ASEAN Engineering Journal

Full Paper

COMPARISON AND EVALUATION OF ENERGY-EFFICIENT CLOUD COMPUTING TECHNIQUES WITH LOAD BALANCING APPROACHES

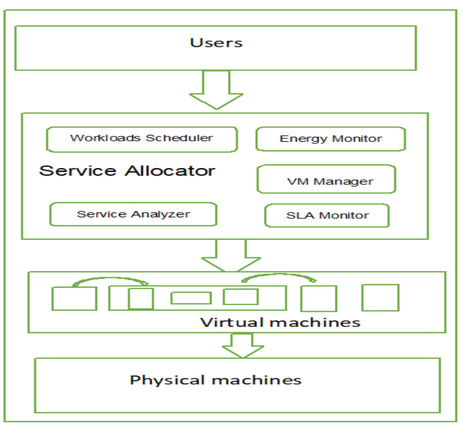
Rudresh Shah*, Suresh Jain, Kailash Chandra Bandhu

Department of Computer Science & Engineering, Medi-Caps University, Indore, India

Article history
Received
30 July 2023
Received in revised form
13 November 2023
Accepted
26 November 2023
Published online
31 May 2024

*Corresponding author
rudresh.shah@medicaps.ac.in
P

Graphical abstract



Abstract

The advent of Cloud Computing has revolutionized the IT landscape by offering computing resources as a service, similar to conventional utilities like electricity. This paradigm shift has made cloud computing a cornerstone of the contemporary digital economy, attracting substantial focus from both academic and industrial sectors. Its unique pay-as-you-go model provides customers with on-demand resource availability, enhancing operational flexibility. However, this convenience is offset by the growing energy demands of cloud data centers, which not only escalate operational expenses but also contribute to environmental degradation through increased carbon footprints. To combat these issues, Green cloud computing has been introduced, striving for energy-efficient and sustainable operations. This involves employing strategies that minimize energy consumption and resource utilization through the application of energy-conscious algorithms. Although numerous algorithms based on server consolidation have been proposed to optimize energy use in cloud environments, they often lack uniform evaluative comparisons and vary in performance due to differing experimental conditions. This variance presents a challenge in selecting the most effective algorithm tailored to specific needs. This study aims to provide a nuanced analysis of existing energy-efficient algorithms, assisting researchers in identifying the algorithm that best suits their requirements. We undertake an exhaustive comparison of various algorithms, examining their architecture, modelling approaches, and performance metrics. These algorithms are then implemented and tested under uniform conditions using the CloudSim toolkit. Our findings offer an in-depth comparative analysis of these algorithms, shedding light on their respective advantages and shortcomings. Additionally, we delve into a thorough discussion of each algorithm's features and their implications for cloud computing environments.

Keywords: Cloud Computing, Virtual Machine Consolidation, SLA, QoS, Data Centre.

1.0 INTRODUCTION

The rise of cloud computing has changed how customers use IT resources by giving them access to computers as a new utility, like gas or electricity. Since its inception, cloud computing has experienced rapid growth, establishing itself as a cornerstone of the modern economy. Government, academic institutions, and business companies are all major cloud computing users who have embraced and benefited from it greatly. Additionally, the quick establishment of new businesses, ease of worldwide commercial expansion, speedy advancement of research, and encouragement of the development of diverse applications and models are all made possible by cloud computing. Furthermore, cloud service providers offer a wide array of cloud services, granting users the convenience of pay and use on-demand access to resources. [1]–[3].

Cloud data centers make up the cloud computing infrastructure. Currently, a number of cloud service providers, notably Google, Amazon and Microsoft, have built sizable cloud data centers to meet consumer demand for resources and services. Cloud data centers must be operational around-the-clock in order to guarantee availability and dependability. Most data centers today are 300–4500 square meters in size and house 100 to 1000 of physical machines. The daily consumption of a data center can reach 30,000 KWh. By 2020, it is predicted that data centers will use 50 coal-fired power plants' worth of energy, or 140 million kWh. Additionally, it is predicted that the carbon footprint will account for 2-3% of world emissions [4]. Energy-based approaches are necessary in cloud data centers to reduce the significant energy usage and carbon emissions [5]–[7].

To increase energy efficiency, virtual Machine consolidation, which is the development of a running computer with an applications and operating system, is a crucial part of eco cloud data centers [2]. VM consolidation is the technique through which virtual machines can be transferred from one host to other hosts without effect the user requests. One of the most popular energy-based methods to lower cloud data centers' energy consumption has been recognized as virtual machine consolidation [5], [8]. Physical machines that are idle can be moved to the low-power mode or switch off as VMs are compressed onto fewer PM through consolidation [9].

Numerous energy-based methods based on Virtual Machine consolidation have been presented. Consolidating virtual machines has shown to be a successful strategy for lowering data center consumption energy. These algorithms seek to minimize energy use while maintaining other requirements, such as Service Level Agreement violations. It is challenging to carry out tests in a large-scale environment and duplicate results because of the uncontrolled network traffic. Running tests with a certified simulation toolset is therefore a valid and sensible method. The use of a simulation toolkit to create a large-scale environment and get repeatable findings is simple [10].

CloudSim [11] is the most popular cloud data center simulation toolkit currently available. Resource scheduling policies, workloads, VM and PM are all supported by CloudSim for cloud data canter's systems and behavior, respectively. The resource provisioning model is also generic, allowing for easy and minimal effort expansion. Users from hundreds of research institutions and universities have been drawn to these appealing characteristics, and along with CloudSim, some further extended simulators like Aneka [12]; NetworkCloudSim[13]; and CloudAnalyst [14]; have been created.

The aim of this paper is to compare different cutting-edge energy-based algorithms used in data centers in order to thoroughly assess the issue of escalating energy uses. The following facts in particular serve as motivation for this article:

• Understanding current Virtual machine-based energy-based methods for cloud data centers is required and in demand.

• The proposed algorithms' advantages and disadvantages were not thoroughly analyzed because they were tested in a variety of situations and configurations.

• The need to choose the most appropriate algorithm based on various priorities.

Employing the CloudSim toolbox, we conduct experiments using a variety of well-known consolidation VM-based energy-based strategies in this work. The algorithms considered for evaluation are modern energy-based ones that have shown great performance in energy. These are the primary contributions of this work:

• Presenting a cross-sectional look at the investigated energy-based consolidation VM-based strategies, which perform incredibly well in the cloud computing sector.

• Introducing a single analytical framework based on simulation that is built on CloudSim and allows for the fair and impartial evaluation and comparison of energy-based VM consolidation tactics.

• Outlining the advantages and disadvantages of the algorithms that were looked at in order to suggest solutions for specific circumstances.

The section of the paper is organized as follows: In Section 2, we give a summary of consolidation of virtual machines based on energy-based algorithms for cloud environments. In Section 3, It presents the investigated algorithms. The review algorithms' modelling is covered in Section 4, and Section 5 provides a summary of the metrics used by the investigated algorithms. The performance comparison of the examined algorithms is shown in Section 6. Finally, recommendations and trends for further research are provided.

2.0 LITERATURE REVIEW

Cloud-based energy-based algorithms In a few studies, classifications or surveys on consolidation of virtual machines based on energy-based methods for data centers have been undertaken. In their assessment on resource managing in the cloud computing, Mansouri et al. [15] covered Virtual machine consolidation-based energy-based methods from the perspective of the cloud management system. A thorough analysis of energy-based scheduling methods in clouds was published by Kaur et al. [16] who also examined several consolidation-based methodologies, such as the VM consolidation approach [5], [17]. Without putting much emphasis on energy efficiency, Gill [5] established a classification for VM consolidation-based methods and proposed a classification for eco-friendly cloud environments.

A review on virtual machines allocation in data centres from the perspectives of optimization methods and problem modelling was introduced by Mann et al. [18]. The similarities and differences of the examined VM migration methods are noted in [19] . Survey on the server consolidation and VM migration framework for cloud data centres. But instead of assessing the performance in experimental settings, these surveys and taxonomies concentrated on high level comparisons of Virtual machine consolidation-based energy-based techniques. Our work builds on earlier research by evaluating cutting-edge algorithms both from a modelling standpoint and through investigational comparisons. It also evaluates the positives and negatives of the examined algorithms to make recommendations for future study in relevant fields[20]. One application that allows for the use of several web services is cloud computing. These services consist of servers, databases, storage, and apps, among others. One of the things we are worried about in cloud computing is the heavy energy usage. Cooling is also required because customers are receiving computational services and a lot of heat is being produced. There's a lot of energy consumption. The large-scale energy use leads to the generation and release of more carbon dioxide into the atmosphere. The increased carbon emissions from data centers have the potential to cause adverse effects on the environment, such as global warming and other climate impacts. As a result, reducing the amount of energy used is necessary to improve the system's productivity and sustainability. Therefore, consuming less energy is necessary [39][40].

Algorithms for energy efficiency based on VM consolidation Modified Best Fit Decreasing is an energy-conscious data center resource allocation technique that Beloglazov et al. [21] devised (MBFD). Their goal is to lower data center energy usage while maintaining SLA. A probabilistic-based energy-based scheduling policy was proposed by Mastroianni et al. [22]. A comprehensive virtual machine scheduling system, developed by Li et al. [23], is capable of reducing all data center energy use, including cooling and processing energy. An approach energy based on learning automata and Service level agreement efficient VMS consolidation in cloud data centers was introduced by Ranjbari et al. [24]. To decrease the energy usage of data centers, Farahnkian [25]; introduced a revolutionary dynamic VM consolidation approach based on ACO. It is challenging to evaluate the effectiveness of various consolidation of virtual machines -based energy-based procedures because these approaches are not contrasted and evaluated collectively.

Unbalanced use of multidimensional resources in physical servers in the cloud computing environment causes resource fragmentation, which results in inefficient use of resources and energy waste in data centers. High energy consumption and poor quality of service (QoS) in resource management are important issues that need to be resolved due to inefficient resource utilization. We provide a load balancing approach based on virtual machine consolidation, which balances the multi-dimensional resource utilization in physical machines (PMs) with the goal of lowering energy usage and service level agreement (SLA) violations in data centers. In order to minimize needless virtual machine (VM) migrations brought on by sporadic load fluctuations, we first offer a load state classification algorithm for PMs with load irregularity that takes into account both current and future loads. Next, we suggest a selection model for migratable virtual machines (VMs) that is based on resource weight. This model minimizes resource fragmentation resulting from load imbalance by choosing suitable VMs for migration based on multi-dimensional resource utilization. In order to ensure load balancing of the destination PMs following VM placement, we lastly build a VM placement algorithm based on resource fitness and load correlation to deploy VMs on the best destination PMs. We conduct simulated trials in resource contexts that are heterogeneous, bottleneck, and homogeneous. According to experimental results, LBVMC outperforms other tactics in terms of overall performance and a reduction in energy usage and SLA breaches[41]-[44].

Using MBFD as the reference point, the effectiveness of various algorithms has been evaluated. In order to demonstrate the performance comparison, In this work, we thoroughly contrast various methods and rate them utilizing the same setups. M Sohani et al. [26] carried out of the earliest works in which energy management was used at the data center level. A method for reducing power usage in a mixed cluster of computing nodes supporting different web apps has been proposed by the authors in this study. The main approach employed to minimize power consumption involves reducing the number of active physical nodes and powering off idle nodes. However, workload consolidation can potentially impact the performance of applications, making it essential to carefully balance the trade-off between power efficiency and performance. SLAs provide requirements for execution time and application throughput in order to guarantee dependable QoS. The suggested algorithm decides when to turn on and off nodes to reduce overall power consumption while maintaining performance expectations by periodically assessing the load on resources (disc storage, CPU and network interface). The apps are responsible for managing the actual load balancing, which is not handled by the system. The method runs on a physical machine, creating a SPF and possibly slowing down system performance. One node can only be added or removed at a time using this method, and the writers have also noted that the reconfiguration procedures take many time, which may further contribute to the system's poor response in large-scale situations. The suggested method can be used in setups with fixed SLAs and several applications running at the same time.

The issue of energy-based management of same resources in internet hosting centers has been examined by Chase [27] . The main issue is determining each application's resource needs for the level of request demand it is facing and effectively allocating resources. The writers have used a low-cost framework to address this issue, in which services "bid" for resources based on quantity and quality. As a result, it is possible to negotiate SLAs that balance the charge of resource utilization with the advantage obtained from using this resource. The system keeps a running collection of servers that have been chosen to handle requests for service. The network switches are dynamically changed when necessary to switch the active servers. By moving idle servers to energy-saving modes, energy consumption is decreased. There is "noise" in the load data because the system is built to manage web workloads.

The suggested method has served as a basis for various research on power efficient resource allocated on the data center and is appropriate for multi-application setups with changing SLAs. Contrary to [28], the system simply manages the processor and does not take other resources. Additionally, the latency caused by turning on and off nodes is not included. The resources managing algorithm is quick while the workload is unchanging, but it becomes quite expensive when the workload undergoes considerable changes, according to the authors. Additionally, same to [28], different software configurations are not managed, a problem that can be fixed by using virtualization technology. With a single web application environment, program-assisted load balancing and fixed SLAs,[29] investigation looked at the issue of power-efficient resource management. Switching on and off the power of compute nodes and frequency scaling and dynamic voltage are two power-saving approaches used, as in [30]. (DVFS). The fundamental aim of the policy is to select the optimal number of nodes, calculate the total CPU frequency required to give the desired reaction time, and set a proportionate frequency for each node. However, the duration of the transition when a node's power is switched is not taken into account. It is anticipated that the system will only run one application, and, similar to [31], load balancing will be managed by an external system. The centralized approach reduces scalability and produces an SPF. Contrary [31] to, the resource utilization data are not approximated contempt the changeable nature of the workload, which could lead to judgments that are not as efficient due to fluctuations. It has never been done previously, but M. Xu, W. Tian, and R. Buyya [32] explored power managing strategies in the context of VM data centers. In addition to consolidating virtual machines and scaling up hardware, the authors have implemented a new power-saving method using resource software scaling. The goal is to simulate hardware scaling by giving a virtual machine a reduced amount of resource time utilizing the scheduling functionality of the Virtual Machine Monitor (VMM). In this paper, to demonstrate how these algorithms perform against one another, we compare them in-depth and assess them using the same configurations.

3.0 OVERVIEW OF THE EXAMINED METHODS

Cloud providers aspire to revolutionize the design of future data centers by creating networks of software services encompassing application logic, user interfaces, databases, and hardware. Users are empowered to deploy applications and access these services from any geographical location on demand, offering competitive pricing based on the desired Quality of Service (QoS) levels [33]. The architecture for enabling Load balancing and energy-based service allocated in a cloud computing infrastructure in Figure 1[34]. There are generally four key parties involved:

- Users: From any location in the world, Cloud users or brokers can submit requests for service. The distinction between users of deployed services and Cloud users should be noted. Using a web application as an example, a consumer may be a business that deploys it. The workloads provided by the application can vary depending on how many users visit it.
- Allocator Service: serves as the interface between the Cloud environments and users. To facilitate energy-based resource management, it is necessary for the following elements to interact:
    a) Energy Monitor: Observes the energy use of physical and virtual machines and gives the VM manager with this data so they may allocate resources in an energy-based way.
    b) Workloads Scheduler: Allocate requests to virtual machines and establishes resource for the VMs that have been allotted. Additionally, if a customer has requested the scaling functionality, it decides when to remove or add virtual machines to meet on demand.
    c) Virtual Machines Manager: This component monitors how much resource each VM uses and decides when and where to condense VMs based on their actions. It needs the SLA and energy data from SLA Monitor and Energy Monitor to accomplish this goal.
    d) Service Analyzer: Interprets and evaluates a given request's service before accepting it. Therefore, it requires the most recent data on energy and load from Energy Monitor and VM Manager, respectively.
    e) SLA Monitor: It keeps track of how system operations affect SLA. When the system's energy usage is reduced, it can also signify performance limitations.

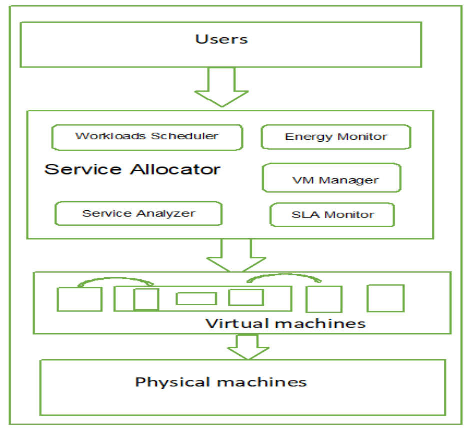


Figure 1. VM consolidation-based energy-based scheduling.

- Virtual machines: An application are set up and run on virtual machines. The initial situation and VM migration phases allow the VMs to be controlled in accordance with the incoming workloads. VMs are initially allocated to actual computers using the initial placement mechanism. The placement can be adjusted using a VM consolidation method based on workloads, and as a result, unneeded machines can be momentarily put into low-power mode or switch off. On a physical machine, more than one virtual machine can be dynamically started and halted in response to incoming requests, enabling administrators to set up different resource partitions on the same physical machine to satisfy different service request requirements. Many virtual machines can run programmes from different operating systems simultaneously on a physical machine. By constantly moving virtual computers among actual machines, this is achieved.

- Physical Machines: To fulfil user requests, the environment provides physical computers to provision resources. In order to meet service demand, the underlying physical computer servers offer the hardware for producing virtualized resources.

Table 1. Comparative analysis of Algorithms based on parameter

Algorithm	Application	Environment Operation	Function Objective	Energy Component	Power System
MBFD [21]	Dynamic workloads (web service)	Distributed and mixed	efficient energy consumption	Memory and CPU	Linear
GRANITE [23]	Heterogeneous	Distributed	To examine the relationship between server CPU temperature and airflow temperature	memory, Cooling, storage, CPU and network	Linear
LAOD [24]	Dynamic workloads	same	number of virtual machines migrations, SLA V and optimize energy consumption,	CPU	Linear
ACS [25]	Memory and CPU workload	Mixed and Distributed	To examine the relationship between VM migrations, QoS and energy consumption	CPU and Memory	Linear

We carefully choose cutting-edge VM consolidation-based methods for our comparisons and evaluations based on the following standards:

The algorithms can serve as a representative of a group of algorithmic techniques and were published in well-known publications or conferences, which improves the point for comparison. The algorithms were put into practice in CloudSim, where they can be swiftly assessed to ensure equivalent evaluation outcomes. The algorithms should have been evaluated against the same standard in order to make them comparable.

Table 2 Comparative analysis of Algorithms based on parameter

Algorithm	Scheduling Mechanism	Workloads	Advantages	Disadvantages
MBFD [21]	Dynamic Consolidation (Proactive)	Mixed	Optimize energy consumption and SLA violation rate	There is a need of holistic resource management
GRANITE [23]	Scheduling using 2D computational fluid dynamics models (Proactive)	Google Data center	Reduce total datacenter energy (cooling and computing)	To improve accuracy, 2D CFD model can be extend to 3D CFD model
LAOD [24]	Based on Learning automata (proactive)	CPU utilization	Improved CPU utilization and reduced SLAV and energy uses	Under-utilization of resource is not considered
ACS [25]	Based on Ant colony optimization (reactive)	Mixed	R VM migrations and reduced energy	To further cut down on power usage, it is possible to evaluate how VM movement affects network bandwidth.

We will describe the overview of the algorithms we looked into in the following subsections. Table 1 and 2 compares the algorithms under investigation according to a variety of factors, including operating environment, application type, scheduling mechanism, function objective, scheduling criteria, workloads, stated benefits and drawbacks of each algorithm.

3.1 Modified Best Fit Decreasing

MBFD (Modified Best Fit Decreasing) [21] tries to lower data centres' energy use while maintaining SLA. It is solved by approaching the VM initial placement step as a bin packing issue. The goal of MBFD is to place virtual machines on hosts with the smallest possible increase in energy consumption. During the phase of virtual machine consolidation, where the algorithm optimises the distribution of virtual machines through consolidation for greater efficiency energy, the target host is also chosen. The unpredictability of task growths—where certain virtual machines are likely to host overly provision programmes while others operate effectively—served as the basis for the suggested strategy. In cloud data centres, unbalanced workloads waste a lot of resources and reduce performance. This work's proactive VM consolidation can be used with a variety of workloads [35].

Many energy-based VM consolidation algorithms have been tested with MBFD as their baseline. To boost the effectiveness of this method, several new algorithms have been developed. This method has the benefit of being simple to construct and taking interface between SLA violation and lower energy use. The drawback is that recent study, which is an addition, complements it by taking into account holistic resource management. The particular data structures that are utilized to keep track of the list of bins and their available space determine how time-consuming the bin packing stage is. The total time complexity for the bin packing step can be O(n^2) in a basic implementation where the best-fit bin is found using a linear search, where n is the number of elements.

3.2 Greedy based Scheduling Algorithm minimizing Total Energy

A comprehensive virtual machine scheduling system called GRANITE (GReedy based scheduling Algorithm miNImizing Total Energy) [23] is capable of reducing the amount of energy used for computation and cooling in data centres overall. The models in this work are based on the computer room air conditioners (CRACs), which are the only cooling devices. Greedy algorithms are used by GRANITE to carry out the initial placement of the VMs and dynamic migration, which are based on server and cooling model assumptions. They presumptively believe it is possible to forecast user resource demands. Across all VMs in GRANITE, the greedy approach is used to select the host with the smallest increase in overall energy after placement during the initial placement phase. CRAC will change if the CPU temperature rises beyond the threshold. The programme seeks to strike a balance between cooling energy usage and workloads during the dynamic consolidation VMs stage. In addition to determining a dynamic temperature threshold, the GRANITE also verifies status of host. If the increase temperature of above the host's threshold, a set of virtual machines will be relocated to another host. The same greedy method used for the initial placement is utilised to select the target host for migration [36].

The algorithms used in cloud data centres can be seen in GRANITE and take into account the total management of energy. While taking cooling power into account, the method's basic concept is similar to that of the MBFD algorithm, producing scheduling results that are more precise and comprehensive. This research has the advantage of combining server status with data centre temperature control to produce precise energy-based scheduling. Utilizing a 3-D computational dynamics model as opposed to a 2-D one, however, might further increase the model accuracy. In the worst scenario, the sorting phase would account for the majority of the Greedy-based Scheduling Algorithm's total time complexity, or O(n log n).

3.3. Learning Automata Overload Detection

The LAOD (Learning automata overload detection) [24] method uses learning automata to VM consolidate in cloud data centres in a way that is both energy and SLA efficient. The suggested approach takes user resource demands into account when predicting overloaded hosts. The suggested technique intends to reduce data centre energy usage by avoiding overloaded hosts and power off idle hosts.

By forecasting hosts' CPU utilization based on past resource usage, overload detection by learning automata improves the VMs consolidation. One automaton with three actions reducing utilization of CPU, maintaining utilization of CPU and increasing utilization of CPU is installed in each virtual machine. The three acts have equal probability at the start. In each iteration, the automata choose any one of the possible actions based on likelihood. In addition, if the automata made the right decision, the action will be rewarded in the next iteration; otherwise, it will be punished. The learning automata are used to calculate the host's estimated VM utilization. If the host shows the overloaded, the virtual machines will be transferred, and other virtual machines under load will not be permitted to migrate to the host. The BFD (Best Fit Decreasing) algorithm underlies the transferred destination [37]. The outcomes of the simulation demonstrate that prediction based on learning can lower the energy uses of data centres.

LAOD is a characteristic approach that optimizes VM consolidation by using learning techniques. By taking into account the dynamic prediction for resource usage, this work improves on previous research. The drawback, however, is that this study only addresses the underutilized circumstances and anticipates the overcrowded ones. Learning automata-based algorithms are typically employed in dynamic and adaptable environments. The number of steps or iterations needed for the algorithm to converge to an optimal or nearly optimal solution can be used to analyse the temporal complexity of these algorithms.

3.4 Ant Colony System

Consolidation of VMs and Ant colony optimization are the foundations of Ant Colony System (ACS) [25], a meta-heuristic online optimization method that aims to find a solution that is close to ideal. Its goal is to strike a compromise between performance-related QoS and energy use, VM migration frequency, and number. In this method, the writers design the energy-based virtual machines consolidation as a multiple goal optimization problem to maximize several measurements at once. The essential components, such as probabilistic decision rules and pheromone update rules, are defined in order to use ACO. If a solution has more trails pheromone, the likelihood of installing the VM on the host rises. Global and local pheromone updating rules are also included in ACS and are used in each iteration. Iteratively, each time an ant moves, the local pheromone is updated. Once the migration process' global pheromone update is finished, only the dominant location will be preserved following the local building of each ant's solution. Until the maximum number of iterations have been reached, the process is repeated [38].

ACS is a group of meta-heuristic algorithms that have been put forth to balance various goals. According to simulation results, the suggested technique can VM migrations and lower energy use while maintaining Quality of service. Observing how VM movement affects the network might further improve performance. Overall, the scheduling procedure based on energy depicted in Fig. 1 is followed by all of the algorithms under investigation. The examined algorithms use multiple methods to optimise the placement of virtual machines. With the exception of EcoCloud [39], the majority of the algorithms under consideration focus on the initial placement utilising a probabilistic method. Through modelling cooling energy usage and VM performance decline, GRANITE takes energy and performance into account simultaneously. The optimized consolidation solutions are discovered by ACS using a meta-heuristic method. The product of the number of iterations (generations) and the time complexity of solution construction is frequently used to represent the time complexity of ACO. The total time complexity can be roughly expressed as O(G * N), where G is the number of iterations and N is the issue size, if the solution construction phase is linear and the number of iterations is a user-defined constant.

4.0 MODELING AND ARCHITECTURE

The evaluated algorithms are discussed in this section from the modelling and architecture viewpoints, and the complexity of the algorithms is also considered.

4.1 Modified Best Fit Decreasing

Architecture: The CPU, disc storage, RAM and network interfaces are the primary determinants of power consumption by nodes in cloud data centers. The CPU more energy utilize as compared to other system resources, hence in this work we concentrate on controlling and using it effectively. Additionally, the relationship between overall system load and CPU utilization is often linear. Broker, VMs, PMs and service allocator, are the four essential elements of the green cloud architecture. The user enables the user interface account to send workloads request and related quality of services requests from any environmental distribution allocation. Resources are virtualized using hardware infrastructure powered by PM. Utilizing DVFS, VMs are combined to dynamically meet workload demands. The eco service allocator integrates the manager of virtual machines and energy management to allocate resources to user in accordance with their runtime execution requirements.

Model: The power model for this research is defined in equation (1).

$$Q(v) = m.Q_{max} + (1 - m) . Q_{max} . v \qquad (1)$$

where Qmax is greatest power usage while the server is utilized; m is the minimal power consumption of a server that isn't in use; and v CPU utilization. equation (2) defines the value of a PM's energy consumption C. v(t), is a time-dependent quantity on CPU utilization.

$$C = \int_{t2}^{t1} P(v(t))\, dv \qquad (2)$$

C= where t1 and t2 is the start and end time of Task T.

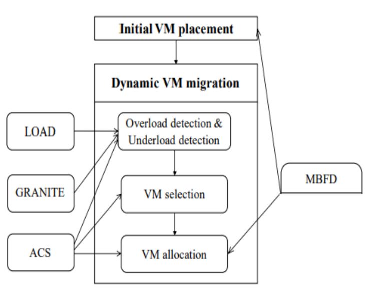


Figure 2. Scheduling based Competitive analysis of Examined algorithms

4.2 Learning Automata Overload Detection

Architecture: Four subcomponents of the system design, include a global Manager, a user portal, a platform in between locally based manager and user portal, locally based manager, which are managed by a single global manager.

Model: The energy consumption for this research project is calculated using equation (1), which depicts the linear model for CPU usage and power.

4.3 Greedy based Scheduling Algorithm Minimizing Total Energy

Architecture: It consists of three supporting elements, namely administrator for workload, scheduling, and cooling. The workload administrator is in charge of the tasks that Users submit and the scheduling processes that are carried out in accordance with their requirements. The scheduling administrator arranges for the execution of the resources of workloads while enhancing data performance minimizes and focuses energy consumption Cooling. Manager keeps data centers at a comfortable temperature and conserves cooling energy through dynamic VM placement migration with effectiveness.

Model: A linear power model is defined in equation (3) to determine the combined processing and cooling energy used by data centres to power their operations.

$$E= E_{computing} + E_{cooling} \qquad (3)$$

Where E is energy.

4.4 Ant Colony System

Architecture: In this architecture, there are two different global and Local agents are among the agents. Local representative is used to monitor host resource usage and address the host status detection subproblem. The Global agent is in charge of overseeing and maximising taking benefit of the ACO-based to position the VM algorithm.

Model: The linear power model is used by ACS as the MBFD in equation 2.

We see that the examined algorithms differ from one another. primarily emphases on the stages of consolidation of VM, and we demonstrate Figure 2 focal points 2. The VM consolidation, as described, two segments that make up the majority of the process: the beginning dynamic VM migration and VM placement. in addition to the overloads and under-used detection, the VM migration.

The inclusion of VM allocation and VM selection. LAOD, on the other hand, concentrates on detection of overloads by learning-based usage prediction automated system to enhance VM overload detection. The CPU temperature must be taken into account, and the overload detection process is improved by GRANITE. ACS enhances every stage in dynamic consolidation of VMs. LiRCUP [29] is used to make forecasts. If servers overloaded than find the close to ideal choices for VM allocation and selection by using the Ant Colony System. To sum up, all of the investigated algorithms use layered architecture from an architectural standpoint. Layers can generally be divided into three types. The resource provisioner, at the base layer, provides virtual or physical resources. The energy comes from the middle layer. effective scheduling, which manages the virtual machine, and a list of scheduling algorithms that use less energy. Users' requests and optimization goals are set up at the layer towards the top.

All algorithms follow the modelling perspective's the linear model. All algorithms based on energy components incorporate the MBFD, and CPU energy consumption ACS consider the memory component, while GRANITE makes use of a broader complete model that takes into account cooling, networking, and storage.

Algorithm complexity analysis: Analysis of the difficulty of the MBFD, GRANITE, and Heuristic algorithms underlie LAOD, and their complexity each of them is P X Q, where P is how many VMs there are Based and Q is the total number of hosts. The ACS complexity is dependent on a meta-heuristic with iterations. M X N X A X I, where A is the quantity of ants that are present simultaneously, I is the number of iterations used in building their migration plans.

5.0 METRICS

Energy consumption is the main indicator to be assessed in order to achieve the aim of energy efficiency. However, the algorithms also balance other criteria, such SLA violations, against other data, like as energy consumption. In this section, we go over the metrics that were used in our examined energy-based algorithms. Keep in mind that while the examined algorithms employ some comparable measures, they also use some additional metrics.

In this article, we'll discuss the metrics used in these algorithms and point out how they differ from one another. Table 3 lists the algorithms and the accompanying adopted measures for each one. Energy efficiency metrics Total amount of energy used: It is the entire amount of energy used by equipment in data centres. It is developed from the energy model of equation (2). Number of servers running: It displays the number of servers that were active during the observation period. More sleep mode servers can be shifted to low-power mode by reducing the value. SLAV percentage metrics [21]: The ratio of service level agreement breaches to all events processed during the time period is what determines the percentage. When a certain VM does not receive the requested quantity of MIPS, the SLA is broken. Moving VMs takes time: The quantity of migrations that the algorithm's VM scheduling procedure has caused. Typical SLA violation Performance deterioration is caused by the average CPU performance that was not allocated to a program when requested.

In conclusion, we can see that a number of measures, including SLA violation, active hosts, average SLA violations, VM migrations and total energy, have been selected for evaluations by more than one algorithm. making our We examine these measures in the section on performance evaluations since they are more similar from a metrics standpoint.

Table 3. Metrics adopted based Comparative analysis

Metrics	Optimization Objective	Algorithms
SLA violation percentage	Minimization	ACS, LOAD, GRANITE, MBFD
Virtual machine migrations	Minimization	LAOD, ACS
Active Host	Minimization	GRANITE
Consumption of Energy	Minimization	ACS, LOAD, GRANITE, MBFD
Average SLAV	Minimization	ACS, LAOD

6.0 PERFORMANCE ASSESSMENTS

In this section, on the basis of different performance metrics and two traces, tests for the four well-known and examined algorithms are run in order to compare the effectiveness of the algorithms under consideration. As baselines, we also include one algorithm that is offered by CloudSim [40]–[42], which controls overload detection based on dynamic threshold.

6.1 Experiments Settings

Each host has two Processing cores with MIPS of 1980 or 2760, 8 GB RAM, and 512 GB of storage, depending on the host. We use the HP ProLiant ML110 G4 or G5 power model, which was utilized in [21]and [25]. For VM setups, four different VMs with Million Instructions Per Second of 600, 1100, 1600, and 2600 are taken into consideration. The no. of virtual machines for each category is chosen at random.

We conduct a number of experiments using fictitious workloads first, and then we use the actual workload data from the CoMon project given by PlanetLab [43] to replicate a genuine cloud data centre. Data on the CPU usage of hundreds of virtual machines assigned to servers spread over more than 500 locations worldwide is included in the workload. Additionally, for 10 days, data is gathered every five minutes, simulating the workload in a genuine cloud system.

Energy use, virtual machines Migration, SLAV, and the Active Hosts are the four measures we choose to assess how well these algorithms perform. We selected these indicators because, as we covered in Section V, they have gained widespread adoption and are utilised in a number of algorithms. Due to the page restriction, we analyse SLAV rather than the average number of SLA breaches because it better represents SLA violations.

6.2 Implementation Details

For the updated learning automaton in LAOD, the penalty and reward parameters, a and b, are both set to 0.1. Although we lack the training set that would enable us to calculate the starting pheromone level, we nevertheless employ the parameters from the initial ACS investigation. As a result, we set P to be the total no. of virtual machines and M to be the no. of under-utilized servers. The original paper's configuration parameters are used with MBFD and GRANITE.

6.3 Synthetic Workloads

The experiments using simulated workloads, used in MBFD and ACS to show performance under such workloads, are the focus of the assessments' first portion.

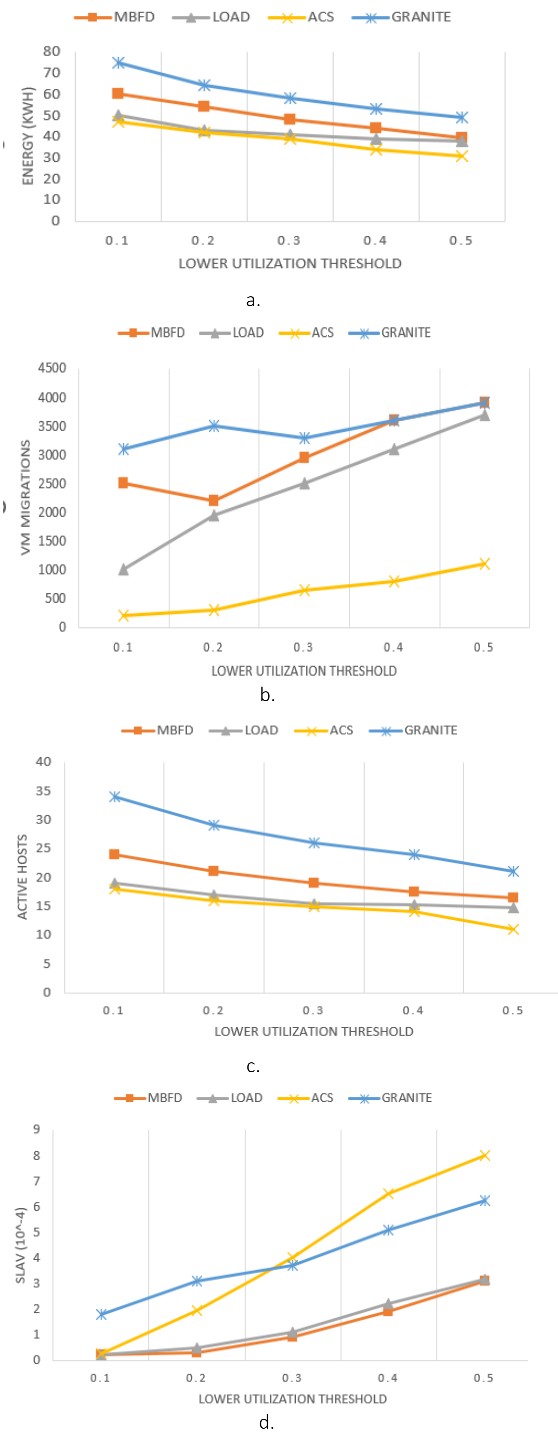


Figure 3. Using Synthetic workloads comparative analysis of algorithms based on performance (ration 1:1 VM and PM) a. Consumptions of Energy b. Virtual Machine Migrations c. Active hosts d. SLA violation

We generate the same number of physical machines and virtual machines while altering the lower utilization criterion in order to identify when a host is underutilised [44]. And with a 0.1 increment, we changed the threshold from 0.1 to 0.5. The difference between the higher and lower utilization thresholds is set at 0.4. We configure the utilization threshold interval as stated in [45]. We randomly construct workloads for the tests under each configuration, run them, and then repeat 10 times. The fifty set as number of PMs and VMs.

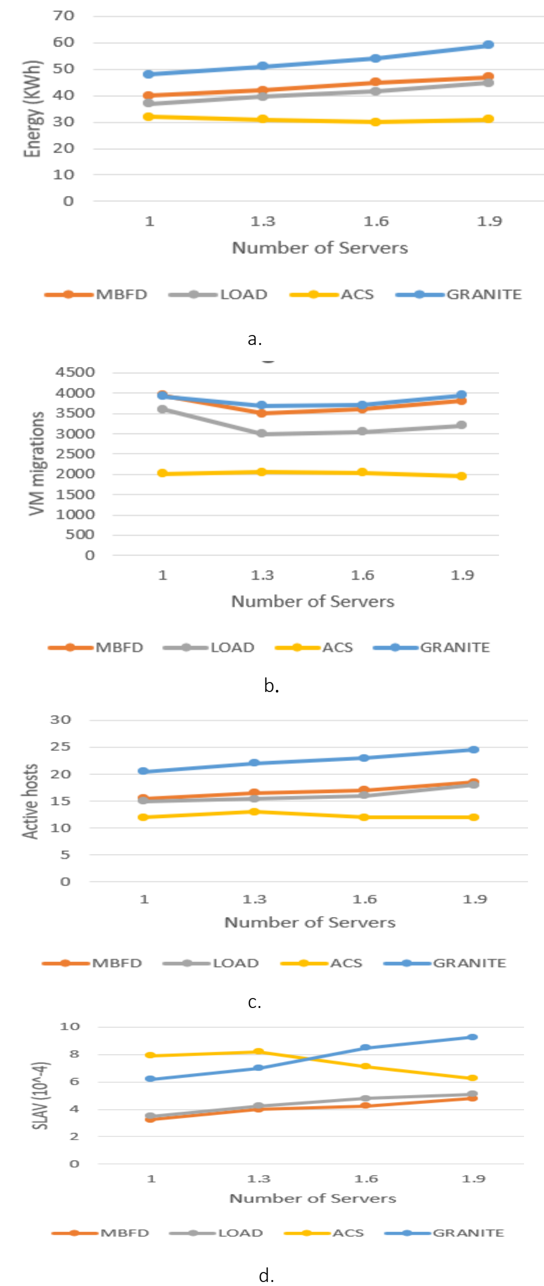


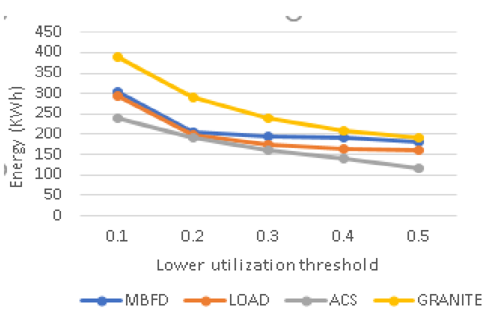
Figure 4. Using Synthetic workloads comparative analysis of algorithms based on performance (setting varying rations of PMs and VMs number are 1:1, 1:1.3, 1:1.6, and 1:1.9) a. Consumptions of Energy b. Virtual Machine migrations c. Active hosts d. SLA violation

The outcomes of the studies with simulated workloads are displayed in Figure 3. According to the findings, all methods may achieve reduce energy consumption, more SLA violation, and less virtual machines migrations with higher values of lower utilization threshold. To be more precise, ACS outperforms MBFD in terms of energy usage by lowering power by 21.1%. Compared to other techniques, ACS needs substantially fewer migration times—less than 600. The fact that ACS has the fewest active hosts is what allows it to achieve the best energy efficiency. These methods work better when we set to value 0.1 threshold for SLA violation comparison.

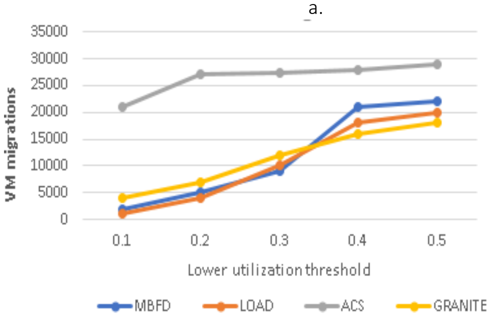
We set 0.5 value for utilization threshold and run this process 10 times to show the change of results, which are Setting 0.5 lower utilization threshold value can obtain the performance for all methods. It is clear that ACS can reduce the typical number of VM migrations to 2020.5, while still achieving the highest energy consumption performance with 31.3 kWh.
We continue to use the lower utilization threshold of 0.5 while simultaneously maintaining the following system ratios: 1:1, 1:1.3, 1:1.6, and 1:1.9, to explore the effects of varied numbers of PMs and VMs. The trials are carried out ten times for each ratio, with the results displayed in Figure 4. We can see that as increases the number of virtual machines, so does the energy use. The most energy-based algorithm is ACS, which uses between 21.2-34.4% less energy than MBFD when the ratio is 1:1 and 1:1.9, respectively. The best outcomes in VM migrations are obtained using ACS. Although the researched algorithms can reduce energy more than GRANITE, they also experience more SLAV.

6.4 PlanetLab Workloads

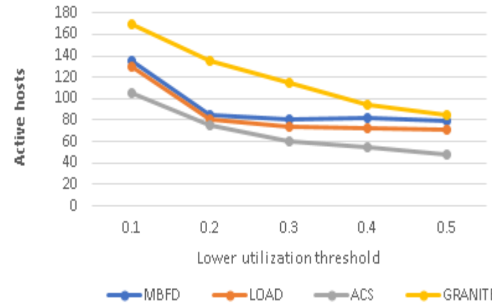
We also run tests with workloads from PlanetLab to show how well the algorithm works with real-world data. The lower CPU utilization threshold is adjustable between 0.1 and 0.5, and a fixed 0.4-second gap separates the low threshold from the high threshold. The configured number of hosts is 900, and PlanetLab traces are used to determine the number of virtual machines. The average results from 10 experiment runs, each having a PlanetLab trace for a day, Figure 5 are displayed.



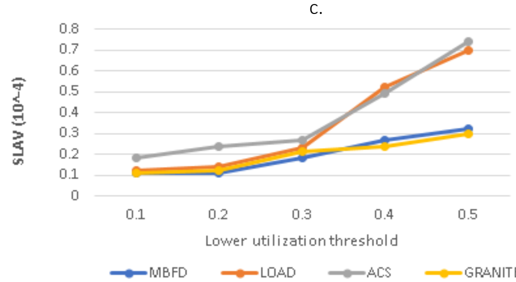
a.



b.



c.

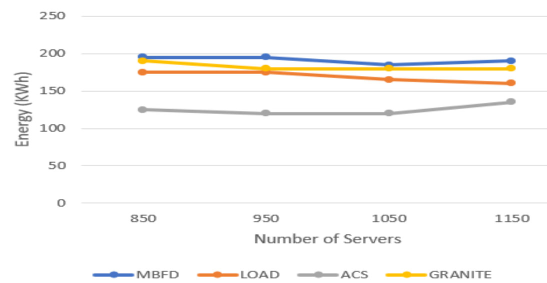


d.

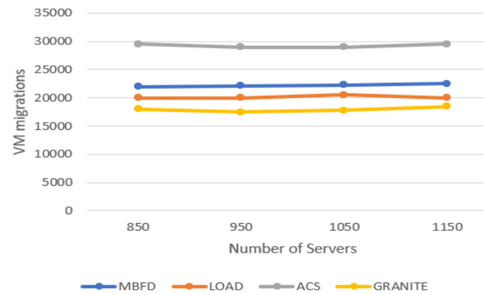
Figure 5. Using PlanetLab workloads comparative analysis of algorithms based on performance a. Consumptions of Energy b. Virtual machines migrations c. active hosts d. SLA violation

Energy usage is compared in Figure 5a, where more power is consumed when the lower utilization threshold is exceeded. In comparison to other algorithms, ACS uses the least amount of energy (148.7 kWh) when the lower utilization criterion is 0.4. Due to the fact that GRANITE keeps more servers active than MBFD, it uses more energy. In Figure 5b, the numbers of VM migrations are contrasted. Compared to MBFD, LAOD reduces migrations by 12.4% while improving the number of migrations. With a rise in the lower usage threshold, GRANITE produces greater outcomes. Figure 5c compares the active hosts, and Ant colony System can power off a maximum number of hosts. The comparison of the SLAV percentage is shown in Figure 5d, and when the lower utilization threshold rises, so does the SLA violation percentage. As shown in the graphic, MBFD perform better on this parameter than LOAD and ACS. GRANITE keeps the rate of SLAV low and so ensures the calibre of services.

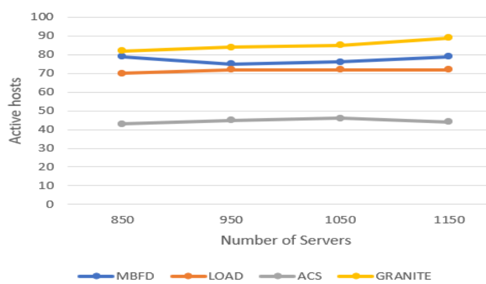
To demonstrate the variation in performance results, we repeatedly ran 10 trials with the lower utilization threshold set to 0.5 and the higher utilization threshold set to 0.9. The results are given in Figure 6. As shown in Figure 6a, ACS and LAOD outperform other baselines in terms of energy consumption. The ACS uses an average of 125.2 KWh of energy while the GRANITE and MBFD use more than 185 KWh. The results for the active hosts are shown in Figure 6c, and Ant Colony System can produce the greatest outcomes with an average of 48 hosts. The comparison of virtual machines migrations is shown in Figure 6b, and GRANITE emphases on optimising this metric by lowering the number of virtual machines migrations to be under 20000. VM migrations are decreased while using LAOD as opposed to MBFD and GRANITE. The comparison of SLA violation percentages is shown in Figure 6d, and ACS performs worse in decreasing SLA violations with 0.71 104, although saving more energy.



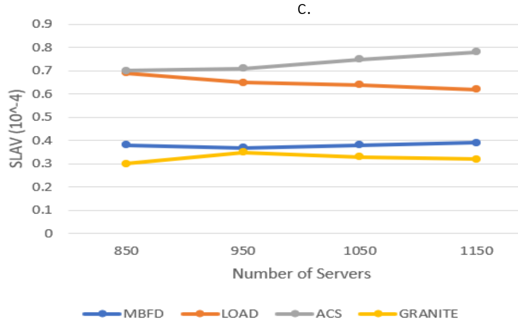
a.



d.



c.



d.

Figure 6. Using PlanetLab workloads comparative analysis of algorithms based on performance (varied number of servers with 850, 950, 1050 and 1150) a. Consumptions of Energy b. Virtual Machines migrations c. active hosts d. SLA violation

In conclusion, it is clear that ACS, which has the fewest active hosts of the two workloads, typically achieves the best energy efficiency. When addressing PlanetLab workloads, MBFD outperforms competing methods in terms of fewer VM migrations and SLA breaches. Since heuristic techniques only search a small portion of the available solution space, As can be shown, ACS (a meta-heuristic algorithm) uses less energy than heuristic algorithms.

As the original consolidation of virtual machine based on energy-based method for data centres, Modified Best Fit Decreasing has gained popularity due to its simplicity and effectiveness. Even if the speed of Modified Best Fit Decreasing has been surpassed by subsequent algorithms, the fundamental idea of Modified Best Fit Decreasing has been mentioned, for example in LAOD and GRANITE where the SLAV and energy are optimised in accordance. GRANITE is suggested in the circumstance were optimising more energy consuming components than simply the PMs is the goal. When the network is the system's bottleneck, for example, MBFD is an excellent choice because it can drastically reduce the number of VM migrations. If future resource demand can be properly forecast, LAOD can perform well. Therefore, it is ideal if the system has enough historical data on resource utilization or if resource usage exhibits a consistent trend, like Wikipedia.

7.0 CONCLUSIONS AND FUTURE WORK

This research delves into five advanced energy-efficient techniques for cloud data centers, centered on VM consolidation. In cloud computing, VM consolidation is a more important factor in improving energy utilization. We explore all the algorithms to improve energy utilisation based on VM consolidation. We explore these algorithms through various lenses, including their foundational principles, architectural frameworks, mathematical models, and computational complexity. Implemented in CloudSim, these techniques are tested using both synthetic and PlanetLab traces, demonstrating their proficiency in reducing energy consumption while balancing other metrics like SLAV and VM migrations.

Future research directions, based on our findings, could include:

- Investigating dynamic threshold configurations for energy consumption, as threshold utilization settings significantly impact energy use.
- Broadening the focus beyond CPU as the primary energy model component, to explore the interplay between energy consumption and other resources, such as networking.
- Evaluating these techniques using contemporary workloads from sources like Google and Alibaba to further validate their effectiveness.

Acknowledgement

The authors like to acknowledge Medi-Caps University, Indore, India for providing research facilities to the first author of this paper during this work.

References

[1] R. Buyya et al.,2018. A Manifesto for Future Generation Cloud Computing, *ACM Computing Surveys*. 51(5): 1-38. doi: 10.1145/3241737. https://doi.org/10.1145/3241737

[2] R. Yadav, W. Zhang, K. Li, C. Liu, M. Shafiq, and N. K. Karn,2018. An adaptive heuristic for managing energy consumption and overloaded hosts in a cloud data center, *Wireless Networks*. 26(3): 1905-1919, doi: 10.1007/s11276-018-187. https://doi.org/10.1007/s11276-018-1874-1

[3] S. S. Gill et al.,2019. Transformative effects of IoT, Blockchain and Artificial Intelligence on cloud computing: Evolution, vision, trends and open challenges, *Internet of Things*.8: 100-118. DOI: https://doi.org/10.1016/j.iot.2019.100118

[4] Brian Lavallée 2014.Data center energy: Reducing your carbon footprint-data center knowledge. http://www.datacenterknowled

ge.com/archives/2014/12/17/ undertaking challenge-reduce-data-center-carbon-footprint.

[5] S. S. Gill and R. Buyya, 2018. A Taxonomy and Future Directions for Sustainable Cloud Computing, *ACM Computing Surveys*,51(5): 1-33, doi: 10.1145/3241038. https://doi.org/10.1145/3241038

[6] R. Ghosh, S. P. R. Komma, and Y. Simmhan, 2018. Adaptive Energy-Aware Scheduling of Dynamic Event Analytics Across Edge and Cloud Resources, *in 18th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID), IEEE*, doi: https://doi.org/10.1109/CCGRID.2018.00022

[7] O. Soualah, M. Mechtri, C. Ghribi, and D. Zeghlache,2017. Energy Efficient Algorithm for VNF Placement and Chaining, *in 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID), IEEE,* doi: https://doi.org/10.1109/CCGRID.2017.84

[8] R. Yadav, W. Zhang, K. Li, C. Liu, and A. A. Laghari,2021. Managing overloaded hosts for energy-efficiency in cloud data centers, *Cluster Computing*. 24(3):2001-2015, doi: https://doi.org/10.1007/s10586-020-03182-3

[9] M. Xu and R. Buyya,2019. Brownout Approach for Adaptive Management of Resources and Applications in Cloud Computing Systems, *ACM Computing Surveys,*52(1): 1-27, doi: https://doi.org/10.1145/3234151

[10] W. Tian, M. Xu, A. Chen, G. Li, X. Wang, and Y. Chen,2015. Open-source simulators for Cloud computing: Comparative study and challenging issues, *Simulation Modelling Practice and Theory.* 58: 239-254, doi: https://doi.org/10.1016/j.simpat.2015.06.002

[11] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. De Rose, and R. Buyya, 2011.CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms, *Software: Practice and Experience*, 41(1): 23-50, doi: https://doi.org/10.1002/spe.995

[12] Vecchiola C., Chu X., and Buyya R.,2009 Aneka: a software platform for.Net-based cloud computing, Advances in Parallel Computing, pp. 267-295.

[13] S. K. Garg and R. Buyya, 2011. NetworkCloudSim: Modelling Parallel Applications in Cloud Simulations, *in Fourth IEEE International Conference on Utility and Cloud Computing, IEEE,* doi: https://doi.org/10.1109/UCC.2011.24

[14] B. Wickremasinghe, R. N. Calheiros, and R. Buyya,2010. CloudAnalyst: A CloudSim-Based Visual Modeller for Analysing Cloud Computing Environments and Applications, *in 24th IEEE International Conference on Advanced Information Networking and Applications, IEEE.* doi: https://doi.org/10.1109/AINA.2010.32

[15] Y. Mansouri, A. N. Toosi, and R. Buyya,2017. Data Storage Management in Cloud Environments, *ACM Computing Surveys*. 50(6): 1-51, doi: https://doi.org/10.1145/3136623

[16] T. Kaur and I. Chana,2015 Energy Efficiency Techniques in Cloud Computing, *ACM Computing Surveys*. 48(2): 1-46, doi: https://doi.org/10.1145/2742488

[17] A.-C. Orgerie, M. D. de Assuncao, and L. Lefevre,2014. A survey on techniques for improving the energy efficiency of large-scale distributed systems, *ACM Computing Surveys*,46(4): 1-31, doi: https://doi.org/10.1145/2532637

[18] Z. Á. Mann,2015 Allocation of Virtual Machines in Cloud Data CentersA Survey of Problem Models and Optimization Algorithms, *ACM Computing Surveys*.48(1): 1-34, doi:. https://doi.org/10.1145/2797211

[19] R. W. Ahmad, A. Gani, S. H. Ab. Hamid, M. Shiraz, A. Yousafzai, and F. Xia,2015. A survey on virtual machine migration and server consolidation frameworks for cloud data centers, *Journal of Network and Computer Applications*,52: 11-25, doi: https://doi.org/10.1016/j.jnca.2015.02.002

[20] M. Sohani and S. C. Jain,2018 State-of-the-Art Survey on Cloud Computing Resource Scheduling Approaches, in Advances in Intelligent Systems and Computing, *Springer Singapore*. 629-639. doi: https://doi.org/10.1007/978-981-10-7386-1_53

[21] A. Beloglazov, J. Abawajy, and R. Buyya,2012 Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing, *Future Generation Computer Systems*. 28(5): 755-768, doi: https://doi.org/10.1016/j.future.2011.04.017

[22] C. Mastroianni, M. Meo, and G. Papuzzo,2013. Probabilistic Consolidation of Virtual Machines in Self-Organizing Cloud Data Centers, *IEEE Transactions on Cloud Computing*.1(2): 215-228, doi: https://doi.org/10.1109/TCC.2013.17

[23] X. Li, P. Garraghan, X. Jiang, Z. Wu, and J. Xu,2018 Holistic Virtual Machine Scheduling in Cloud Datacenters towards Minimizing Total Energy. *IEEE Transactions on Parallel and Distributed Systems*. 29(6): 1317-1331, doi: https://doi.org/10.1109/TPDS.2017.2688445

[24] M. Ranjbari and J. A. Torkestani,2018. A learning automata-based algorithm for energy and SLA efficient consolidation of virtual machines in cloud data centers. *Journal of Parallel and Distributed Computing*. 113: 55-62, doi: https://doi.org/10.1016/j.jpdc.2017.10.009

[25] F. Farahnakian et al.,2015 Using Ant Colony System to Consolidate VMs for Green Cloud Computing, *IEEE Transactions on Services Computing* 8(2): 187-198. doi: https://doi.org/10.1109/TSC.2014.2382555

[26] M. Sohani and Dr. S. C. Jain,2021. Threshold based VM Placement Technique for Load Balanced Resource Provisioning using Priority Scheme in Cloud Computing. *International journal of Computer Networks &amp Communications* 13(5): 1-18, doi: https://doi.org/10.5121/ijcnc.2021.13501

[27] Moore J., Chase J., and Ranganathan P.,2006. Weatherman: automated, online, and predictive thermal mapping and management for data centers, *in International Conference on Autonomic Computing.*

[28] R. Yadav, W. Zhang, H. Chen, and T. Guo,2017. MuMs: Energy-Aware VM Selection Scheme for Cloud Data Center, *in 28th International Workshop on Database and Expert Systems Applications (DEXA), IEEE.* doi:. https://doi.org/10.1109/DEXA.2017.43

[29] F. Farahnakian, P. Liljeberg, and J. Plosila,2013. LiRCUP: Linear Regression Based CPU Usage Prediction Algorithm for Live Migration of Virtual Machines in Data Centers, *in 39th Euromicro Conference on Software Engineering and Advanced Applications, IEEE*. doi: https://doi.org/10.1109/SEAA.2013.23

[30] D. Meisner, B. T. Gold, and T. F. Wenisch,2009. PowerNap, *ACM SIGPLAN Notices*. 44(3): 205-216, doi: https://doi.org/10.1145/1508284.1508269

[31] L. Brochard,V. Kamath 2019 Power Consumption of Servers and Workloads. *Wiley*. 65-86, doi: https://doi.org/10.1002/9781119422037.ch4

[32] M. Xu, W. Tian, and R. Buyya,2017 A survey on load balancing algorithms for virtual machines placement in cloud computing, *Concurrency Computing*. 29(12): e4123, doi: https://doi.org/10.1002/cpe.4123

[33] Buyya R., Beloglazov A., and Abawajy J.,2010 Energy-efficient management of data center resources for cloud computing: a vision, architectural elements, and open challenges, *in International Conference on Parallel and Distributed Processing Techniques and Applications, Las Vegas, USA.*

[34] Buyya R., Yeo. C.S., Venugopal S., Broberg J., and Brandic I.,2009 , Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing, *Future Generation Computer Systems*,6: 599-616. https://doi.org/10.1016/j.future.2008.12.001

[35] A. Beloglazov, J. Abawajy, and R. Buyya,2012 Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing, *Future Generation Computer Systems*. 28(5): 755-768, doi: https://doi.org/10.1016/j.future.2011.04.017

[36] X. Li, P. Garraghan, X. Jiang, Z. Wu, and J. Xu,2018 Holistic Virtual Machine Scheduling in Cloud Datacenters towards Minimizing Total Energy, *IEEE Transactions on Parallel and Distributed Systems.* 29(6): 1317-1331, doi: https://doi.org/10.1109/TPDS.2017.2688445

[37] M. Ranjbari and J. Akbari Torkestani,2018 A learning automata-based algorithm for energy and SLA efficient consolidation of virtual machines in cloud data centers. *Journal of Parallel and Distributed Computing*. 113: 55-62, doi: https://doi.org/10.1016/j.jpdc.2017.10.009

[38] G. Narendrababu Reddy and S. Phani Kumar,2019 Modified ant colony optimization algorithm for task scheduling in cloud computing systems, *in Smart Innovation, Systems and Technologies, Springer Science and Business Media Deutschland*

GmbH. 357-365. doi: https://doi.org/10.1007/978-981-13-1921-1_36

[39] Venkatesh Rajath, C. N. G.2021 Efficiency in Cloud Computing. *International Journal of Engineering Research & Technology (IJERT)*. 10(6): 687-689.

[40] Satya Sobhan Panigrahi, Bibhuprasad Sahu, Amrutanshu Panigrahi, Sachi Mohanty,2021 Green Cloud Computing: An Emerging Trend of GIT in Cloud Computing. *In book Green Engineering and Technology*. 11 doi: https://doi.org/10.1201/9781003176275-13

[41] W. Yao, Z. Wang, Y. Hou, X. Zhu, X. Li, and Y. Xia,2023 An energy-efficient load balance strategy based on virtual machine consolidation in cloud environment, *Future Generation Computer Systems.* 146: 222-233, doi: https://doi.org/10.1016/j.future.2023.04.014

[42] M. Imran, M. Ibrahim, M. S. U. Din, M. A. U. Rehman, and B. S. Kim, 2022. Live virtual machine migration: A survey, research challenges, and future directions. *Computers and Electrical Engineering*. 103: 108297, doi: https://doi.org/10.1016/j.compeleceng.2022.108297

[43] X. Xu, X. Zhang, M. Khan, W. Dou, S. Xue, and S. Yu, 2020 A balanced virtual machine scheduling method for energy-performance trade-offs in cyber-physical cloud systems.*Future Generation Computer Systems,* 105: 789-799, doi: https://doi.org/10.1016/j.future.2017.08.057

[44] W. Wei, K. Wang, K. Wang, H. Gu, and H. Shen,2020. Multi-resource balance optimization for virtual machine placement in cloud data centers. *Computers & Electrical Engineering*. 88: 106866, doi: https://doi.org/10.1016/j.compeleceng.2020.106866

[45] A. Beloglazov, J. Abawajy, and R. Buyya,2012. Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing. *Future Generation Computer Systems*. 28(5): 755-768, doi: https://doi.org/10.1016/j.future.2011.04.017

[46] X. Li, X. Jiang, K Ye,2014. Virtual Machine Scheduling Considering both Computing and Cooling Energy. *IEEE International Conference on High Performance Computing and Communications*: 244-247.

[47] Z. Xiao, W. Song, Q. Chen, 2013. Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment, *IEEE Transactions on Parallel and Distributed Systems*. 24(6): 1107-1117. DOI: https://doi.org/10.1109/TPDS.2012.283

[48] Z. Song, X. Zhang, C. Eriksson,2015. Data Center Energy and Cost Saving Evaluation. Proc. *The 7th International Conference on Applied Energy:* 1255-1260. DOI: https://doi.org/10.1016/j.egypro.2015.07.178

[49] Misra, Sudip, P. Vamsi Krishna, K. Kalaiselvan, Vankadara Saritha, and Mohammad S. Obaidat. 2014. *Learning automata-based QoS framework for cloud IaaS Network and Service Management, IEEE Transactions.* 11(1): 15-24. DOI: https://doi.org/10.1109/TNSM.2014.011614.130429

[50] Arianyan Ehsan, Hassan Taheri, and Saeed Sharifian,2015. Novel energy and SLA efficient resource management heuristics for consolidation of virtual machines in cloud data centers. *Computers & Electrical Engineering.* 47: 222-240 DOI: https://doi.org/10.1016/j.compeleceng.2015.05.006