration Policy in Cloud Systems[58]	×	✓	×	×	Xen	✓
Experimental Analysis of Task-based Energy Consumption in Cloud Computing Systems[37]	✓	×	×	×	VMware	×
Optimizing Energy Consumption with Task Consolidation in Clouds[42]	×	✓	×	×	✓	✓
Communication and migration energy aware task mapping for reliable multiprocessor systems[87]	✓	✓	×	×	×	✓

Hypervisor

Unlike in systems with a dedicated operating system, in the Cloud hardware-related instructions have to be handled by a VMM, which mediates between a guest operating system and the hardware that runs it. Instead of full virtualisation, para-virtualisation is sometimes used with a hypervisor responsible for the shared access to the hardware while the operating systems hosted contain virtualisation-aware code. This approach obviates the need for any recompilation or instruction trapping because the operating systems themselves cooperate in the virtualisation process. A typical para-virtualisation product is Xen [12], while VMware [11] provides full virtualisation with a VMM. The maximum number of VMs allowed per host can affect the results of provisioning approaches. Table2 shows the maximum number of VMs when working with a specific hypervisor. Where more VMs can be deployed on a host, shorter queuing delays are expected. A smaller number of VMs per host indicates potential improvement in the performance. VMware and Xen are used in a large number of studies as these hypervisors support live VM migration.

Table -2 Maximum numbers of VMs in common hypervisors

Hypervisor	Virtual machines per host
VMware vSphere ESXi 5.0	512
XenServer 6.0	75
Hyper-V 2008 R2	384
Oracle VirtualBox V4	128

RELATED WORK

In a study by Rimal et al [88] the cloud services provided by big companies including Amazon, force.com and google are covered. It is an exclusive review of the available service providers and the way it is being carried out and the features that are supported. It is a valuable research in giving insights about the way Cloud services are offered to the public. Cloud monitoring tools in a study by Fatema et al [89] intend to monitor the resource utilisation and system performance. Available monitoring tools are described and their characteristics and desirable features are discussed. A review of available energy-aware simulation software and test beds are gathered by Sakellari and Loukas [90] and resource management in the Cloud is described and analysed in [91] where the challenges related to different perspectives of resource management is discussed. One of the strength of the study is in its identified challenges in each section and comparative nature of the text. In 2011 Beloglazov and Buyya [92] surveyed energy-aware studies in the Cloud in a book chapter. It includes a review of a set of energy related studies. Although, our survey includes other objectives related to the energy consumption and is shorter than earlier mentioned survey. The load balancing and re-scheduling is more broadly covered in our study and challenges are identified.

IDENTIFIED CHALLENGES

There are multiple instances in the literature where resource provisioning refers to how the processing power is shared among VMs. The effect of memory and bandwidth limitation and how they affect are disregarded, arguing that they can be added to the system at any time. It indicates that these studies reported their results according to that assumption. A more comprehensive way of simulation, when considering other resources as well as processing units can lead to a more compatible result for simulations to the measured values.

A study might assume that one application will be deployed on a VM or there is the option of having multiple applications on one VM. It depends on how the system setup is and whether it's related to the correlation between jobs. Correlated jobs in the system require certain flow of executions where start of a job will be on termination of another. This can be disregarded by arguing that the group of these jobs can be presumed to be a single unrelated job to the others. Nevertheless, it effects the response/execution time and SLA (depending on the definition) whether the related jobs are deployed on different machines or waiting for the end of the other job.

There is a gap in the energy-aware studies that can measure different objectives when these attributes are altered. That is, to consider CPU, memory and bandwidth requirements of applications while changing the settings where jobs can run independently or interconnected. Interconnection between jobs can be of different type and complexity. VMs run one or multiple jobs will also impact the system that can be studied in these scenarios. They can include variety of application types, CPU-intensive, memory-intensive, bandwidth-intensive or non-intensive.
Considering a combination of these assumptions, scheduling in the Cloud is an online problem as there is not comprehensive information available about job's resource requirements. Their resource requirements can vary in their life cycle. Relying on the approximations and prior knowledge simplifies the Cloud provisioning but will come short in some instances of real practice. Online scheduling ideas are needed to be developed and tested on a variety of test beds. Test beds and evaluation workloads are not standard and any application set can potentially be used which makes comparison difficult.

The way scheduling happens might make some nodes in the system overloaded or under loaded. The reason a node in the system is detected as under-/overloaded is its utilisation which violates a predetermined threshold. The optimal threshold value is not investigated. Nonetheless, a diverse set of jobs are expected to be used for testing as different strategies work differently according to the test bed used. Then, the policies to choose a VM for migration and where to take it to assure a balanced load or save energy. There is a need for an energy model that can be adapted by simulation packages to provide a comparison platform. Simulation packages such as CloudSim intend to provide it. However, the updates released consist of formulae and real system values that are sometimes incompatible.

Also, energy-aware heuristic multi-objective algorithms are commonly used in the Cloud. However, the result of multi-objective algorithms is a Pareto set. How a member of this set is selected is commonly considered to be random or averaged over all objectives. Averaged values can potentially lead to the selection of a member that is significantly high in one or two objectives (usually related objectives such as energy and temperature) and considerably unsatisfactory on others. This area can be better covered when joined with studies related to heuristic algorithms.

Given all the assumptions and the needs for further development, it is important to note that even the best strategy in a given test bed may come short in another and the other way around. Each developed algorithm, strategy and policy is expected to be optimal in a certain circumstances. A study that looks into switching between these alternatives seems to achieve the most in the long term run of system. An adaptive provisioning is more likely to handle a complex system than static ones.

CONCLUSION

The Cloud resource provisioning is an area yet to be fully investigated. It plays an important role in providing reliable service in a satisfactory level regarding energy consumption. The relation between energy consumption and other sometimes conflicting objectives in the Cloud make the research in this area of high importance and complexity.

In this paper, details of the studies related to energy aware Cloud provisioning are reviewed to provide insights. The aim was to describe multiple aspects and assumption of the studies to identify the differences. Details include the way energy consumption is obtained and the differences it makes, the energy saving strategies and the effect of hypervisor on the reported results. Also, other goals that might be simultaneously considered at the time of experiment and their impact on and from energy consumption are discussed. We noted that there are details in the experiments that can potentially change the results of the algorithms. This includes but is not limited to the way energy consumption is collected. Formulae and rule-based approximations give a rough estimate of the real energy consumption but makes simulation possible. On the other hand power meters attached to the hardware might read different values according to where they are connected and how often the meter is being read. Power meters circuits also add to the energy consumption that is being measured. Therefore, the more often the power is read the more the power meter energy will be added and less reading gives less accuracy to the power reading as there are less reading points.

Research in the energy-aware Cloud provisioning consists of studies related to scheduling and re-scheduling of the incoming requests. It includes the initial assignment to the VMs as well as the mapping to the physical machines. There are differences in a way it is carried out where the size of VMs varies or the system does/doesn't allow assigning more than one request to a VM. There might be interconnections between requests that adds to the complexity of the experiment. VMs running interconnected tasks might be deployed on a same physical machine or not.

The other challenging issue is the system with overloaded nodes. It is of high importance to detect overloaded nodes and solve it as it increases the chance of hardware failure and longer task completion time due to the wasted time for overheads such as context switching. Nevertheless, the notion of prescribing a threshold for overloaded node detection might cause frequent and unnecessary switching. And, the threshold is solely based on the CPU utilisation whereas memory and a network link might be overloaded and cause the same problem of lateness in the system. A general formula that averages the load on CPU, memory and network also comes short in detecting a situation where a resource is terribly overloaded and causes the rest of the resource not to have anything to do. Either way an overloaded part of the system is detected, policy/ies are needed to solve the problem.

Solving the overloaded node problem is often through migrating VMs from overloaded machine to another. This process is called re-scheduling. It might be with the aim of balancing the load in the Cloud. It's important to decide which VM should be migrated and where it should be taken to as it impacts the effect of the migration. VM selection policies and host selection policies are investigated solely or in conjunction. Noting that the type of hypervisor determines the maximum number of VMs on a physical machine, scheduling and re-scheduling algorithms can vary in experiments and change the outcome even though it is neglected by some authors. The number of VMs on a physical machine affects the real portion of the resources each VM receives which then effects the resource utilisation, response time and completion time and depending on the definition, the SLA/QoS. Resource utilisation directly impacts the energy consumption for computation. It is also responsible for the rise of the temperature which in turn requires more energy to cool the system down and adds to the total energy consumption. An energy-aware provisioning strategy aims to minimize the energy but needs to behave in a satisfactory level in case of other criteria in the system. That is, the energy-aware provisioning problem should be formulated in a way that covers the presumed objectives by researchers.

Strategies applied to formulate the Cloud provisioning problem are mainly defined as a trade-off between minimizing energy consumption and other goals as either a simple trade-off or a multi-objective optimisation problem. It might also be described as a bin packing problem where available resources are the bin that accommodates applications/application's VMs. Regarding the problem formulation a solution strategy is developed. Solutions are categorized as deterministic and heuristic.

After building the assumptions of the experiment, defining the problem and finding a solution, it is of high importance to evaluate it. Evaluation process can be a comparison to the best or worst case energy consumption in the given system or the algorithm suggested in another study or even a random selection policy. We aimed to provide information regarding the comparisons between studies and to help researchers to find the suitable benchmark for their evaluation. The review and analysis in this paper have provided an overview of available provisioning strategies and their characteristics to support researchers in finding the research gap. It also helps to detect the possible deviations from the reported results. Identified challenges are to suggest new trends and paths that requires further investigation. We consider this review as a reference and a basis for further research work.

REFERENCES

[1] P Johnson and T Marker, Data Centre Energy Efficiency Product Profile, Pitt Sherry, *Rep to Equip energy Effic Comm Aust Gov Dep Environ Water*, Herit Arts, **2009.**
[2] J Koomey, *Growth in Data Center Electricity use 2005 to 2010*, Oakland, CA Anal Press **2011**vol 1, p 2010**.**
[3] R Brown et al, *Report to Congress on Server and Data Center Energy Efficiency: Public law* 109-431, Lawrence Berkeley Natl Lab, **2008.**
[4] J G Koomey, *Estimating Total Power Consumption by Servers in the US and the World*, Report **2007.**
[5] P Padala, K G Shin, X Zhu, M Uysal, Z Wang, S Singhal, A Merchant and K Salem, Adaptive Control of Virtualized Resources in Utility Computing Environments, *ACM SIGOPS Operating Systems Review*, **2007,** vol 41, no 3, p. 289-302**.**
[6] C Pettey, Gartner Estimates ICT Industry Accounts for 2 Percent of Global CO_2 Emissions, *Gartner Press Release*, **2007.**
[7] J M Kaplan, W Forrest, and N Kindler, *Revolutionizing Data Center Energy Efficiency*, **2008.**
[8] S N T Chiueh and S Brook, A survey on Virtualization Technologies, *RPE Rep*, **2005**, p. 1-42.
[9] J Carolan, S Gaede, J Baty, G Brunette, A Licht, J Remmell, L Tucker, and J Weise, Introduction to Cloud computing Architecture, *SUN Microsystems Inc*,**2009**, p.1-40.
[10] R P Goldberg, Survey of Virtual Machine Research, *Computer (Long Beach Calif)*, **1974**, vol 7, p. 34-45.
[11] Vmw Inc, - Jul-**2013.**
[12] 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, **2003**, vol 37, p. 164-177.
[13] A Whitaker, M Shaw, and S D Gribble, *Denali: Lightweight Virtual Machines for Distributed and Networked Applications*, **2002.**
[14] A Kivity, Y Kamay, D Laor, U Lublin, and A Liguori, {KVM}: the Linux Virtual Machine Monitor, *Proceedings of the Linux Symposium*, **2007**, vol 1, p. 225-230.
[15] R Uhlig, G Neiger, D Rodgers, A L Santoni, F C M Martins, A V Anderson, S M Bennett, A Kagi, F H Leung and L Smith, Intel Virtualization Technology, *Computer (Long Beach Calif)*, **2005**, vol 38, p. 48-56.
[16] W Voorsluys, J Broberg and R Buyya, Introduction to Cloud Computing, *Cloud Comput Princ Paradig*,**2011**, p. 2-44.
[17] S U Khan and I Ahmad, A Cooperative Game Theoretical Technique for Joint Optimization of Energy Consumption and Response Time in Computational Grids, *IEEE Trans on Parallel Distrib Syst*, **2009**, vol 20, p. 346-360.
[18] R Subrata, A Y Zomaya and B Landfeldt, Cooperative Power-Aware Scheduling in Grid Computing Environments, *J Parallel Distrib Comput*,**2010**, vol 70, no 2, p. 84-91.

[19] P Ranganathan, P Leech, D Irwin and J Chase, Ensemble-level Power Management for Dense Blade Servers, *ACM SIGARCH Computer Architecture News*, **2006**, vol 34, no 2, p. 66-77.
[20] J O Kephart, H Chan, R Das, D W Levine, G Tesauro, F L Rawson III and C Lefurgy, Coordinating Multiple Autonomic Managers to Achieve Specified Power-Performance Tradeoffs, ICAC,**2007**, vol 7, p 24.
[21] K H Kim, R Buyya and J Kim, Power Aware Scheduling of Bag-of-Tasks Ap.lications with Deadline Constraints on DVS-enabled Clusters, *CCGRID*,**2007**, vol 7, p. 541-548.
[22] Q Tang, S K S Gupta and G Varsamopoulos, Energy-Efficient Thermal-Aware Task Scheduling For Homogeneous High-Performance Computing Data Centers: A Cyber-Physical Approach, IEEE Trans on Parallel Distrib Syst, **2008**, vol 19, no 11, p. 1458-1472.
[23] H Viswanathan, E K Lee, I Rodero, D Pompili, M Parashar and M Gamell, Energy-Aware Application-Centric {VM} Allocation For {HPC} Workloads, IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum, **2011**, p. 890-897.
[24] Y C Lee and A Y Zomaya, Minimizing Energy Consumption for Precedence-Constrained Applications using Dynamic Voltage Scaling, 9th IEEE/ACM International Symposium on Cluster Computing and the Grid,**2009**, p. 92-99.
[25] D Filani, J He, S Gao, M Rajappa, A Kumar, P Shah and R Nagappan, Dynamic Data Center Power Management: Trends, Issues, and Solutions, *Intel Technol J*, **2008**, vol 12, no 1.
[26] C Lefurgy, X Wang, and M Ware, Server-Level Power Control, Fourth International Conference on Autonomic Computing, **2007** ICAC'07, p 4.
[27] D Kliazovich, P Bouvry and S U Khan, {GreenCloud}: a Packet-Level Simulator of Energy-Aware Cloud Computing Data Centers, *J Supercomput*, **2012**, vol 62, no 3, p. 1263-1283.
[28] D Kliazovich, P Bouvry and S U Khan, Simulating Communication Processes in Energy-Efficient Cloud Computing Systems, 1st IEEE International Conference on Cloud Networking, **2012**, p. 215-217.
[29] T Issariyakul and E Hossain, An Introduction To Network Simulator {NS2} *Springer*, **2012.**
[30] S H Lim, B Sharma, G Nam, E K Kim, and C R Das, {MDCSim}: A Multi-Tier Data Center Simulation, Platform, *IEEE International Conference on Cluster Computing and Workshops*, *CLUSTER'09*, **2009**, p. 1-9.
[31] R N Calheiros, R Ranjan, C A F De Rose and R Buyya, {CloudSim}: A Novel Framework for Modeling and Simulation of Cloud Computing Infrastructures and Services, *arXiv Prepr arXiv09032525*, **2009.**
[32] R N Calheiros, R Ranjan, A Beloglazov, C A F De Rose and R Buyya, {CloudSim}: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms, *Softw Pract Exp*, **2011**, vol 41, no 1, p. 23-50.
[33] K Kurowski, J Nabrzyski, A Oleksiak and J Weglarz, Grid Scheduling Simulations with {GSSIM}, *International Conference on Parallel and Distributed Systems*, **2007**, vol 2, p. 1-8.
[34] S Bkak, M Krystek, K Kurowski, A Oleksiak, W Pikatek and J Wkaglarz, {GSSIM}--A Tool For Distributed Computing Experiments, *Sci Program*, **2011,** vol 19, no 4, p. 231-251.
[35] G Kaur, Study of Comparison of Various Cloud Computing Simulators, *2nd National Conference in Intelligent Computing & Communication*, **2013.**
[36] T Goyal, A Singh and A Agrawal, {CloudSim}: Simulator for Cloud Computing Infrastructure and Modeling, *Procedia Eng*, **2012**, vol 38, p. 3566-3572.
[37] F Chen, J Grundy, Y Yang, J G Schneider and Q He, Experimental Analysis of Task-Based Energy Consumption in Cloud Computing Systems, *Proceedings of the 4th ACM/SPEC International Conference on Performance Engineering*, **2013,** p. 295-306.
[38] K H Kim, A Beloglazov and R Buyya, Power-Aware Provisioning of Virtual Machines for Real-Time Cloud Services, *Concurr Comput Pract Exp*, **2011,** vol 23, no 13, p. 1491-1505.
[39] A Gandhi, M Harchol-Balter, R Das and C Lefurgy, Optimal Power Allocation in Server Farms, A*CM SIGMETRICS Performance Evaluation Review*, **2009,** vol 37, no 1, p. 157-168.
[40] C H Lien, M F Liu, Y W Bai, C H Lin and M B Lin, Measurement by the Software Design for the Power Consumption of Streaming Media Servers, *Proceedings of the IEEE Instrumentation and Measurement Technology Conference, IMTC* **2006**, p. 1597-1602.
[41] C H Hsu, S C Chen, C C Lee, H Y Chang, K C Lai, K C Li and C Rong, Energy-aware Task Consolidation Technique For Cloud Computing, IEEE Third International Conference Cloud Computing Technology and Science (CloudCom), **2011**, p. 115-121.
[42] C H Hsu, K Slagter, S C Chen and Y C Chung, Optimizing Energy Consumption with Task Consolidation in Clouds, *Inf Sci (Ny)*, **2014,** vol 258, p. 452-462.
[43] J Torres, D Carrera, K Hogan, R Gavaldà, V Beltran and N Poggi, Reducing Wasted Resources to Help Achieve Green Data Centers, *IEEE International Symposium on Parallel and Distributed Processing*, **2008**.
[44] N Bobroff, A Kochut and K Beaty, Dynamic Placement of Virtual Machines for Managing Sla Violations, *10th IFIP/IEEE International Symposium on Integrated Network Management*, **2007,** IM'07, p. 119-128.
[45] J D Moore, J S Chase, P Ranganathan and R K Sharma, Making Scheduling 'Cool': Temperature-Aware Workload Placement in Data Centers, *USENIX Annual Technical Conference, General Track*, **2005,** p. 61-75.

[46] K H Kim, A Beloglazov and R Buyya, Power-Aware Provisioning of Cloud Resources for Real-Time Services, *Proceedings of the 7th International Workshop on Middleware for Grids, Clouds and e-Science*, **2009.**

[47] D Zhu, D Mosse and R Melhem, Power-Aware Scheduling for {AND/OR} Graphs in Real-Time Systems, *IEEE Trans on Parallel Distrib Syst*, **2004**, vol 15, no 9, p. 849-864.

[48] H Aydin, R Melhem, D Mossé, and P Mejía-Alvarez, Power-Aware Scheduling For Periodic Real-Time Tasks, *IEEE Trans Comput*, **2004,** vol 53, no 5, p. 584-600.

[49] M Mao, J Li and M Humphrey, Cloud Auto-Scaling with Deadline and Budget Constraints, 11th *IEEE/ACM International Conference on Grid Computing (GRID)*, **2010**, p. 41-48.

[50] M Mao and M Humphrey, Auto-scaling to Minimize Cost and Meet Application Deadlines in Cloud Workflows, I*nt. Conf. for High Performance Computing, Networking, Storage and Analysis*, **2011**, p 49.

[51] X Fan, W D Weber and L A Barroso, Power Provisioning for a Warehouse-Sized Computer, *ACM SIGARCH Computer Architecture News*, **2007,**vol 35, no 2, p. 13-23.

[52] 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 Comput*, **2009**, vol 12, no 1, p. 1-15.

[53] E N M Elnozahy, M Kistler and R Rajamony, Energy-Efficient Server Clusters, Power-Aware Computer Systems, *Springer*, **2003**, p. 179-197.

[54] W Feng, X Feng and R Ge, Green Supercomputing Comes of Age, *IT Prof*, 2008, vol 10, no 1, p. 17-23.

[55] C Clark, K Fraser, S Hand, J G Hansen, E Jul, C Limpach, I Pratt and A Warfield, Live Migration of Virtual Machines, *Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation*, **2005**, vol 2, p. 273-286.

[56] L D Babu and P V Krishna, Honey bee behavior inspired load balancing of tasks in cloud computing environments, *Appl Soft Comput,* **2013**

[57] A Singh, M Korupolu and D Mohapatra, Server-Storage Virtualization: Integration and Load Balancing in Data Centers, *ACM/IEEE Conference on Supercomputing*, **2008**, p 53.

[58] A Sallam and K Li, A Multi-objective Virtual Machine Migration Policy in Cloud Systems, Comput J, **2013**

[59] T Wood, P J Shenoy, A Venkataramani and M S Yousif, Black-box and Gray-box Strategies for Virtual Machine Migration, *NSDI*, **2007,** vol 7, p. 229-242.

[60] W Tian, Y Zhao, Y Zhong, M Xu and C Jing, A dynamic and integrated load-balancing scheduling algorithm for Cloud datacenters, *IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*, **2011,**p. 311-315.

[61] Z Hongyuan, Z Liang and W Jiaqi, Design and Implementation of Load Balancing in Web Server Cluster System, *J Nanjing Univ Aeronaut Astronaut*, **2006,**vol 3, p 15.

[62] A Verma, P Ahuja and A Neogi, pMapper: Power and Migration Cost Aware Application Placement in Virtualized Systems, *Springer*, **2008,** p. 243-264.

[63] A Beloglazov, J Abawajy and R Buyya, Energy-Aware Resource Allocation Heuristics for Efficient Management of Data Centers for Cloud Computing, *Futur Gener Comput Syst*, **2012,** vol 28, no 5, p. 755-768.

[64] 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, C*oncurr Comput Pract Exp*, **2012,** vol 24, no 13, p. 1397-1420.

[65] H Ghanbari, B Simmons, M Litoiu and G Iszlai, *Feedback-based Optimization of a Private Cloud*, *Futur Gener Comput Syst*, **2012,** vol 28, no 1, p. 104-111.

[66] A Borodin and R El-Yaniv, Online Computation and Competitive Analysis, *Cambridge University Press Cambridge*, vol 53,**1998.**

[67] X Wang, Y Wang and Y Cui, A New Multi-Objective Bi-Level Programming Model for Energy and Locality Aware Multi-Job Scheduling in Cloud Computing, *Futur Gener Comput Syst*, **2014,** vol 36, p. 91-101,

[68] R Nathuji and K Schwan, Virtualpower: Coordinated Power Management in Virtualized Enterprise Systems, *ACM SIGOPS Operating Systems Review*, **2007,** vol 41, no 6, p. 265-278.

[69] R Raghavendra, P Ranganathan, V Talwar, Z Wang and X Zhu, No Power Struggles: Coordinated Multi-Level Power Management for the Data Center, *ACM SIGARCH Computer Architecture News*, **2008,** vol 36, no 1, p. 48-59.

[70] R Das, J O Kephart, C Lefurgy, G Tesauro, D W Levine and H Chan, Autonomic Multi-Agent Management of Power and Performance in Data Centers, *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems: Industrial Track*, **2008,** p. 107-114.

[71] S Srikantaiah, A Kansal and F Zhao, Energy Aware Consolidation for Cloud Computing, *Proceedings of Power Aware Computing and Systems Conference*, **2008,** vol 10.

[72] M Y Lim, F Rawson, T Bletsch and V W Freeh, {PADD}: Power Aware Domain Distribution, 29th IEEE *International Conference on Distributed Computing Systems*, **2009,** p. 239-247.

[73] S K Garg, C S Yeo, A Anandasivam and R Buyya, Energy-Efficient Scheduling of HPC Applications in Cloud Computing Environments, *arXiv Prepr arXiv09091146*, **2009.**

[74] B Li, J Li, J Huai, T Wo, Q Li and L Zhong, Enacloud: An Energy-Saving Application Live Placement Approach for Cloud Computing Environments, *IEEE International Conference on Cloud Computing,* **2009**, p. 17-24.

[75] L Liu, H Wang, X Liu, X Jin, W B He, Q B Wang and Y Chen, GreenCloud: A New Architecture for Green Data Center, *6th International Conference on Autonomic Computing and Communications* , **2009,** p. 29-38.

[76] R Buyya, A Beloglazov and J Abawajy, Energy-Efficient Management of Data Center Resources for Cloud Computing: A vision, Architectural Elements, and Open Challenges, *arXiv Prepr arXiv10060308*, **2010.**

[77] H Mi, H Wang, G Yin, Y Zhou, D Shi and L Yuan, Online Self-Reconfiguration with Performance Guarantee for Energy-Efficient Large-Scale Cloud Computing Data Centers, *IEEE International Conference on Services Computing (SCC)*, **2010,** p. 514-521.

[78] N B Rizvandi, J Taheri, A Y Zomaya and Y C Lee, Linear combinations of DVFS-enabled Processor Frequencies to Modify the Energy-Aware Scheduling Algorithms, *10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (CCGrid)*, **2010,** p. 388-397.

[79] G Jung, M A Hiltunen, K R Joshi, R D Schlichting and C Pu, Mistral: Dynamically Managing Power, Performance and Adaptation Cost in Cloud Infrastructures, *30th IEEE International Conference on Distributed Computing Systems (ICDCS)*, **2010,** p. 62-73.

[80] J Xu and J A B Fortes, Multi-Objective Virtual Machine Placement in Virtualized Data Center Environments, *IEEE/ACM Int'l Conference on & Int'l Conference on Cyber, Physical and Social Computing (CPSCom)in Green Computing and Communications (GreenCom)*, **2010,** p. 179-188.

[81] J Li, J Peng, Z Lei and W Zhang, An Energy-Efficient Scheduling Approach Based on Private Clouds, J *Inf Comput Sci*, **2011,** vol 8, no 4, p. 716-724.

[82] Y Shi, X Jiang and K Ye, An Energy-Efficient Scheme for Cloud Resource Provisioning Based on CloudSim, I*EEE International Conference on Cluster Computing (CLUSTER)*, **2011,** p. 595-599.

[83] E Feller, L Rilling and C Morin, Energy-aware Ant Colony Based Workload Placement in Clouds, *12th IEEE/ACM International Conference on Grid Computing*, **2011,** p. 26-33.

[84] Y C Lee and A Y Zomaya, Energy Efficient Utilization of Resources in Cloud Computing Systems, *J Supercomput*, **2012,** vol 60, no 2, p. 268-280.

[85] Y Gao, H Guan, Z Qi, Y Hou and L Liu, A Multi-Objective Ant Colony System Algorithm for Virtual Machine Placement in Cloud Computing, *J Comput Syst Sci*, **2013.**

[86] C-M Wu, R S Chang and H Y Chan, A Green Energy-Efficient Scheduling Algorithm using the Dvfs Technique for Cloud Datacenters, *Futur Gener Comput Syst*, **2014,** vol 37, p. 141-147.

[87] A Das, A Kumar, and B Veeravalli, Communication and Migration Energy Aware Task Mapping for Reliable Multiprocessor Systems, *Futur Gener Comput Syst,* **2014,** vol 30, p. 216-228.

[88] B P Rimal, E Choi and I Lumb, A Taxonomy and Survey of Cloud Computing Systems, Fifth Joint International Conference on INC, IMS and IDC, **2009,** p. 44-51.

[89] K Fatema, V C Emeakaroha, P D Healy, J P Morrison and T Lynn, A Survey of Cloud Monitoring Tools: Taxonomy, Capabilities and Objectives, *J Parallel Distrib Comput,* **2014,** vol 74, no 10, p. 2918-2933.

[90] G Sakellari and G Loukas, A Survey of Mathematical Models, Simulation Approaches and Testbeds used for Research in Cloud Computing, *Simul Model Pract Theory*, **2013,** vol 39, p. 92-103.

[91] S S Manvi and G K Shyam, Resource Management for Infrastructure as a Service (IaaS) in Cloud Computing: A survey, *J Netw Comput Application*, **2014,**vol 41, p. 424-440.

[92] A Beloglazovet al, A Taxonomy and Survey of Energy-Efficient Data Centers and Cloud Computing Systems, *Adv Comput,* **2011,** vol 82, no 2, p. 47-111.