International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 12 | Dec – 2025 DOI: 10.55041/ISJEM05227
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

A Migration-Free Energy-Efficient VM Allocation Method for Data Centers

Archana Kulkarni ,Asmita Marathe ,Ditixa Mehta ,Sony Jha
Assistant Professor *Computer Science & Engineering*
(Cyber Security)
Thakur College of Engineering &Technology
Mumbai, Maharashtra
Archana.kulkarni@tcemumbai.in
asmitamarathe@tcetmumbai.in
ditixa.mehta@tcemumbai.in
sony.jha@tcetmumbai.in

Growing operational costs and high energy consumption in data centers have increased the need for smart resource allocation and effective energy management. This paper introduces an energy-efficient Virtual Machine (VM) allocation technique that eliminates the need for VM migration while maintaining stable system performance. The proposed approach, named Energy Efficient Virtual Machine Allocation without Migration (EEVMA), aims to lower energy usage, reduce costs, and avoid performance degradation.EEVMA integrates resource optimization techniques, predictive analysis, and real-time workload monitoring to place VMs on servers based on workload patterns, server utilization, and power consumption models. By distributing workloads efficiently and consolidating tasks onto fewer servers, the approach minimizes energy wastage and improves overall efficiency. To validate the effectiveness of EEVMA, we compare its performance with existing VM allocation methods, focusing on metrics such as energy reduction, cost efficiency, and system stability. Experiments using real workload traces and performance indicators show that EEVMA enhances energy savings and service quality without the drawbacks associated with VM migration. The results indicate that data centers can achieve improved sustainability, reduced operational expenses, and better resource utilization through the proposed migration-free allocation strategy.

Keywords— Cloud Computing, Energy Efficiency, Resource Management

I. Introduction

Cloud computing is a paradigm that involves delivering on-demand computing resources, such as servers, storage, and applications, over the internet. While cloud computing offers various benefits, including scalability, flexibility, and cost savings, it also consumes a significant amount of energy.

The energy consumption in cloud computing arises from several factors:

1. Data Centers: Cloud service providers operate large-scale data centers to host and manage the infrastructure required for cloud services. These data centers consume substantial amounts of electricity to power the servers, cooling systems, networking equipment, and other supporting infrastructure.

2. Server Utilization: Efficient utilization of servers is crucial for energy efficiency. In traditional data centers, servers often operate at low utilization levels, resulting in wasted energy. Virtualization and resource consolidation techniques in cloud environments aim to improve server utilization and reduce energy waste.

3. Cooling and Power Distribution: Data centers require cooling systems to maintain optimal temperature levels for the servers. Cooling infrastructure consumes a significant amount of energy. Power distribution systems, including uninterruptible power supplies (UPS) and power conditioning equipment, also contribute to energy consumption.

4. Network Infrastructure: Cloud computing relies on a vast network infrastructure for data transmission between users and data centers. Networking equipment, such as routers, switches, and data transmission lines, consume energy during data transfer and communication.

Addressing energy consumption in cloud computing is crucial for reducing environmental impact and optimizing resource usage. Researchers and industry professionals are actively exploring various strategies, including:

1. Energy-efficient Hardware: Designing and deploying energy-efficient servers and networking equipment can significantly reduce energy consumption in data centers.

2. Virtual Machine Management: Techniques like dynamic VM consolidation and load balancing help optimize server utilization, minimizing the number of active servers and their associated energy consumption.

3. Renewable Energy Sources: Incorporating renewable energy sources, such as solar or wind power, into data center operations can reduce reliance on fossil fuels and decrease the carbon footprint of cloud computing.

ISSN: 2583-6129
DOI: 10.55041/ISJEM05227

4. Energy-aware Scheduling: Smart scheduling algorithms can allocate workloads and resources in a manner that minimizes energy consumption, considering workload characteristics and server availability.

Efforts are ongoing to improve the energy efficiency of cloud computing through advancements in hardware, software algorithms, and sustainable practices. By optimizing energy consumption, cloud providers can reduce their environmental impact and contribute to a more sustainable IT ecosystem.
Virtual machine (VM) management techniques play a vital role in optimizing server utilization and reducing energy consumption in cloud environments. Two common techniques used for energy-efficient VM management are dynamic VM consolidation and load balancing.

1. Dynamic VM Consolidation: Dynamic VM consolidation involves migrating VMs across physical servers to consolidate their workloads. Consolidating virtual machines (VMs) onto fewer live servers will enable the low-power or shutdown of the remaining servers. Energy savings can be achieved by increasing server usage overall through the consolidation of virtual machines. In order to make wise migration decisions, virtual machine consolidation strategies take into account variables including workload demand, resource utilization, and performance needs.

2. Load Balancing: In order to maximize resource efficiency and prevent overwhelming particular servers, load balancing seeks to divide the workload equally among several servers. Load balancing strategies guarantee that workloads are distributed evenly, preventing certain servers from being underutilized while others are overworked. Even workload distribution improves server efficiency and can reduce energy consumption. Load balancing algorithms consider factors such as server capacity, resource utilization, and network conditions to allocate and distribute workloads effectively.

Both dynamic VM consolidation and load balancing techniques are typically implemented using intelligent algorithms and monitoring mechanisms. These techniques continuously analyze the resource demands of VMs, the performance of servers, and the overall system state to make informed decisions about VM migrations and workload distribution. The goal is to achieve better resource utilization, reduce the number of active servers, and consequently minimize energy consumption in the cloud environment.

Cloud providers may maximize resource utilization, enhance energy efficiency, and save operating costs by dynamically combining virtual machines and distributing the workload. These methods are crucial parts of cloud environments' energy-efficient resource provisioning.

II. RELATED TO WORK

Numerous studies have suggested that the best way to improve energy economy in servers is to run their processing units at full capacity, however doing so degrades performance. [1,3,4].

Another strategy in the literature is to specify a static ideal utilization threshold for each resource type, such as CPU, RAM, bandwidth, and so on, in order to address performance degradation brought on by operating at maximum capacity [6].

However, because the resource need in the future is not taken into account, a static threshold may result in machines being turned on or off unnecessarily. [7].

Another strategy in the literature is server consolidation, which uses virtual machine migration to lower the number of physical computers that are in use. However, from the standpoint of the service provider as well as the service user, VM migration and server consolidation strategies result in low throughput and energy overheads. [10,13].

A real-world workload is performed to assess the method using Google cluster data. [14].

As in the prior study, the method is developed and run on CloudSim, a widely used cloud simulator. With a primary focus on IaaS-related tasks, CloudSim offers multiple policies for VM allocation and migration. But the allocation strategies that are offered have static workloads. [15].

In this paper, objective is how to reduce the massive amount of energy consumption in cloud computing data center. To address this issue, many power-aware virtual machine (VM) allocation and consolidation approaches are proposed in exiting system to reduce energy consumption efficiently by Dynamic utilization and threshold value [22].

In this paper, the main goal was to provide the CSP a versatile scheduling with an optimization framework aiming to maximize the efficiency of energy in meeting user deadlines. Greedy algorithm is well suited for the heterogeneous environments of cloud resources that have a dynamic behavior [23].

The dynamic and unpredictable character of the cloud computing environment is discussed in this study, which makes it difficult to address the work scheduling problem. Therefore, in order to improve resource usage, a good task scheduling approach should be created and implemented in the Cloud Broker in order to both satisfy the QoS requirements set by Cloud users and conduct good load balancing among virtual machines. [21].

In this paper, the objective is to achieve the optimal balance between the host utilization and the energy consumption for cloud data centers [20].

In this paper, an energy-efficient scheduling method for virtual machine reservations is proposed and objective is to provide an optimal solution with the minimum number of job migrations [25].

III. OBJECTIVES

The proposed approach involves developing an algorithm that optimizes resource allocation in virtualized data centers without the need for migration. The algorithm considers workload monitoring, performance and cost metrics, server consolidation, load balancing, predictive analytics, energy efficiency, and cost optimization. The algorithm is evaluated, validated, and deployed in the data center environment, with continuous monitoring and adjustments based on real-time workload characteristics and energy efficiency goals.

In summary, the objective is to optimize energy consumption and costs in data centers while maintaining or improving service performance to meet the expectations of both regulatory requirements and user satisfaction.

IV. EXISTING ALGORITHM

The goal of Energy Efficient VM Allocation (LAA) is to maximize server count in virtualized data centers to accommodate dynamic workloads without requiring VM relocation. [24]. The algorithm achieves this by efficiently allocating resources based on workload predictions and periodic comparisons.

Here is an overview of the LAA algorithm:

1. Workload Prediction: LAA incorporates a Holt Winters-based prediction module to forecast future workload demands [24]. This module analyzes historical workload data and identifies patterns and trends to predict the future workload requirements. By anticipating the workload, the algorithm can make proactive decisions about resource allocation.

2. Periodic Server Comparison: LAA periodically compares the predicted number of required servers with the currently active servers. This step allows the algorithm to assess the adequacy of the existing server capacity in meeting the workload demands. The comparison serves as a basis for determining whether any adjustments are necessary.

3. Energy-Performance Trade-off: LAA balances energy efficiency and performance by considering the predicted workload and the number of active servers. The algorithm acknowledges the inverse relationship between energy consumption and performance. It aims to minimize energy consumption while ensuring that the available servers can handle the workload efficiently.

4. Resource Management: Based on the periodic comparisons and the energy-performance trade-off, LAA makes decisions regarding resource provisioning. It determines whether additional servers need to be activated or whether existing servers can be powered down or put into low-power states. The algorithm aims to optimize server utilization while meeting the workload requirements.

The proposed algorithm, LAA, avoids the need for VM migration by focusing on efficient resource management and workload prediction. By optimizing the number of servers without migration, LAA reduces energy consumption and improves resource usage in virtualized data centers.

Tests carried out using actual workload traces from Google Cluster demonstrate that LAA performs better than the Local Regression-Minimum Migration Time (LR-MMT) method offered by Cloud Sim, resulting in a 45% decrease in the amount of energy required to finish a task.

V. APPROACH

The proposed approach, Energy Efficient Virtual Machine Allocation without Migration (EEVMA), is designed to assign virtual machines (VMs) to physical servers in a data center in a way that minimizes energy consumption and operational costs while avoiding VM migration and maintaining stable system performance. EEVMA leverages real-time workload monitoring, predictive resource analysis, and dynamic allocation strategies to ensure efficient utilization of server resources. The data center is assumed to consist of heterogeneous physical machines with varying CPU and memory capacities, along with known power consumption characteristics. Workload requests, represented as VMs, arrive dynamically with specific resource requirements and lifetimes, which may vary over time. Each server's energy usage is modeled based on its utilization level, capturing the power consumed at idle, low, and high loads, as well as the cost associated with powering servers on or off.

The framework of EEVMA consists of multiple functional modules. A monitoring module continuously tracks server utilization and VM resource consumption over fixed intervals. The workload or VM admission module receives incoming VM requests along with their resource requirements and expected runtime. The allocation decision module determines the best server for each incoming VM based on current server utilization, available resources, and the power consumption model. If an active server can accommodate the VM without exceeding a predefined utilization threshold, it is assigned there using a heuristic that minimizes energy increase and resource wastage. If no suitable active server is available, a new server is powered on to host the VM. Once a VM is placed, it remains on the assigned server until its termination, ensuring that no migration occurs. Additionally, a server management module manages powering servers on or off based on the workload to further optimize energy consumption.

EEVMA aims to efficiently consolidate workloads onto fewer servers while balancing resource utilization to reduce energy

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 12 | Dec – 2025 DOI: 10.55041/ISJEM05227
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

wastage. By avoiding VM migrations, the approach eliminates the overhead and potential performance degradation associated with moving VMs between servers. The allocation strategy can also incorporate short-term workload prediction to prevent frequent power-on/off cycles, which further improves energy efficiency.

To evaluate the performance of EEVMA, we compare it with existing VM allocation strategies that may involve migration or simple allocation heuristics. The evaluation uses realistic server configurations, power models, and publicly available workload traces to simulate dynamic VM arrivals. Key performance metrics include total energy consumption, number of active servers, resource utilization efficiency, server on/off transitions, and system performance indicators such as response time or SLA violation rates. Experiments are conducted under varying workload intensities, including low, medium, and high, to assess the stability and effectiveness of the proposed method. The results are expected to demonstrate that EEVMA achieves reduced energy consumption, lower operational costs, and improved resource utilization without any performance degradation caused by VM migration, providing a sustainable and efficient solution for modern data centers.

The following actions must be taken in order to optimize the number of servers in virtualized data centers to satisfy the demands of dynamic workloads without requiring migration:

1. Workload Characterization: Begin by analyzing and characterizing the dynamic workloads that the virtualized data center is expected to handle. This involves understanding the patterns, trends, and fluctuations in the workload demands over time.

2. Workload Prediction: Utilize techniques such as statistical analysis, machine learning, or time series forecasting methods to predict future workload demands. This can be based on historical workload data, input from monitoring systems, or other relevant data sources. The goal is to forecast the workload in advance, enabling proactive resource allocation decisions.

3. Resource Utilization Monitoring: Continuously monitor the resource utilization levels within the virtualized data center. This includes monitoring CPU, memory, storage, and network utilization across the active servers. Real-time monitoring provides insights into the current resource demands and helps identify underutilized or overloaded servers.

4. Dynamic Resource Allocation: Make well-informed judgments about resource allocation based on workload projections and resource use tracking. Without using virtual machine migration, the objective is to maximize the number of active servers to satisfy workload needs. This may entail turning on more servers during peak demand and turning off servers during off-peak times.

5. Load Balancing: Implement load balancing mechanisms to distribute the workload evenly across the active servers. Load balancing algorithms consider factors such as server capacity, resource utilization, and network conditions to allocate and redistribute workloads effectively. This ensures that no server is overloaded while others remain underutilized.

6. Energy-Performance Trade-off: Consider the trade-off between energy consumption and performance. Determine the optimal number of servers that strikes a balance between energy efficiency and meeting workload requirements. This involves periodic comparisons between predicted and active server numbers to adjust the server capacity as needed.

7. Validation and Evaluation: Validate the methodology through experimentation and evaluation using real-world workload traces or simulated scenarios. Compare the performance of the proposed methodology against other approaches, considering metrics such as energy consumption, resource utilization, workload response time, and system stability.

By following this procedure, virtualized data centers can optimize their resource usage, meet dynamic workload requirements, and reduce energy consumption without relying on VM migration. The given approach leverages workload prediction, resource monitoring, dynamic resource allocation, load balancing, and the energy-performance trade-off to achieve these objectives.

VI. CONCLUSION

In summary, our proposed work introduces the Look-ahead Energy Efficient VM Allocation (LAA) algorithm, which aims to optimize server allocation for incoming workloads without the need for migration [24]. To evaluate LAA, we compare its performance with the LR-MMT algorithm using real-world workload traces from the Google cluster.

During the evaluation, we consider parameters such as energy consumption reduction, system performance, and cost. We analyze the extent to which LAA reduces energy consumption compared to LR-MMT, assess its impact on system performance metrics like response time and throughput, and evaluate its cost-effectiveness in terms of server maintenance, electricity expenses, and cooling requirements.

The evaluation covers various workload types, including CPU-intensive tasks, networking components, I/O devices, and storage, ensuring a comprehensive assessment of LAA's effectiveness across different scenarios.

Ultimately, based on the evaluation results, we can draw conclusions about LAA's ability to reduce energy consumption, its impact on system performance, and its cost-effectiveness compared to LR-MMT. These findings will provide insights into the practicality and benefits of using LAA as an energy-efficient resource allocation algorithm for virtualized data centers without migration.

REFERENCES

[1] Cao J, Wu Y, Li M (2012) Energy efcient allocation of virtual machines in cloud computing environments based on demand forecast. In international conference on grid and pervasive computing. Springer, Berlin, pp 137–151.

[2] Garg SK, Yeo CS, Anandasivam A, Buyya R (2011) Environment-conscious scheduling of HPC applications on distributed cloud-oriented data centers. J Parallel Distrib Comput 71(6):732–749.

[3] Stillwell M, Schanzenbach D, Vivien F, Casanova H (2009) Resource allocation using virtual clusters. In 2009 9th IEEE/ACM international symposium on cluster computing and the grid. pp 260–267. K. Elissa, "Title of paper if known," unpublished.

[4] Xu L, Seng S, Ye X (2012) Multi-objective optimization based virtual resource allocation strategy for cloud computing. In 2012 IEEE/ACIS 11th international conference on computer and information science. pp 56–61.

[5] Xu L, Seng S, Ye X (2012) Multi-objective optimization based virtual resource allocation strategy for cloud computing. In 2012 IEEE/ACIS 11th international conference on computer and information science. pp 56–61.. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[6] Li X, Qian Z, Chi R, Zhang B, Lu S (2012) Balancing resource utilization for continuous virtual machine requests in clouds. In Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012 Sixth International Conference on IEEE. pp 266–273.

[7] Galloway J, Smith K, Carver J (2012) An empirical study of power aware load balancing in local cloud architectures. In Information Technology: New Generations (ITNG), 2012 Ninth International Conference on IEEE. pp 232–236.

[8] Dhiman G, Marchetti G, Rosing T (2010) Vgreen: A system for energyefcient management of virtual machines. ACM Transactions on Design Automation of Electronic Systems (TODAES) 16(1):6

[9] Geronimo GA, Werner J, Weingartner R, Westphall CB, Westphall CM (2014) Provisioning, resource allocation, and DVFS in green clouds. International Journal on Advances in Networks and Services, 7(1 & 2):108–117.

[10] Garg SK, Toosi AN, Gopalaiyengar SK, Buyya R (2014) SLA-based virtual machine management for heteroggeneous workloads in a cloud data center. J Netw Comput Appl 45:108–120.

[11] Verma A, Ahuya P, Neogi A (2008) pMapper: power and migration cost aware application placement in virtualized systems. ACM/IFIP/USENIX international conference on distributed systems platforms and open distributed processing. Springer, Berlin, Heidelberg, 243– 264.

[12] Gmach D, Rolia J, Cherkasova L, Kemper A (2009) Resource pool management: Reactive versus proactive or let's be friends. Comput Netw 53(17):2905–2922.

[13] Srikantaiah S, Kansal A, Zhao F (2008) Energy aware consolidation for cloud computing. Proceeding of the 2008 conference on power aware computing and systems.

[14] Google cluster-usage traces: format and schema. https://github.com/ google/cluster-data. Published May 6, 2013. Updated October 17, 2014. Accessed June, 2015.

[15] Calheiros RN, Ranjan R, Beloglazov A, De Rose CAF, Buyya R (2011) CloudSim: a toolkit for modelling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. Soft PractIcal Exp 41(1):23–50.

[16] Tian W, He M, Guo W et al (2018) (2018) On minimizing total energy consumption in the scheduling of virtual machine reservations. J Netw Comput Appl 113:64–74.

[17] Zhang Y, Cheng X, Chen L, Shen H (2018) Energy-efficient workloads scheduling heuristics with multi-constraints in virtualized clouds. J Grid Computer 2018:1–17.

[18] Gawali MB, Shinde SK (2018) Task scheduling and resource allocation in cloud computing using a heuristic approach. J Cloud Computing 7(1):1–16.

[19] Chen H, Liu G, Yin S, Liu X, Qiu D (2018) Erect: energy-efficient reactive scheduling for real-time workloads in heterogeneous virtualized clouds. J Computer Sci 28:416–425.

[20] An Heuristics based Dynamic Power-Aware Resource Allocation for Cloud Computing, IJCSMC Journal 2018.

[21] An Energy and Deadline Aware Scheduling Using Greedy Algorithm for Cloud Computing, Pradeep Venuthurumilli ,2019,IIETA.

[22] Look-ahead energy efficient VM allocation approach for data centers, Journal of Cloud Computing (2022) 11:11

[23] On minimizing total energy consumption in the scheduling of virtual machine reservations, Science Direct,Volume 113, 1 July 2018, Pages 64-74.

[24] E. I. Elsedimy, M. Herajy, and S. M. M. Abohashish, "Energy and QoS-aware virtual machine placement approach for IaaS cloud datacenter," Neural Computing and Applications, vol. 37, pp. 2211–2237, Jan. 2025. DOI: 10.1007/s00521-024-10872-1

[25] P. Zhang, J. Gao, J. Liu, and L. Tan, "EVMC: An Energy-Efficient Virtual Machine Consolidation Approach Based on Deep Q-Networks for Cloud Data Centers," Electronics, vol. 14, no. 19, article 3813, 2025. DOI: 10.3390/electronics14193813

[26] "Scalable energy-aware VM allocation on cloud data centers through mathematical programming models," Future Generation Computer Systems, vol. 174 (to appear Jan. 2026), 2025. DOI: 10.1016/j.future.2025.108011