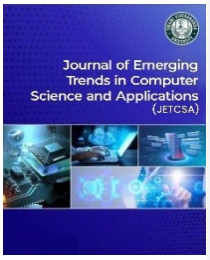
Journal of Emerging Trends in Computer Science and Applications (JETCSA), Volume 1 (Issue 2: May-August, 2025): Pages 08-16; Published on: September 10, 2025.



ISSN: XXXX-XXXX Awaiting for Approval (Online)
Journal of Emerging Trends in Computer Science and Applications(JETCSA)
Contents available at: *https://www.swamivivekanandauniversity.ac.in/jetcse/*

A review on task-scheduling algorithms in the cloud computing towards energy efficiency

Mrs. Sukriti Santra[1*]

[1]*Department of Computer Science and Engineering, Swami Vivekananda University, Barrackpore-700121, WB, INDIA*

ABSTRACT

The growing dependence on cloud computing, fueled by technological advancements, has led to a surge in energy consumption across cloud servers. Efficient resource scheduling within the clouddata centers can reduce energy consumption. This paper provides a comprehensive overview of various task-scheduling algorithms, encompassing traditional methods, heuristics, metaheuristics, and hybrid approaches. It delves into the challenges associated with workload distribution, resource allocation, and the dynamic nature of cloud environments. The significance of incorporating QoS parameters, such as deadline constraints and energy efficiency, is highlighted. Furthermore, the paper discusses the limitations and improvements of notable algorithms, such as Min-min and Max-min, and explores recent advancements in agent-based and heuristic scheduling methods.

Keyword: Cloud computing, cloud data centers, resource scheduling, green computing, energy efficiency.

I. INTRODUCTION

Cloud computing is a service-centric model providing on demand access to computing resources, transforming IT pro- visioning and enabling global, distributed access from any device [1]. Cloud service providers are categorized into three layers such as Software-as-a-Service (SaaS), Infrastructure-as-a-Service (IaaS), and Platform-as-a-Service (PaaS). IaaS defines virtualized resources like on-demand storage. PaaS provides a higher level of abstraction, making it easily programmable and allowing users to create applications without concerning themselves with specific hardware requirements. SaaS enables users to access software over the internet, freeing them from the responsibilities of software maintenance. In essence, IaaS provides virtualized storage, PaaS offers a programmable platform, and SaaS delivers software

applications over the internet, relieving users of maintenance tasks [2]. Cloud computing serves as a versatile platform, providing access to resources, facilitating scalability, and enhancing operational efficiency for various application and services. Indeed, catering to a growing demand for cloud services involves substantial energy consumption, contributing to carbon emissions. The environmental impact highlights the importance of adopting sustainable practices in the cloud computing to mitigate adverse effects on the environment. To address environmental impact of cloud server energy consumption, various task-scheduling algorithms have been proposed. The aim of these algorithms are used to optimize resource utilization, reduce more energy consumption, and ultimately contribute to a more sustainable and ecofriendly cloud computing environment. The rise in the digital economy is concomitant with data centers' increasing energy usage. Due to the environmental impact of these high-energyconsuming entities, there is a focus on energy saving and the emission reduction. As a result, improving cloud data centers and the energy efficiency management has become a major topic of research. Researchers are working very hard to develop useful measurements and approaches for assessing energy efficiency in order to accomplish this main objective [3].

In this comprehensive survey paper, we have endeavored to encompass task-scheduling methodologies/algorithms within the realm of the cloud computing, with a specific focus on promoting energy efficiency.

Section 2 describes a literature survey encompassing various task-scheduling algorithms. Section 3 deliveries into diverse task-scheduling approaches, while Section 4 offers insights from an energy efficiency perspective. Section 5 concludes the research paper and outlines avenues for the future research.

II. LITERATURE SURVEY

Several papers have been carefully looked at regarding task-scheduling in cloud computing, especially focusing on improving energy efficiency. In one study, Li Mao and others [4] tackled the changing challenges in cloud computing by creating two algorithms: time-aware and energyaware. These are meant for scheduling tasks in environments with different types of resources. Their algorithm, called ETMCTSA, combines both methods, letting users choose between energy savings and performance based on their needs. They tested this with simulations and found it worked better than other methods In another study, Nirmal Kr.
Biswas and others [5] came up with a new way to reduce energy use and avoid breaking service agreements in cloud data centers. Their approach uses a new linear regression model, balances the load on hosts, and decides where to place virtual machines. This could help create smarter and more sustainable systems for future smart cities. Author [6] and others introduced an algorithm called EERS that focuses on both energy efficiency and system reliability in cloud environments. This method works to save energy and keep the system running smoothly by using five different sub-algorithms addressing task dependencies, communication costs, sub-make span definition, cluster-VM mapping, and slack reclamation. Evaluation on the Workflow Sim simulator with

realworld scientific workloads demonstrates EERS outperforming existing approaches in energy efficiency and reliability optimization. In another work JK Jeevitha et. al [7] proposed Shortest Round Vibrant Queue (SRVQ) algorithm, a combination of Shortest Job First, Round Robin, and Vibrant Quantum, significantly reduces waiting times in the scheduling process and minimizes starvation. Through the collaborative use of DVFS and SRVQ, the research achieves a noteworthy 45% improvement in energy efficiency and a substantial 33% enhancement in Quality of Service (QoS) performance com- pared to existing algorithms. Author Sarita Simaiya et. al [8]in their paper introduces the ”EEPSA” method, a novel energy efficiency priority scheduling system for cloud computing that emphasizes pre-emptive scheduling and considers both optimal fit and system availability in routing requests to processing servers. The experimental analysis demonstrates the effectiveness of EEPSA in minimizing overall energy utilization compared to existing methods, evaluating parameters such as energy usage, number of tasks migrated, and processing time. In another study, H. Momeni and their team [9] introduced EaRTs, an energy-aware method for scheduling tasks in real-time cloud computing applications. Their approach uses virtualization and resource consolidation to save energy, use resources more efficiently, and ensure tasks are completed on time. The method includes four different algorithms that work together to achieve better performance in terms of meeting deadlines, using resources effectively, and saving energy compared to other similar scheduling methods. In another research, Mohan Sharma and colleagues [10] developed an energy-efficient task scheduler that works independently. It uses supervised neural networks to reduce the total time tasks take, lower energy use, decrease execution overhead, and minimize the number of active servers in the cloud. The neural network was trained using a dataset created with a genetic algorithm and achieved 99.9% accuracy. This shows much better performance than existing methods like Genetic Algorithm, MIN-MIN heuristic, and energyefficient task schedulers based on linear regression, especially in both heavily and lightly loaded cloud environments.

III. DIFFERENT TASK-SCHEDULING TECHNIQUES

Cloud processes scheduling is a way to manage resources in the best possible manner depending on service quality parameters either by a heuristic or meta-heuristic type of algorithm. To solve problems such as workload imbalance, over-provisioning, and under-provisioning, which lead to cost minimization and resource usage optimization, it is very important to have a provisioning algorithm of great efficiency [11]. The task scheduling in the cloud environment can be performed by static and dynamic methods. The algorithms used for task-scheduling can be put into groups namely traditional scheduling, heuristic scheduling, meta-heuristic scheduling, and hybrid scheduling algorithms [12].

A) Traditional scheduling

Traditional ways of scheduling tasks in cloud computing often use either static or dynamic methods. In static scheduling, tasks are given to resources at the start, and they don't change based on what happens later. Dynamic scheduling, on the other hand, changes as the tasks are being run, so it can handle things like delays or new problems that come up. Both kinds of scheduling try to

make the best use of available resources, get tasks done as quickly as possible, and follow the quality-of-service standards. Examples of these traditional methods include Round Robin, First Come First Serve, and Priority Scheduling.

Heuristic scheduling

Heuristic task-scheduling methods in the cloud computing rely on rules of thumb and approximation techniques to make decisions quickly, especially in situations where finding an optimal solution is computationally complex. These heuristics aim to efficiently allocate tasks to virtual machines while considering factors like resource availability, load balancing, and minimizing task completion time.

Various heuristic techniques are employed in the cloud computing for task-scheduling; each tailored to specific optimization goals.

Min-Min and Max-Min Algorithm

The Min-min scheduling algorithm minimizes expected completion time by iteratively assigning tasks to resources in two phases, calculating and selecting the task with the overall minimum expected completion time. Though it tries to distribute the load evenly, the system often ends up selecting smaller tasks first when it executes the scheduling for the first time [13].

On the other hand, the Max-min algorithm, which is often referred to as a distributed system approach, sets the main focus on the tasks that have the longest time of execution. It matches these tasks with resources that have the lowest total execution time, and on it goes until all the tasks have been scheduled, thus, sparing the resources to be used in the best possible way.

The Max-min algorithm allocates the tasks to the resources on the basis of a matrix of expected completion times. It chooses the tasks with the longest expected completion times and the minimum execution times so that resources can be allocated in an efficient manner.

FCFS, SJF, and RR algorithm

Task-scheduling in the cloud computing is vital to making the best use of available resources and achieving good performance. This is where FCFS comes in handy; it assigns the highest priority to the task which came first in the queue, although it may result in some inefficiencies such as the convoy effect. One of the ways to increase efficiency is through SJF whereby the shorter tasks get prioritized, however, a correct burst time prediction is imperative. Round Robin assigns a certain amount of fixed time capacity to each task thus making the system fair and ensuring that a long task is not able to monopolize the resource, on the other hand, a higher total turnaround time may be the sacrifice. The decision of the scheduling algorithm depends on factors such as workload characteristics and system goals, with hybrid or adaptive approaches often used to address the dynamic nature of cloud environments.

Bin Packing

Cloud computing bin packing scheduling is the management of resources through which you assign tasks that are necessary for use of the resources that are available to you in such a way that there are minimum wastage and maximum utilization. Just like packing items into bins, this method seeks to get a perfect combination of tasks on the virtualized resources for example virtual machines so as to raise the level of the resource's use and save money in the operations. Different algorithms, namely First Fit, Best Fit, and Worst Fit, are used for task allocation to resources. The efficient use of bin packing is very important in the cloud environment to guarantee the utilization of resources in an effective way, which leads to being cost-effective and having a positive overall system.

Deadline based scheduling

Deadline-based scheduling algorithms in the cloud computing prioritize task execution based on specific time constraints or deadlines. The primary goals include meeting service level agreements (SLAs), minimizing the risk of missing deadlines, optimizing resource utilization, and enhancing overall system performance. These algorithms assign priorities to tasks based on their deadline requirements and dynamically allocate resources to ensure timely completion. Factors like task urgency, resource availability, and system load are considered for effective scheduling decisions. Efficient deadline-based scheduling is crucial for time- sensitive applications, contributing to improved Quality of Ser-vice (QoS) and resource optimization in the cloud environments. In [14]authors paper introduces DEESA, a Deadline-based Energy Efficient Scheduling Algorithm, designed for task and VM scheduling in the cloud computing. Through dynamic queue classification of tasks and virtual machines, the proposed algorithm demonstrated in the cloud Sim simulator effectively minimizes energy consumption and make span time, surpassing the performance of existing scheduling algorithms.

QoS Based Scheduling

Quality of Service, or QoS, based scheduling is very important for handling the growing need to use resources efficiently in distributed and cloud computing systems. The main goal is to improve how tasks are scheduled so that available resources are used well without making any part of the system too busy. Good task scheduling helps cloud systems work better and gives users a better experience. Many different methods have been suggested for QoS-based task scheduling. One example is the QoS-guided Min-Min heuristic, which uses a flexible scheduling approach along with QoS guidelines. Other methods use fixed-priority techniques like Rate Monotonic and Deadline Monotonic, which help prioritize tasks according to their urgency. There are also QoS-aware methods that group tasks by features like user type, task type, size, and how quickly they need to be completed. These approaches show how important it is to consider QoS in scheduling tasks to make better use of resources and satisfy the varied needs of users in cloud environments.

Agent credit-based scheduling algorithm

Agent Credit-Based Scheduling Algorithm is an innovative method of task scheduling that involves intelligent agents to achieve better results in Quality of Service (QoS) parameters. A. Singh and colleagues came up with the Autonomous Agent-Based Load Balancing Algorithm

(A2LB) where the idea of agents assisting the distribution of resources for upcoming tasks is introduced which in turn leads to lowering of response and execution times besides making the system more scalable. T. Thomaset al. came up with the credit-based algorithm to resolve some issues related to the min-min algorithm. This approach estimates the average length of the tasks and allocates credits depending on how much a particular task deviates from the average. After that, the tasks with the highest credit scores are executed on the resources, which results in minimizing the total time required for all tasks as well as optimal use of resources. The agent credit-based scheduling algorithm permits the agents to make decisions on the fly, thus allowing for increased QoS in cloud computing environments.

Meta-heuristic Based Scheduling

Metaheuristic algorithms have good characteristics, e.g. independence of problems, fast searching in search spaces to find reasonable solutions to NP-complete problems, and nondeterministic, approximate characteristics. The meta-heuristic strategies, heuristics and randomization, are a versatile approach that can apply to a variety of industries with demonstrated high performance and adaptability to different applications. In the cloud computing environments, where task-scheduling requires exploration of large search spaces for optimal or near-optimal solutions, meta-heuristic algorithms are an integral consideration. Due to their non-deterministic and randomized nature, meta-heuristics are well suited for NP-hard optimization problems. Utilizing a meta-heuristic strategy in the cloud environments allows for easy and efficient solutions for NP-complete problems—which in turn means an efficient acquirement of the sub-optimal solution. [12]. Ways in which meta-heuristic algorithms are different is described below.

Particle Swarm Optimization Algorithm

Particle swarm optimization (PSO) is a highly effective heuristic algorithm for task-scheduling in the cloud computing environments. Like most swarm-based methodologies, PSO's functionality is fundamentally based on the manner in which particles behave in a search space, to optimize criteria including cost, workflow and resource usage. As an example, Pandey et al. in their work on cost optimization [Pande et al.] and Juan et al. with respect to the optimization of the execution of tasks in the cloudstorage [Juan et al.], just to name a few, have examined how the PSO method allocates the tasks to improve. On top of that, Gomathi and Krishnasamy's use of the hybrid PSO algorithm gave a very effective way of making good use of the available resources. The article that Alkayal et al. wrote as part of their research is an illustration of the nature of the problem being solved i.e. by using PSO to optimize multi-distance, waiting was minimized and throughput as well as resource utilization were maximized. Moreover, even the hybrid PSO work as per the example of Oriabi Dordaie and Jafari Navimipoour [2021] keeps on bringing out more flexibility in those configurations that face PSO challenges in the cloudtaskscheduling, thereby emphasizing the function of PSO in arriving at the efficient and effective solutions for complicated cloud computing environments.

Ant Colony Optimization Algorithm

Ant Colony Optimization (ACO) has evolved to be a very effective metaheuristic algorithm, which is widely used for the task-scheduling optimization in the cloud computing. Ant Colony Optimization (ACO), which takes its inspiration from the most perfect ant colony that can resolve

the most complex problems in a very efficient way, has been an excellent algorithm for scheduling the tasks in the grid and cloud computing. One can see the same just in nature where ants use ACO to achieve their goal, whereas few artificial ants adopt the same strategy to find the best task assignments in the whole spec of solution space. The pheromone trails lead iterative solution construction, which allows the algorithm to come pretty close to the best or even exact task schedules. ACO is capable of being the cause of the minimization of make span, solving of the dynamic changes in the cloud environments, and capabilities of multi- objective optimization goals such as execution time, cost, and resource utilization. Also, the decentralized and adaptive character of ACO greatly complements the distributed and dynamic features of cloud computing, thereby giving ACO an important role in task-scheduling that is most efficient in complicated cloud environments.

Genetic Algorithm Based Task-scheduling Algorithm

Genetic Algorithms (GAs) are cited as one of the most successful optimization methods that can be used to evaluate the problem of scheduling of tasks in the cloud computing. The use of a chromosome representation for possible task schedules enables GAs to implement genetic operations such as crossover and mutation to sequentially improve a population of schedules. The goodness of each schedule is assessed on the grounds of chosen goals, for example, the reduction of total completion time or the better utilization of resources. GAs is able to show their flexibility in changing environments; thus, they are perfect for cloud computing situations that have different workloads and resources that are not always available. Research pieces have been published on the effectiveness of the Genetic Algorithm (GA) in task-scheduling which has resulted in several performance metrics having been improved.

IV. CONCLUSION AND FUTURE WORK

Task-scheduling optimization is the major point that directly influences system efficiency as serviceability of the cloud, besides, it has been the topic of numerous papers in which the choice of various scheduling algorithms, such as traditional, heuristic, metaheuristic, and hybridapproaches, has been discussed to solve workload distribution, resource utilization along with meeting QoS requirements. All algorithm's virtues are its flaws, and a considerable amount of research is still mainly focused on the development of adaptive, energy-saving and scalable solutions that can meet requirements of cloud environments that keep changing. Moreover, the adoption of QoS factors like deadline conditions and energy saving, for example, has become the issue of guaranteeing that users' satisfaction and environmental sustainability, on the whole, are accomplished. First, the future work in this direction should be focused on new algorithms that include the use of machine learning, security aspects, and the changing scenario of cloud technologies. Second, the link of task-scheduling with future technologies such as edge and quantum computing could provide different cloud environmental ways to access the resources and thus be able to perform at higher levels. In general, the problem of cloud task-scheduling is still lively and the development that adapts to changes and challenges of cloud computing is still open for a wide range of possibilities.

REFERENCES

[1] A. Sunyaev and A. Sunyaev, "Cloud computing," *Internet Computing: Principles of Distributed Systems and Emerging Internet-Based Technologies*, pp. 195–236, 2020.

[2] A. Amini Motlagh, A. Movaghar, and A. M. Rahmani, "Task scheduling mechanisms in cloud computing: A systematic review," *International Journal of Communication Systems*, vol. 33, no. 6, p. e4302, 2020.

[3] S. Long, Y. Li, J. Huang, Z. Li, and Y. Li, "A review of energy efficiency evaluation technologies in cloud data centers," *Energy and Buildings*, vol. 260, p. 111848, 2022.

[4] L. Mao, Y. Li, G. Peng, X. Xu, and W. Lin, "A multi-resource task scheduling algorithm for energy-performance trade-offs in green clouds," *Sustainable Computing: Informatics and Systems*, vol. 19, pp. 233–241, 2018.

[5] N. K. Biswas, S. Banerjee, U. Biswas, and U. Ghosh, "An approach towards development of new linear regression prediction model forreduced energy consumption and sla violation in the domain of green cloud computing," *Sustainable Energy Technologies and Assessments*, vol. 45, p. 101087, 2021.

[6] R. Medara and R. S. Singh, "Energy efficient and reliability aware workflow task scheduling in cloud environment," *Wireless Personal Communications*, vol. 119, no. 2, pp. 1301–1320, 2021.

[7] J. Jeevitha and G. Athisha, "A novel scheduling approach to improve the energy efficiency in cloud computing data centers," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 6639–6649, 2021.

[8] S. Simaiya, V. Gautam, U. K. Lilhore, A. Garg, P. Ghosh, N. K. Trivedi,and A. Anand, "Eepsa: Energy efficiency priority scheduling algorithm for cloud computing," in *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*. IEEE, 2021, pp. 1064– 1069.

[9] H. Momeni and N. Mabhoot, "An energy-aware real-time task scheduling approach in a cloud computing environment," *Journal of AI and Data Mining*, vol. 9, no. 2, pp. 213– 226, 2021.

[10] M. Sharma and R. Garg, "An artificial neural network-based approach for energy efficient task scheduling in cloud data centers," *Sustainable Computing: Informatics and Systems*, vol. 26, p. 100373, 2020.

[11] M. Kumar, S. C. Sharma, A. Goel, and S. P. Singh, "A comprehensive survey for scheduling techniques in cloud computing," *Journal of Network and Computer Applications*, vol. 143, pp. 1–33, 2019.

[12] E. H. Houssein, A. G. Gad, Y. M. Wazery, and P. N. Suganthan, "Task scheduling in cloud computing based on meta-heuristics: review, taxonomy, open challenges, and future trends," *Swarm and Evolutionary Computation*, vol. 62, p. 100841, 2021.

[13] B. Kanani and B. Maniyar, "Review on max-min task scheduling algorithm for cloud computing," *Journal of emerging technologies and innovative research*, vol. 2, no. 3, pp. 781–784, 2015.

[14] S. K. Grewal and N. Mangla, "Deadline based energy efficient scheduling algorithm in cloud computing environment," in *2021 Fourth Inter- national Conference on Computational Intelligence and Communication Technologies (CCICT)*. IEEE, 2021, pp. 383–388.

[15] A. A. Khan and M. Zakarya, "Energy, performance and cost-efficient cloud datacenters: A survey," *Computer Science Review*, vol. 40, p. 100390, 2021.