The Journal of Supercomputing
https://doi.org/10.1007/s11227-020-03165-6

An approach toward design and development of an energy-aware VM selection policy with improved SLA violation in the domain of green cloud computing

Riman Mandal[1] · Manash Kumar Mondal[1] · Sourav Banerjee[2] · Utpal Biswas[1]

Abstract
With the rapid demand for service-oriented computing in association with the growth of cloud computing technologies, large-scale virtualized data centers have been established throughout the globe. These huge data centers consume power at a large scale that results in a high operational cost. The massive carbon footprint from the energy generators is another great issue to deal global warming. It is essential to lower the rate of carbon emission and energy consumption as much as possible. The live-migration-enabled dynamic virtual machine consolidation results in high energy saving. But it also incurs the violation of service level agreement (SLA). Excessive migration may lead to performance degradation and SLA violation. The process of VM selection for migration plays a vital role in the domain of energy-aware cloud computing. Using VM selection policies, VMs are selected for migration. A new power-aware VM selection policy has been proposed in this research that helps in VM selection for migration. The proposed power-aware VM selection policy has been further evaluated using trace-based simulation environment.

Keywords Green cloud computing · Energy efficiency · Virtualization · Live migration · VM consolidation · VM selection policy

✉ Sourav Banerjee
mr.sourav.banerjee@ieee.org; souravacademia@gmail.com

Riman Mandal
mandal.riman@gmail.com

Manash Kumar Mondal
manashkumarmondal@gmail.com

Utpal Biswas
utpal0172@gmail.com

[1] Department of Computer Science and Engineering, University of Kalyani, Kalyani, Nadia, India

[2] Department of Computer Science and Engineering, Kalyani Government Engineering College, Kalyani, Nadia, India

Published online: 23 January 2020

1 Introduction

Nowadays, service-oriented architectures [1] are new trends. This is the heart of cloud computing [2, 3]. These services can be like computing as a service, platform as a service, storage as a service, software as a service, infrastructure as a service, and so on. The demand for these services is increasing day by day, as the customers, known as cloud users, have to spend currency for the time and unit they have availed the services [4]. Another reason behind the popularity of this type of architecture is that the cloud users do not need to establish their setup, so the setup cost can be minimized for the cloud users.

The increasing demand for these services is served by virtualized heterogeneous cloud data centers dispersed over the globe. To provide continuous power supply, these large data centers require an enormous amount of uninterrupted electricity supply. This requirement is fulfilled by the power grids provided by the local electricity company. To ensure uninterrupted power supply in the case of large-scale electrical failure, data centers must have one or two backup power supply units powered by gas, diesel or solar. Generating these colossal amounts of electricity elevates the aggregation of carbon footprint in the environment. It also raises the cost of computation for the data center authorities as with increasing load on the data centers, and the energy requirements are also increasing. In the year 2017, roughly 416 terawatts of electricity, that contributed around 3% of the planet's overall energy generation, had been consumed by the data centers all over the world [5]. Surprisingly, this amount is 40% more than the annual energy consumption of the UK [5].

The number of cloud users increases exponentially. Eventually, the service requests are increasing at a large scale. To provide stabilized user satisfaction, cloud service providers are building new data centers. The electricity demands required to provide power these data centers are also increasing. To fulfill these power requirements, fossil fuels are used, which eventually increase the carbon footprints in the environment. Data centers consist of thousands of computing resources, network equipment, and large cooling units. Hardware is not the only reason for the data center's energy consumption, but also the inefficient use of computing resources adds up to the total power consumption.

Applying virtualization [6] to the physical servers, virtual machines (VMs) are created to serve cloud users' requests. By efficient resource management and VM consolidation with live migration [7–9], the energy requirements of data centers can be minimized. VM consolidation [10–13] is a technique to minimize the number of active physical servers by migrating VMs from underutilized servers to other servers and then shutting underutilized servers off. Migration is also required when a physical server is experiencing overutilization. A physical server is overutilized when the host is overly crowded with VMs and has no free resource. VMs may require additional resources over time. Additional resource overloaded physical servers cannot serve additional resource requests. So the VM has to work within that limited resource and that will contradict with SLA agreement with cloud users. Due to this reason, some VMs need to be migrated into moderately loaded physical servers. These migrations are called live migration as the VMs are kept alive while they are in the migration process so that the cloud users can have uninterrupted service. Too much of migration can hamper service level agreement (SLA) [14] between service providers and cloud users and aggregate energy consumption and operating cost [9].

The VM selection process is an integrated part of the consolidation and VM migration. This is a process of creating a VM pool by selecting VMs from overloaded physical servers or hosts. These VMs will be migrated to other hosts. The selection of VMs for migration is one of the important tasks in VM consolidation. By choosing appropriate VMs from overload physical servers or hosts, the number of migrations can be controlled. This will result in reduced operational cost and energy consumption without minimal violating of SLA. In this article, we have proposed a power-aware VM selection policy. This research work considers the framework proposed by Beloglazov et al. [9]. This proposed policy selects VMs from overloaded hosts to migrate them to other hosts while creating a balance between energy consumption and SLAV. The other goal of the policy is to minimize number of VM migrations. The proposed policy has been implemented in CloudSim 4.0 Simulator [15] and analyzed with some existing VM selection policies [9, 16–18]. The proposed policy has shown significant improvement over these existing policies in terms of energy consumption, SLAV and other performance metrics.

The remaining portion of the paper is organized as follows. A brief literature survey is recorded in Sect. 2. Later, in Sect. 3 the proposed power-aware VM selection policy is described with proper pictorial representation of the work. Experiment results and comparison analysis are shown in Sect. 4.2. Lastly, the conclusion and future directions are discussed in Sect. 5.

2 Literature review

Implementing a cloud computing model with 100% energy efficiency is impossible. Excessive energy efficiency may lead to the violation of SLA established between service providers and cloud users. There should have been a trade-off between energy consumption and SLAV [9]. In 2009, Liu et al. presented GreenCloud a new cloud computing architecture [19]. The objective of the design was to minimize the energy consumption in data centers. Development of a green cloud computing involves creating green data centers, where carbon footprints will be minimized and energy efficiency will be maximized [20]. But the concept is not limited only to the IT equipment of the data center. All other environmental elements like cooling units, buildings, other hardware units should be energy efficient.

Since 2009, researchers have invested more and more time for reducing carbon footprint of data centers by proving energy-efficient solutions to the problem while maintaining the quality of service (QoS) [21] in the process. There are two different approaches to achieve green cloud computing: non-technical solutions and technical solutions [20]. Non-technical approaches may include the use of renewable energy sources, efficient cooling techniques without wasting electricity on cooling equipment. The technical solutions involve virtualization and VM consolidation with live migration [10–13], dynamic voltage and frequency scaling (DVFS) [22], thermal management techniques [10], energy-aware VM scheduling [23], etc.

In [22, 24–27], authors have successfully devised different techniques based on DVFS, to solve the problems regarding energy consumption in cloud computing. DVFS method is used to adjust the voltage and frequency of the CPU dynamically depending on the load on the server. Lower voltage and frequency will result in less power consumption. But lowering the CPU frequency may hamper the CPU performance. It is a negative impact of DVFS. Many energy-efficient techniques work in thermal-unaware situations. In [10, 28–32], authors have implemented efficient resource management techniques keeping thermal awareness in mind. These works have a significant impact on energy consumption and resource utilization.

Virtualization and VM consolidation is one of the efficient techniques to tackle the problem of energy consumption and resource utilization. In Fig. 1, a modern architecture of the cloud system with virtualization and VM consolidation is shown. Cloud users request for services to the service providers. Global resource managers map these services to individual VMs that are running on physical servers. Later in the process, underutilized servers are turned off by migrating all VMs from those servers on to moderately utilized servers. This architecture helps in reducing the energy consumption of underutilized servers. On the other hand, some servers might get overloaded on the arrival of migrated VMs. Additive resource requests by existing VMs can also make a server overloaded. Overloaded servers may increase SLAV. So some VMs need to be migrated from overloaded servers as well. VM selection policies play a crucial role in selecting VMs from these overloaded servers.

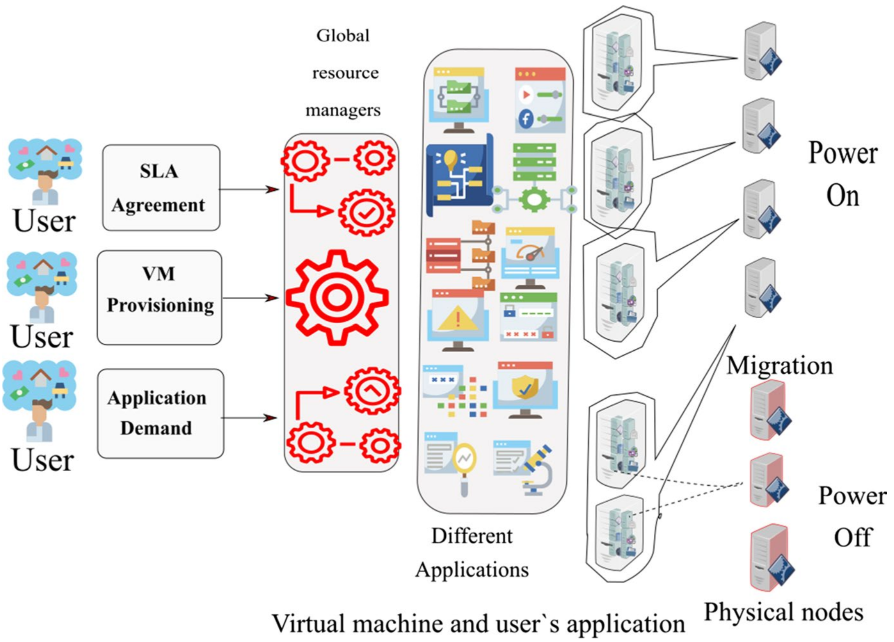


Fig. 1 The cloud system view with VM consolidation

These selected VMs are migrated to other hosts. Efficient VM selection techniques are required for optimized energy consumption and SLAV.

Using limited lookahead control, the power management problem has been defined in a virtualized assorted data center environment [33]. Their objective was to minimizing energy consumption and SLAV and maximizing profit. Applying Ksalman filter [34], they tried to predict the upcoming states and work essential reallocation. But their proposed model has worked only on a simulation-based model, not in real infrastructure like Amazon EC2. Additionally, the optimizer controller has taken too much time for less amount of nodes.

Srikantaiah et al. [35] have investigated that consolidating the workload might aggravate performance degradation because of the high utilization of resources in a virtualized system. They have discovered that the consumption of energy for each transaction outcomes a "U" shapes curve besides determining the spot which marked optimal utilization. To deal with these multiple resources, they have proposed a heuristic algorithm for that multi-dimensional bin packing issues.

In the context of a power-efficient VM allocation problem, Cardosa et al. [36] have proposed a solution. They have defined three parameters in Xen's VM environment as min, max, and shares. CPU is shared among various virtual machines depending on these three parameters. In a virtualized data center along with the virtualization technique, the VM consolidation mechanisms have given excellent results in terms of energy consumption and resource utilization in a data center.

Jung et al. [37, 38] explored the problems in the dynamic consolidation of VMs. They applied live migration on multilayer web applications while controlling SLAV. They used bun packing and gradient search techniques to develop a VM placement policy. The migration controller tackles the issue of the rearrangement of VMs to fulfill SLA. They have used two resource management schemes [39]. The local manager works on individual hosts and manages power consumption. The global manager collects information from the local managers and decides the placements of VMs.

Beloglazov and Buyya [9] have proposed a cloud framework based on VM consolidation and live migration. Their system is divided into three phases: load detection, VM selection, and VM placement. The load detection phase categorizes all hosts into three sets as overloaded hosts, moderately loaded hosts, and underloaded hosts. All the VMs from underloaded hosts are added to a migration pool. In the VM selection phase, some VMs from overloaded hosts are added to the migration pool. The VMs in the migration pool are allocated to hosts using VM placement policies. The authors have proposed four VM selection policies such as minimum migration time (MMT), maximum correlation (MC), minimum utilization (MU), and random selection (RS). MMT selects VMs with minimum migration time for migration. The MC selects maximum correlated VMs. Underutilized VMs are selected by the MU VM selection policy. The RS selects VMs randomly.

Nadjar et al. [40] have proposed some algorithms for reducing energy consumption and the number of migrations. In terms of VM selection, they have proposed policies like select a VM with minimum CPU utilization (MinU) or select a VM for migration that will experience less estimated SLA (MSV). Using the benefit of the health parameter of each host, an upper and lower threshold (static threshold) scheme for overload and underload detection is computed. Furthermore, the maximum load placement (ML)

[41] policy migrates a virtual machine (VM) to a host that will not be overloaded after migration. Tejha et al. [42] have proposed a genetic algorithm-based VM consolidation technique with the help of minimum migration time (MMT) VM selection policy. They have used two statistical methods IQR and LRR which have reduced SLAV and improved QoS. Thaman et al. [43] proposed a selection policy that depends on consolidating the performance across the hosts. It selects a virtual machine (VM) that minimizes and controls the variance of remaining load levels of hosts in a data center.

Li et al. [44] have proposed the Bayesian network-based VMs consolidation (BN-VMC) method. They have divided the task into three different stages. At the first stage, they have considered nine different parameters to propose a Bayesian network-based estimation model (BNEM) for dynamically migrating VMs. This method worked like other Bayesian networks (BN) having the excellent potentiality of probability estimation and probabilistic reasoning computation. Based on specific load patterns and migration probability, the VMs are deployed in various physical nodes. In the second part, the BN-VMC method has carried overload detection of hosts with the help of estimated overload probability. The VM migration happens at the third stage of BN-VMC. The author claims that their method has a load-balancing capability as well.

Rahmani et al. [45] have proposed an initial placement algorithm to map VMs to physical hosts based on the response of new requests. It allocates VMs to physical nodes with minimal overloading probability, called minimum overloaded probability (MOP). They have also proposed a VM reallocation algorithm called correlation-aware placement (CAP) algorithm for the placement purpose of each VM based on correlation efficient. Zhou et al. [46] have presented a design that is based on historical utilization data of the host. Braiki et al. [47] proposed a VM reallocation policy based on Fuzzy-Logic to improve energy consumption and resource utilization. To achieve these two objectives, they used a multi-objective best-fit-decreasing (BFD) policy.

Yadav et al. [48] have proposed an energy-aware VM selection policy to minimize the number of VM migrations. Base on the robust regression model, they have proposed Gradient descent-based regression (Gdr) and maximize correlation percentage (MCP) methods for overload detection. They have designed a bandwidth-aware selection policy (Bw) for overload networks. To efficiently manage energy consumption, Moghaddam et al. [49] have proposed a load-balancing framework. They have considered the CPU utilizations of each host element and find the linear correlation of CPU usage of VMs. It has claimed that their policy reduces around 66% SLAV. A three-stage framework has been proposed by Arockia et al. [50] to make an energy-efficient cloud environment. In the first stage, host loads are detected based on SLAV. A VM selection policy has been proposed depending on CPU utilization in the second stage. They have proposed a VM placement policy.

An adaptive three-threshold energy-aware algorithm has suggested by Zhou et al. [16]. In that algorithm, they have focused on reducing energy consumption and SLAV. Depending on the load, they have classified the hosts into four different classes. These four classes are heavily loaded hosts, moderately loaded hosts, light loaded hosts, and little loaded hosts. They have proposed five different VM selection policies based on the three-threshold energy-saving algorithm (TESA). The minimization of migrations based on TESA (MIMT) is one of the best among the five solutions for VM selection they have proposed. These algorithm uses CPU utilization and thresholds calculated using TESA to minimize the total number of migrations. They claimed that MIMT is one of the best solutions to reduce energy consumption and improving the energy efficiency of the cloud system. An energy-aware VM selection algorithm known as maximum utilization minimum size (MuMs) had been proposed by Yadav et al. [17]. Considering CPU utilization and VM RAM size, MuMs selects VMs which has very high utilization though the VM is smaller in size than other VMs in the host. MuMs considered a ratio between CPU utilization and VM size as a key parameter. Akhter et al. [18] developed a VM selection policy to select VMs for migrations. They consider the VMs with maximum migration time (MxMT) for migration. They claimed to have improved energy consumption by 19%. MxMT may decrease energy consumption, but this algorithm is bound to suffer from a high degree of SLAV.

3 The proposed system model

The basic framework that has been used to model an energy-efficient cloud computing environment is proposed by Beloglazov et al. [9] which is a extension to the original CloudSim [15] simulation toolkit. CloudSim offers energy-efficient resource provisioning for cloud computing. It also offers implementation of Quality of Service (QoS) in terms of Service Level Agreements.

The system consists of N_{Host} number of heterogeneous hosts and N_{VM} number of heterogeneous virtual machines (VMs) that are used to serve requests from different cloud users as per demand. Figure 2 gives the brief view of system model. Data center broker (DCB) working as a global manager receives job requests from different cloud users and processes and allocates those jobs to different VMs within the available hosts. The virtual

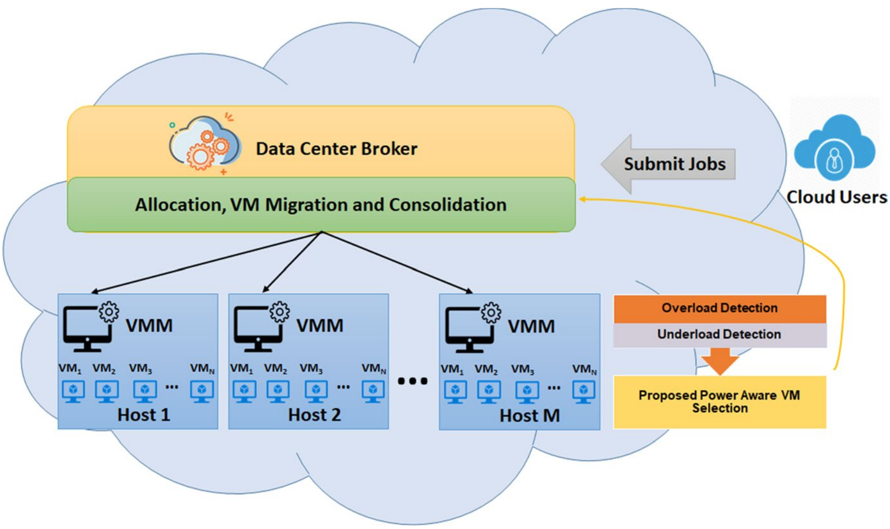


Fig. 2 System model

machine manager (VMM) resides in each host and continuously monitors the execution states and CPU utilization of VMs. It reconfigures VMs as per the resource requirements. Whenever an overloaded or underutilized host is detected by VMMs, it reports back to the DCB. The DCB issues a command for VM migration and consolidation. The VMMs are also responsible for selecting VMs that are needed to be migrated from one host to another host (VM selection phase). Whenever a host is being overloaded (overload detection phase), some of the VMs are needed to be migrated from the host. The selection of VMs for migration is known as VM selection problem. A host, being underutilized will be turned off after completely migrating all the VMs from that host. Migration of VMs form overload hosts can ensure that future demands of VMs can be fulfilled. VM selection is considered to be an important job. Efficient VM selection can minimize the number of migrations. That will help in minimizing the energy consumption as well as SLAV. The data center broker migrates those selected VMs to different active hosts if possible. Otherwise, a new host is being used to accommodate the remaining VMs. In this research work, the local regression robust (LRR) method [9, 51] is used to detect host overload.

3.1 Proposed power-aware VM selection policy

It is shown in Fig. 2 that the proposed power-aware VM selection policy has been implemented for selecting VMs from overloaded hosts. These selected VMs will be migrated to some other hosts. A wrong selection may lead to aggregation of the number of migrations, that alone may hike the cost and overall energy consumption of the data center. Live migration has negative impact on the cost and energy consumption as shown by Beloglazov in [9].

Algorithm 1: Proposed Power Aware VM selection policy

```
Input: host_list = List of Active Hosts
Output: VMsToMigrateList = List of VMs needs to be Migrated
start
foreach host, in Host_list do
    if isHostOverloaded(host) then
        vm_list ← host.allocatedVMs()
        selectedVM ← null
        uMax ← Double.MinValue()
        foreach vm in vm_list do
            if !(isInMigration(vm)) then
                u ← currentVMUtilization / vm.allocatedResource
                if uMax < u then
                    uMax ← u
                    selectedVM ← vm
                end
            end
        end
        VMsToMigrateList.add(selectedVM)
    end
end
return VMsToMigrateList
stop
```

Figure 3 depicts a flow diagram of the proposed power-aware VM selection policy. Also, The proposed power-aware VM selection policy is presented in Algorithm 1. The process starts with a list of hosts say, *host_list* as the list of all active hosts and returns a list of VMs say, *VmsToMigrateList* as the list if all VMs that need reallocation. For all the overload hosts say, $host_i$ in *host_list* a *vm_list* is prepared containing all the VMs allotted to $host_i$. The utilization of all the VMs in *vm_list* is calculated using Eq. 1.

$$u_i = \frac{currentVMUtilization}{AllocatedResource_{vm_i}} \tag{1}$$

where vm_i is one of the VMs from vm_list and u_i is the utilization of vm_i. The VM say the selected VM is selected from vm_list which has max utilization using Eq. 2.

$$SelectedVM = \max_{\forall vm_i \in vm\_list} \left\{ u_i \right\} \tag{2}$$

This *selectedVM* is added to the *VmsToMigrateList*. The process continues until all the overloaded VMs are checked. After checking all the overload hosts, the algorithm stops and returns *VmsToMigrateList* for reallocation.

4 Performance evaluation

4.1 Experimental setup

The proposed power-aware VM selection policy is implemented to decrease the number of migrations. The basic idea is to relocate the VM that uses the maximum resource in a overloaded host. Relocating or migrating the resource-hungry VM will ensure that the minimum number of migrations is required to readjust the load on overloaded hosts. It also provides space for VMs that are being relocated from underutilized hosts. These factors together will help in reducing overall energy consumption of a data center. Keeping overloaded hosts under control and minimizing number of migration will also help in reducing overall SLAV.

The IaaS cloud architecture is supposed to be a cloud environment capable of providing a view of unlimited computing resources to the customers. For this reason, it is necessary to analyze the performance of the proposed power-aware VM selection policy on a extensive virtualized cloud data centers. Nonetheless, on any real infrastructures, it is quit strenuous to repeatedly conduct and test large-scale experiments due to the cost factor and reconfiguration of the system for every instances of the experiment. Accordingly, to overcome this limitation and ensuring experiments in repeatable manner, cloud simulation approaches have been considered, as a better approach for evaluating the proposed power-aware VM selection policy.

For the experiment of large-scale virtualized cloud applications, CloudSim toolkit [9, 15] is one of the promising simulation tools. In comparison with the other cloud simulation tools, it provides a better modeling of virtualized cloud

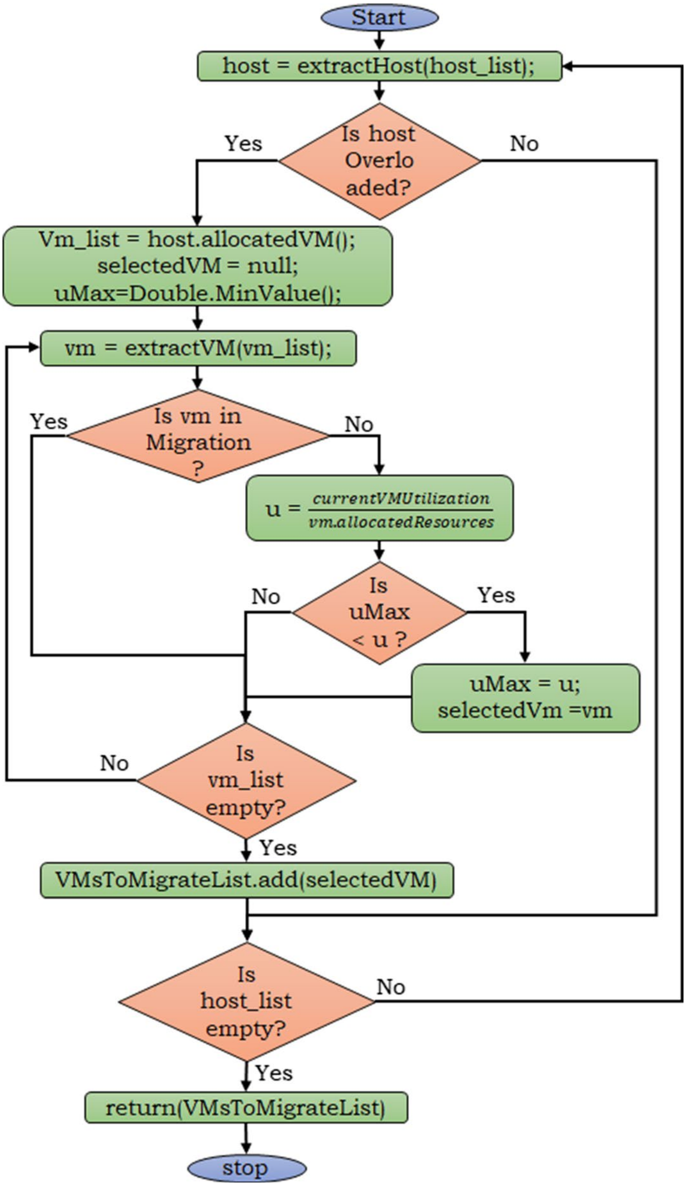


Fig. 3 Flowchart of proposed power-aware VM selection policy

architecture that supports scalability and dynamic resource management and provisioning along with the capability to simulate green cloud applications.

A data center containing 800 heterogeneous physical hosts has been used for the simulation purpose. Two different types of hosts have been considered in the process, HP Pro Liant ML 110 G5 powered by Intel Xeon3075 clocked at 2660

MHz and HP Pro Liant M L 110 G4 powered by Intel Xeon3040 clocked at 1860 MHz each having 4 GB of RAM. The MIPS (millions instruction per second) has been used to map the processing capability or the frequency of the physical servers' CPU. That is, the HP Pro Liant ML 110 G5 has 2 cores each having 2660 MIPS, and HP Pro Liant M L 110 G also has 2 cores each with processing capability of 1860 MIPS. The power consumption characteristics of these two hosts are given in Table 1. Allocated network bandwidth for each host is 1 GB/s.

For the VM instances, standard Amazon EC2 [52] types of VM instances are followed with small tweaks. Four different types of VM instances have been created. Characteristics of those VM instances are given in Table 2. Depending of the requirement of the workloads, one of these four VM instances is created on the physical hosts.

It is very much essential to carry out the experiment based on workload traces that are generated from real systems, so that the simulation-based approaches can be more acceptable for evaluations. The workload traces used in experiment are collected from a infrastructure monitoring program conducted by PlanetLab, known as CoMon [53] project. Around 1516 data instances are considered for the experiment that are collected from several VMs from data centers located all over the world in a interval of 5 min. Each VM is assigned to one of the workload traces depending on the characteristics of workload.

4.2 Results and analysis

The performance of the proposed energy-aware VM selection policy has been considered in terms of the existing classic VM selection algorithms like maximum correlation (MC) [9], minimum migration time (MMT) [9], minimum utilization (MU) [9], random selection (RS) [9], maximum utilization minimum size (MuMs) [17], minimum migrations based on TESA (MIMT) [16], maximum migration

Table 1 Power consumption of hosts at different loads

Host	Power consumption (in Watts) at different load on hosts										
	100%	90%	80%	70%	60%	50%	40%	30%	20%	10%	0%
G5	135	133	129	125	121	116	110	105	101	97	93.7
G4	117	114	121	108	106	102	99.5	96	92.6	89.4	86

Table 2 Characteristics of VM instances

VM instances	MIPS	Cores	RAM (in MB)	Bandwidth (in Mbits/s)	Storage (in MB)
Microinstance	500	1	613	100	2500
Small instance	1000	1	1740	100	2500
Extra-large instance	2000	1	1740	100	2500
High-CPU medium instance	2500	1	870	100	2500

time (MxMT) [18]. These policies are already discussed in Sect. 2. As far performance metrics, number of migrations (NM), number of shutdowns (NS), SLA time per active host (T_{SLAH}), performance degradation due to migration ($PerfDeg_M$), SLA violation ($Violation_{SLA}$), total energy consumption ($Total_{Energy}$), energy and SLA violation (ESV) and average execution time ($Avg_{ExcTime}$) are used. The metrics are categorized into four different categories based on the characteristics: migration-based metrics (NM and NS), SLA-based metrics (T_{SLAH}, $PerfDeg_M$ and $Violation_{SLA}$), energy-based metrics ($Total_{Energy}$ and ESV) and computation-based metrics ($Avg_{ExeTime}$). In subsequent subsections, all these performance metrics will be discussed with proper analysis.

4.2.1 Analyzing migration-based metrics

Live migration is a process of VM consolidation that can be used to maintain SLA in cloud task processing. But excessive migration also has some negative impact. NM is the total number of migrations performed throughout the experiment process. Excessive live migration can aggregate the total energy consumption as well as migration cost because in a live migration, the VM is being migrated when it is still in service and until the migration is completed the cost and energy from running the same VM in two simultaneous hosts will be added to the total energy consumption and execution cost. Excessive number of migration can also affect the network bandwidth as most of the bandwidth will be used to migrating VMs. Migration also influences another important factor that is number of shutdowns (NS) performed through out the execution. It also specifies that some hosts are repeatedly turned on and off that will only aggregate the energy consumption and cost. Table 3 presents the result set based on migration-based metrics.

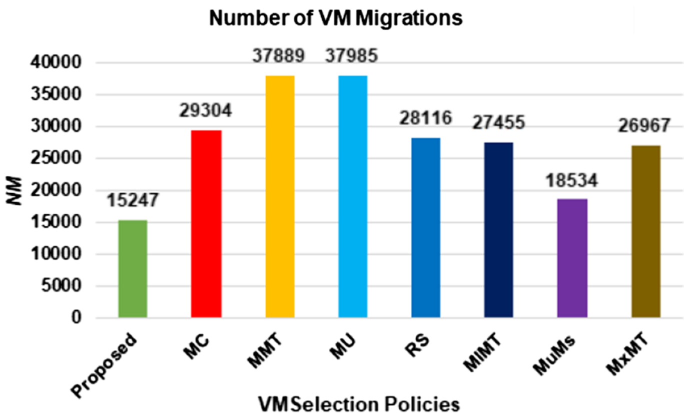
Figure 4 represents a comparative analysis of NM and NS, respectively. In Fig. 4a, x-axis denotes VM selection policies and y-axis denotes the number of migrations. The points represents the number of VM migrations obtained using a particular VM selection policy from the x-axis. In Fig. 4b, the number of host shutdowns is shown in y-axis, while x-axis denotes different VM selection Policies. It is evident from these figures that in terms of NM and NS, the proposed power-aware VM selection policy outperforms the existing VM selection Policies.

4.2.2 Analyzing SLA-based metrics

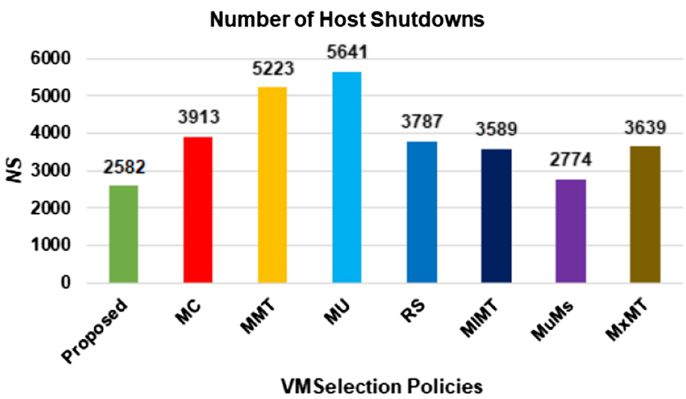
For any cloud computing environment, ensuring quality of service (QoS) is crucial. QoS requirements are established through a service level agreements (SLAs) between

Table 3 Analyzing migration-based metrics

VM selection policy	NM	NS
Proposed	3046	775
MC	29,304	3913
MMT	37,889	5223
MU	37,985	5641
RS	28,116	3787



(a) Comparison Based on Number of Migrations



(b) Comparison Based on Number of Shutdowns

Fig. 4 Analyzing migration-based metrics

the provider and subscribers. SLAs ensure the user level satisfaction in terms of response time, throughput, availability, etc., of the service they are getting from the user. These characteristics can be varied for different types of applications. Beloglazov et. al. [9] provided a measurement to SLA violation for an IaaS. SLA violation ($Violation_{SLA}$) can be identified using Eq. 3.

$$Violation_{SLA} = T_{SLAH} \times PerfDeg_{M} \tag{3}$$

- $Violation_{SLA}$ is the SLA violation in percentage.
- T_{SLAH} is the percentage of time an host has experienced 100% CPU utilization while it is in active state.
- $PerfDeg_{M}$ is the degradation of system performance due to migrations.

It can be observed that when a host is being utilized to its maximum capability, the VMs are bound with the amount of resources available at that host, so these VMs cannot utilize more resources if required. That limits the service provided by the VM, so T_{SLAH} is considered as one of the important metrics to SLA violation. Whenever a VM is being migrated, it is evident to have some degree of performance degradation. T_{SLAH} and *PerfDeg M* can be evaluated using Eqs. 4 and 5.

$$T_{SLAH} = \frac{1}{N_{Host}} \sum_{i=1}^{N_{Host}} \frac{T_{FU_i}}{T_{Active_i}} \tag{4}$$

- N_{host} is the number of hosts.
- T_{FU_i} represents time ith host was at its 100% utilization, that may lead to SLA violation.
- T_{Active_i} is the total time ith host was in the awaken state.

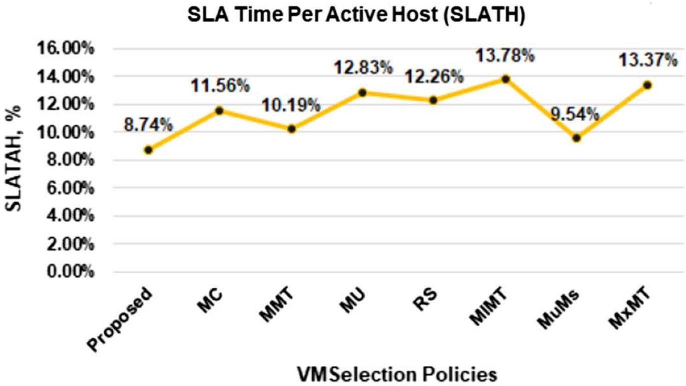
$$PerfDegM = \frac{1}{N_{VM}} \sum_{j=1}^{N_{VM}} \frac{Deg_j}{MIPS_j} \tag{5}$$

- N_{VM} is the number of VMs
- Deg_j is the performance degradation of jth VM due to its migration, considering 10% of MIPS.
- $MIPS_j$ is the total MIPS requested by the jth VM in its lifespan.

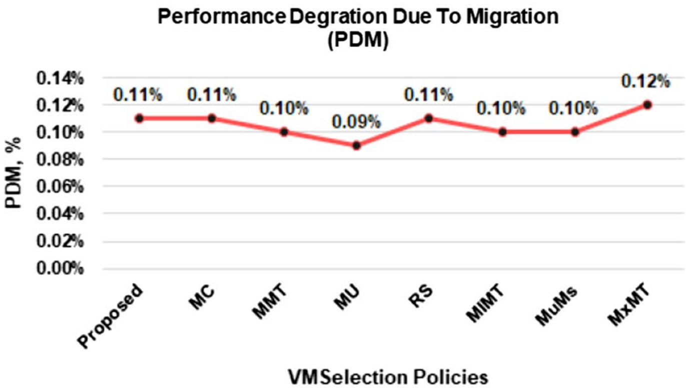
In Table 4, a comparative result set of all these SLA-based metrics is listed comparing the proposed power-aware VM selection policy with existing ones obtained from simulation. Alongside, Fig. 5 highlights the analysis based on SLA-based metrics. Figure 5a compares different T_{SLAH} values (y-axis) obtained from different VM selection Policies (x-axis). From this figure, we can say the proposed VM selection policy outperforming existing policies having the lowest T_{SLAH}. Figure 5b presents comparison based on $PerfDeg_M$, while x-axis denotes different policies and y-axis denotes performance degradation in percentage. Though in terms of $perfDeg_M$ the proposed energy-aware VM selection policy struggles a bit

Table 4 Analyzing SLA-based metrics

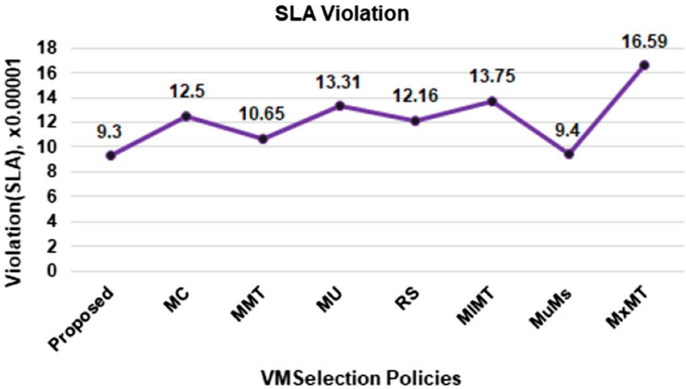
VM selection policy	T_{SLAH} (%)	$PerfDeg_M$ (%)	$Violation_{SLA} \times 10^{-5}$,
Proposed	5.29	0.01	0.3
MC	11.56	0.11	12.5
MMT	10.19	0.10	10.65
MU	12.83	0.09	13.31
RS	12.26	0.11	12.16



(a) Comparison Based on T_{SLAH}



(b) Comparison Based on $perfDef_M$



(c) Comparison Based on $Violation_{SLA}$

Fig. 5 Analyzing SLA-based metrics

compared to some of the existing VM selection policies, the overall SLA violation is improved. In Fig. 5c, a comparison based on overall SLA violation ($Violation_{SLA}$) is shown, where x-axis denotes $Violation_{SLA}$ and y-axis denotes VM selection Policies. The proposed power-aware VM selection policy has the minimum SLA violation compared to other VM selection Policies.

4.2.3 Analyzing energy-based metrics

Energy consumption is one the most important criteria in the modern computing environment. We have used Table 1 to model the energy consumption in terms of host loads. Sometimes, energy consumption can be reduced drastically if we are allowed to hamper the SLA requirements. But that is not optimized solution. There will be optimization between SLA violation and energy consumption, as they negatively correlate each other. Beloglazov et al. [9] proposed a metric known as ESV (energy and SLA violation) represented in Eq. 6.

$$\mathrm{ESV} = Total_{Energy} \times Violation_{SLA} \tag{6}$$

The objective is to minimize $Violation_{SLA}$ as well as $Total_{Energy}$. Table 5 shows simulation results related to energy-based metrics. In Fig. 6, energy-based metrics are used to evaluate the performance of the proposed power-aware VM selection policy in terms of existing ones. From Fig. 6a, it is clear that the proposed power-aware VM selection policy has minimum power consumption. On the other hand, Fig. 6b represents ESV on the x-axis and VM selection Policies on the y-axis and points represents energy and SLA for each VM selection policy. The proposed power-aware VM selection policy outruns other existing policies in this regard.

4.2.4 Analyzing computation-based metrics

Considering computational-based metrics, average execution time ($Avg_{ExcTime}$) of workload traces has been evaluated for all participating VM selection Policies. In Table 6 $Avg_{ExcTime}$ is shown corresponding to VM selection Policies. Figure 7 is used to draw the comparison graph of $Avg_{ExcTime}$ by plotting $Avg_{ExcTime}$ in y-axis and VM selection Policies in x-axis. The proposed power-aware VM selection policy has minimum $Avg_{ExcTime}$ over other VM selection Policies.

Table 5 Analyzing energy-based metrics

VM selection policy	$Total_{Energy}$ in kWh.	ESV $\times 10^{-3}$
Proposed	132.33	4.7
MC	147.16	5.7
MMT	164.26	5.5
MU	176.12	8.3
RS	145.39	6.3

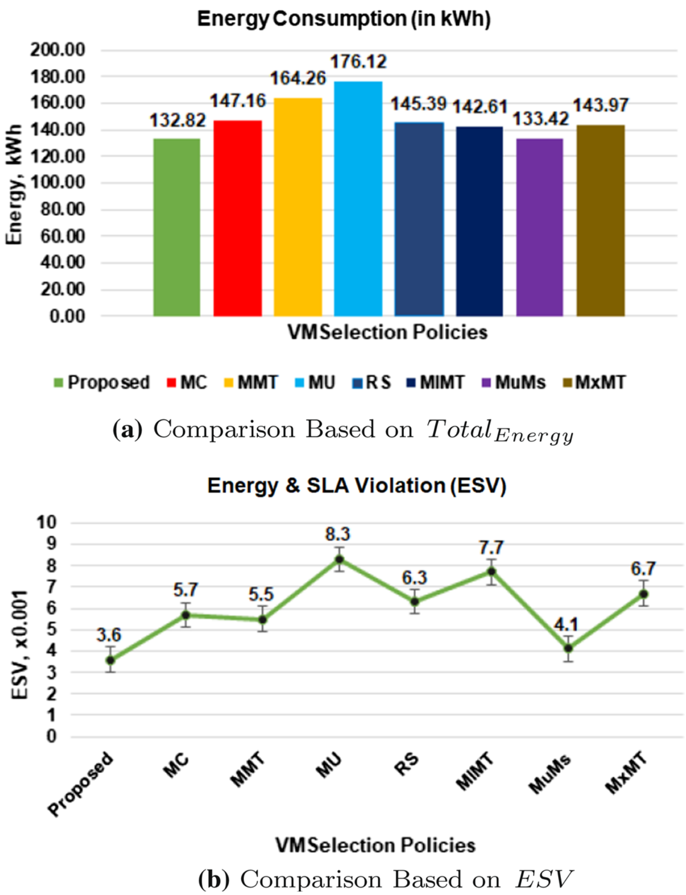


Fig. 6 Analyzing energy-based metrics

Table 6 Analyzing computation-based metrics

VM selection policy	$Avg_{ExcTime}$
Proposed	132.33
MC	147.16
MMT	164.26
MU	176.12
RS	145.39

5 Conclusion and future scope

In recent years, due to global warming and ever increasing carbon footprint, it is very much necessary to control and reduce energy consumption of data centers throughout the globe. The energy consumption also affects the volume of carbon

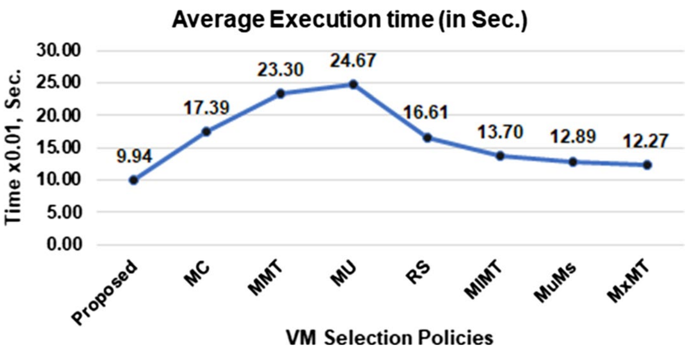


Fig. 7 Comparison based on $Avg_{ExcTime}$

emissions. Also, providing better services to cloud users in an efficient manner it is essential to lower the SLA violation along with less power consumption. A proposed power-aware VM selection policy has been depicted in this paper that helps in energy-efficient VM selection with minimum SLA violation. The proposed power-aware VM selection policy has been modeled and simulated on a simulation environment with lots of flexibilities and constrains. This research work is now being planned for the Internet of Things (IoT)-assisted green cloud computing environment for the future work.

References

1. Erl T (1900) Service-oriented architecture: concepts, technology, and design. Pearson Education India, Noida
2. Fox A, Griffith R, Joseph A, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A, Stoica I (2009) Above the clouds: a berkeley view of cloud computing. Dept. Electrical Eng. and Comput. Sciences, University of California, Berkeley, Rep. UCB/EECS, vol 28(13), p 2009
3. Zhang Q, Cheng L, Boutaba R (2010) Cloud computing: state-of-the-art and research challenges. J Internet Serv Appl 1(1):7–18
4. Buyya R, Yeo CS, Venugopal S, Broberg J, Brandic I (2009) Cloud computing and emerging it platforms: vision, hype, and reality for delivering computing as the 5th utility. Future Gener Comput Syst 25(6):599–616
5. Marashi A (2019) Power hungry: the growing energy demands of data centers. https://www.vxchnge.com/blog/power-hungry-the-growing-energy-demands-of-data-centers. Accessed 25 July 2019
6. Barham P, Dragovic B, Fraser K, Hand S, Harris T, Ho A, Neugebauer R, Pratt I, Xen AW (2003) The art of virtualization. In: Proceedings of the 19th ACM Symposium on Operating Systems Principles
7. Ferreto TC, Netto MA, Calheiros RN, De Rose CA (2011) Server consolidation with migration control for virtualized data centers. Future Gener Comput Syst 27(8):1027–1034
8. Corradi A, Fanelli M, Foschini L (2014) Vm consolidation: a real case based on openstack cloud. Future Gener Comput Syst 32:118–127
9. Beloglazov A, Buyya R (2012) 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 24(13):1397–1420
10. Yavari M, Rahbar AG, Fathi MH (2019) Temperature and energy-aware consolidation algorithms in cloud computing. J Cloud Comput 8(1):1–16

11. Zhou Z, Abawajy J, Chowdhury M, Hu Z, Li K, Cheng H, Alelaiwi AA, Li F (2018) Minimizing sla violation and power consumption in cloud data centers using adaptive energy-aware algorithms. Future Gener Comput Syst 86:836–850
12. Xiao H, Hu Z, Li K (2019) Multi-objective vm consolidation based on thresholds and ant colony system in cloud computing. IEEE Access 7:53441–53453
13. Ranjbari M, Torkestani JA (2018) A learning automata-based algorithm for energy and sla efficient consolidation of virtual machines in cloud data centers. J Parallel Distrib Comput 113:55–62
14. Mirobi GJ, Arockiam L (2015) Service level agreement in cloud computing: an overview. In: 2015 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT). IEEE, pp 753–758
15. Calheiros RN, Ranjan R, Beloglazov A, De Rose CA, Buyya R (2011) Cloudsim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. Softw Pract Exp 41(1):23–50
16. Zhou Z, Hu Z, Li K (2016) Virtual machine placement algorithm for both energy-awareness and sla violation reduction in cloud data centers. Sci Program 2016:15
17. Yadav R, Zhang W, Chen H, Guo T (2017) Mums: Energy-aware vm selection scheme for cloud data center. In: 2017 28th International Workshop on Database and Expert Systems Applications (DEXA). IEEE, pp 132–136
18. Akhter N, Othman M, Naha RK (2018) Energy-aware virtual machine selection method for cloud data center resource allocation. arXiv preprint arXiv:1812.08375
19. Liu L, Wang H, Liu X, Jin X, He WB, Wang QB, Chen Y (2009) Greencloud: a new architecture for green data center. In: Proceedings of the 6th International Conference Industry Session on Autonomic Computing and Communications Industry Session. ACM, pp 29–38
20. Radu L-D (2017) Green cloud computing: a literature survey. Symmetry 9(12):295
21. Ardagna D, Casale G, Ciavotta M, Pérez JF, Wang W (2014) Quality-of-service in cloud computing: modeling techniques and their applications. J Internet Serv Appl 5(1):11
22. Beloglazov A, Buyya R (2010) Energy efficient resource management in virtualized cloud data centers. In: Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing. IEEE Computer Society, pp 826–831
23. Goiri I, Julia F, Nou R, Berral JL, Guitart J, Torres J (2010) Energy-aware scheduling in virtualized datacenters. In: 2010 IEEE International Conference on Cluster Computing. IEEE, pp 58–67
24. Hsu S-F, Lee C-C, Hwang S-W, Chen CH (2005) Highly efficient top-emitting white organic electroluminescent devices. Appl Phys Lett 86(25):253508
25. Sharma V, Thomas A, Abdelzaher T, Skadron K, Lu Z (2003) Power-aware QoS management in web servers. In: RTSS 2003. 24th IEEE Real-Time Systems Symposium. IEEE, pp 63–72
26. Deng Q, Meisner D, Ramos L, Wenisch TF, Bianchini R (2011) Memscale: active low-power modes for main memory. In: ACM SIGARCH Computer Architecture News. ACM, vol 39, pp 25–238
27. Deng Q, Meisner D, Bhattacharjee A, Wenisch TF, Bianchini R (2012) Multiscale: memory system DVFS with multiple memory controllers. In: Proceedings of the 2012 ACM/IEEE International Symposium on Low Power Electronics and Design. ACM, pp 297–302
28. Chaudhry MT, Ling TC, Manzoor A, Hussain SA, Kim J (2015) Thermal-aware scheduling in green data centers. ACM Comput Surv (CSUR) 47(3):39
29. Banerjee A, Mukherjee T, Varsamopoulos G, Gupta SK (2010) Cooling-aware and thermal-aware workload placement for green HPC data centers. In: International Conference on Green Computing. IEEE, pp 245–256
30. Yun B, Shin KG, Wang S (2011) Thermal-aware scheduling of critical applications using job migration and power-gating on multi-core chips. In: 2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications. IEEE, pp 1083–1090
31. Tan H, Ranka S (2014) Thermal-aware scheduling for data parallel workloads on multi-core processors. In: 2014 IEEE Symposium on Computers and Communications (ISCC). IEEE, pp 1–7
32. Van Damme T, De Persis C, Tesi P (2018) Optimized thermal-aware job scheduling and control of data centers. IEEE Trans Control Syst Technol 27(2):760–771
33. Kusic D, Kephart JO, Hanson JE, Kandasamy N, Jiang G (2009) Power and performance management of virtualized computing environments via lookahead control. Cluster Comput 12(1):1–15
34. Bishop G, Welch G et al (2001) An introduction to the kalman filter. In: Proceeding of SIGGRAPH, Course 8(27599–23175), p 41

35. Srikantaiah S, Kansal A, Zhao F (2008) Energy aware consolidation for cloud computing. In: Proceedings of the 2008 Conference on Power Aware Computing and Systems. USENIX Association, pp 10–10
36. Cardosa M, Korupolu MR, Singh A (2009) Shares and utilities based power consolidation in virtualized server environments. In: 2009 IFIP/IEEE International Symposium on Integrated Network Management. IEEE, pp 327–334
37. Jung G, Joshi KR, Hiltunen MA, Schlichting RD, Pu C (2008) Generating adaptation policies for multi-tier applications in consolidated server environments. In: 2008 International Conference on Autonomic Computing. IEEE, pp 23–32
38. Jung G, Joshi KR, Hiltunen MA, Schlichting RD, Pu C (2009) A cost-sensitive adaptation engine for server consolidation of multitier applications. In: Proceedings of the 10th ACM/IFIP/USENIX International Conference on Middleware. Springer, p 9
39. Nathuji R, Schwan K (2007) Virtualpower: coordinated power management in virtualized enterprise systems. In: ACM SIGOPS Operating Systems Review. ACM, vol 41, pp 265–278
40. Nadjar A, Abrishami S, Deldari H (2017) Load dispersion-aware vm placement in favor of energy-performance tradeoff. J Supercomput 73(4):1547–1566
41. Nadjar A, Abrishami S, Deldari H (2015) Hierarchical VM scheduling to improve energy and performance efficiency in IAAS cloud data centers. In: 2015 5th International Conference on Computer and Knowledge Engineering (ICCKE). IEEE, pp 131–136
42. Theja PR, Babu SK (2016) Evolutionary computing based on QoS oriented energy efficient VM consolidation scheme for large scale cloud data centers. Cybern Inf Technol 16(2):97–112
43. Thaman J, Singh M (2017) SLA conscious VM migration for host consolidation in cloud framework. Int J Commun Netw Distrib Syst 19(1):46–64
44. Li Z, Yan C, Yu X, Yu N (2017) Bayesian network-based virtual machines consolidation method. Future Gener Comput Syst 69:75–87
45. Rahmani S, Khajehvand V, Torabian M (2019) Burstiness-aware virtual machine placement in cloud computing systems. J Supercomput pp 1–26
46. Zhou X, Li K, Liu C, Li K (2019) An experience-based scheme for energy-SLA balance in cloud data centers. IEEE Access 7:23500–23513
47. Braiki K, Youssef H (2019) Fuzzy-logic-based multi-objective best-fit-decreasing virtual machine reallocation. J Supercomput pp 1–28
48. Yadav R, Zhang W, Kaiwartya O, Singh PR, Elgendy IA, Tian Y-C (2018) Adaptive energy-aware algorithms for minimizing energy consumption and sla violation in cloud computing. IEEE Access 6:55923–55936
49. Moghaddam SM, Piraghaj SF, O'Sullivan M, Walker C, Unsworth C (2018) Energy-efficient and sla-aware virtual machine selection algorithm for dynamic resource allocation in cloud data centers. In: 2018 IEEE/ACM 11th International Conference on Utility and Cloud Computing (UCC). IEEE, pp 103–113
50. Arockia RA, Arun S (2019) Virtual machine consolidation framework for energy and performance efficient cloud data centers. In: 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN). IEEE, pp 1–7
51. Cleveland WS (1993) Visualizing data. Hobart press, New Jersey
52. Amazon EC2 Instance Types (2019) https://aws.amazon.com/ec2/instance-types/. Accessed 2 Feb 2019
53. Park K, Pai VS (2006) Comon: a mostly-scalable monitoring system for planetlab. ACM SIGOPS Oper Syst Rev 40(1):65–74