International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-6, Issue-6)

Research Article

June
2017

A Comprehensive Survey on SLA Compliant Energy Aware Resource Allocation in Cloud Datacenters

Anitha R
Associate Professor, Dept of CSE,
The National Institute of Engineering,
Mysuru, Karnataka, India

Dr. C Vidya Raj
Professor, Dept of CSE,
The National Institute of Engineering,
Mysuru, Karnataka, India

Abstract:

Cloud Computing has achieved immense popularity due to its unmatched benefits and characteristics. With its increasing popularity and round the clock demand, cloud based data centers often suffer with problems due to over-usage of resources or under-usage of capable servers that ultimately leads to wastage of energy and overall elevated cost of operation. Virtualization plays a key role in providing cost effective solution to service users. But on datacenters, load balancing and scheduling techniques remain inevitable to provide better Quality of Service to the service users and maintenance of energy efficient operations in datacenters. Energy-Aware resource allocation and job scheduling mechanisms in VMs has helped datacenter providers to reduce their cost incurrence through predictive job scheduling and load balancing. But it is quite difficult for any SLA oriented systems to maintain equilibrium between QoS and cost incurrence while considering their legal assurance of quality, as there should not be any violations in their service agreement. This paper presents some state-of-the-art works by various researchers and experts in the arena of cloud computing systems and particularly emphasizes on energy aware resource allocations, job scheduling techniques, load balancing and price prediction methods. Comparisons are made to demonstrate usefulness of the mechanisms in different scenarios.

Keywords: Cloud Computing, Energy-Aware Resource Allocation, VM Migration, Load Balancing, Price Prediction

I. INTRODUCTION

Cloud computing came in as a game changer for modern IT enterprises. Instead of constructing and maintaining an in-house IT infrastructure, the cloud computing enabled trouble-free outsourcing of IT infrastructure to third party service providers for their computational needs [1]. The clouds refer to a set of large scale out-sourcing hardware infrastructure, known as datacenters, which has the ability to accommodate several thousands of servers. More often, these servers will be run with the help of virtualized operating systems and thus, they are known as virtual machines (VMs). Thus, these cloud based datacenters supports wide array of system application and offers a concept of unlimited computing resources to its users which are later billed as per usage or subscription plans. Although these cloud based server provide numerous benefits to its users through its virtualized infrastructures, they are prone to few overhead and maintenance costs. Even though, the cloud datacenters have several in built management techniques, like on-demand resource scheduling, energy management, etc, they suffer from few issues which restricts their wide spread acceptance [2],[3].

Essentially, the cloud based datacenters must guarantee superior Quality of Service (QoS) and reliability in order to make them commercially successful in the field of computing infrastructure. The assurance and guarantees made by these cloud datacenters are recorded on paper, known as Service Level Agreements (SLAs). The SLA are quite essential and it must be properly documented in order to infuse confidence and trust among its clients and users [4]. For this reason, the vendors might over-provision the resources to comply with the SLA requirements and provide back up for critical demands. This resource provisioning is very crucial, as it dictates how effectively the vendor follows their regulations and commitments towards the users by maintaining efficient and effective use of the underlying infrastructure. The idea of efficient resource provisioning is to allocate each service request dynamically to a set of minimal computing resources which are sufficient for SLA requirement by keeping the resources free so that more virtual machines can be installed.

The provisioning decisions must be able to adjust to variations in resource handling, and react effortlessly to unexpected flow of incoming requests. Therefore, the task of partitioning the server datacenter resources among the different applications becomes critical. On the contrary, the servers tend to operate on low utilization level which prompts to misuse of computing resources during low-load scenarios.

This practice of over-provisioning of resources may cause additional expenses in terms of hardware maintenance such as server administration, power, cooling, etc [5]. Several scholars and engineers have made an attempt to tackle these problems with the help of dynamic resources allocation through virtualization. Nevertheless, the main objective was laid on resource scheduling with respect to any specific type of SLA or application, for instance bandwidth, transactional capacity, etc. Despite the fact that the complex computing applications are gradually developing into an integral part of the datacenter, the SLA with various kinds of applications and their progression still needs evolvement.

Currently, the majority of the datacenters operate numerous variety of applications on discrete VMs without including any measures to know about the various SLA requirements, like deadline, bandwidth, to name a few. This type of ignorance will cause under-utilization of available resources which will increase management complications. Conversely, the cloud vendor nowadays provides regulated and restricted performance or QoS guarantees. For example, the Amazon EC2, provides guarantee on the accessibility of resources, but not the performance of VMs [6],[7]. Hence, the scheduling is most important criteria when it comes to research area of enhancing the performance of the data centers. The primary objective of scheduling is to assigns the incoming tasks to suitable resources which can achieve one or more directives, as per regulation and guidelines. The category of problems relating to the Scheduling in data centers belongs to NP-hard problem. As the solution space of the data centers area quite cast, finding an optimal solution usually consumes large amounts of time. On the other hand, live migration of VMs can be expensive, as it needs more network bandwidth and power. Furthermore, the live migration might also affect the VM availability, momentarily. In a particular scenario, where more than one VMs requires migration, proper care must be taken to opt for a wise scheduling technique, as it can help to reduce the effect of migration on infrastructure and quality of service [8].

Since 1996, the cost of power, energy and cooling systems for the data centers has gone through the roof by over 800%. And moreover, the expenses are ever increasing without any hope of cost reduction [9]. When it comes to large-scale datacenters of present and future infrastructure services, the energy efficiency is one of the most significant challenge. Regardless of the profitable and commercial benefit of lower energy expenditure on the users' side, the cloud vendors must face numerous challenges, like meeting the high performance expectations of clients, maintaining a balance between favorable energy-efficiency and complying with the escalating demand of the users.

With the help of virtualization, the earlier generation of energy-efficient Cloud computing methodologies have presented several techniques to dynamically vary the size of the server pools on the basis of the oncoming demand [10],[11]. Some of the research works [12],[13] have suggested to prolong these techniques with enriched live-migration and server activation policies. These tactics have helped in lowering violations related to Service Level Agreement (SLA). On the other hand, these techniques don't consider the prospective inefficiencies at a fine grained level like the operating cost created by the rivalry for resources in virtualized environments [14]. Subsequently, the performance boost and energy-efficiency claimed by the vendors might be considerably weakened under practical scenarios.

Numerous management tactics and approaches have been established to efficiently lower the energy utilization under various factors. In spite of this, they don't help most of the modern age's datacenters as they are built on the grounds of virtualization technologies. Today's virtual machine technology can proficiently monitor and administer the server association, and enhance the overall productivity by reducing the energy consumption in data center. The implementation of an energy related computing system can deal with the problem concerning with the efficiency and energy consumption. The overall energy consumption of any computing resources is directly related the amount of load applied by the application and requests [15].

Several researchers have conducted various studies to estimate the cost of datacenters which offers cloud based services. The results from these studies have indicated that about 40% of the total remunerated cost of the datacenter is spent on power related areas, such as power supply, cooling, electricity utility, among others [16]. The energy consumption in datacenters has been raising radically over the past few years. It now accounts for over 2% of total power usage in the USA alone [17]. This has forced to improve energy efficiency of the datacenters, which has become a higher priority in cloud computing.

Figure 1 illustrates the breakdown of peak power consumption with respect to various components of a 2012 Google datacenter. It is apparent that the highest portion of power is consumed by the CPU [18]. Hence, more importance id given to lowering operating voltage of CPU which can reduce power consumption. But the problem is, the energy consumption of the CPU is not related to their utilization levels [19]. When fully utilized, the CPU produce high performance per watt, and consequently, the energy efficiency rapidly descents with low CPU utilizations. Figure 2 demonstrates how the performance and power is influenced by various CPU utilization levels. The y-axis represents power value, and x axis represents CPU utilization levels. To be specific, the power consumption during idle state of CPU is (denoted by P(0%)) is more than one half of the maximum power (denoted as P(100%)). This is due to noticeable power leakage in the nodes.

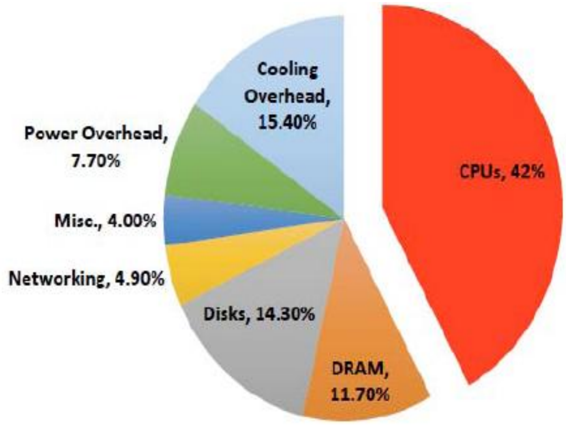

Figure 1: Breakdown of Peak Power Consumption in Various Components of Google Datacenter

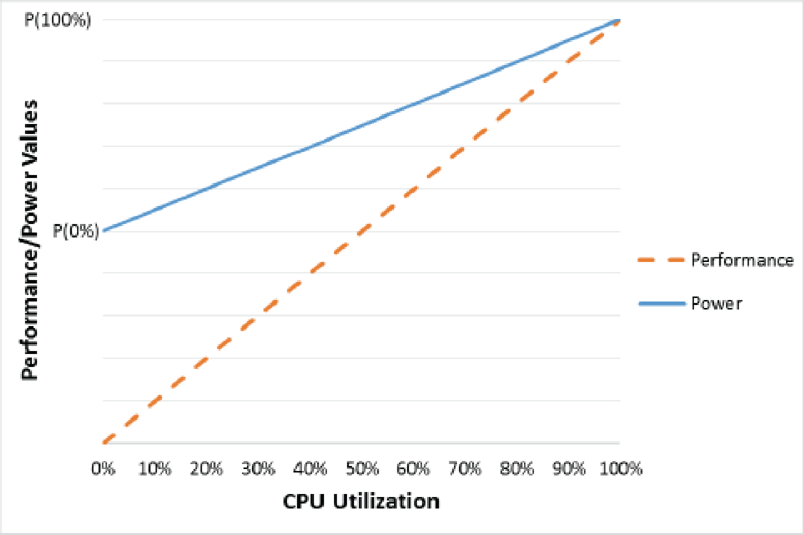

Figure 2: Representation of CPU Performance vs CPU Utilization

Energy efficient datacenters tend to consume less electricity due to better utilization of resources, and hence result in reducing overall monitory cost. Price prediction methods are of great use in forecasting overall cost of operation while performing energy aware scheduling. Various algorithms such as ANNs, SVMs or ARIMA models can be applied to predict approximate cost incurring in jobs scheduling and VM migrations over a span of request processing. In this paper, we provide an extensive survey and comparative analysis of various energy aware resource provisioning techniques, VM scheduling algorithms in SLA compliant cloud systems and price prediction methods.

1.1. Virtualization

The virtualization is central intelligence of cloud computing environment. This technique can be translated as generalization of computer resources. It undergrounds the physical characteristics of the resources from the consumers and gives them the perspective for utilizing them as logical ones [20]. By virtualize, we mean that a single physical resource can be exposed as multiple virtual resources or multiple physical resources can be exposed as a single virtual resource. Resource can be anything like server, an OS, an application, or a storage device. The main aim of virtualization is to efficiently utilize the limited IT resources by making use of many idle resources [21].

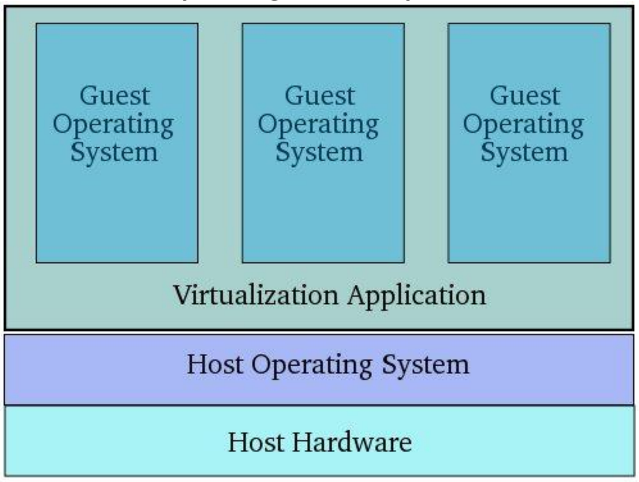

Figure 3**Error! No text of specified style in document.**: Virtualization System

In the architecture, one physical machine is branched into multiple virtual machines. Host OS/VMM (Virtual Machine Monitor) helps virtual machines by allocating necessary resources. The guest OS in turn provide the platform to perform required tasks by corresponding virtual machines and also providing the high level of abstractions such as file access and network support to applications running on the virtual machines [21]. The fact is, virtual machine (VM) serving as logical machine have almost the same architecture as a real host machine, along with an operating system running in it [22].

Benefits of Virtualization:

- Provides great flexibility for both Users and Administrators.
- Allows users to create, save, copy, share, migrate, read & modify, & rollback the execution state of machine.
- Cost is reduced.

- Lower Energy consumption.
- Reduced CO2 emission.

1.2. Energy Saving Strategies on VMs

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 [23] and extended in Mao and Humphrey [24], 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 [25], [26]. 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 [27] and are commonly used.

II. LITERATURE

Nathuji et al. [28] have proposed some techniques to address the with the isolated and self-reliant operations in virtual machines in an online power management infrastructure. These techniques were intended to manage various power management approaches pertaining to virtual machines. The power management policies are deployed with the help of virtual power, which represent the soft versions of the hardware power state. They also provide tools for channels, Virtual Power Management (VPM) state, mechanisms, rules, etc, and helps in mapping the soft power state with the actual variations in the virtualized resource.

There can be certain situation when one physical machine(PM) will be overloaded due to heavy service requests. In such case, the load must be dynamically transferred to different physical machine, which has lower load, without any interruption to the users. This particular technique is termed as migration or virtual machine(VM) migration, in which, the VM will be shifted from one PM to another PM [29][30].

X. Wang et al. [31] introduced a new technique which manages power and energy in datacenters with the help of sustainable energy resources and transforming it to green-aware. The main objective of this research is to make use of green energy as much as possible without hampering the Quality of Service of the application which are running in the datacenters. With the help of energy-aware techniques, the authors have devise an overall optimizing problem, and it is solved by using combined approach of heuristics and statistical searching. Upon analyzing the simulation results, it can be seen that the green energy utilization is greatly enhanced and therefore, it can help the vendors to gain better profits.

Singh A K, Shaw SB [32] propositioned a new technique for VM migration based on forecasting method. In this scheme, the system will check during the server and resource overload, and decides whether migration is required or not. This decision is based on the present as well as future load. The future load is predicted by using time-series forecasting method.

From various sources, it is evident that electricity price forecasting has overwhelmed the research community. On the contrary, few research book and journal suggest that the studies conducted so far is not well established yet. After filtering through various publications, only three books can be found on price prediction.

1. Shahidehpour et al. [33] describes the fundamentals concept of electricity pricing and forecasting based on neural networks. Performance evaluation was conducted by considering volatility, price formation and exogenous variables.
2. Weron R [34] postulates an outline of modeling approaches which focuses on practical applications of statistical approaches for day-ahead forecasting using ARMAX, GARCHtype, ARMA-type, regime-switching, etc. Furthermore, discussion is laid on interval forecasts and it extends to quantitative stochastic models for derivatives pricing with the help of Markov regime-switching and jump-diffusion models.
3. Zareipour [35] (in the author's Ph.D. Thesis from 2007) illustrates a review of linear time series models, such as, ARX, ARIMA, ARMAX and nonlinear models like, neural networks, regression splines, etc. with the help of these techniques, hourly prices of Ontario power market are forecasted.

Amjady and Hemmati [36] justified the necessity of review problems related to EPF, short-term price forecasts, and suggested few proposals for the purpose of such predictions. They claim that time series techniques such as, ARIMA, AR, GARCH are usually effective only in the regions where the frequency of the data is low, like weekly patterns. Additionally, they promoted the implementation of artificial intelligence and other hybrid approaches like, fuzzy regression, neural networks, cascaded architecture of neural networks, fuzzy neural networks and committee machines, etc., as these techniques are proficient enough to track the hard nonlinear behaviors of load and price signals on hourly basis. Amjady et. al. [37] conducted a brief review on EPF methods and later shifted the attention towards techniques based on artificial intelligence, especially hybrid forecast engines and feature selection techniques.

Prevailing research and developments has provided credible solutions pertaining to several characteristic drawbacks of the automatic Service Level Agreement (SLA) assurance in cloud datacenter. For instance, Ref [38] propositions a dynamic cloud resource adaptation schemes, built on the basis of SLA with the help of popular autonomic system framework, known as MAPE loop [39]. The objective of this scheme is laid on lowering the response time of the multi-tier Web application with respect to service level objective, and at the same time, reducing resource utilization. In

Ref [40], Feedback control theory was introduced for adaptive management of virtualized resources in cloud datacentres. The objective of this adaptive feedback controller is to dynamically tweak and modify the multiple virtualized resources. With such advanced techniques, the issues relating to SLA assurance and self-management was effectively addressed. With the help of dynamic provisioning and placing of the VMs, efficiency of resource utilization is also enhanced. On the downside, these researches failed to address the problems with the energy utility efficiency factor and energy consumption of the physical infrastructure.

As described in paper [41], about 30% of energy spent in data centre, is on maintaining the idle times. Therefore, efficiency of resource and energy utilization can be enhanced by considering the EC factor. Paper [42] presented few techniques to enrich energy efficiency by using Dynamic Voltage/ Frequency Scaling (DVFS) technologies. With this technique, the servers and other hardware in datacentre can be shut down when not in use, or the operating voltage can be scaled down for more efficiency. In another research, the authors suggested that tweaking and making certain modifications to VM density and shutting down the idle servers can improve the efficiency, which can be done by using green cloud framework and power-aware VM scheduling approach [43]. The authors in paper [44] presented a Dynamic Power Management (DPM) schemes on the basis of MAPE-K, which uses local and global control loops. The local loops correlates to each servers in the service center, where as the global loops combines whole service centre. On the downside, the DPM scheme can be used for CPU and hard disk of the physical server, but not the energy utility of the cloud data center, without violating the SLA agreements of the users. The authors in paper [45] have considered issues with the SLA assurance and proposed a technique to reduce the energy consumption of the cloud datacenter by constantly merging, transferring and reallocating the VMs and turn off idle nodes and devices with respect to present objective and service level policies.

In paper [46], an energy efficient virtual machine architecture was proposed for cloud data center. The energy utility of data center can be spontaneously operated and moved in accordance with the VM migration policies and server consolidation regulations. These VM centric policies cane help in lowering the total energy utility of the datacenter, efficiently. The shortcoming of these papers is little upsetting as it does not consider EC management and cloud application level SLA assurance. Generally, the cloud service vendors utilize the resources and deploy VMs for implementing cloud based applications. This helps in offering independent business services through several clusters of VM groups. The overall performance of the cloud based applications are influenced by the workload and performance of individual VMs, even though they incorporate EC management. Therefore, it is essential to manage the overall EC of the cloud based application without causing any violations to application level SLA. The limitations with the SLA assurance schemes based on the VM energy management has forced the researchers to adopt new techniques. The EC control at the cloud application level and service quality assurance measures comprises of scaling the VMs with respect to induced workload, dynamic scaling of VM CPU and Memory, augmenting mapping relation between cloud application and servers, bestowing to energy efficiency metrics of the server.

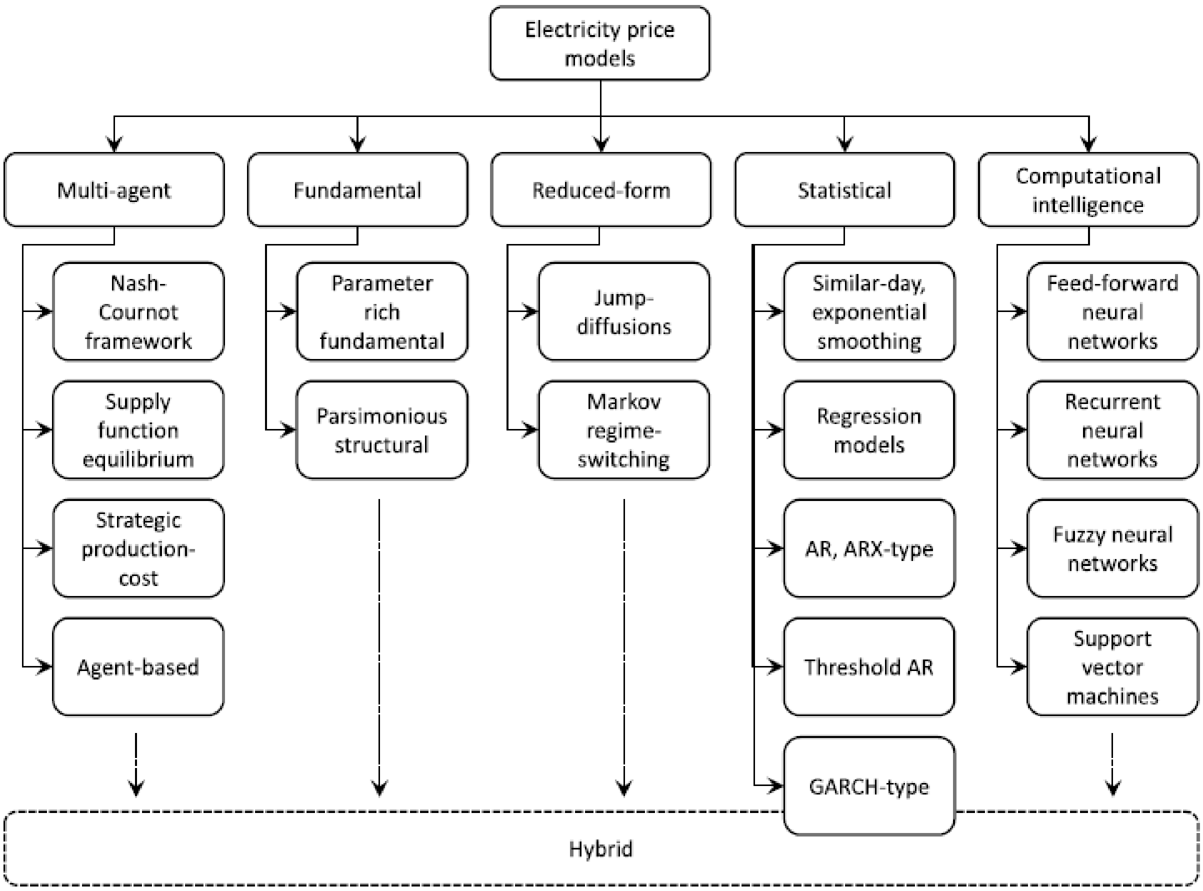

Figure 4: Taxonomy of Electricity Price Modeling Schemes

III. SIGNIFICANCE OF RESOURCE SCHEDULING

Resource scheduling is a process which was designed in several computers based infrastructure domains, for example, grid computing, operating systems, hosting, datacenter management, etc. Resource scheduling is a complex set of mechanisms, which involves scheduling of several activities and allocating the available resources in a cost-effective manner. The scheduling is accomplished through a set of optimal algorithms that resourcefully allocates physical or virtual resources to applications of the users, without impending any additional overhead or cost. The hardware and software resources will be scheduled the application running in cloud, on service request or demand basis. In modern virtualization framework, the VMs will rented and they will be deployed on specific physical hardware which can efficiently utilize the underlying resources.
Proper time and order must be followed during resource allocation to achieve optimal reliability, availability and serviceability. It must not overlook the capacity of available resources, while adhering to the Service Level Agreement (SLA) and providing quick response to application request with minimal permissible latency.
One of fundamental principal of scheduling is the load balancing. Load balancing assures that the given physical is not overloaded, but the load is distributed to other available resources for better performance and to avoid overloading a single hardware. An unbalanced load will cause latency in completion times. The objective of load balancing is to schedule the applications on hosts and monitor them to discover any imbalances in the load.
Babu and Krishna [47] presented a load balancing technique based on Honey Bee behavior (HBB-LB). This technique splits the processors into three groups, namely under loaded, balanced and overloaded. It detects the overloaded servers and it removes the VMs, which will be allotted to any other server of under loaded group. The source server will be moved to balanced group, based on some predefined criteria.

3.1 Need of Resource Scheduling in Cloud

There are several benefits of resource scheduling, and it is important to achieve better efficiency and performance. Some of the essential advantages of proper resource scheduling are,

1) There are no restrictions to location and connection medium.
2) Since, the application and data are hosted our internet, they can be accessed anywhere in world, anytime, thus adhering to on-demand service policy and access.
3) The fundamental advantage of resource provision is that, the users are not required to deploy or install software or hardware to host the computing resource and it doesn’t require constant monitoring and adjusting according to the service request.
4) The clients and users can be benefited by cutting down on deploying hardware and software as they can be rented out cheaply via Infrastructure-as-a-Service (IaaS) cloud model.
5) The vendors of cloud computing can share their resources over the internet during times of resource shortage through resource pooling and avoiding starvation.

3.2 Constraints of Resource Scheduling in Cloud

1) As clients obtain the hardware and resources on rent basis, the true ownership of the resources remain with the vendor. Hence, physical access to the servers are not possible to users, and they won’t have any control over it.
2) There is a risk of data transfer if the users decide to relocate from one datacenter to another. The data migration from current location to new location is very troublesome.
3) User cannot connect optional peripheral devices to the cloud. Some additional application requires administrative access.
4) Since the data is transferred through internet, there is a chance of phishing attacks or malware attacks which is a threat to the security.

In order to overcome the above issues and several other problems, many researchers and engineers have made attempts to make cloud computing secure and hassle-free. Over the past few years several techniques were developed for the same. These techniques considered different parameter to enhance cloud based services. The table 1 below describes the comparison various existing techniques for cloud computing.

Table 1: Comparison of Existing Energy-Aware Scheduling Techniques

Authors and Year	**Techniques Used**	**Parameters**	**Advantages**	**Disadvantages**
Yue Gao Ming Hsieh, Gupta S.K., Yanzhi Wang [48] 2014	The fault tolerant cloud scheduling framework with two phases: Static scheduling and dynamic scheduling.	Application Index, Slack and Replication factors.	CSP boosts high error coverage and fault tolerance assurance while reducing global energy costs under deadline limits of users.	No guarantee to meet the deadlines. Compatibility among the VMs of the same machine is not assured.
Youwei Ding, Xiaolin Qin, Liang Liu, Taochun Wang,	Energy efficient scheduling scheme with EEVS in support of DVFS.	Amount of virtual machines and ratio of Performance and	Total energy consumption of the cloud is reduced. Optimal performance	Not suitable for network-intensive or I/O-intensive VMs. The power usage and

[49] 2015		Power.	power ratio is boosted	execution time and are overlooked.
Zhen Xiao, Weijia Song , Qi Chen, [50] 2013	Green Computing with Hotspot Migration. Skewness Algorithm to estimate Server usage and resource allocation.	Temperature of the hot spot, overloaded resources in server and hot threshold for the resources.	Better server utilization with overload deterrence. Green computing saves energy.	Since the assessment of resource allocation in VMs are based on the predicted demands, the efficiency depends on the accuracy of the forecast and offset from real time requests.
Mehiar Dabbagh, Bechir Hamdaoui, Mohsen Guizani, Ammar Rayes [51] 2015	VM placement with Workload prediction & Consolidation, Resource Over commitment.	Number of Requests and Energy saved over a period of time	Under-utilization issues of resources are addressed by Resource over commitment, which can also conserve energy.	Complex resource management schemes are needed to reduce energy. Resource over commitment results in PM overload.
Riddhi Patel, Hitul Patel, Sanjay Patel, [52] 2015	Energy Aware Best Fit Decreasing (EABFD) algorithm.	Percentage of SLA violations, Number of VM migrations and Energy consumption against Policy.	Considerable reduction in Energy consumption and EABFD with MAD RS policy achieves better optimization of QoS.	Only simulation model is available in Cloud Sim toolkit. Real time deployment is needed.
Ashkan Paya and Dan C. Marinescu, [53] 2015	Energy-aware Scaling Algorithm with Load Balancing and energy-optimal operation routine	Regime of operation versus Number of servers. Low-cost versus high-cost application scaling.	Energy is saved by putting Idle and low-loaded servers to sleep states.	Overhead evaluation is vague. Balancing the computational efficiency and SLA violations is needed

The table illustrates a comparative investigation of several existing resource scheduling schemes in cloud computing by different researchers. The objective is to accomplish automated resources reallocation while emphasis is laid on the optimal performance of datacenters whilst conserving energy. Numerous parameters are considered, like number of VMs, CPU utilization, execution time, cost, service request, available resources, etc. The evaluation suggests that the cloud service providers prefer dynamic resource provisioning with energy aware scheduling as they can help achieve more profit while adhering to increasing demand of the users. It also helps in lowering response time and compliance to Service Level Agreements (SLA).

With several power saving schemes and resource allocation techniques, the cloud computing service is more profitable to the vendors and the users can benefit by less expenses on implementing and maintaining an in-house infrastructure. Although the existing techniques provide better performance and optimal power saving, there are still few drawbacks with each schemes and requires more centralized approach.

IV. COMPARISON

While a single resource scheduling and power saving schemes cannot deliver total performance, a distributed approach is made to control energy consumption and resource allocation across various parameters. Depending on the type of approach and technology, a survey is made on some of the popular technique and comparison is made to bring out the strengths of each scheme. The scheduling schemes are distributed into,

1. Scheduling based on service-level agreement
2. Scheduling based on energy conservation
3. Scheduling based on virtualization
4. Scheduling based on cost-effectiveness

4.1 Comparison of various scheduling schemes based on service-level agreement

Table 2: Comparison of SLA based Scheduling Schemes

Authors	**Year**	**Technique**	**Metrics**	**Environment**	**Strengths**
Lu, K., Yahyapour, R., Wieder, P., Yaqub, E., Abdullah, M., Schloer, B., &	2016	Self-governing SLA management with resource conciliation and allocation.	Energy Efficiency, rate of SLA violation, Cost	CloudSim	Simulation results with extra CPU power of 100 MIPS have shown better result. The average rate of SLA violation is reduced to 10.5%, ROI is decreased to

Kotsokalis, C. [54]					0.68%.
Rajabi, A., Faragardi, H. R., & Yazdani, N. [55]	2013	Online energy conscious resource scheduling framework	rate of SLA violation, Cost, Energy efficiency	not mentioned	Energy consumption and SLA violation rate of is lowered by 21% and 16% better than the genetic algorithm.
Nayak, D., Martha, V. S., Threm, D., Ramaswamy, S., Prince, S., & Fatimberger, G. [56]	2015	Adaptive Scheduler based on negotiation.	Violation rate of SLA.	Hadoop	The analytical study in Hadeep have delivered better Resource Distribution, Priority Basis Allocation, Resources Availability, Guaranteed Service and Negotiation
Yaqub, E., Yahyapour, R., Wieder, P., Jehangiri, A. I., Lu, K., & Kotsokalis, C. [57]	2014	Capacity planning solution for Service Consolidation with compliance to SLA in OpenShift PaaS.	rate of SLA violation, Cost, Energy efficiency	not mentioned	Simulated Annealing provides robust means of implementing most of the policies with yielding algorithms.
García, A. G., Espert, I. B., & García, V. H. [58]	2014	A novel SLA based architecture with WS-Agreement policy for dynamic allocation and management of Cloud resources with automated provision, scheduling,	rate of SLA violation, Cost	OpenNebula, jLinpack	Delivers maximum efficiency at minimum cost under heterogeneous environment
Farokhi, S., Jrad, F., Brandic, I., & Streit, A. [59]	2014	Automatic service selection for multi-cloud environment while adhering to SLA as claimed by SaaS providers.	Service ranking	Java based simulator	This approach selects best plausible set of services satisfying SLA parameters. It helps in maintaining SLA claims of the SaaS providers.

Table 2 depicts the comparison of various scheduling schemes based on service-level agreement. The cloud vendors and their clients will discuss and build a Service Level Agreement (SLA), which principally summarizes the service requirements and assurance by vendor in delivery of a service with proper quality. Violation of the SLA is very critical issue as it is likely to make the users/clients dissatisfied with the service and sooner or later their level of gratification and preference towards the vendor will decline.

One of the noticeable violation of SLA is delayed service or responses which miss the deadline of the requests of the users'. In paper [1], the authors have presented few ways to lower the deadline miss rate; they are,

1. Designing of resource scheduling problem on the basis of mixed Integer Linear Programming (ILP) model.
2. Design and modeling of the Analytics as a Service (AaaS) policy.
3. An algorithm to monitor admission control and resource allocation.

In paper [60], the propositions of the authors can be listed into:

1. A novel domain specific language which is adept to illuminate the QoS-based SLA schemes in accordance with the cloud services.
2. A general approach based on control theory for managing SLA of cloud service.
3. Implement the proposed language and control scheme to ensure SLA in several case studies, for instance, e-commerce service, MapReduce service, locking service, etc.

With the aim of moderating the negative influence of SLA violations and to provide remarkable proactive resource scheduling schemes, the authors in paper [61], focused on tow hidden characteristics of the users, which are, will to pay for service and will to pay for assurance. Moreover, learning automaton based holistic approaches were introduced for approximation of these characteristics.

4.2 Comparison of various scheduling schemes based on energy conservation

The need of Cloud infrastructure is very high and there is no sign of reduction of this need. Due to this ever growing demand of cloud datacenters, the energy consumption has intensely increased, which has raised an alarm and levitated the anxiety of government and service providers [62], [63]. Upon closer investigation of previous literature, it is implied that servers and other hardware in numerous datacenters typically utilize only 30% of their full capacity [64]. Therefore, including an energy reduction schemes during the times of low utilization is a primary goal and hence, it formed an area of interest of most of the researchers.

Table 3: Comparison of Various Scheduling Schemes based on Energy Conservation

Author	Year	Technique	Evaluation Environment	Hardware	Strengths
N.J. Kansal, I. Chana [65]	2015	A Firefly Optimization technique to reduce energy usage, migrations and host saving.	Cloudsim	Memory of host and Bandwidth.	44.39% & 72.34% of Considerable reduction in energy consumption & migration by over 45% and 72%, respectively. Host saving at 35% improvement.
X. Wang et al., [66]	2015	Using green energy for sufficient power usage.	C#, DotNET	Virtualized environment with 40 server and CPU capacity 1500MIPS	Solar energy improves utilization of green energy. Service provider gained high profits. The green energy from solar energy resulted in low power consumption.
Y. Ding et al. [67]	2015	Using deadline constraint to boost processing and reduce energy consumption.	Not mentioned	4 units of PC and server processor	8% improvement in processing capacity and energy consumption reduce by over 20%.
M. Gaggero, L.Caviglione [68]	2016	Using Predictive Control for Energy-Aware Consolidation to achieve balance between SLA violations & Energy saving.	MATLAB	CPU and Memory	Increased power saving by over 61% and SLA violations reduced by 6.6 %.
Shaw SB, Singh AK. [69]	2015	Proactive and reactive hotspot detection technique to reduce the amount of VM migration, SLA violations and energy usage.	cloudsim	Virtualized environment with 800 PMs (HP ProLiant ML110) and 1050 VMs.	Energy consumption reduced by 35%, SLA violations reduced by 64%.

The algorithm presented by Shaw SB et. al. [69] is based on forecasting technique for VM migration. In this scheme, the algorithm checks if the host is overloaded and decided whether there is need for migration of VM. The migration of VM relies on the applied load at present and in the future. The future load is estimated by using time-series forecasting method. The time series forecasting method uses previously obtained values to compute and estimate future values. And in the next step, Exponential Smoothing is applied for final forecasting.

The firefly algorithm was introduced by N.J. Kansal, I. Chana[65], which is an energy-aware technique VM migration. This algorithm calculates the minimum criteria for energy consumption and the VM which satisfy this condition, will be migrated from current active node to another node which consumes the minimal energy. With the help of this technique, the datacenters can conserve more energy and achieve better resource utilization.

The Dynamic Voltage and Frequency Scaling (DVFS) was proposed by Y. Ding et al.[67]. It is a novel energy efficient scheduling algorithm with EEVS support, for addressing the VMs with deadline constraints. This algorithm is based on performance-power ratio (PPR) of the PM and it focuses on optimal frequency of PM required to process the VM. The PM with high PPR signifies high computation capacity and low peak power, and hence VMs can be assigned to this PM to conserve energy.

4.3 Comparison of various scheduling schemes based on virtualization

Table 4: Comparison of Virtualization based Scheduling Schemes

Authors	Year	Technique	Performance Metrics	Environ ment	Strengths
Xu, X., Hu, H., Hu, N., & Ying, W. [70]	2012	Conglomeration of random allocation strategy, Sequence allocation scheme and greedy allocation algorithm for dynamic	Makespan	CloudSim	Minimal Makespan for the greedy strategy, maximum for random strategy and moderate for sequence allocation schemes.

		Virtual machine allocation.			
Quang-Hung, N., Nien, P. D., Nam, N. H., Tuong, N. H., & Thoai, N. [71]	2013	Genetic Algorithm	Energy efficiency and Resource utilization.	CloudSim	energy consumption is saved by over 13% in comparison to the other scheduling algorithms.
Saraswathi, A. T., Kalaashri, Y. R. A., & Padma-vathi, S [72]	2015	Priority based job allocation in VMs	Application execution time	CloudSim	Lower complexity in processing all the jobs in queue, in contrast to deploying new VMs.
Garbacki, P., & Naik, V. K. [73]	2007	Using existing resource capacity, workload requirements and provisioning policies for efficient management of VMs.	Execution time and Throughput,	Harmony grid environm ent	Provides up to 20% more services than the heuristics, which are able to fulfil same amount of requests.
Sahal, R., & Omara, F. A [74]	2014	VM configuration based scheduling.	Cost and Makespan.	CloudSim	The results illustrates the issues influencing configuration of VM in order to handle the preferences of the users and vendors.
Wang, X., Liu, X., Fan, L., & Jia, X. [75]	2013	Design and development of the autonomic and decentralized mechanisms for Dynamic Virtual Machine (DVM) management	Energy efficiency and Rate of SLA violation.	C#n DotNET	DVM strategy lower overall power consumption, which leads to more than 20% energy savings

As seen earlier, the Virtualization technology permits various applications to be allotted on the single Physical Machine in logically isolated VMs. In cloud environment, the Virtual Machine allocation and migration is a challenging task. The principal idea behind VM allocation is to map the VM to a relevant PM with an objective to enhance the performance by saving energy, which can boost the revenues of the vendor [76]. A Virtual Computing Laboratory framework model was developed with the notion of private cloud by expanding the open source IaaS solution Eucalyptus [77]. Using this technique as foundation, a mapping algorithm based on rules was developed for VMs. The algorithm aimed towards atomizing provisioning between VMs and resources of datacenter. The migration of Virtual Machine is of highest significance in implementing resource management schemes for enhancement of performance parameters like utilization of resource, energy consumption and QoS. These parameters are critical for any VM migration in terms of server downtime and network usage, as discussed in ref [78].

In a novel model based on linear programming along with an automatic strategy for VM migration [79], the results illustrated the scope of implementation of this scheme to discover which VM needs to be migrated. The decision is based on several parameter that decides which physical server should host the VM whilst reducing overall operational and migration costs. A Linear Programming formulation and heuristics to manage VM migration was introduced which provides greater priority to the VMs with uniform capacity [80]. In order to obtain proper comparison between this migration management schemes with other popular eager-migration solutions, TU-Berlin and Google data center workloads were used for simulations. The results proved to be fruitful by reducing the amount of migrations with least penalty in the physical servers, given that the migration of VMs with steady capacity is not considered.

Saraswathi et. al [72] proposed Priority based job allocation in VMs, while considering application execution time as fundamental performance metric. In another research, Sahal R et. al [74] proposed VM configuration based scheduling for efficient VM migration, while laying more focus on cost and makespan. Both these algorithms were simulated in Cloudsim. The result achieved were slightly different due to variation of performance metric. The former technique achieved lower complexity in operating all the job request, the latter one made efficient configuration VM which meets the users' requirement. Both these technique uses different approaches but they both equally contribute to enhancing the resource utilization and provide better scheduling.

Some authors [78] have made an attempt to determine suitable time for VM migration, which explores the application level data, in combination with the hi-tech system level metric. The result of this research has produced an application-oriented live migration model. Various experiment were run on real time applications to stimulate the VM live migration. The result have shown a noteworthy reduction in the network operating cost up to 41% and the live migration time was reduced to 63%.

4.4 Comparison of various scheduling schemes based on cost-effectiveness

In cloud computing environment, the service providers emphasize on reducing the resource rental costs while still adhering to workload and service demands of the users. Some techniques such as, reducing application makespan [81], optimized VM placement [74], dynamic resource renting schemes [82] can provide economic and financial benefits to both vendors and users. A cost-aware resource scheduling strategy works on the basis of trade-off between computation, storage and bandwidth [83].

Table 5: Comparison of Scheduling Schemes based on Cost-Effectiveness

Authors	Year	Technique	Performance Metrics	Environment	Strengths
Zhang, X., Liu, C., Nepal, S., Pandey, S., & Chen, J. [84]	2013	Identification of potential data sets which requires encryption for privacy of the users	Cost	OpenStack and Hadoop, U-Cloud with KVM virtualization.	The cost encryption of potential data sets is at least 40% lower than the cost suffered in encrypting all data sets for confidentiality.
Mei, J., Li, K., Ouyang, A., & Li, K. [85]	2015	A holistic double renting system for service vendors	Cost	Analytical solution	The results indicate that the proposed technique is better than Single-Quality-Unguaranteed (SQU) renting scheme with respect to revenue and quality of service.
Sandholm, T., Ward, J., Balestrieri, F., & Huberman, B. A [86]	2015	A QoS-based renting scheme for batch jobs in a multi-vendor OpenStack cluster	Cost	OpenStack Sahara environment	Revenue of cloud service vendors is improved by approximately 40% in comparison to a fixed per-node-period pricing model.
Shi, W., & Hong, B. [87]	2011	Enhanced GA to handle data-intensive services	Execution time and Cost.	MATLAB	Proposed GA shows steady boost in execution time in constrast to to Mixed Integer Programming scheme with the upsurge in number of distinct services.
Li, C., & Li, L. [88]	2013	Enhancing the objectives of cloud users and vendors with efficient resource scheduling scheme.	Resource Utilization and Cost.	Simulation (Environment not mentioned)	Result exhibit 16% more utilization than DRP exponential pricing and 12% more utilization than DRP linear Pricing [28]
Zhou, A., Wang, S., Sun, Q., Zou, H., & Yang, F. [82]	2013	A dynamic virtual resource renting scheme which can dynamically adjust the virtual resource according to price distribution and task priority.	Cost	Java based simulator	The average rental cost with the proposed method is lower and average profit of the vender is higher, in comparison to other traditional revenue based algorithms.

Some authors have formulated few approaches to reduce the rental costs while not compromising on the QoS [89], which are given below:

1. An Online Cost-efficient Scheduling (OCS) uses priority function to estimate the priority of the service requests. In this scheme, the priority function is based on the projected VM speed required to respond and process a request. Therefore, faster VM will be assigned with higher priorities requests.
2. The Dynamic Resource Planning (DRP) will turn off unused and unwanted VM instances prior to the next billing cycle. The decision is based on prediction made by using everyday usage patterns of resources.
3. The CCS works on the principle of workload partitioning, which breakups the batch jobs into a large amount of time slots in accordance with the available resource capacity and instantaneous pricing.

V. CONCLUSION

Cloud Computing, with its on-demand and shared model, it has offered cheap and competent solution to small organizations and individuals. Large benefits to service consumers often bring large risks to service providers. Highly overloaded datacenters barely provide expected QoS to the consumers. With the help of information presented in literature survey, the challenges and hurdle faced by the cloud datacenters, such as customer satisfaction, energy consumption, profitability, etc, are highlighted. Merely load balancing techniques or VM migration techniques can't benefit both service providers and users. It's inevitable for any SLA compliant system to maintain agreement in any circumstances. Some of the most popular techniques which can address the issues of the cloud computing vendor are described. A comparative analysis is made on these research papers to provide a basic outline of the pros and cons of various techniques. With the help of this data, further improvements can be attained in datacenters and cloud environments.

REFERENCES

[1] Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal, James Broberg, Ivona Brandic, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility", Future Generation Computer Systems 2009

[2] Simon Ostermann, Alexandru Iosup, Nezih Yigitbasi, Radu Prodan, Thomas Fahringer and Dick Epema, "A Performance Analysis of EC2 Cloud Computing Services for Scientific Computing", Cloudcomp, LNICST, 2010.

[3] Mladen A. Vouk, "Cloud Computing – Issues, Research and Implementations", Journal of Computing and Information Technology - CIT 16, 2008

[4] Rajkumar Buyya, Chee Shin Yeo, and Srikumar Venugopal, "Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities", The University of Melbourne, Australia, 2005

[5] Kim J-K, Siegel HJ, Maciejewski AA, Eigenmann R. Dynamic resource management in energy constrained heterogeneous computing systems using voltage scaling. IEEE Trans Parallel Distrib Syst 2008.

[6] Goiri I, Oregui LM, Garcia-Rodriguez A. Use of chitosans to modulate ruminal fermentation of a 50:50 forage-to-concentrate diet in sheep. J Anim Sci. 2010.

[7] Nathuji R, Kansal A, Ghaffarkhah A. Q-clouds: managing performance interference effects for qos-aware clouds. In: Proceedings of the 5th European conference on Computer systems (EuroSys 2010). Paris, France; 2010.

[8] V. Kherbache, E. Madelaine, and F. Hermenier, "Planning Live-Migrations to Prepare Servers for Maintenance," in Euro-Par: Parallel Processing Workshops. Springer, 2014.

[9] IBM, "The Benefits of Cloud Computing - A New Era of Responsiveness, Effectiveness and Efficiency in IT Service Delivery," Dynamic Infrastructure, July 2009.

[10] L. Barroso and U. Holzle, "The Case for Energy-Proportional Computing," in Computer, Vol. 40, No. 12, pp. 33-37, 2007.

[11] L. Barroso, J. Clidaras, and U. Holzle, "The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines," Second Edition, Morgan & Claypool Publishers, 2013.

[12] R. Buyya, et al., "Energy-Efficient Management of Data Center Resources for Cloud Computing: A Vision, Architectural Elements, and Open Challenges," in Proc. of the 2010 International Conference on Parallel and Distributed Processing Techniques and Applications, Las Vegas, NV, USA, 2010.

[13] J. L. Berral, et al., "Towards energy-aware scheduling in data centers using machine learning," in Proc. of the 1st International Conference on Energy-Efficient Computing and Networking, Passau, Germany, 2010.

[14] M. Hauck, et al., "Towards Performance Prediction for Cloud Computing Environments based on Goal-oriented Measurements," in in Proc. CLOSER, 2011

[15] L. A. Barroso and U. Holzle, "The case for energy-proportional computing," Computer, vol. 40, no. 12, pp. 33–37, 2007.

[16] A. Greenberg, J. Hamilton, D. Maltz, and P. Patel, "The Cost of a Cloud: Research Problems in Data Center Networks," in ACM SIGCOMM Computer Communications Review, Vol. 39, No. 1, pp. 68-73, 2008.

[17] Jonathan Koomey, "Growth in Data Center Electricity Use 2005 to 2010," A Report by Analytical Press, Completed at the Request of The New York Times, 2011.

[18] L. Barroso, J. Clidaras, and U. Holzle, "The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines," Second Edition, Morgan & Claypool Publishers, 2013.

[19] L. Barroso and U. Holzle, "The Case for Energy-Proportional Computing," in Computer, Vol. 40, No. 12, pp. 33-37, 2007.

[20] Yunfa Li, Wanqing Li, Congfeng Jiang, "A Survey of Virtual Machine System: Current Technology and Future Trends" in 2010 Third International Symposium on Electronic Commerce and Security, Pg 332-336.

[21] Sandeep Kaur, Prof. Vaibhav Pandey, "A Survey of Virtual Machine Migration Techniques in Cloud Computing", ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online), Computer Engineering and Intelligent Systems, Vol.6, No.7, 2015.

[22] Divya Kapil, Emmanuel S. Pilli and Ramesh C. Joshi, "Live Virtual Machine Migration Techniques: Survey and Research Challenges" In 2013 3rd IEEE International Advance Computing Conference (IACC), Pg 963-969.

[23] 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.

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

[25] X Fan, W D Weber and L A Barroso, Power Provisioning for a Warehouse-Sized Computer, ACM SIGARCH Computer Architecture News, 2007
[26] 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.
[27] E N M Elnozahy, M Kistler and R Rajamony, Energy-Efficient Server Clusters, Power-Aware Computer Systems, Springer, 2003, p. 179-197
[28] R. Nathuji, et al., "Exploiting Platform Heterogeneity for Power Efficient Data Centers," in Proc. of the IEEE International Conference on Autonomic Computing Washington, DC, USA 2007, pp. 5-15.
[29] Pankajdeep Kaur,Anita Rani, “Virtual Machine Migration in Cloud Computing”, International Journal of Grid Distribution Computing, Vol. 8, No.5, (2015), Pg 337-342.
[30] Bharti Wadhwa, Amandeep Verma, “Energy and Carbon Efficient VM Placement and Migration Technique for Green Cloud Datacenters” In 2014 Seventh IEEE International Conference on Contemporary Computing (IC3), Pg 189-193.
[31] X. Wang et al.,A green-aware virtual machine migration strategy for sustainable datacenter powered by renewable energy, Simulat. Modell. Pract. Theory (2015),
[32] Shaw SB, Singh AK. Use of proactive and reactive hotspot detection technique to reduce the number of virtual machine migration and energy consumption in cloud data center. Comput Electr Eng (2015),
[33] Mohammad Shahidehpour, H.Yamin, and Zuyili, “Market Operations in Electric Power Systems: Forecasting, Scheduling and Risk Management”. Wiley, New York, 2002.
[34] R. Weron, “Forecasting wholesale electricity prices: A review of time series models, in "Financial Markets: Principles of Modelling, Forecasting and Decision-Making", Institute of Mathematics and Computer Science, Wrocław University of Technology, 2008.
[35] H. Zareipour, K. Bhattacharya, C.A. Canizares, “Electricity market price volatility: the case of Ontario” 2007
[36] Amjady, N. and Hemmati, M. “Energy Price Forecasting—Problems and Proposals for Such Predictions” IEEE Power and Energy Magazine, 2006
[37] D Huang, H Zareipour, WD Rosehart, N Amjady, “Data mining for electricity price classification and the application to demand-side management”, IEEE Transactions on Smart Grid 3 (2), 2012
[38] Hien Nguyen Van, Frederic Dang Tran, Jean-Marc Menaud, “SLA aware Virtual Resource Management for Cloud Infrastructures,” IEEE Ninth International Conference on Computer and Information Technology, 2009.
[39] Waheed Iqbal, Matthew N. Dailey, David Carrera, “SLA-Driven Dynamic Resource Management for Multi-tier Web Applications in a Cloud,” 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, 2010.
[40] Qiang Li, Qinfen Hao, Limin Xiao, Zhoujun Li, “Adaptive Management of Virtualized Resources in Cloud Computing Using Feedback Control,” The 1st International Conference on Information Science and Engineering, 2009.
[41] Uddin, M., Rahman, A, “Server Consolidation: An Approach to Make Data Centers Energy Efficient & Green,” Int. Journal of Scientific & Engineering Research, Vol. 1, Issue 1 (2010).
[42] Von Laszewski. G., Lizhe Wang, Younge. A.J, Xi He, “Power-aware scheduling of virtual machines in DVFS-enabled clusters,” Cluster Computing and Workshops, 2009. CLUSTER ‘09. IEEE International Conference. 2009. pp:1-10.
[43] Younge. A.J, von Laszewski. G, Lizhe Wang, Lopez-Alarcon, S, Carithers, W, "Efficient Resource Management for Cloud Computing Environments," Green Computing Conference, 2010 International. 2010. pp:357-364.
[44] Ioan Salomie, Tudor Cioara, Ionut Anghel, Daniel Moldovan, “Energy Aware Adaptation Methodology for Improving the Service Centers Energy Efficiency,” unpublished.
[45] Anton Beloglazov, Rajkumar Buyya, “Energy Efficient Allocation of Virtual Machines in Cloud Data Centers,” 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, 2010.
[46] Kejiang Ye, Dawei Huang, Xiaohong Jiang, Huajun Chen, Shuang Wu,”Virtual Machine Based Energy-Efficient Data Center Architecture for Cloud Computing: A Performance Perspective,” GREENCOMCPSCOM '10 Proceedings of the 2010 IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing, 2010.
[47] L D Babu and P V Krishna, Honey bee behavior inspired load balancing of tasks in cloud computing environments, *Appl Soft Comput*, 2013
[48] Yue Gao Ming Hsieh, Gupta, S.K., Yanzhi Wang “An Energy-Aware Fault Tolerant Scheduling Framework for Soft Error Resilient Cloud Computing Systems”, IEEE 2014.
[49] Youwei Ding, Xiaolin Qin, Liang Liu, Taochun Wang, “Energy efficient scheduling of virtual machines in cloud with deadline constraint”, Science Direct 2015.
[50] Zhen Xiao, Weijia Song , Qi Chen, “Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment”, IEEE 2013.
[51] Mehiar Dabbagh, Bechir Hamdaoui, Mohsen Guizani, Ammar Rayes, “Towards Energy-Efficient Cloud Computing: Prediction, Consolidation, and Over commitment”, IEEE 2015
[52] Riddhi Patel, Hitul Patel, Sanjay Patel, “Quality of Service Based Efficient Resource Allocation in Cloud Computing”, IJTRE 2015

[53] Ashkan Paya and Dan C. Marinescu:, "Energy-aware Load Balancing and Application Scaling for the Cloud Ecosystem", IEEE 2015.
[54] Lu, K., Yahyapour, R., Wieder, P., Yaqub, E., Abdullah, M., Schloer, B., & Kotsokalis, C. Fault-tolerant Service Level Agreement lifecycle management in clouds using actor system. Future Generation Computer Systems, 54, 2016,
[55] Rajabi, A., Faragardi, H. R., & Yazdani, N. Communication-aware and energy-efficient resource provisioning for real-time cloud services. In Computer Architecture and Digital Systems (CADS), 2013 17th CSI International Symposium, IEEE. 2013,
[56] Nayak, D., Martha, V. S., Threm, D., Ramaswamy, S., Prince, S., & Fatimberger, G. Adaptive scheduling in the cloud—SLA for Hadoop job scheduling. In Science and Information Conference (SAI), IEEE. 2015
[57] Yaqub, E., Yahyapour, R., Wieder, P., Jehangiri, A. I., Lu, K., & Kotsokalis, C. Metaheuristics-based planning and optimization for sla-aware resource management in paas clouds. In Proceedings of the IEEE/ACM 7th International Conference on Utility and Cloud Computing, 2014
[58] García, A. G., Espert, I. B., & García, V. H. SLA-driven dynamic cloud resource management. Future Generation Computer Systems, 31, 2014,
[59] Farokhi, S., Jrad, F., Brandic, I., & Streit, A. HS4MC–Hierarchical SLA-based Service Selection for Multi-Cloud Environments. In: CLOSER 2014
[60] Serrano D, Bouchenak S, Kouki Y, de Oliveira Jr FA, Ledoux T, Lejeune J, Sopena J, Arantes L, Sens P. SLA guarantees for cloud services, In Future Generation Computer Systems. In: Future Generation Computer Systems, 2016
[61] Morshedlou, H., & Meybodi, M. R. Decreasing impact of sla violations: a proactive resource allocation approach for cloud computing environments. Cloud Computing, IEEE Transactions, 2 (2), 2014
[62] Dong, D., & Herbert, J. Energy efficient vm placement supported by data analytic service. In Cluster, Cloud and Grid Computing (CCGrid), 13th IEEE/ACM International Symposium, 2013
[63] Uddin, M., Darabidarabkhani, Y., Shah, A., & Memon, J. Evaluating power efficient algorithms for efficiency and carbon emissions in cloud data centers: A review. Renewable and Sustainable Energy Reviews, 51, 2015,
[64] Barroso, L.A., & Hölzle, U. The case for energy-proportional computing. Computer 12, 2007
[65] N.J. Kansal, I. Chana, "Energy-aware Virtual Machine Migration for Cloud Computing - A Firefly Optimization Approach" ,J Grid Computing(2015).
[66] X. Wang et al.,A green-aware virtual machine migration strategy for sustainable datacenter powered by renewable energy, Simulat. Modell. Pract. Theory (2015).
[67] Y. Ding et al.,Energy efficient scheduling of virtual machines in cloud with dealine constraint, Future Generation Computer Systems(2015),
[68] M. Gaggero, L.Caviglione "Predictive Control for Energy-Aware Consolidation in Cloud Datacenters" , In IEEE Transactions On Control Systems Technology, Vol. 24, NO. 2, March 2016
[69] Shaw SB, Singh AK. Use of proactive and reactive hotspot detection technique to reduce the number of virtual machine migration and energy consumption in cloud data center. Comput Electr Eng (2015)
[70] Xu, X., Hu, H., Hu, N., & Ying, W. Cloud task and virtual machine allocation strategy in cloud computing environment. In Network Computing and Information Security, 2012.
[71] Quang-Hung, N., Nien, P. D., Nam, N. H., Tuong, N. H., & Thoai, N. A genetic algorithm for power-aware virtual machine allocation in private cloud. In Information and Communication Technology 2013.
[72] Saraswathi, A. T., Kalaashri, Y. R. A., & Padmavathi, S. Dynamic resource allocation scheme in cloud computing. Procedia Computer Science, 47, 2015,
[73] Garbacki, P., & Naik, V. K. Efficient resource virtualization and sharing strategies for heterogeneous grid environments. In Integrated Network Management, IM'07. 10th IFIP/IEEE International Symposium, 2007,
[74] Sahal, R., & Omara, F. A. Effective virtual machine configuration for cloud environment. In Informatics and Systems (INFOS), 9th International Conference, IEEE. 2014, December, PDC-15.
[75] Wang, X., Liu, X., Fan, L., & Jia, X. A decentralized virtual machine migration approach of data centers for cloud computing. Mathematical Problems in Engineering, 2013.
[76] Li, W., Tordsson, J., & Elmroth, E. Virtual machine placement for predictable and time-constrained peak loads. In Economics of Grids, Clouds, Systems, and Services, 2011
[77] Ezugwu, A. E., Buhari, S. M., & Junaidu, S. B. Virtual machine allocation in cloud computing environment. International Journal of Cloud Applications and Computing (IJCAC), 3 (2), 2013
[78] Baruchi, A., Toshimi Midorikawa, E., & Netto, M. A. Improving Virtual Machine live migration via application-level workload analysis. In Network and Service Management (CNSM), 10th International Conference, IEEE. 2014
[79] Park, J. G., Kim, J. M., Choi, H., & Woo, Y. C. Virtual machine migration in self-managing virtualized server environments. In Advanced Communication Technology, ICACT 2009. 11th International Conference, IEEE. 2009,
[80] Ferreto, T. C., Netto, M. A., Calheiros, R. N., & De Rose, C. A. Server consolidation with migration control for virtualized data centers. Future Generation Computer Systems, 27 (8), 2011
[81] Rodriguez, M. A., & Buyya, R. Deadline based resource provisioning and scheduling algorithm for scientific workflows on clouds. Cloud Computing, IEEE Transactions on, 2 (2), 2014

[82] Zhou, A., Wang, S., Sun, Q., Zou, H., & Yang, F. Dynamic Virtual Resource Renting Method for Maximizing the Profits of a Cloud Service Provider in a Dynamic Pricing Model. In Parallel and Distributed Systems (ICPADS), International Conference, IEEE. 2013,

[83] Yuan, D., Cui, L., Liu, X., Fu, E., & Yang, Y. A Cost-Effective Strategy for Storing Scientific Datasets with Multiple Service Providers in the Cloud, 2016.

[84] Zhang, X., Liu, C., Nepal, S., Pandey, S., & Chen, J. A privacy leakage upper bound constraint-based approach for cost-effective privacy preserving of intermediate data sets in cloud. Parallel and Distributed Systems, IEEE Transactions, 24 (6), 2013

[85] Mei, J., Li, K., Ouyang, A., & Li, K. A profit maximization scheme with guaranteed quality of service in cloud computing. Computers, IEEE Transactions, 64 (11), 2015

[86] Sandholm, T., Ward, J., Balestrieri, F., & Huberman, B. A. QoS-Based Pricing and Scheduling of Batch Jobs in OpenStack Clouds, 2015

[87] Wang, L., Shen, J., Luo, J., & Dong, F. An improved genetic algorithm for cost-effective data-intensive service composition. In Semantics, Knowledge and Grids (SKG), Ninth International Conference, IEEE. 2013.

[88] Li, C., & Li, L. Efficient resource allocation for optimizing objectives of cloud users, IaaS provider and SaaS provider in cloud environment. The Journal of Supercomputing, 65 (2), 2013

[89] Malawski, M., Juve, G., Deelman, E., & Nabrzyski, J. Algorithms for cost-and deadline-constrained provisioning for scientific workflow ensembles in IaaS clouds. Future Generation Computer Systems, 48, 2015.