ISSN: 0974-5572
IJCTA
International Journal of
CONTROL THEORY
AND APPLICATIONS
UGC APPROVED JOURNAL
ज्ञान-विज्ञान विमुक्तये
INTERNATIONAL SCIENCE PRESS
Gurgaon, Haryana (India)

International Journal of Control Theory and Applications

ISSN : 0974-5572

© International Science Press

Volume 9 • Number 45 • 2016

Energy Efficient Allocation of Resources in Cloud: State of the Art Survey

Harjit Kaur[a] **Harshpreet Singh**[a] **NishaSethi**[b] **and Rajneesh Randhawa**[c]

[a]*School of Computer Science and Engineering, Lovely Professional University Phagwara, India*

E-mail: mannirakhra27@gmail.com, harshpreet.17478@lpu.co.in

[b]*School of Computer Applications, Lovely Professional University Phagwara, India*

E-mail: nisha.18349@lpu.co.in

[c]*Department of Computer Science, Punjabi University Patiala, India*

E-mail: drrajneeshrandhawa@gmail.com

Abstract : Increase in the performance of the cloud datacenter is one of the main concerns of this paper. With the increase in the performance parameter, the completion of the tasks would be done quickly leading to the decrease in the delay. The energy consumption of the datacenter is one of the major problems in the cloud environment. Thus power management in the cloud datacenter is the main focus in this paper. The variety of techniques adopted by various authors are reviewed. The main parameters that are considered in this paper besides energy are Quality-of-Service (QoS), performance, load-balancing and the cost related to the cloud service providers (CSP). The paper covers a wide variety of energy-efficient algorithms which are used in the cloud environment to curb the consumption of energy in the cloud environment. Two types of allocation strategies are discussed: the placement of the tasks onto the virtual machines (VM) and the allocation of the virtual machines onto the physical machines.

Keywords : *Cloud computing, virtual machine, physical machine, performance, energy-efficiency, resource allocation, performance, QoS.*

1. INTRODUCTION

With the enhancement of processing and storage know-hows and the achievement of the internet, the cloud resources are becoming more economical, prevailing and universally available than ever before[1]. This technological development has steered the new apprehension known as Cloud Computing. In Cloud Computing the resources are considered as the general utilities that can be leased and released through the internet by the users in the on-demand and pay-as-you-go manner. The responsibility of the cloud providers in the cloud environment is divided into two main tasks: the infrastructure provider and the service provider. The infrastructure providers manage the cloud platforms by lending out the resources in an on-demand and pay-as-you-go basis. The service providers lease the resources from different infrastructure providers to quench the demands of the end-users [1].

There is a lot of hype and the confusion around the globe regarding the definition of the cloud computing. Due to the absence in the standardization of cloud computing definition, the NIST (National Institute of Standards and Technology) definition is considered as the accepted one[1]. The NIST definition of cloud computing:[2] "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (*e.g*., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".

The cloud environment has changed the approach the computing infrastructure is used[3]. Some of the features like the elasticity (cloud services are provisioned and released elastically), metered services (the users are charged by the quantity of resources they have used), scalability (the resources can be dynamically allocated to the tasks which are executed by the *VM* hence the services can be scaled for the business organizations) and accessibility (the cloud services are ubiquitous in nature. Hence are available everywhere around the globe via the internet facility) has made the cloud computing looked-for.

There are three foremost types of services and four deployment models that are provided by the cloud. The services are SaaS (Software as a Service), PaaS (Platform as a Service) and IaaS (Infrastructure as a Service). The deployment models are Public Cloud, Private Cloud, Community Cloud and Hybrid Cloud. The Hybrid Cloud is the combination of Public, Private and Community Cloud. The acceptance and deployment of platforms provided by clouds have many striking benefits like reliability, Quality of Service and robustness[4].

The services provided by the cloud are:

1. **Software as a Service (SaaS):** Business values are provided by the application layer to the customers. The transparency (for the allocation and execution of VMs) from the customers is maintained in an effective way. Consumers are able to access the applications as a service. The applications are executed on the infrastructure that are managed by the vendors of the SaaS.

2. **Platform as a Service (PaaS):** It concentrates on providing the high level services. The computing platform and/or the solution stacks with the operating system, database, web-server and the platform are provided to the customers for the successful execution of the jobs. The only thing that customers have to manage are the applications that is deployed on the cloud.

3. **Infrastructure as a Service (IaaS):** Various types of resources needed for computing the jobs or the tasks are provided to the customers like the storage, network and computational resources. The resources help the customers to deploy their software including operating system easily on the cloud.

There are various deployment models for cloud:

1. **Public Clouds:** They are managed by the third parties. The main benefit is that they are larger than the private clouds, thus the organizations can shift their data from their private clouds to the storage offered by the public clouds, even on the temporary basis. They provide many services on demand to the customers and are cheaper than other deployment models.

2. **Private Clouds:** They are the set of standardized computing resources owned by the organization. Private clouds are used for the security purposes. These types of clouds are managed by the organizations itself. They provide the limited amount of services to be used by the client.

3. **Hybrid Clouds:** It has the features of both the private and the public cloud that are bounded together but are unique at the same time.

4. **Community Clouds**: The infrastructure is share amongst the different organizations, all belonging to the same community with a shared common goal amongst them. This infrastructure is either managed internally by the community itself or externally by the third party organization.

Virtualization is the chief technology because of which the cloud computing is possible. Virtualization leads to the efficient utilization of the computing power, which is one of its desirable features[5]. The cloud datacenter is benefited from the Virtualization technique, which allows the resources of the solitary server to be divided into various autonomous implementation environments installed on the Virtual Machine[6]. This results in provisioning the Virtual Machine on the less number of Physical Machines, thus providing an increase in the CPU utilization, reduction in both the cost (hardware and the operational) and the physical space, and an increase in the availability and flexibility of the cloud services.

1.1. Cloud Broker

It is a third party union or an individual that acts as an intercessor between the cloud service provider and the cloud service consumer. It provides the single interface for dealing with multiple *CSPs*. Its main aim is to save the time of purchaser by negotiating different services that are available across multiple cloud vendors. The service request is handled by the cloud service broker that employs the broker policy to grant the services to the purchaser and then sends the requests for the computation of the task to the appropriate datacenter. The cloud broker is the one who is responsible for the allocation of Virtual Machines to the customers for the completion of their task.

1.2. Physical Machine

Physical Machines (PM) are the physical hardware having their own CPU, memory, storage and networking requirements. Physical machines can have any number of Virtual Machines depending on their configuration and constraints put up by the hypervisor. So, there are different levels of virtualizations on physical hardware.

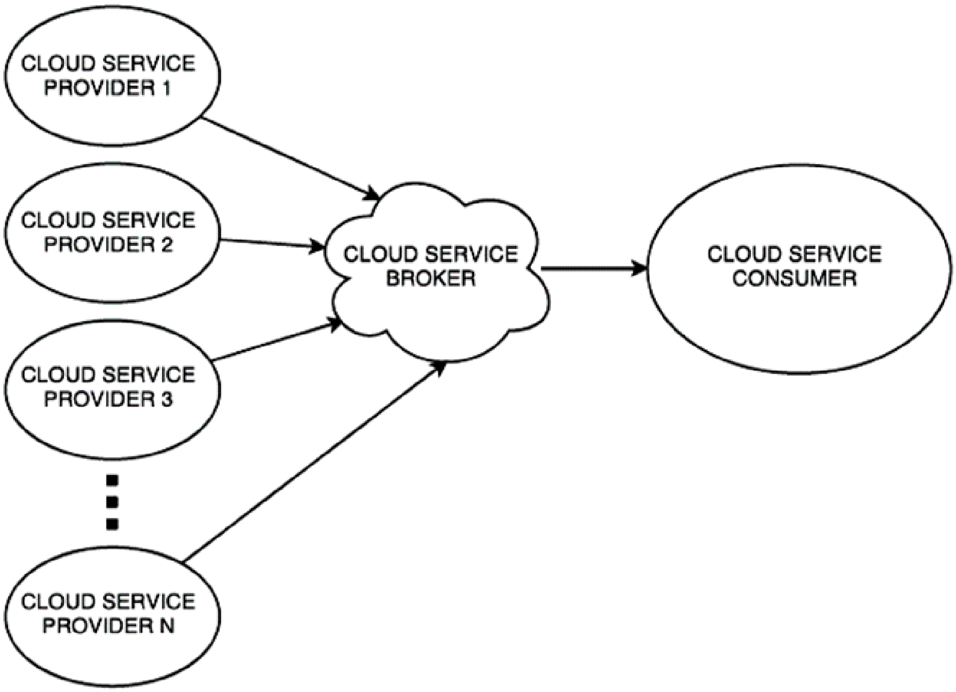

Figure 1: Cloud Broker

Physical machines are viewed as the package and the VM's are viewed as the boxes. The main aim of the cloud computing is packing the boxes in the packages. Both the *PMs* and the *VMs* have their own resources like CPU, networking, storage and memory.

Let us consider a cloud service provider as CSP and the datacenter as D which are available to the CSP as depicted in Figure.1. Let$d_i \in$ D be the subset of datacenters in and $pm_j \in$ PM be the subset of the Physical Machines PM which includes both on (active) and off (switched off) machines. The parameters of the $pm_j \in$ PM are[7] :

1. Cores $c_r \in$ C : Count of processor cores.
2. RAM $r_r \in$ R: in GB/s.
3. Cpu_Capacity $cc_r \in$ CC the processing power of the CPU core (million instructions per second (MIPS)).
4. Network_Bandwidth $nb_r \in$ NB: the bandwidth requirement of the PM.

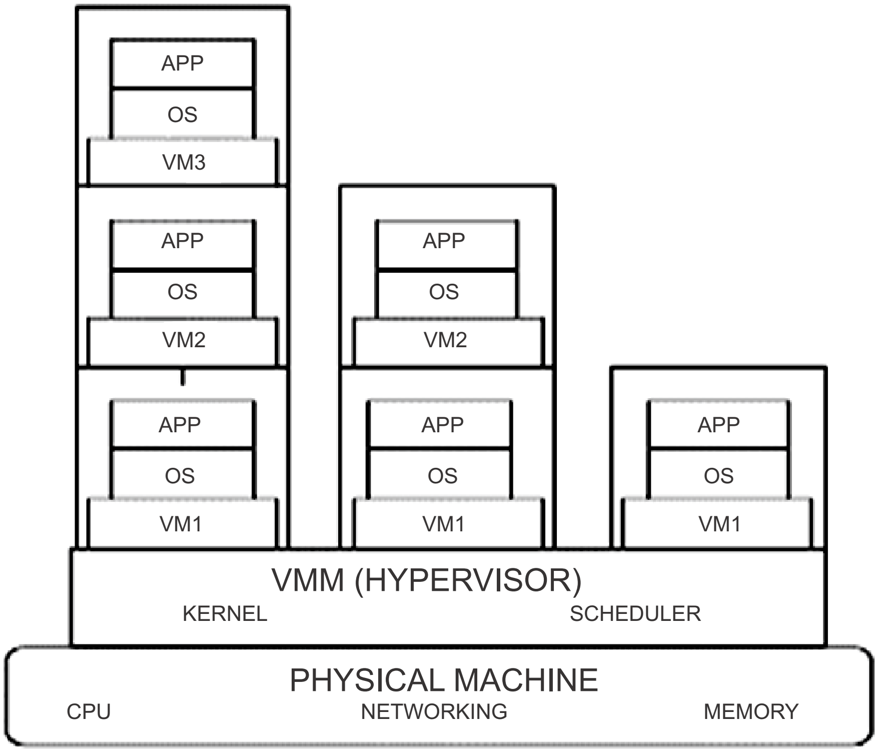

Figure 2: Overview of Datacenter

The machine that manages and schedules the Virtual Machine on the Physical Machines is known as hypervisor or the Virtual Machine Manager *(VMM)* as represented inFigure.2. The *VMM* acts as the intermediary between the *PMs* and the various *VMs*. *VMM* is known as the kernel of the *PM*.

1.3. Resource Allocation

The allocation of resources in the cloud datacenter is an integral and an on-going process. The allocation of resources in the cloud datacenter are done on the basis of the requirements of the customers taken by the *CSPs*. The *CSP* first allocates the servers to the *VM* tenants, then the allocation of the network bandwidth is done, later on fault-tolerance strategies are introduced in the *VM*[8].

The power of Virtualization and cloud computing has resulted in the decoupling of application and services because of which there is an increase in the performance of *VMs*. The decoupling has also improved the allocation of the resources on the *VMs* in an on-demand fashion. Hence, it provides the flexibility to the system as a whole.

VM allocation to the *PM* is an important part of the resource allocation strategy in the cloud environment. Allocation of resources is done according to the changing needs of the *VM* leading to the elastic nature of the allocation strategy, hence enhancing the flexibility feature of the cloud environment[9]. During the placement of *VM* to the *PM* various resources like CPU, bandwidth, servers, RAM, storage capacity, etc. are used. Firstly, the resources are allocated to the *VMs* according to the requirements listed in the Service-Level-Agreement *(SLA)* and then the *VMs* are treated as resources which are further allocated to the *PMs* in the cloud. Both the computational and networking resources are taken into consideration when the allocation of the resources are done either to the *VM* or to the *PMs*[10]. The load on the cores of the Virtual Machines need not be the same for all the *VMs* soavailable.

The resources need to be allocated in an efficient manner so that the energy-efficiency of the *VM* and the *PM* is improved, in-turn leading to an increase in the performance and benefits in the cost to both the *CSPs* and to the customers.

1.4. Virtual Machine Allocation

VM allocation is one of the most important challenges faced by the researchers in the cloud environment. The placement of *VM* onto the *PM* (also known as the servers or the hosts) is known as Virtual Machine provisioning or Virtual Machine allocation. The approaches for the allocation of the *VM* have different purposes: initial placement, throughput maximization, consolidation[11] Service-Level-Agreement *(SLA)* satisfaction versus provider operating cost minimization, etc.[1213].

Firstly, the allocation of the *VM* to the hosts is done after the proper mapping of the resources like CPU utilization, network bandwidth, operating system, storage and application requirements. Secondly, the mapping of the resources, the checking of the availability of the resources and the capacity of the *PM* in order to host the *VM* is done. This also includes the consideration of the constraints set up by the hypervisor for the Physical Machine to host the definite number of Virtual Machines.

The strategy for the allocation of the *VM* to the *PM* can be static or dynamic in nature. In the dynamic allocation, the VM's are allocated to the hosts dynamically i.e. according to the changing need of the customers about the resources. Every time the needs for the resources by the customer changes, correspondingly the allocation of *VM* onto the *PM* is done again according to the mapping and availability criteria[6]. The dynamic allocation strategies are the need of the hour because the requirements of the customers keep on changing with time. In the static allocation, the algorithm is employed to search the *PM* which can host the *VM*. The static allocation once done cannot be changed with the changing needs of the resources.

The placement of the VM onto the PM is the main objective of this report. The data center comprises a set of m Virtual Machines say $vm = \{vm_1, vm_2 \ldots.vm_m\}$ and the set of n Physical Machines say $= \{pm_1, pm_2, \ldots, pm_n\}$. Each vm andpm has their own resources like CPU cores, RAM, MIPS (millions instruction per second), size and network bandwidth say c, r and b respectively. Then the resource set of vm vm_j *i.e.* $\{c_j, r_j, m_j, s_j, b_j\}$ and the resource set of PM be pm_k *i.e.* $\{c_k, r_k, m_k, s_k, b_k\}$ The placement of the vm_j on the pm_k is done only if the decision variables say a_{jk} and b_k are true *i.e.* $a_{jk} = 1$ and $b_k = 1$ where $0 < j < m$, $0 < k < n$, a_{jk} = mapping of the *vm* on the *pm* on the bases of the resources and b_k = the hosting done by the pm for the various *vm* on the basis of load balancing and availability factors.

2. LITERATURE REVIEW

This section of the papers presents a brief literature review of various VM allocation methods. The papers are reviewed keeping energy, cost, load balancing, Quality of service and performance for allocation approaches as parameters of evaluation.

Yousefian & Zadnavin[14] proposed the method for scheduling the *VMs* which enhances the resource utilization and hence increase the inflow of revenue of the *CSPs*. The scheduling mechanism automate the cost based on scheduling *VMs* on the hosts. At the time of scheduling the low-cost servers have the highest priority and are selected for hosting the *VMs*. The final cost of computation can be varied as the *VMs* can be placed onto different servers. The break-even cost calculation method is used to calculate the costs for providing the services. The method calculates the cost for each and every request for the placement of the *VMs* onto the servers. This method increased the income of the service provider by 8% and the resource utilization by 3% as compared to the traditional approaches. The energy utilization of the scheduling mechanism was not considered by the authors and is referred as a further scope.

Mann n.d.[15] analyzed the Minimization of Migration (MM) and the Modified Best Fit Decreasing (MBFD) methods for the consolidation of VM's in the cloud datacenter in order to utilize less amount of energy hence leading to energy efficient consolidation. The comparison of the MM and MBFD done in order to find out which algorithm suits best according to the specific problem. The main objective of this paper is to consolidate the *VM* onto the minimum number of hosts while overloading the *PMs* on which the *VMs* are placed. MM is optimal in regards to the energy consumption when compared to the MBFD. If the capacity of the hosts are different then the MM proves to be optimal as compared to the MBFD regarding the number of available hosts. If the capacity of all the hosts are same then both MM and MBFD behaves optimally. The bin-packing method has a considerable impact on the placement of the *VM* on hosts. The detailed study of the bin-packing method is not done and referred as a future scope.

Kruekaew & Kimpan[16] proposed the ant-bee-colony for the scheduling of *VMs* onto the *PMs* in the cloud environment. According to the proposed method, the collection of *VMs* is considered to be a dynamic resource pool. The routing service request the servers on the basis of the Cloud Management Policies based and the server's load capacity. Scheduling is done based on the load capacity of the servers. The proposed method increased the performance of the scheduling strategy and balanced the work load on the servers resulting in the minimization of the data processing time of the cloud servers. The preemptive scheduling with allocation of heterogeneous task in the cloud was not considered and is referred as a future scope.

Computing[17] proposed an optimized method for the allocation of the requested *VMs* using autonomic computing which maximized the cloud provider's revenue. The proposed method deals with the self-configuring, self-healing, self-optimizing and self-protecting computer systems. It deals with the low level-tasks of the computer system for which it is not possible for the humans to make decisions. The servers are organized as racks, which are further organized as clusters, which are then organized into datacenters. The result of the paper shows that the revenues of the providers can be increased if the proper resource demands of the application are already known before hand, so that the requested *VMs* are allocated closely onto the related servers.

Kumar et al. [18] proposed the modified best-fit algorithm that emphasized the power consumption problem in cloud datacenter. Best-fit method includes the global and local variables. The local variables includes the information about the *VMs* and the global variables collect the information from the local variables to maintain the overview of the resource utilization of the system. The resources like CPU load, RAM and network throughput are the main considerations in the algorithm. This method resulted in the minimization of the power consumption by putting restrictions on the power and data and by efficiently placing the *VMs* onto the *PMs.*

Li et al[19] proposed an algorithm named EAGLE for the energy efficient *VM* placement in the cloud environment with the balanced and optimized used of resource utilization in the datacenter. The proposed method focus on the resource utilization of datacenter in order to increase the performance of datacenter and to lower the cost of data processing. EAGLE can balance the multi-dimensional resource utilization, minimize the number of active servers, and thus lowers the power consumption in the datacenter. The basic idea is to balance the tradeoff between the balancing of the multi-dimensional resource utilization and lowering the number of *PMs* when placing the *VMs* at each time slot. The server with the optimal suitability is selected for the *VM* hosting.

Kumar & Ramachandra[20] proposed a Genetic Algorithm (based on time and energy consumption) and the Dynamic Voltage Scaling (DVS) and Dynamic Voltage Frequency Scaling (DVFS) (for the conservation of power in the datacenter) for optimizing the consolidation of *VMs* on the servers. Energy consumption with respect to the consolidation of *VMs* on the servers includes two parts: part 1 includes the request by the users for the provisioning of *VM* and the placement of *VM* on the servers, part 2 includes the energy considerations regarding the placement of *VMs* on the servers. The proposed methodology resulted in the balancing of load on the servers thus lowering the energy consumption.

N & Hemalatha[21] proposed the reservation method for the efficient placement of *VM* in cloud datacenter. The algorithm RTBBE (Reservation Technique for BIN BECK Entropy) is combined with PR (Polynomial Regression) for the overload detection on the servers. The MUR (Minimum Utilization Rank) method is used for the *VM* selection. The problem of dynamic *VM* consolidation is divided into four sub-problems which are the host under load detection, host overload detection, *VM* selection and the *VM* placement. The PR uses the *VM* Table which consists of the resources like CPU Utilization, Memory, Cost and Power. From the available resource information the PR selects the suitable the *VM*. The MUR takes into consideration the Failure Rate, Maximum Uptime and Minimum per- *VM* resets for the ranking of the *VMs*. The proposed method lowered the energy consumption up to 21.30 kWh when PR is used in combination with MUR.

Sharma & Singh[22] developed an algorithm for the optimization of power consumption and the balancing of load in cloud datacenter. The algorithm for the energy saving (Green Computing) includes various techniques like Round Robin algorithm, Equally Spread Current Execution Algorithm (ESCE) and Throttled Load Balancing Algorithm (TLB). The algorithm for the overall energy consumption in the cloud datacenter includes five components like Power monitoring system, Power optimizing algorithm, Power consumption calculator, Reconfigurable temporal policy of workload management and Feedback loop. The methodology includes first, the creation of Broker, User base and Datacenter. Secondly, the *VM* allocation and migration policies are implemented and the overload detection of the host is done. After the overload detection *VM* Migration policy is again implemented and the new migration map is developed. The proposed algorithm resulted in the less consumption of energy as compared to the Round Robin, Throttled Load Balancing and Equally Spread Current Execution algorithm. Because of the decrease in the number of migrations of *VMs*, the energy consumption in the data center is also reduced. Thus the algorithm promotes Green Computing.

Ajith Singh & Hemalatha[23] proposed the hierarchical clustering technique for the placement of *VM* in the datacenter based on CPU, speed and memory considering each cluster as a single resource. The Bees algorithm states that the cloudlets, *VMs* and the servers are all heterogeneous in nature. The request of task processing is different for different *VMs*. So the allocation of *VM* is done based on different resource requirements like size, memory, storage, etc. For the efficient execution of task the load balancing algorithms must be considered so that the *VM* are created on the servers with minimum time thus achieving the services requested to the user by the brokers. The proposed clustering approach resulted in the ease to search for the resources that are available thus reducing the migration response time.

Engineering et al. [24]proposed the DVFS (Dynamic Voltage Frequency Scaling) method for the provisioning of *VM* in the cloud environment. DVFS method is used to lower the energy and power consumption of microprocessor. The power consumption can be reduced by lowering the operating frequency and the energy consumption can be reduced by lowering the supply voltage. To reduce both the power and energy consumption, the lowering of both the operating frequency and the supply voltage is required. Reducing the operating frequency by half reduces the power consumption of the system in turn allowing the task to be completed by deadline. Reducing the supply voltage by half reduces the energy consumption of the system in turn preventing any further increase in the execution time. This results in lowering the power and energy consumption without compromising the performance. The MinPower*VM* Provisioning algorithm is proposed for the placement of *VM* in the cloud. This algorithm returns the price to the broker. The price includes the services provided by the

VM for the completion of the task if it can. The provisioning policy selects the *VM* that returns the minimum costs for completing the tasks. The policy so proposed provides 33.62% better results in terms of the power and energy consumption as compared to the other non-power aware policies. The proposed policy did not consider the soft deadline service and is left for the future work.

Computing[25] proposed Dynamic Round Robin algorithm for the balancing of load in the cloud environment. The main objective of the algorithm is to balance the load of the *VMs*. The cloudlet's long length is analyzed with respect to the bandwidth of the host. The load on the *VM* is analyzed by varying both the long length of the cloudlet and the values of the bandwidth of the host, both of these variables are varied within the range of 10000 to 40000. The main objective is to achieve the desired performance while forming the analysis on the host bandwidth and the *VM* task length. The main focus is on the load-balancing parameter of the *VMs* in the heterogeneous datacenter. This paper also includes the detailed survey of other cloud computing methodologies related to the homogeneity and heterogeneity of the cloud datacenter along with the process of migration and scaling. The proposed approach tested through simulation on CloudSim has worked out well in terms of increase in both the performance and increase in server utilization.

Quang-Hung et al. [26]proposed the Genetic Algorithm for the static power-aware *VM* allocation in the private cloud environment. The problem with the static *VM* allocation is that the *VM* needs certain processing elements like million instructions per second (mips), the start time of *VM* (t) and the finish time (t + d) where d is the deadline. Each physical machine can host any *VM* and there is no limit on the amount of resources that are available. The power consumption at time t is directly proportional to the resource utilization by this model. This states that the Power Model for the system proposed is proportional to the utilization of the resources. The main objective of the system is to minimize the power consumption by maximizing the fulfilment of the resource requirements of the *VMs*. The genetic algorithm is applied to the proposed model for the allocation of *VMs* to the server. The inclusion of the Genetic Algorithm resulted in minimizing the energy consumption by 130% as compared to the baseline scheduling algorithm. The computational time in the Genetic Algorithm is not considered and is left for the future work.

Tesfatsion et al. [27] proposed an energy efficient system by integrating the hardware and software techniques for datacenter management. The main objective is to increase the performance and energy saving while defining the ideal configuration. In order to meet the demands of the application, the controller constraints are attuned automatically based on an online learned system model. Various scaling techniques for the management of the system power are considered such as vertical scaling (it includes the ability of *VMM* to change the configuration of *VMs* on the basis of resource workload), horizontal scaling (it includes the ability of the *VMM* to add or to remove the *VMs* from the datacenter) and hard-power scaling (DVFS technique results in the reduction of power to a great extend when applied to the clusters of *VMs* and *PMs*). The work so proposed resulted in the decrease of energy in the datacenter by 34% compared to the traditional approaches which include the change in the count of cores, the change of *VM* and the change in the frequency standards of CPU while meeting the demands of high performance.

Singh et al[28] proposed a pre-copy mechanism for reusing the memory of the *VM* in cloud environment. During the migration of *VM* from one host to another the memory images of *VMs* are kept on the source hosts from which the *VMs* are being migrated. The problem arises when the number of migration increases leading to an increase in the *VM* image which causes memory starvation. The objective is to increase the utilization of resources by reducing the size of the memory image of *VM* on the source host. The proposed method states that the when the *VM* are migrated back to the source hosts then, instead of transferring whole data back to the source host, only the new data added to the *VM* should be migrated. The previous data could be retrieved easily from the memory image of the *VM* stored on the source host. This method resulted in reusing the existing memory image hence leading to the increase in performance. Additionally, prior to the migration of *VM*, the current state of that *VM* is also stored for reducing the data transfer. The memory starvation problem is also dealt by using the probability phenomenon. The data that has the maximum probability of being updated and the data that has minimum or zero probability of being updated are not needed because the data cannot be reused further. The proposed method resulted in the memory management. It leads to an increase in utilization of resources by reducing the problem of memory starvation. The use of post-copy method for the VM migration is not considered and is left as the future scope.

Arianyan et al. [29] proposed the Novel energy and the SLA efficient resource management heuristics for the consolidation of *VMs* in the cloud data centers. An increase in the IT services has resulted in the establishment of hungry data centers. This in turn leads to the increase in pressure on the cloud providers to lower down both the energy consumption and the CO_2 emission in the cloud environment. To aid this problem, the energy efficient consolidation method is considered in cloud data centers. The study resulted in the improvement of on-line resource allocation process in two ways. First, Enhanced Optimization policy is proposed for the on-line resource management procedure. This policy suggests the gathering of *VM* from the overloaded as well as under loaded hosts and listing them in the *VM* migration list. After *VM* listing the novel heuristics are used for solving the on-line resource allocation problem. Secondly, The TOPSIS power and *SLA* aware allocation (TPSA) policy is used as the novel heuristics for solving the off-line and on-line resource allocation problem. The scores are calculated based on the least power, maximum number of resource availability, the number of *VMs* hosted by *PM*, the correlation between the already hosted *VM* and to be hosted *VM* and the migration delay based on hosting the *VM* on *PM*. The policy considers the available capacity of *PM* as the determining condition for the under-load detection of *PM*. The method so proposed resulted in the 99.9% reduction in electronic support measure (ESM) metric. The implementation of the proposed policy using real cloud infrastructure management products and the algorithms for the other resource management process is not considered and is left as a future work.

Esfandiarpoor et al. [30] proposed the structure-aware algorithms for the online *VM* consolidation for the cloud datacenter energy reduction by considering the structural features like racks, network topology of datacenter in the cloud. The cooling devices and the network structure are considered for the consolidation of *VM* on the host so that less number of racks and routers are used without the violation of *SLA*. The main objective is to reduce the total power consumption in the cloud environment. The consideration of both the million-instructions-per-second (MIPS) and the memory requirement resulted in increasing the computing capability of the *VMs*. The two basic models are considered which are *VM* analysis (which determines the mips and the memory of each *VM* that is being arrived and running on the hosts) and *VM* placement (which considers the resource demands and the current status of the datacenter). Our Proposed Modified Best Fit Decreasing (OBFD) algorithm considers mips as a parameter for sorting the *VM* in the decreasing order instead of the CPU utilization. The proposed method resulted in the reduction of energy consumption of hosts by 2.5%, network equipment by 18.8% and the cooling system by 28.1%.

Gao et al.[31]proposed the Multi-Objective Ant Colony System for the placement of *VMs* in the cloud environment. The main objective is to find the pareto-set (non-dominant solution) which will lead to better resource utilization and improvement in the energy efficiency. The *VM* placement is associated with the multi-dimensional packing methods. The dimensions which are focused in this paper are CPU and memory. The solution of the proposed method is compared to other methods like min-max ant system, bin-packing algorithm, multi-objective genetic and two single-objective algorithms. The comparison results showed that the proposed algorithm is more effective and efficient in terms of resource utilization and power consumption.

Ilkhechi et al. [13] proposed the Network-Aware *VM* placement in cloud datacenters with multiple traffic-intensive components. The scenario of the *VMs* performance is dependent on the capability of the infrastructure to meet the growing traffic demands. The metrics considered for the performance of the *VM* is named “satisfaction”. The main objective is to find the best assignment of the set of *VMs* onto the set of *PMs* in the scenario by maximizing the proposed metrics. The proposed method is based on single-cloud environment in which the online controller is designed which decides how many instances of the services are allowed to execute on the particular *PM*. Two approaches are considered in this paper namely Greedy-Based approach and Heuristic-Based approach. In Greedy-Based approach the decision is taken to place one *VM* on at least one *PM* by expecting that the summation of satisfaction for all the *VM* after assigning to the *PM* will be nearly maximum. In Heuristic-Based approach the best decision is considered for the placement of *VM* on the *PM* and also for the *VM* that are not allocated on the *PM*. When both the approaches are applied on the 15 sink *PMs*, the approaches showed better results in terms of performance as compared to when these approaches are applied in the datacenter network. Simulating and applying similar results on the similar scenarios with different limitations and assumptions is left as a future work.

Khani et al. [32] proposed the distributed mechanism for the dynamic consolidation of *VM* in heterogeneous data centers of big cloud providers. The main objective is to map the *VMs* on the *PMs* in such a way that in a way minimizes the total cost by reducing the total power consumption of the datacenter. The power consumption is considered as a part of the total cost. Game Theory is considered as the base approach for the consolidation of the *VMs*. According to the proposed theory, each *VM* has the rights to choose any *PM* (which is capable enough in terms of resources to host the *VM*) on which it will be allocated. Two main issues are considered like the Nash Equilibrium (when the game reaches the constant point after the certain number of *VM* migration) and the Convergence Time (the time taken to reach the Nash Equilibrium). The non-cooperative sequential gaming approach is considered for the power consumption in the datacenter. The main contributions related to the paper are that the game always converges to the Nash Equilibrium and the upper bounds are introduced on the convergence time. This was the first work done on the Game Theory based on the heterogeneous *PMs*. The results of the proposed method with the consideration of power consumption are analogous with the regional results. The parallel gaming approach and the multidimensional resource management environment which includes multiple resources like memory, storage, and network are not considered in this paper and is left as the future work.

Luo et al. [33] proposed the hybrid shuffled frog leaping algorithm (SFLA) for energy-efficient dynamic consolidation of *VM* in cloud datacenter. The main objective is to reduce both the operational cost and the power consumption by employing the SFLA in the cloud environment. The problem of dynamic allocation of *VM* to the PM cannot be solved by the exact method, so the modified SFLA and the enhanced optimization are considered in the paper. SFLA is motivated by the imitation of natural creatures. Two search models, the individual meme evolution (considered as meme carrier in ethnic group) and the global search for the information exchange (in the whole meme population) are considered. The random function is used for the generation of frog population. Then the fitness function is applied on the generated population and the proper fitness value is calculated. Based on the fitness value, the frogs are arranged in decreasing order and are then arranged in groups. The global search capability besides the local search is also taken into consideration. The leaping version phenomenon of frogs is considered according to which the regions outside the groups for obtaining the optimal solution. With the availability of thousands of *PMs* in the cloud datacenter, the heuristic-based approach is proved to provide fast results as compared to the traditional methods. The proposed modified (SFLA) increased the searching capability. Comparable to other resource management techniques, the proposed technique resulted in the excellent performance, better Quality-of-Service (QoS) and an increase in the conservation power resources in the cloud environment. Dynamic prediction of services and the use of multi-resource environment is not considered and is left as the future work.

Raycroft et al. [34]proposed the performance bounded energy efficient *VM* allocation in global cloud. The key objective is to evaluate the effect of *VM* allocation strategies on the energy consumption. Many *VM* allocation strategies like Round Robin, Striping, Packing, Load-Balancing (free CPU count), Load-Balancing (free CPU ratio), Watts per Core, and Cost per Cores are simulated and their effect on the performance is evaluated. In this paper, large-scale websites especially Reddit.com along with the Amazon's EC2 datacenters are simulated. The clusters of *PMs* are considered in different geographical positions around the globe. The Reddit.com is shifted entirely on the Amazon EC2 *VM* instances. The website will be hosted on a number *VMs*. The dissemination of the *VM* into various clusters will be the focus to study the dynamic server loads. The three steps involved in the simulation process are firstly, the allocation of load-handling *VM* is done by considering the current amount of load in each cluster to divide the load-handling *VMs*, secondly, using the scheduling policy the mapping of both the load-handling *VMs* and the cluster set of loaded *VMs* is done to the hosts and lastly, the scheduling policy is used to map the hosts to the remaining *VMs* and the *VMs* that are not mapped, their mapping is done against all the hosts and against any cluster. Each iteration takes 24 hours for analyzing the performance of scheduling strategy in the cloud. With respect to energy, the Watts per Core resulted as the best energy saving strategy which consumed 24.9 kW of energy. With respect to cost the Watts per Core and the Cost per Core were the best methods. Stripping and load balancing methods performed well in regards to the CPU load. The modification in the algorithm to minimize the SLA violation is left as the future work.

Gautam & Bansal[35] proposed a Round-Robin Scheduling algorithm for the balancing of the load in cloud datacenters based on the cloudlets that are received and then to consolidate the cloudlets onto the *VMs* in the datacenter. The author analyzed the frequent work load patterns and on the basis of these patterns made a prediction about the availability of resources on the *VMs*. The method so proposed resulted in the minimization of cost and time for the submission of tasks to the cloud environment. Further work can be done relating to the cost metrics for the method proposed.

Portaluri & Giordano[36] proposed the genetic algorithm for the allocation of tasks in the cloud datacenter by increasing the energy efficiency in the datacenter. The computational and networking resources required for the task allocation are considered. Two objectives that are dealt with are the minimization of the make span (completion time of the task) and the minimization in the power consumption of the servers and the switches used in the networking. The first objective is obtained by the use of multi-objective genetic algorithm for the optimal allocation of the tasks in the datacenter. The pareto set (non-dominant) is found that improves the existing solution. The ranks and fitness values are assigned to the sets and the comparison between the pareto set and the existing sets is done. The one with the highest fitness value is chosen. The pareto genetic algorithm is also able to deal with the constraints and unfeasible solutions also. The second objective of power management in the datacenter is handled with the use of DVFS (reduces power by decreasing the operating voltage and frequency) and the dynamic power shutdown (which lowers down the power consumption of the servers by putting the servers in the sleep mode to the minimum possible value) methodologies. These two methods are easily applied on the static consideration of the model. The proposed method resulted in the fixed provision of large number of isolated chores on the similar single-core servers within the identical datacenter and with a quadratic time complexity. It is achieved with the dedicated open-source genetic multi-objective framework called jMetal.

Xie et al. [37] proposed the energy efficient heuristic approach to reduce the energy consumption in the datacenter. The main objective is to allocate the *VMs* onto the servers in such a way that the resource requirements of the *VMs* are met efficiently and the power consumption of the datacenter can be kept to the minimum level as possible. The objective is achieved by the allocation of as many *VMs* as possible to the active servers and not onto the power saving servers where the energy consumption is 0. The servers that are considered in this paper are non-homogenous in nature that is each server has its own separate resource capacity, power intake and changeover cost. Heuristic algorithm assigns the *VMs* to the *PMs* in the growing order of their starting time. The consideration of the subset of servers having the sufficient resources to fulfill the *VMs* requirement is done. From this lot the server is selected and *VM* is assigned to that server. For the cost involved in the energy consumption two aspects the cost of running the *VM* and the cost to keep the server in an active mode are considered. The paper considers energy efficient low transition cost servers, the efficient allocation of the resources to the *VMs* and the subset of the server is considered such that it reduces the incremental energy cost. The servers with the less resource capacity will munch less power as compared to the large resource capacity, so the author proposed to allocate *VMs* onto the servers with the less resource capacity. The method resulted in the significant lowering of the energy consumption as compared to the first fit power saving model.

Li et al. [38] proposed the innovative technique named EnaCloud for the live placement dynamically with energy concern in the cloud datacenter. EnaCloud is analogous to the bin packing method. The main objectives of the paper are the minimization of the number of servers for the placement of the applications on the servers and the reduction of the migration time involved in the remapping of the resources to the servers. The tasks arrival is considered to be dynamic in nature. The tasks are allocated onto the active servers and the effort is taken to not activate the power down servers. When one application is finished with the jobs being allocated to it, it is made to go into the hibernate mode, this in turns results in the migration of the applications to the active servers by proper remapping so as to switch the servers to the sleep mode. The workload arrival, workload migration and the workload resizing are considered. The workload resizing includes workload inflation (it effects the performance of other task assigned to the same server) and workload deflation (this would release some resources, and result in the idleness and hence the wastage of energy). The proposed approach is implemented in the iVIC platform. This resulted in the feasible live placement of the VMs dynamically. The multi-resource consideration is left as the future work.

Ren et al. [39] proposed the multi-objective evolutionary game theoretic framework named Cielo with the DVFS in considerations for the adaptive and stable allocation of applications in the cloud datacenter. Cielo supports the cloud providers in the allocation of the resources to the application and then the allocation of the application in the cloud datacenter according to the operational conditions. The consideration of the DVFS to balance the response time performance, resource utilization and power consumption is done in the paper. The main objectives CPU allocation, bandwidth allocation, response time and power consideration are achieved successfully in this paper. Cielo theoretically proves the steady deployment of task in turn providing the equilibrium result under the given operational constraints. Cielo maintains the number of application that are needed to be allocated, different strategies are applied on each application for searching the location of and the resource allocation for the *VMs*. Cielo beats ongoing renowned heuristics in the quality, steadiness and computational cost of application placement in the cloud datacenter.

Beaumont et al. [40] proposed the approximation and heuristic algorithms for lowering down the energy consumption when allocation the tasks to the *VMs* in the cloud datacenter. The main job of the cloud providers in providing the services to the customers include looking for the count of necessary resources, their clock frequency and the allocation of the instances onto the machines. The main objective is to satisfy all the types of constraints involved in the assignment of task onto *VMs* while minimizing the energy consumption of used resources. DVFS technique is used for lowering the energy consumption by adapting the frequency needs to the instances available. To prove the approximation ratio, two techniques are used. Firstly, for the lower bound given the reliability conditions of the services and the failure constraints for the given machine, the least number of the services and instances at a specified energy level is needed. Secondly, for the allocation schemes for the upper bound two aspects homogenous and step are considered. For the homogenous aspect the min-replication strategy is used, which puts upper bound on the number of instances that can be assigned onto the machine. For the step solution, authorization of one unit step is done. The proposed work resulted in the optimal solution even when the number of machines and instances to be assigned to the machine increases. The use of small scale instances by using different approximation techniques is left as a future work.

Beloglazov et al. [41] proposed the heuristics for the allocation of resources in an energy efficient manner in the cloud datacenter. Green computing is the strategy used in this paper for minimizing the operational cost and reducing the energy consumption. The green cloud architecture includes the deployment of applications onto the *VMs* on demand at the competitive prices and the QoS constraints. For the power management in the cloud datacenter, the DVFS technique is used. The relation between the DVFS and the servers is linear in nature because minimum number of states can set to the operational frequency and the specified voltage. Three crucial issues which are extreme power cycling of the server could decrease the reliability, turning off the resources in an active environment is dangerous and ensuring SLA results in the difficulties in the to the performance parameter of the applications are considered. Various method for the VM placement, VM selection and minimization of migrations (using modified best-fit decreasing method) are proposed and are implemented in the CloudSim environment. Main parameter that is considered is the performance metrics. The proposed work resulted in the decrease of the threshold energy utilization and energy consumption in the cloud datacenter.

Beloglazov & Buyya[42] proposed the *VM* consolidation in the cloud datacenter in an energy efficient manner by the use of modified best-fit decreasing (MBFD) algorithm. The main objective is to continuously allocate *VMs* on *PMs* by leveraging the *VM* migration which would result in switching of the idle *PMs* thus reducing the power consumption of the *PMs*. The chief parameter so considered is QoS. The best-fit algorithm is modified in a way that first the VMs are arranged in the decreasing order of their current utilization and then each *VM* is allocated to the *PM* that reduces their energy consumption. The migration of the *VMs* is done in two steps, firstly the selection the *VMs* is done that need to be migrated onto another *PM*, and secondly the selected *VM* are allocated to the *PM* according to the MBFD algorithm. The single threshold heuristic is considered in which an upper limit on the CPU utilization of the *PM* is imposed. For the energy management in the cloud datacenter, DVFS technique is used which regulates the voltage and frequency of the CPU conferring to the current consumption. The proposed work resulted in the reduction of operational and establishment cost along with the decrease in the energy consumption in the cloud environment.

Dhiman et al. [43] used the multitier software system named vGreen for the energy proficient management of *VMs* in the cloud environment. vGreen is lightweight and has slight runtime overhead. It is based on client-server model. The method so proposed focuses on the consolidation of *VMs* with heterogeneous characteristics on the same *PM*. This will result in an increase in the performance of the *PM* and decrease in the overall energy consumption. The vGreen architecture includes the cluster of *PMs* hosting the *VMs*. Each pm in the cluster is referred to as the client. The server manages and schedules the *VMs* based on the policy adopted by the server. The policy so adopted is based on the CPU utilization, memory-per-cycles (MPC) and inter-process communication (IPC). The up gradation of policy is done dynamically by the clients which hosts various VMs. The vGreen algorithm works on four aspects which are MPC balance, IPC balance, utilization balance and DVFS for the reducing the energy consumption the cloud datacenter. The proposed methodology resulted in the increase in performance parameter by 100% and the reduction in the energy by 55% compared to the state-of-art scheduling and power supervision techniques.

Dupont et al. [44] proposed the effective, flexible and energy-aware framework for the allocation of *VMs* onto the *PMs*. The framework so proposed is different from the existing energy-aware heuristics as it assigns the *VM* to the *PM* under the proper consideration of the constraints so imposed using the constraint-programming (CP). The CP framework extends the entropy management and considers the *VM*-repacking scheduling algorithm (VRSP) for the scheduling of the *VMs* so as to allocate them onto the *PMs* by taking into consideration the placement and resource constraints. The timeout can be specified on the entropy, which then stop to solve after the timeout expires. If no timeout is specified then the entropy computes and return the reconfiguration with the best possible solution according to the power and resource restrictions. For the power management purpose, the "power calculator" is used. This component computes the power consumption to the fine-grain so possible for each and every part of the datacenter when fed with the datacenter's physical and dynamic fundamentals. The proposed work resulted in the decrease of energy consumption by 18% and allocation solution for 2800 *VMs* in 700 *PMs* divided into two clusters was found within 1 minute.

For et al. [45] proposed Dynamic Round-Robin methodology for the provisioning of the resources onto the cloud servers. The energy efficiency of the cloud datacenter is increased by the live migration of *VM* to other servers, so as to shut down the non-active servers, thus enabling the power management. The chief objective is to reduce the count of active *PMs* that hosts the *VMs*. The Dynamic Round-Robin methodology is adopted for the provisioning of *VMs* onto the *PMs* in an energy-efficient manner, which in turn switches the non-active *PMs* to the sleep mode. The Dynamic Round-Robin includes two rules. Firstly when *VMs* assigned to the *PM* have completed their task, then except the *VMs* running on that *PM*, no other *VM* will be hosted by that particular *PM*. This would result in shutting down that particular *PM* when all the *VMs* on it has completed their jobs and that *PM* would be in the "retirement" mode. Secondly, if the time taken by the *VMs* assigned the "retired" *PM* is long then those *VMs* would be forced to migrate to another *PM*. This would result in shutting down the *PMs*. The results convinced that the *VM* migration and Dynamic Round-Robin technique provided the feasible solution for the energy management in the cloud datacenter. The monitoring of the QoS parameter of the *VM* is left as the future work.

Hatzopoulos et al.[46] addresses the problem of energy-efficient task allocation in the cloud datacenter. The existence of a time-varying network energy price and the randomness and time variation of provisioned power by the renewable energy source (RES) is considered in this paper. The chief objective is to reduce the cost that is paid by the cloud service providers to the main network by maintaining the necessary power in operating the execution of *VMs*. The power management in the cloud environment is taken care by the consideration of the RES. The RES cut the want on the main power grid and if they are aptly misused, they can lead to a major cost reduction. The advantage of RES is that it need the extra set-up cost in the initial deployment phase, once the RES is deployed, the operational cost and the provisioned energy cost of RES is minimal. For the allocation of the task to the *VM*, the *CSP* either allocates the *VM* onto the high processing servers that could finish the task before deadline or allocate the *VMs* on the low processing servers that has the capability to finish the task before deadline. The low power-consuming servers are considered which reduces the voltage and frequency so used. The proposed RES model achieved the *VM* scheduling and placement with the aim to burden the *PMs* with the resulting power consumption that harmonized the RES pattern. The consideration of the potential multi-tenancy and other convincing phenomenon is left as the future work.

Beloglazov[47] proposed the algorithm for the dynamic placement of *VMs* in the cloud datacenter. The energy-efficient placement of *VMs* considers the system where the energy of computational resources is proportional to the workload on the *VMs*. This approach is partially implemented with the aid of DVFS technology. In DVFS the voltage and frequency adjusts according to the current consumption thus resulting in 30% reduction in the power consumption of the cloud servers CPUs. Two requirements which are the handling of the heterogeneous workloads and the maintenance of the QoS constraints are taken care of. The main objectives that are required to be achieved includes the definition of workload-independent QoS constraints, the time when *VMs* would be migrated to other servers, the selection of *VMs* to be migrated, the selection of hosts where the migrated *VMs* would be hosted, when and which *PMs* would be put to active or sleep mode and the designing of the algorithm for the placement of *VMs* onto the *PMs*. An optimal online deterministic algorithm is considered for the placement of *VMs* onto the *PMs*. The algorithm so used has resulted in improving the quality of their deterministic counterparts. For the migration of *VMs* the heuristic method is considered which includes host under load and overload detection, *VM* selection and placement.

Kliazovich et al. [48] proposed the DENS (datacenter energy-efficient network-aware scheduling) method. The proposed methodology improves the trade-off between job assignment (to reduce the count of computing servers) and circulation of traffic arrangements (to avoid hotspots in the cloud datacenter network). Thus a balance is maintained between the job-performance, traffic demands and the energy consumption in the cloud datacenter. The DENS method lowers the energy consumption by selecting the best-fit computing resources for executing the jobs based on the level of the load on the servers and the communication potential of the cloud datacenters. DENS developed the hierarchical topology analogous to the state-of-art topologies of the cloud datacenter. Thus the paper summarizes the role of communication fabric in cloud datacenter and for the same proposed the DENS methodology. The DENS method is able to maintain the QoS level constraint as desired by the end-user with the slight increase in the energy consumption. DENS is in use in the existing and upcoming datacenter architectures.

La et al. [49] used two strategies namely knapsack problem and the evolutionary computation heuristic for the placement of *VM* in the cloud datacenter in an energy-efficient manner. The main work is to place as many *VMs* as possible onto fewer *PMs* because the energy consumption of servers is low when they are operations at nearly 100% of their utilization. The objectives of this paper are the information retrieval (the information about *VMs* placement are limited to CPU utilization, the resources that are considered are network resources along with the CPU utilization parameter), under-load and overload detection, *VM* selection and placement method. The knapsack problem is used for the placement of the *VMs* onto the *PMs*. This includes the consideration of multidimensional-bin-packing (MDBP) methodology. Besides MDBP the iterated-knapsack method is used for multiple resources and for its less complex nature as compared to MDBP method. The evolutionary computation method includes the evaluation of the fitness function and based on this value the *VMs* are placed onto the *PMs*. Both the strategies used have resulted in the reduction of energy consumption in cloud datacenter from 40.33% to 92.21% as compared to the methodologies that does not consider the energy-efficient as the main parameter.

Li et al. [50] proposed the optimal strategy in scheduling the resources for economizing the energy in the extended *VM* system. In an extended *VM* system, the count of physical resources that the *VMs* can schedule at any time are less than the total count of system resources. In an extended system, the physical resources are managed by the solo image administration module and its equivalent *VM* monitor. When different physical resource request arrives for the different *VMs*, then *VMs* need to schedule these heterogeneous task. If the workload of these tasks on the *VMs* is beyond the load level, then those task need to be migrated to another *VM* where these tasks can be assigned to the *VMs* having sufficient resources for the proper task execution. Two rules are stated in the paper with their corresponding conditions which are the minimization of the frequency that each *VM* need to process the task (to maintain the balance among the various *PMs*) and the minimization of the frequency that each *VM* are required to migrate their task (the count of the *VM* task migration should be as low as possible). To decide the optimal solution of the resource scheduling model, the author proposed an algorithm. The algorithm consists of the following steps, firstly the random selection of a group initial solution is done, called a group initial value. Secondly, the search for the ideal result in the global scope of domain is done. At last, the repetition of the execution of search process is done till the results are fulfilled with the finish condition. This method resulted in the increase in providing the services to the task in the cloud datacenter.

Murshed et al. [51] addresses the problem of energy-efficient consolidation of *VMs* in the cloud datacenter by the use of Ant-Colony-Optimizing (ACO) meta-heuristic along with the balance in the use of computing resources. ACO methodology is based on the foraging behavior of the insects. The Ant-Colony-System (ACS), a later version of ACO considers each *VM* assigned to *PM* (VM-to-PM) as a component of the solution. To each component of the solution, pheromone levels are related to the VM-to-PM assignments thus representing the interest of assigning the *VM* to that *PM*. The computation of the heuristic values are done for the VM-to-PM assignments dynamically. Based on the computed value favorability of placing the *VM* on the particular *PM* is represented in terms of both the general and stable resource utilization of the *PM*. Both the local and the global best solution are considered for the proper allocation of *VM* onto the *PMs*. The proposed meta-heuristic resulted in outperforming existing method of *VM* consolidation in terms of both the energy consumption and the resource utilization. The consideration of network-aware resources in the cloud infrastructure is not considered in this paper and is left as the future work.

Lee et al. [52] proposed the Proactive Thermal-aware placement of *VMs* in High-Performance-Cloud (HPC) datacenters. The paper deals with the conflicting objectives of increasing both the energy-efficiency and resource utilization thus avoiding the thermal hotspots and guaranteeing the QoS constraint as desired by the end-user. This conflicting objective in turn is replaced by the joint contemplation of numerous pairwise tradeoff. The thermal awareness which is introduced in this paper predicts the upcoming temperature of the cloud datacenter and seizures the jaggedness in both heat removal and generation. The three main focus points of the paper includes the placement of *VMs*, the minimization of the energy consumption and the improvement in the efficiency of heat extraction. The proactive approach includes various layers like the cross-layer (requires continuous analysis and feedback from various other layers), application-layer (provides information related to the resource consumption of the applications), hardware-resource-layer (equipped with the sensors that provide information related to the utilization, fan-speed and operating frequency), environment-layer (consists of heterogeneous sensing infrastructure of cloud datacenter) and virtualization-layer (has ability to allocate and manage the *VMs*). The VMAP approach resulted in the increase of energy saving by 9%, and 35% than best-fit and "cool job" (state- of-the-art reactive thermal-aware solutions), respectively. The consideration of the joint optimization of duty cycle for the VMAP allocation and cooling is left as the future work.

Velayudhan Kumar & Raghunathan [53] presented the adaptive approach for the placement and consolidation of *VMs* in an energy-efficient manner in the cloud datacenter. It takes into account the consideration of the heterogeneous nature of the servers, the servers sleep (low power mode) state, latency in the state transition and other related thermal controls which maintains the datacenter within the operational temperature. The main aim is to reduce the energy ingestion with tolerable level of performance. The proposed system model comprises of various modules like monitor-module (monitors and tracks various servers present in cloud datacenter), provisionary-module (it assigns the *VMs* to the best servers in terms of energy consumption), heterogeneity-controller-module (categorizes the servers in terms of their performance and power parameters), optimizer-module (to reduce both the utilization of servers and the count of active servers), server-power-state-transitioner-module (to transit between the states of the server) and the thermal coverage and integrator module (it dynamically computes the thermal temperature of the servers). For the provisioning of *VM,* the *VMs* are placed onto *PMs*, the identification of *VMs* is done that needs to be migrated, the selection of *PM* that can host the migrated *VM* is done and then the replacement of *VM* is done.

Masoumzadeh & Hlavacs[54] proposed the Fuzzy Q Learning (FQL) for the energy-efficient distributed dynamic *VM* consolidation in cloud datacenter. The dynamic consolidation of *VM* includes the host overload and under-load detection, the *VM* selection and placement tasks. The main objective of the paper is to dynamically consolidate the *VMs* by optimizing the energy-performance balance online. The objective is achieved by optimizing the mentioned tasks included in the dynamic consolidation of *VMs* through the implementation of the FQL approach. The FQL is the extension to Q leaning and is a well-known reinforcement-learning (RL) technique. In RL the rules are stated and corresponding to the rules the actions are stated. The actions carry some weight with them which changes after each learning phase. Based on the weight value of the rule the action is taken. The agent chooses the *VM* according to the requirements so listed and hen places the selected *VM* onto the *PM*. This whole process is learned and is done according to the rules and actions so listed. The approach resulted in yielding far better results with respect to the energy-performance balance in cloud datacenters as compared to the state-of-art methodologies.

Quan et al[55] addresses the problem related to the power consumption in the cloud datacenter. Thus the optimizing algorithms are proposed for the energy-efficient resource allocation in the cloud environment. The rudimentary idea is reshuffling the allocation in a way that saves energy. The paper uses power consumption models related to the server power consumption which includes processor (the power consumption of servers varies linearly with the CPU utilization), memory (depends on the count of installed modules related to the memory), hard disk (computes the accessing and startup power consumed by the hard disk), mainboard (the power consumption by the mainboard), server power (power consumption by the modules of the server), network power consumption (computes the power consumption of the switches involved in the network traffic) and datacenter power consumption (computes the energy consumption of the *VM* consolidated onto the *PM*). Two optimizing algorithms are proposed in the paper which are F4G-CS (Traditional single algorithm) (when new VM comes for the placement, the energy-efficient server is selected for hosting the *VM*) and F4G-CG (Cloud global optimization algorithm) (this algorithm migrates the *VM* from the low load servers to the high load servers and then switching the non-active servers to the sleep mode which would result in saving the energy in the cloud environment). The improvement in the performance parameter when the datacenter is loaded with the old heavy loaded servers is not considered and is left as the future work.

Table 1
Comparison Table

Research Papers	*Technique*	*Allocation*	*Objective*	*Migration*	*Scheduling*	*Experiment*
Yousefian & Zadnavin2015[14]	Novel Scheduling Method	VM on PM	Cost + Resource Utilization	-	✓	CloudSim
Zoltan Adam Mann [15]	Modified Best Fit Decreasing	VM on PM	Energy	✓	-	-
Kruekaew & Kimpan2014[16]	Ant-Bee Colony	VM on PM	Performance + Load Balancing	-	-	CloudSim
Aldhalaan & Menasce 2014[17]	Autonomic Computing	VM on PM	Cost	-	-	MatLab
Kumar et al. 2014[18]	Modified Best-Fit	VM on PM	Energy	-	-	-
Li et al. 2013[19]	EAGLE	VM on PM	Energy + Cost + Performance	-	-	CloudSim
Kumar & Ramachandra 2014[20]	Genetic Algorithm	VM on PM	Energy	-	-	C++
N & Hemalatha 2014[21]	Reservation Technique for BIN BECK Entropy	VM on PM	Energy	-	-	CloudSim
Sharma & Singh2014[22]	Round Robin + Throttled Load Balancing + Equally Spread Current Execution	VM on PM	Energy + Load Balancing	✓	-	CloudSim
Ajith Singh & Hemalatha 2013[23]	Ant-Bee Colony	Task on VM	Energy	-	-	-
Hetal Joshiyara 2013[24]	Dynamic Voltage Frequency Scaling (DVFS)	VM on PM	Energy + Cost	-	-	CloudSim
Gulati & Chopra2013[25]	Dynamic Round Robin	VM on PM	Load Balancing + Performance	-	-	CloudSim
Quang-Hung et al. 2013[26]	Genetic Algorithm	VM on PM	Energy	-	-	CloudSim
Tesfatsion et al. 2014[27]	Dynamic Voltage Frequency Scaling (DVFS) + Elastic Scaling Approach	Task on VM	Energy + Performance	-	-	CloudSim
Singh et al. 2015[28]	Pre-Copy Mechanism	VM on PM	Memory	✓	-	CloudSim
Arianyan et al. 2015[29]	TOPSIS + TPSA	VM on PM	Energy	-	-	CloudSim
Esfandiarpoor et al. 2015[30]	Our Proposed Modified Best Fit Decreasing	VM on PM	Energy	-	-	CloudSim

Research Papers	*Technique*	*Allocation*	*Objective*	*Migration*	*Scheduling*	*Experiment*
Gao et al. 2013[31]	Multi-Objective Ant Colony	VM on PM	Energy + Resource Utilization	-	-	CloudSim
Ilkhechi et al. 2015[13]	Greedy-Based + Heuristic-Based Approach	VM on PM	Performance	-	-	CloudSim
Khani et al. 2015[32]	Game Theory	VM on PM	Energy + Cost	-	-	CloudSim
Luo et al. 2014[33]	Hybrid Shuffled Frog Leaping	VM on PM	Energy + Cost	-	-	CloudSim
Raycroft et al. 2014[34]	Round-Robin, Stripping, Packing	VM on PM	Energy + Cost + Performance	-	-	CloudSim
Gautam & Bansal2014[35]	Round-Robin Scheduling	Task on VM	Load Balancing	-	✓	NetBeans
Portaluri & Giordano 2014[36]	Genetic Algorithm	Task on VM	Energy + Computation	-	-	CloudSim
Xie et al. 2013[37]	Heuristic Algorithm	VM on PM	Energy	-	-	CloudSim
Li et al. 2009[38]	EnaCloud	VM on PM	Energy	✓	-	iVC
Ren et al. 2014[39]	Evolutionary Game theory CIELO	VM on PM	Energy	-	-	CloudSim
Beaumont et al. 2013[40]	Heuristic Algorithm + DVFS	VM on PM	Energy	-	-	CloudSim
Beloglazov et al. 2012[41]	Heuristic Algorithm + DVFS	Task on VM + VM on PM	Energy + Cost	✓	-	CloudSim
Beloglazov & Buyya2010[42]	Modified Best Fit Decreasing	VM on PM	Energy + Cost	✓	-	CloudSim
Dhiman et al. 2010[43]	Multitier Software System named vGreen	VM on PM + management of VM	Energy + Performance	✓	✓	Xen
Dupont et al. 2012[44]	VM Repacking Scheduling Algorithm (VRSP)	VM on PM	Energy + Performance	✓	✓	CloudSim
For et al. 2012[45]	Dynamic Round-Robin	VM on PM	Energy + Load Balancing	✓	-	CloudSim
Hatzopoulos et al. 2013[46]	Renewable Energy Source (RES)	VM on PM	Energy + Cost	-	✓	CloudSim

Research Papers	*Technique*	*Allocation*	*Objective*	*Migration*	*Scheduling*	*Experiment*
Beloglazov 2013[47]	Dynamic Voltage Frequency scheduling (DVFS)	VM on PM	Energy	✓	-	OpenStack NEAT
Kliazovich et al 2013[48]	Datacenter Energy-Efficient Network Aware Scheduling (DENS)	Task on VM	Energy + Performance	-	✓	GreenCloud
La et al. 2014[49]	Iterated knapsack + Evolutionary Computation Heuristic	VM on PM	Energy	-	-	Python
Li et al. 2012[50]	Optimal Strategy for Economizing Energy in Extended VM System	Task on VM	Energy	-	✓	Xen
Murshed et al. 2014[51]	Ant-Colony-Optimizing (ACO) Algorithm	VM on PM	Energy + Resource Utilization	-	-	CloudSim
Lee et al.2012[52]	VMAP	Task on VM	Energy + Resource Utilization + Quality of Service(QoS)	-	-	CloudSim
Velayudhan Kumar & Raghunathan 2015[53]	StaticPPMMax Algorithm	VM on PM	Energy + Cost	-	-	CloudSim
Masoumzadeh & Hlavacs 2013[54]	Fuzzy Q-Learning (FQL)	VM on PM	Energy + Performance	✓	-	CloudSim
Quan et al. 2011[55]	F4G-CS (Traditional single algorithm) + F4G-CG (Cloud global optimization algorithm)	VM on PM	Energy + Resource Utilization	-	-	CloudSim

The comparison of the techniques adopted by various authors is done inTable.1. Besides this, the table also includes the experimental setup adopted by various authors. The migration and scheduling approaches considered while allocation is also considered as a parameter of evaluation.

3. CONCLUSION AND FUTURE WORK

The paper's main goal was to review all the related papers on the basis of certain parameters like cost, performance, load balancing and quality-of-service. The detailed comparison is done in the tabular format. The result of the review is that various techniques are considered for the allocation strategy with the main focus on the energy consumption in the cloud datacenter. The allocation strategies include two allocation strategies firstly, the allocation of the task or the job to the *VM* and secondly, the allocation of the *VM* onto the *PM*.

The future scope of this paper includes the consideration of the multi-resource framework besides the energy consideration for the allocation strategies. The neural training of the *VMs* can be done, so that *VMs* are self-trained to be allocated to the appropriate *PMs* according to the set requirements and constraints listed in the *SLA* document.

REFERENCES

[1] Zhang Q, Cheng L, Boutaba R. Cloud computing: state-of-the-art and research challenges. J Internet Serv Appl [Internet]. 2010;1(1):7–18.

[2] Nist. The NIST Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology. Nist Spec Publ [Internet]. 2011;145:7.

[3] Shahzad F. State-of-the-art Survey on Cloud Computing Security Challenges, Approaches and Solutions. Procedia Comput Sci [Internet]. Elsevier Masson SAS; 2014;37:357–62.

[4] Randles M, Lamb D, Odat E, Taleb-Bendiab a. Distributed redundancy and robustness in complex systems. J Comput Syst Sci [Internet]. Elsevier Inc.; 2011;77(2):293–304.

[5] Lee S, Prabhakaran RPV, Ramasubramanian V, Uyeda KTL, Wieder U. Validating Heuristics for Virtual Machines Consolidation. ResearchMicrosoftCom [Internet]. 2010;81–97.

[6] Xu J, Fortes J a B. Multi-objective Virtual Machine Placementin Virtualized Data Center Environments. 2010;

[7] Mann ZÁ. A taxonomy for the virtual machine allocation problem. Int J Math Model Methods Appl Sci. 2015;9:269–76.

[8] Rai A, Bhagwan R, Guha S. Generalized resource allocation for the cloud. Proc Third ACM Symp Cloud Comput - SoCC'12 [Internet]. 2012;1–12.

[9] Huber N, Brosig F, Kounev S. Model-based Self-Adaptive Resource Allocation in Virtualized Environments. 2011;

[10] Ts'epoMofolo R. Heuristic Based Resource Allocation Using Virtual Machine Migration: A Cloud Computing Perspective. Int Ref J Eng Sci [Internet]. 2013;2(5):40–5.

[11] Das S, Kagan M, Crupnicoff D. Faster and Efficient VM Migrations for Improving SLA and ROI in Cloud Infrastructures. Dc Caves [Internet]. 2010;1–7.

[12] Mills K, Filliben J, Dabrowski C. Comparing VM-placement algorithms for on-demand clouds. Proc - 2011 3rd IEEE Int Conf Cloud Comput Technol Sci CloudCom 2011. 2011;91–8.

[13] Ilkhechi AR, Korpeoglu I, Ulusoy Ö. Network-aware virtual machine placement in cloud data centers with multiple traffic-intensive components. Comput Networks [Internet]. 2015;91:508–27.

[14] Yousefian S, Zadnavin AH. Scheduling Virtual Machines in Cloud Computing For Enhancing Income and Resource Utilization. 2015;3(1):389–97.

[15] Zoltan Adam Mann. Rigorous results on the effectiveness of some heuristics for the consolidation of virtual machines in a cloud data center. Futur Gener Comput Syst. 2015;51:1–6.

[16] Kruekaew B, Kimpan W. Virtual Machine Scheduling Management on Cloud Computing Using Artificial Bee Colony. 2014;I:1–5.

[17] Aldhalaan A, Menasce DA. Autonomic allocation of communicating virtual machines in hierarchical cloud data centers. Proc - 2014 Int Conf Cloud Auton Comput ICCAC 2014. 2015;161–71.

[18] Kumar P, Singh D, Kaushik A. Power and Data Aware Best Fit Algorithm for Energy Saving in Cloud Computing. 2014;5(5):6712–5.

[19] Li X, Qian Z, Lu S, Wu J. Energy efficient virtual machine placement algorithm with balanced and improved resource utilization in a data center. Math Comput Model [Internet]. Elsevier Ltd; 2013;58(5-6):1222–35.

[20] Kumar V, Ramachandra GA. Energy Conservation for Datacenters in Cloud Computing using Genetic Algorithms. 2014;3(12):577–83.

[21] N AS, Hemalatha M. Reservation Resource Technique for Virtual Machine Placement in Cloud Data Centre. 2014;7(14):2954–60.

[22] Sharma A, Singh UP. Energy Efficiency in Cloud Data Centers. 2014;11(4):174–81.

[23] Ajith Singh N, Hemalatha M. Cluster based bee algorithm for virtual machine placement in cloud data centre. J Theor Appl Inf Technol. 2013;57(3):1–10.

[24] Hetal Joshiyara. Energy Efficient Provisioning of Virtual Machines in Cloud System using DVFS. 2013;(May):153–6.

[25] Gulati A, Chopra R. Dynamic round robin for load balancing in a cloud computing. Int J Comput Sci … [Internet]. 2013;2(June):274–8.

[26] Quang-Hung N, Nien PD, Nam NH, Tuong NH, Thoai N. A Genetic Algorithm for Power-Aware Virtual Machine Allocation in Private Cloud. 2013;

[27] Tesfatsion SK, Wadbro E, Tordsson J. A combined frequency scaling and application elasticity approach for energy-efficient cloud computing. Sustain Comput Informatics Syst [Internet]. Elsevier Inc.; 2014;4(4):205–14.

[28] Singh G, Behal S, Taneja M. Advanced Memory Reusing Mechanism for Virtual Machines in Cloud Computing. Procedia - Procedia Comput Sci [Internet]. Elsevier Masson SAS; 2015;57:91–103.

[29] Arianyan E, Taheri H, Sharifian S. Novel energy and SLA efficient resource management heuristics for consolidation of virtual machines in cloud data centers. Comput Electr Eng [Internet]. Elsevier Ltd; 2015;

[30] Esfandiarpoor S, Pahlavan A, Goudarzi M. Structure-aware online virtual machine consolidation for datacenter energy improvement in cloud computing. Comput Electr Eng [Internet]. Elsevier Ltd; 2015;42:74–89.

[31] Gao Y, Guan H, Qi Z, Hou Y, Liu L. A multi-objective ant colony system algorithm for virtual machine placement in cloud computing. J Comput Syst Sci [Internet]. Elsevier Inc.; 2013;79(8):1230–42.

[32] Khani H, Latifi A, Yazdani N, Mohammadi S. Distributed consolidation of virtual machines for power efficiency in heterogeneous cloud data centers. Comput Electr Eng [Internet]. Elsevier Ltd; 2015;1–13.

[33] Luo J, Li X, Chen M. Hybrid shuffled frog leaping algorithm for energy-efficient dynamic consolidation of virtual machines in cloud data centers. Expert Syst Appl [Internet]. Elsevier Ltd; 2014;41(13):5804–16.

[34] Raycroft P, Jansen R, Jarus M, Brenner PR. Performance bounded energy efficient virtual machine allocation in the global cloud. Sustain Comput Informatics Syst [Internet]. Elsevier Inc.; 2014;4(1):1–9.

[35] Gautam P, Bansal R. Extended Round Robin Load Balancing in Cloud Computing. 2014;3(8):7926–31.

[36] Portaluri G, Giordano S. A power efficient genetic algorithm for resource allocation in cloud computing data centers. 3rd Int Conf Cloud Netw [Internet]. 2014;58–63.

[37] Xie R, Jia X, Yang K, Zhang B. Energy saving virtual machine allocation in cloud computing. Proc - Int Conf Distrib Comput Syst. 2013;132–7.

[38] Li B, Li J, Huai J, Wo T, Li Q, Zhong L. EnaCloud: An energy-saving application live placement approach for cloud computing environments. CLOUD 2009 - 2009 IEEE Int Conf Cloud Comput. 2009;17–24.

[39] Ren Y, Suzuki J, Lee C, Vasilakos A V., Omura S, Oba K. Balancing performance, resource efficiency and energy efficiency for virtual machine deployment in DVFS-enabled clouds. Proc 2014 Conf companion Genet Evol Comput companion - GECCO Comp ’14 [Internet]. 2014;1205–12.

[40] Beaumont O, Duchon P, Renaud-Goud P. Approximation algorithms for energy minimization in Cloud service allocation under reliability constraints. High Perform Comput (HiPC), 2013 20th Int Conf. 2013;(February):295–304.

[41] Beloglazov A, Abawajy J, Buyya R. Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing. Futur Gener Comput Syst [Internet]. Elsevier B.V.; 2012;28(5):755–68.

[42] Beloglazov A, Buyya R. Energy efficient allocation of virtual machines in cloud data centers. CCGrid 2010 - 10th IEEE/ACM Int Conf Clust Cloud, Grid Comput. 2010;577–8.

[43] Dhiman G, Marchetti G, Rosing TS. VGreen: A system for energy-efficient management of virtual machines. ACM Trans Des Autom Electron Syst [Internet]. 2010;16(1):1–27.

[44] Dupont C, Schulze T, Giuliani G, Somov A, Hermenier F. An energy aware framework for virtual machine placement in cloud federated data centres. Proc 3rd Int Conf Futur Energy Syst Where Energy, Comput Commun Meet - e-Energy '12 [Internet]. 2012;1–10.

[45] For C, Computing C, Sammy K, Shengbing R, Wilson C. Energy Efficient Security Preserving VM Live Migration In Data. J Comput Sci. 2012;9(2):33–9.

[46] Hatzopoulos D, Koutsopoulos I, Koutitas G, Van Heddeghem W. Dynamic virtual machine allocation in cloud server facility systems with renewable energy sources. IEEE Int Conf Commun. 2013;(Icc):4217–21.

[47] Beloglazov A. Energy-Efficient Management of Virtual Machines in Data Centers for Cloud Computing. 2013;(February):1–232.

[48] Kliazovich D, Bouvry P, Khan SU. DENS: data center energy-efficient network-aware scheduling. Cluster Comput. 2013;16(1):65–75.

[49] La APM De, Vigliotti F, Macêdo D. Energy-Efficient Virtual Machines Placement. Sbrc 2014. 2014;17–30.

[50] Li Y, Wan J, Ouyang R, Zhang J, You X. An Optimal Method about Resource Scheduling for Economizing Energy in Extended Virtual Machine System. 2012;3(2):61–74.

[51] Murshed M, Calheiros RN, Buyya R. Virtual Machine Consolidation in Cloud Data Centers Using ACO Metaheuristic. Euro-Par 2014 Parallel Process. 2014;8632:306–17.

[52] Lee EK, Viswanathan H, Pompili D. VMAP: Proactive thermal-aware virtual machine allocation in HPC cloud datacenters. 2012 19th Int Conf High Perform Comput HiPC 2012. 2012;

[53] Velayudhan Kumar MR, Raghunathan S. Heterogeneity and thermal aware adaptive heuristics for energy efficient consolidation of virtual machines in infrastructure clouds. J Comput Syst Sci [Internet]. Elsevier Inc.; 2015;

[54] Masoumzadeh SS, Hlavacs H. Integrating VM selection criteria in distributed dynamic VM consolidation using Fuzzy Q-Learning. 2013 9th Int Conf Netw Serv Manag CNSM 2013 its three collocated Work - ICQT 2013, SVM 2013 SETM 2013. 2013;332–8.

[55] Quan DM, Basmadjian R, Meer H De, Lent R. Energy efficient resource allocation strategy for cloud data centres. 2011;(Iscis).