A Investigative Review on Energy Efficient Utilization in Virtual Machine Allocation

Amit G. Patel[1], Pankaj S. Patel[2], Jayesh A. Patel[3], Komal C. Patel[4], Keta R. Patel[5]

Sr. Lecturer, Dept. of Computer Eng., Swami Sachchidanand Polytechnic, Sankalchand Patel University, Visnagar, India [1]
I/C Head of CE/IT/ICT, Swami Sachchidanand Polytechnic, Sankalchand Patel University, Visnagar, India[2]
Lecturer, Dept. of Computer Eng., Swami Sachchidanand Polytechnic, Sankalchand Patel University, Visnagar, India[3,4,5]

agpatel.sspc@spu.ac.in[1], pspatel.sspc@spu.ac.in[2], japatelce.sspc@spu.ac.in[3],
kcpatelce.sspc@spu.ac.in[4], krpatel.sspc@spu.ac.in[5]

Abstract: This study has carefully analyzed and categorized the objectives, varieties, working principles, benefits, limitations, and simulation tools of existing energy-efficient scheduling algorithms. Our systematic and comprehensive study will enhance the fundamental understanding of energy-efficient techniques for new scholars who wish to delve deeper into the energy sector. This evaluation highlights how crucial it is to distribute virtual machines (VMs) in an energy-efficient way to reduce data centers' environmental impact and running costs. It provides useful information for researchers, practitioners, and policymakers who want to develop and implement long-term strategies for optimizing resource utilization in cloud and edge computing environments. To sum up. Load balancing systems aim to distribute workload evenly across physical servers to prevent resource underutilization and overload situations and reduce energy wastage. Consolidation solutions integrate virtual machines (VMs) onto fewer physical servers in an effort to optimize resource allocation. As a result, less power is used and the server is used more. Migration solutions lower energy consumption while preserving performance levels through the use of dynamic virtual machine reallocation. By considering energy consumption indicators, task characteristics, and resource availability, scheduling algorithms aim to allocate resources as efficiently as feasible.

Keywords: Energy efficiency, Resource management, Cloud computing, Workload consolidation, Virtual machine migration.

.

I. INTRODUCTION

Due to the significant rise in energy consumption brought on by the expanding need for computational power in data centers, energy efficiency has emerged as a critical issue. Virtualization technology, in particular virtual machine (VM) allocation, is crucial for enhancing resource utilization and reducing energy usage in data centers. This paper offers a comprehensive review of research projects and methods aimed at achieving energy-efficient use through virtual machine allocation algorithms.

The demand for processing resources in data centers has expanded considerably due to the speedy development of digital services and the adoption of data-intensive apps. In addition to performance and scalability problems, this growing demand raises significant concerns about energy use and environmental sustainability. Data centers, which are the backbone of modern computing infrastructure, are notorious for using enormous amounts of energy, increasing running costs and releasing carbon dioxide into the sky.

The inherent inefficiencies present in traditional data center architectures have been demonstrated to be mitigated by the use of virtualization technologies, particularly virtual machine (VM) allocation. Software can be isolated from physical hardware and virtualized resources can be dynamically provisioned through the VM allocation process, which enhances resource usage and energy efficiency. Allocating virtual machines (VMs) in a way that optimizes energy efficiency, however, poses a number of challenging issues that require in-depth research and innovative thinking.

This is how the rest of the document is organized: Section II describes the various algorithms used in cloud computing. Section III contains related work. Section V provides the conclusion, and Section IV includes Open Research Issues.

http://doi.org/10.63766/spujstmr.24.000041

II. DIFFERENT CLOUD COMPUTING ALGORITHMS

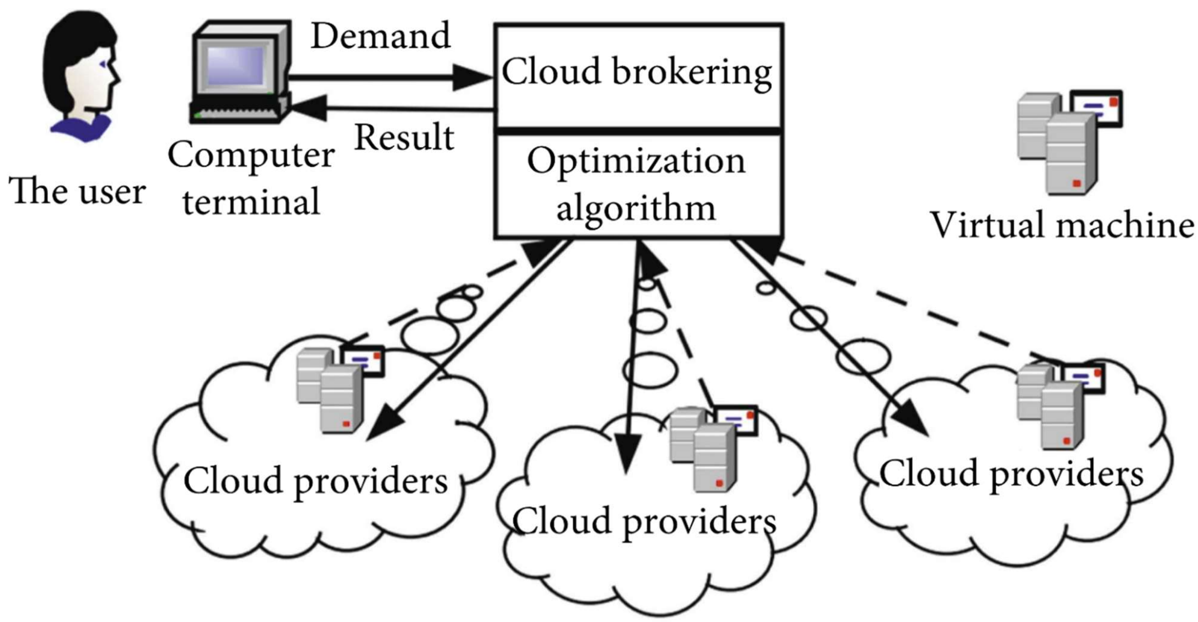

Fig. 1 VM resource allocation in cloud computing.[2]

2.1. Modified Hummock Algorithm [4]

In order to reduce virtual machine energy consumption, the Modified Ant Colony Optimization (MACO) technique was created within a cloud computing framework. In order to improve the quality of the results, MACO, which was initially created to solve the traveling salesman problem, added a solution reconstruction method. MACO successfully reduced the problem of path overlap by making sure that all ants released pheromones and responded to pheromone levels while navigating. Each ant was able to successfully finish its path-finding process thanks to the pheromone management mechanism, which eventually increased convergence rates. The strategy also aimed to lower operating expenses and energy consumption while improving overall performance.

2.2. First Fit Decreasing Algorithm [4]

Items are assigned to the first available bin using the First-Fit Decreasing (FFD) algorithm, which arranges them in a non-increasing order. By optimizing the number of physical machines (PMs), this technique can increase their utilization and lower power consumption. Two variants of FFD—the FFD mean and the FFD median—can be used in conjunction with Vector Bin Packing (VBP) methods for Virtual Machine Placement (VMP) in cloud computing. FFD explains the container Packing Problem (BIP), which involves gathering a container and a collection of items. Virtual machines (VMs) ought to be stacked in a downward array based on size prior to FFD. Each and every virtual machine (VM), beginning with the most important one, must be located in physically accessible machines (PM) that can be kept in good working order. Until not all virtual machines are deployed, the procedure will keep going.

2.3. Exchange Strategies for Multiple Ant Colony System [4]

This method uses a graphical layout to map potentially linked virtual machines. The OEMACS algorithm is used to find the best mapping solution after each virtual machine has been mapped across a single server. Placement, migration, and selection are the three phases of virtual machine consolidation. OEMACS uses the TSP, which offers the quickest route for migration and placement, in accordance with graph theory.

2.4. PSO (Particle Swarm Optimization)-based Scheduling

Cloud computing virtual machine (VM) allocation based on scheduling based on Particle Swarm Optimization (PSO). The objective function should reflect the optimization goal, such as optimizing resource use or reducing energy use, depending on the specific requirements of the cloud computing environment. To achieve optimal performance for a given problem instance, tuning and experimentation may be required to modify parameters like search space bounds, inertia weight, and cognitive and social learning aspects. This technique provides a framework for deploying PSO-based scheduling for virtual machine (VM) allocation in cloud computing settings, facilitating efficient resource and energy consumption management. PSO is a nature-inspired metaheuristic algorithm that mimics the social behavior of bird flocking or fish schooling to find an optimal solution. It is used in cloud computing to optimize VM placement by minimizing energy consumption, cost, or resource wastage. Each particle in the swarm represents a potential solution (a VM placement strategy).Particles move in the solution space based on their own best-known position and the swarm's best-known position. The movement is influenced by two factors: personal experience (pBest) and global experience (gBest) to converge towards an optimal VM allocation strategy. The objective is to optimize a given function, such as minimizing power consumption or maximizing resource utilization.

III. ENERGY-EFFICIENT AND SECURE VM ALLOCATION STRATEGIES

3.1 Dynamic VM Consolidation

Dynamic VM consolidation involves migrating VMs to fewer physical hosts to reduce power consumption while ensuring minimal service disruption. Techniques include:
Threshold-based Consolidation: Uses CPU and memory utilization thresholds to trigger VM migrations.
AI-Based Consolidation: Implements machine learning algorithms to predict VM usage patterns and optimize migration decisions.
Energy-Aware Migration: Considers energy metrics when deciding which VMs to migrate, reducing unnecessary power usage.

3.2 Secure Load Balancing Techniques

Efficient load balancing ensures that VMs are allocated optimally across servers while maintaining security and energy efficiency.
Trust-Based Load Balancing: Assigns workloads based on VM trust scores to prevent compromised nodes from hosting critical applications.
Energy-Aware Load Balancing: Uses power consumption models to allocate resources dynamically.
AI-Driven Approaches: Reinforcement Learning (RL) and Deep Learning (DL) models optimize workload distribution in real-time, ensuring secure resource allocation.

3.3 Attack Mitigation in VM Allocation

Cloud environments face threats such as DDoS attacks, VM escape attacks, and side-channel attacks, requiring proactive mitigation techniques.
Intrusion Detection Systems (IDS): AI-based IDS detects unusual VM behavior in real-time, flagging potential security breaches.
Blockchain-Based Security: Decentralized authentication prevents unauthorized access and tampering with VM allocations.
Isolation-Based Strategies: Using sandboxing and hardware-level isolation (Intel SGX, AMD SEV) to prevent VM escape attacks.

3.4 AI and Machine Learning for Secure and Energy-Efficient VM Optimization

Machine learning enhances energy efficiency by predicting workload demands and optimizing allocation in real time while improving security measures.
Predictive VM Scaling: Uses historical data to anticipate peak demand and preemptively allocate resources.
Reinforcement Learning for Energy Efficiency: Adaptive algorithms learn optimal VM placement strategies over time.

Anomaly Detection: AI-driven methods detect inefficient or compromised VMs consuming excessive power, improving both efficiency and security.

3.5 Green Computing and Sustainable Cloud Practices

Green computing focuses on minimizing energy consumption by leveraging renewable energy and eco-friendly data center designs.
Renewable Energy Integration: Allocates VMs to data centers powered by solar or wind energy. Carbon-Aware Scheduling: Shifts workloads to regions with lower carbon footprints. Energy-Proportional Computing: Uses power-efficient hardware to scale power consumption dynamically.

IV. RELATED WORK

4.1 Power-Aware scheduling [1]

In [1], The authors place a high priority on effective resource allocation in order to reduce energy consumption in cloud-based systems. This study investigates a number of energy-saving techniques and algorithms, such as power-aware scheduling, virtual machine migration, and dynamic workload consolidation. It emphasizes how important it is to incorporate performance metrics and energy efficiency when making decisions about resource management. The suggested methods seek to maintain Quality of Service (QoS) and cost-effectiveness while lowering energy waste and carbon emissions. The results of the study show that methods like power-aware scheduling, workload consolidation, virtual machine migration, and dynamic resource provisioning can efficiently maximize resource use and reduce energy consumption.

Merits- These techniques offer a number of benefits, including reduced expenses, a reduced carbon footprint, and enhanced efficiency. Offer several significant benefits, including reduced expenses, a reduced carbon footprint, and increased overall effectiveness.
Demerits- More attention needs to be paid to challenges including complexity, scalability, and implementation issues.

4.2 MBFD and SI based Grasshopper Technique [2]

In [2], The Modified Best Fit Decreasing (MBFD) algorithm has been used to effectively assign virtual machines (VMs) to the best host. In order to improve VM selection, this study presents a novel Grasshopper Optimization Algorithm (GOA) based on swarm intelligence. The suggested framework provides a reliable way to lower cloud data centers' overall energy usage. The method reduces energy consumption and migration frequency by optimizing resource utilization in virtual machines (VMs) for real-world applications. Furthermore, the GOA ensures the best VM selection while reducing problems associated with SLA violations and improper migrations. The "VM Migration" method entails assigning the virtual machines (VMs) to the appropriate PM at the outset and then requiring their migration in accordance with requirements. The efficiency with which the hosts and virtual machines (VMs) use their resources determines the allocation strategy. To minimize energy consumption during migration, the proposed MBFD method continuously updates the lowest power requirement for allocating or migrating a virtual machine across a host

Merits: By using the MBFD algorithm and GOA, the suggested method outperforms conventional techniques like ABC, E-ABC, and CS in terms of lowering energy consumption, minimizing SLA violations, and lowering the number of migrations.
Demerits: To develop the proposed framework and achieve differentiated performance on a limited number of hosts (10 hosts deploying 103 VMs), the authors plan to conduct additional research in the near future involving a bigger number of hosts and virtual machines (VMs).

4.3 Peak Efficiency Aware Scheduling (PEAS) [3]

In [3], The authors create the measurements of peak power efficiency and optimal usage for actual equipment. We also calculate the minimal number of Compute Resource Units (CRUs) a PM should supply in order to achieve optimal power efficiency through performance quantification using CRUs. Peak Efficiency Aware Scheduling (PEAS) is a new virtual machine consolidation technique. Peak Efficiency Aware Placement (PEAP) and Peak Efficiency Aware Cost-effective Reallocation (PEACR) are the two algorithms that make up PEAS. For each virtual machine, PEAP chooses the best host based on the three previously mentioned ideas. The task of reassigning virtual machines from overworked and unproductive hosts falls to PEACR. Comprehensive tests were conducted on CloudSim to evaluate PEAS's performance in relation to several baseline algorithms, such as TVRSM and PABFD.

Merits: PEAS keeps running computers running at less than maximum capacity whenever possible in order to maximize the power efficiency of servers. A variety of VM placement and migration algorithms and techniques, including TVRSM and PABFD, were used to assess PEAS. According to experimental findings, PEAS efficiently lowers cluster energy consumption, shortens the typical job execution time, and enhances a number of QoS metrics.

Demerits: In edge computing and fog computing, two domains that are expanding quickly and presenting related difficulties, the author plans to investigate the optimization of task scheduling and resource provisioning. In order to facilitate proactive resource allocation and virtual machine consolidation in intricate cloud systems, we will also focus on using machine learning to profile and predict job workloads and measure power consumption at the virtual machine level.

4.4 The OEMACS Algorithm [4]

In [4], The authors discuss the Order Exchange and Migration Ant Colony System's operation and results. The Clouds simulator is used to solve various Virtual Machine Placement issues in various scenarios. Of the nine cases, FFD employs the most physical servers, while OEMACS uses the fewest. Furthermore, it is observed that the VMP through OEMACS is remarkably quick. Other trials conducted in various situations and environments have shown that the OEMACS has the highest number of good findings when compared to other algorithms.

Merits- The OEMACS algorithm has allowed us to save energy by turning off servers that aren't in use. It is also completed very efficiently and in little time.
Demerits- For optimal efficacy, researchers in this field may employ the newest techniques and algorithms.

4.5 Intelligent Mine Blow Optimization (IBMO) technique [5]

In [5], The objective of this study is to develop an artificial intelligence (AI)-driven virtual machine (VM) allocation model for cloud load balancing that prioritizes security and energy efficiency. A key contribution of this work is the introduction of the Intelligent Mine Blow Optimization (IBMO) technique, designed to securely allocate VMs in the cloud. This method also mitigates security threats posed by attacks on virtual machines from various users. The study evaluates the performance of the proposed IBMO-based VM allocation model by comparing it against other approaches using multiple assessment metrics. The analysis of the results demonstrates that the proposed method outperforms existing alternatives. Notably, the suggested approach reduces SLA violations by an average of 50% and achieves a 60% reduction in energy consumption. By significantly lowering energy usage in cloud data centers, this method contributes to advancing green computing. IBMO is an enhanced version of binary multi-objective optimization tailored for discrete problems like VM placement. It extends classical binary optimization techniques by improving the way multiple objectives are handled.The algorithm encodes VM-to-host assignments in binary form.It uses multi-objective optimization techniques to find trade-offs between factors like energy efficiency, load balancing, and QoS (Quality of Service).Improvements over basic binary optimization include dynamic mutation strategies, adaptive weight adjustments, and better exploration-exploitation balance.

Merits- The three primary benefits of the suggested IBMO technique are enhanced exploration, fewer iterations to find the best solution, and global searching capability. When compared to the most recent state-of-the-art techniques, the effectiveness of the suggested secure and energy-efficient virtual machine allocation model is demonstrated by the validation of the optimization-based scheduling mechanism's results in terms of energy consumption, SOA violations, number of PMs used, and execution time during performance analysis.
Demerits- By putting into practice an improved security method for detecting and preventing cloud system attacks, this work can be expanded.

4.6 The Grey Model in Traffic Burst (GMTB)[6]

In [6], The researchers have introduced a resource management system utilizing an arrival model switching mechanism to enhance energy efficiency in cloud data centers. This study examines cloud task characteristics and presents two distinct task arrival models. The Grey Model in Traffic Burst (GMTB) is applied in burst scenarios, whereas the Poisson process-based arrival model is used under normal conditions. To facilitate model switching and detect anomalies, an anomaly detection module is incorporated. The proposed approach includes an effective task arrival switching technique that addresses both workload prediction and anomaly event detection. By leveraging the Poisson process-based arrival model and the GMTB model, this method provides reliable predictions for various cloud task arrival scenarios.

Merits- A virtual machine (VM) migration strategy based on utilization and a Service Level Agreement (SLA)-aware power management policy is proposed to enhance energy efficiency.
Demerits- To ensure the effectiveness of load forecasting and energy-efficient scheduling algorithms in a distributed cluster, thorough testing in a real-world environment is necessary.

4.7 GDR and MPC [7]

In [7], Reducing energy consumption in cloud systems is the main goal of many energy-conscious strategies now in use. Two models are suggested here: the maximum percentage of correlation (MPC) and the gradient regression-based method (GDR). These models can dramatically reduce cloud computing energy consumption while preserving required performance levels when evaluated in a cloud data center using the CloudSim simulator. According to the simulation results, (1) the Gdr host overloaded detection method works better in terms of energy usage than the MCP approach, and (2) it is more beneficial to take CPU, memory, and internet traffic into account when choosing virtual machines (VMs) from an overloaded host.

Merits- The GDR and MPC can significantly lower the energy usage of cloud computing..
Demerits- The SLA breach is a serious problem from the customer's point of view. A more efficient method that lowers the SLA violation can be offered in the future because, in this instance, the SLA violation tends to rise as energy usage decreases.

4.8 Polymorphous Energy Efficient Resource Allocation Approach (PEERA) [8]

In [8], A polymorphous energy efficient resource allocation approach, or PEERA, is proposed to enhance energy and performance measures such as energy consumption, make span, and resource usage. To maximize the dynamic allocation of resources, a method for reducing the amount of time required to complete each activity is described. To shorten the Make span and save energy, an extra operation is added to the algorithm that caches a subset of the best solutions at each iteration. The PEERA has been contrasted with the Particle Swarm Optimization (PSO) method and the Multiverse Optimizer (MVO) algorithm. The findings were computed using standard data sets and the virtual machines were left in their original configurations. With the following settings, CloudSim 3.0 produced the following results: VM = "10, 20, 30, 40, 50, 60" and tasks = "100, 200, 300, 400, 500, 600."

Merits- It is evident that the PEERA yields better results in terms of resource utilization, energy efficiency, and makespan. In addition to optimizing Makespan, it makes intelligent use of resources while using less energy.
Demerits- More research is required on energy efficiency, reducing energy consumption, and more efficient use of resources.

4.9 VMRRU Technique [9]

In [9], The use of the Virtual Machine Energy Resource Request Utilization (VMERRU) Advanced Algorithm is suggested. This method allocates data centers (DC) and virtual machines (VM) based on factors such as resource consumption levels, energy monitoring, input request demands, and matching criteria. As a result, an effective request processing method based on parallel computing is proposed. This method enhances request prediction by utilizing request analysis, consumption estimation, and finally the state of available resources. An appropriate scheduling technique is suggested by utilizing multiple factors and processing them concurrently. The Cloud Simulation The technique is implemented in Java and is simulated using the Cloud Analyst simulator. The simulation results demonstrate the effectiveness of the proposed algorithm in comparison to existing methods such as Round Robin and the Throttled approach for cloud component scheduling. Additionally, the study outlines the algorithm's implementation and the configuration of each component.

Merits- Calculating calculation time, computation cost, and energy utilization determines how efficient the outcomes are.
Demerits- Using the right architecture to store and retrieve data efficiently while working with input data utilization and storage.

4.10 AI model automatic training and deployment scheme based on K8S Architecture [10]

The researchers present an automated platform designed for training and deploying AI models, handling data processing, annotation, model training optimization, and publishing. Built on a cloud-edge architecture, the proposed system can generate customized models based on room environments while ensuring consistency and automation in model training, making it highly suitable for large-scale data center applications. Simulation and experimental results indicate that the system can manage multiple training tasks simultaneously and reduce the training time for a single model by 76.2%. The first experiment compares the time required for AI-based automatic training versus manual training using the same dataset, while the second evaluates the training time for different data volumes in AI-based automatic training.

Merits: this method, Additionally, it supports the execution of tasks concurrently. Thus, this approach can be used to improve the efficiency of model training and deployment. Consequently, the intelligent support provided by this proposed solution can be advantageous to the green data center.
Demerits- Despite efforts to save energy, the full environmental impact, including the manufacturing and disposal of edge devices, should be considered.

4.11 Maximum-Minimum Round-Robin algorithm (MMRR)[11]

In [11], The cloud services were significantly improved by the suggested hybrid strategy of using the Maximum Minimum load balancing algorithm in conjunction with Round Robin.The maximum and minimum constraints that typically prevent jobs with the shortest completion times from being completed on schedule will be resolved by the suggested hybrid. On the other hand, the task that takes the longest to complete will be completed first. Round robin that distributes work without regard to priorityThe data center's processing time is good for both MMRR (0.57ms) and Throttled (0.50ms), but it was worse for Round Robin (1.48ms). With an overall reaction time of 227.26 ms and an 89% cost-effectiveness rating, MMRR outperformed the other algorithms that were assessed. Therefore, in order to increase customer happiness, the study suggests that MMRR be included in cloud services.

Merits - The data center's processing speed is above average, the response time is promising, and using the new method to process a work is less expensive than using other algorithms.
Demerits- Future iterations of this study would use alternative quality of service (QOS) requirements, such as cost, throughput, and delay, for different routing policies. To determine which algorithm would produce the greatest results in the cloud environment, more data centers, users, and algorithms will be deployed.

4.12 Shortest Queue of d with Randomization and Round robin(SQ-RR(d))[12]

In [12], The authors' task is to suggest an alternative algorithm that enhances the results of the traditional power of two choices technique. Additionally, d servers are chosen using the new algorithm, though not entirely at random. The other d 1 servers are picked at random, while one server is chosen in a round robin method. The job is then routed to the server that has the fewest jobs. In the new algorithm, we use round-robin selection in conjunction with randomization to provide static load balancing. According to a theoretical approximation of this method, the new version performs better than the classical solution in every scenario, including systems operating at 99% capacity.

Merits- The theoretical approximation made in the paper is demonstrated by the simulations supplied by the authors. These simulations also demonstrate that for high service rates, there is a 60% greater chance of selecting an empty server.
Demerits- Given the enormous potential and long-term advantages, cloud service providers should also fund research.

4.13 Software Defined Networking (SDN)[13]

In [13], This study introduces a load balancing mechanism designed for Software-Defined Networking (SDN) in a cloud environment, focusing on an HTTP server handling 400 requests per second, totaling 6 million requests. Due to server overload, additional client requests remain unprocessed. Implementing a load balancing system to reassess these parameters after balancing offers a solution. Instead of relying on a single server, traffic can be distributed across multiple servers. With the integration of an HTTP load balancer, HAProxy effectively manages the incoming traffic, ensuring improved response times and preventing request drops, even when 6 million requests arrive simultaneously. This enhances the availability and reduces the latency of HTTP requests.

Merits- It's crucial to emphasize that traffic from the HTTP server can be split up among several servers rather than just one. Furthermore, it was discovered that both availability and latency had increased following the implementation of

the load balancer mechanism in the HTTP server.
Demerits- There are numerous installation and configuration challenges with the OpenStack cloud. Knowing the OpenStack cloud topology is preferable, as is deciding whether to install on a single node or a distributed system where each machine is in charge of one or more modules. Carefully selecting the Linux distribution to be installed on is also important.

4.14 Learning Automata (LA) [14]

In [14], The researchers introduced an energy-efficient scheduling algorithm (LAEAS) based on Learning Automata (LA) theory for managing independent real-time tasks in a cloud environment. Additionally, they developed a scheduling framework utilizing LA to enhance job allocation efficiency. They compare the efficiency levels of the algorithms using the two measures listed below. The first statistic is the total amount of energy that the virtual machines (VMs) use to run their functions. The second performance metric, known as the success ratio (SR), represents the percentage of tasks completed within their deadline compared to the total number of tasks. By adapting to a dynamic cloud environment, the proposed LAEAS algorithm enables the scheduler to make optimal decisions. This learning-based approach enhances the overall performance metrics of the cloud system.

Merits- The cloud environment's diverse resources, dynamic decision-making, application-specific demands, etc., make it easier to schedule work effectively.
Demerits- Additional planning is required to extend our method to the multi-objective real-time work scheduling algorithm on the cloud.

V. OPEN RESEARCH ISSUES

This research paper provides a thorough analysis of energy-efficient virtual machine allocation. This paper identifies and discusses the primary research gaps in this topic and suggests potential future research directions. The following are some worries regarding open challenges

5.1 Dynamic Workload Management: To optimize energy efficiency without compromising performance, investigate strategies for dynamically adjusting virtual machine (VM) allocations in response to workload fluctuations.
5.2 Resource Consolidation Strategies: Advanced resource consolidation methods, including load balancing algorithms, job scheduling, and virtual machine live migration, are used to balance energy consumption and Quality of Service (QoS) standards.
5.3 Heterogeneous Infrastructure Optimization: Examine methods for optimizing energy efficiency in cloud environments that are heterogeneous and comprise different hardware (such CPUs, GPUs, and FPGAs) and software stacks (like virtualization technologies and containers).
5.4 Energy-Aware Task Scheduling: Investigate energy-conscious task scheduling techniques that assign computing jobs to virtual machines (VMs) with lower energy consumption while accounting for dependencies, deadlines, and resource constraints.
5.5 Security and Privacy Concerns: Consider any vulnerabilities resulting from resource sharing and isolation techniques when analyzing the effects of energy-efficient virtual machine allocation strategies on security and privacy.

VI. CONCLUSION

In conclusion, this investigative research has carefully investigated the critical importance of energy-efficient utilization in virtual machine (VM) allocation within cloud computing environments. In order to achieve energy efficiency in virtual machine allocation, we have examined a variety of research initiatives, methodologies, and challenges. From dynamic workload management to machine learning-based prediction models, researchers have looked into a variety of tactics aimed at optimizing resource utilization while lowering energy consumption Furthermore, the evaluation has highlighted how important it is to consider real-time monitoring and control techniques as well as green data center design concepts when aiming for energy efficiency. Ultimately, by pushing the limits of energy-efficient virtual machine allocation, researchers can contribute to the development of more sustainable and environmentally sensitive cloud computing infrastructures. This will ensure that the benefits of cloud computing are experienced without unduly taxing energy resources. Energy-efficient and secure VM allocation is crucial for sustainable and safe cloud computing. This review highlights various strategies, including dynamic consolidation, AI-driven optimization, secure load balancing, attack mitigation, and green computing techniques. While challenges persist, advancements in AI, federated learning, blockchain, and renewable energy integration offer promising solutions.

REFERENCES

[1] Kavitha, J., PSV Srinivasa Rao, and G. Charles Babu. "*Energy Efficient Resource Utilization of Cloud Computing Environments for Deployment Models*." 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS). IEEE, 2023.

[2] Singh Rachhpal, Sarpreet Singh, and Balwinder Kaur. "*Energy-Efficient VM Allocation and Migration Approach Using Swarm Intelligence Algorithm in Cloud Computing Environment*." 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2023.

[3] Lin Weiwei, Wentai Wu, and Ligang He. "*An on-line virtual machine consolidation strategy for dual improvement in performance and energy conservation of server clusters in cloud data centers*." IEEE Transactions on Services Computing 15.2 (2019): 766-777.

[4] Khan Mohammad Shuaib, Satnam Singh Saini, and Pradeep Kumar. "*Ant Colony System for Efficient Virtual Machine Placement in Cloud Computing*." 2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN). IEEE, 2023.

[5] Alhebaishi Nawaf. "*An Artificial Intelligence (AI) based Energy Efficient and Secured Virtual Machine Allocation Model in Cloud*." 2022 3rd International Conference on Computing, Analytics and Networks (ICAN). IEEE, 2022.

[6] Wang, B., Liu, Y., Zhang, F., & Jiang, J. (2022, December). " *Energy Efficient Resource Scheduling in Cloud Computing Based on Task Arrival Model*". In 2022 IEEE Globecom Workshops (GC Wkshps) (pp. 686-691). IEEE.

[7] Smriti Manu, Kumud Shaily, and Danthuluri Sudha. "*Minimization of Energy Consumption in Cloud*." 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT). IEEE, 2023.

[8] P. A. Malla, S. Sheikh and T. A. Teli, "*A Polymorphous Energy Efficient Resource Allocation Approach (PEERA) in Cloud Computing*," 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 920-923.

[9] R. Soni and N. K. Gupta, "*Energy Efficient Cloud Task Scheduling Policy Using Virtual Machine Concept and VMRRU Technique*," 2022 International Conference on Inventive Computation Technologies (ICICT), Nepal, 2022, pp. 296-304.

[10] Li, C., Guo, Z., He, X., Hu, F., &Meng, W. (2023, April). " *An AI Model Automatic Training and Deployment Platform Based on Cloud Edge Architecture for DC Energy-Saving*". In 2023 International Conference on Mobile Internet, Cloud Computing and Information Security (MICCIS) (pp. 22-28). IEEE.

[11] Kazeem Moses, A., Joseph Bamidele, A., Roseline Oluwaseun, O., Misra, S., & Abidemi Emmanuel, A. (2020). " *Applicability of MMRR load balancing algorithm in cloud computing*." International Journal of Computer Mathematics: Computer Systems Theory, 6(1), 7–20. Enhancing the power of two choices load balancing algorithm using round robin policy Felix Garcia-Carballeira

[12] Felix Felix Garcia-Carballeira, Alejandro Calderon, Jesus Carretero, " *The power of two choices load balancing algorithm using round robin policy* ." Springer Science+Business Media, LLC, part of Springer Nature 2020.

[13] Y. A. H. Omer, M. A. Mohammedel-Amin and A. B. A. Mustafa, "Load Balance in Cloud Computing using Software Defined Networking," 2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), Khartoum, Sudan, 2021, pp. 1-6.

[14] S. Sree Priya and T. Rajendran, "*Improved round-robin rule learning for optimal load balancing in distributed cloud systems*" International Journal of System of Systems Engineering 2023 13:1, 83-99.