Algorithms for Service-Based Task Scheduling in Cloud Computing

C Kesava Reddy
Department of Electrical and Computer Science Engineering, Origin College of Engineering, Bangalore
Corresponding Author: kesavareddyecse@gmail.com

To Cite this Article
Kesavareddy, "Algorithms for service-based task scheduling in cloud computing", Journal of Information Technology and Cyber Security Engineering", Vol. 01, Issue 01, July 2025, pp:15-19.

Abstract*: Having focused on the maximum utilization of resources, reducing the execution time, and improving the operational efficiency of the system in general, this paper surveyed and analysed the approaches to the service-based scheduling of tasks in cloud computing environments. The effectiveness of the various scheduling methods, including heuristic-based, metaheuristic-based, and machine learning-based scheduling in dynamic workload and various kind of resources will be explored. The paper highlights critical considerations during the design of an algorithm such as the priority of the tasks, time limit constraint, and affordability. Comparative evaluations have shown that adaptive and smart scheduling techniques dynamically assigning work to appropriate virtual machines are better than traditional static algorithms in terms of scalability, fault tolerance, and performance aspects. Its outcomes allow the provision of enhanced service provision and user satisfaction by providing opportunities regarding the selection and formulation of suitable methods of schedule to meet diverse quality of service (QoS) requirements in clouds infrastructures.*

Keywords*: Task scheduling, Cloud computing, Quality of service, Algorithms*
This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/
CC BY-NC-ND 4.0

I. Introduction

Scheduling of the tasks is vital in the maximization of resource utilization and in ensuring quality of service (QoS) in cloud computing. In service-based task scheduling, we have the tasks assigned dynamically with respect to resource capability, deadline and priority to the available resources. The cloud environments are distributed and diverse, which makes scheduling algorithms hard to design even in the efficient manner. Traditional Static Schedule Methods Often present inefficiencies as it cannot adapt to changing workloads and available resources. To overcome these problems, several algorithms have been designed, with the most common being heuristic, metaheuristic as well as machine learning techniques. The aim of such algorithms is to maximize the performance of a system, load balancing, reduction in execution time and decreasing costs. This introduction discusses the importance of the scheduling algorithms in cloud computing as well as putting in perspective the new inventions that make job handling a breeze in complex, service-oriented cloud systems.

II. QoS in Task Scheduling

Part of the parameters defining service-level agreements (SLAs) between cloud consumers and providers are QoS parameters. The net effect of effective job scheduling results into high user satisfaction and system dependability which ensure that such SLAs are met. The main goal of QoS-based scheduling is to accomplish this by assigning tasks to suitable resources in a shorter execution time and a smaller price without affecting the availability of resources and rising throughput and efficiency. Heuristic methods such as Min-Min, Max-Min and Round Robin employ simple rules that determine assignment of work and base it on certain QoS requirements. Max-Min focuses on longer jobs with respect to time of executions whereas Min-Min picks tasks that take the least time. With these methods, despite their computational ease of usage, one may not get the optimal results in dynamic or complex situations.

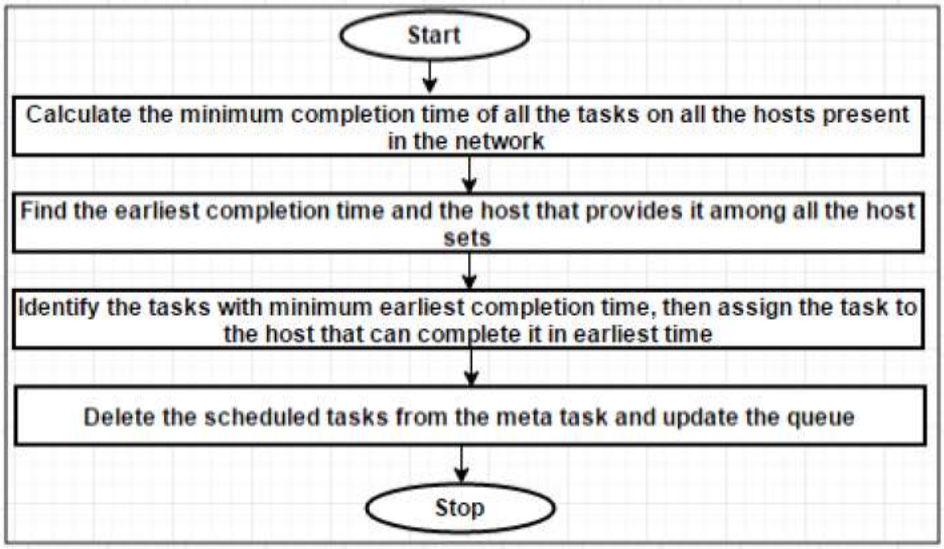

Fig 1: QoS guided grid task scheduling

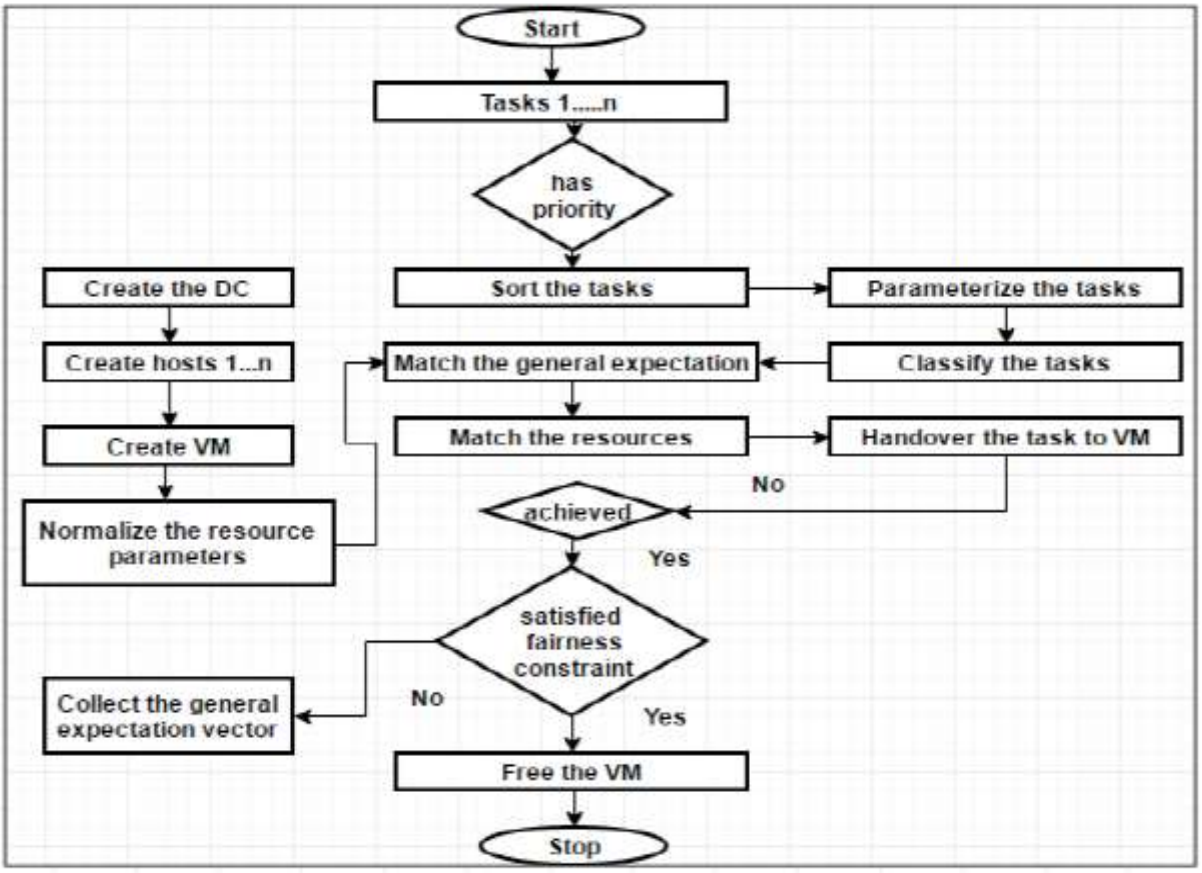

Fig 2: Berger Model for scheduling

They are Simulated Annealing (SA), Ant Colony Optimisation (ACO), Particle Swarm Optimisation (PSO) and Genetic Algorithms (GA). They give better solutions to large-scale multi-objective scheduling problems. As an illustration, GA develops the finest scheduling combinations through generations applying the principles of natural evolution. Whereas metaheuristics may be applied to optimize multiple QoS parameters simultaneously, they are often slower to calculate.

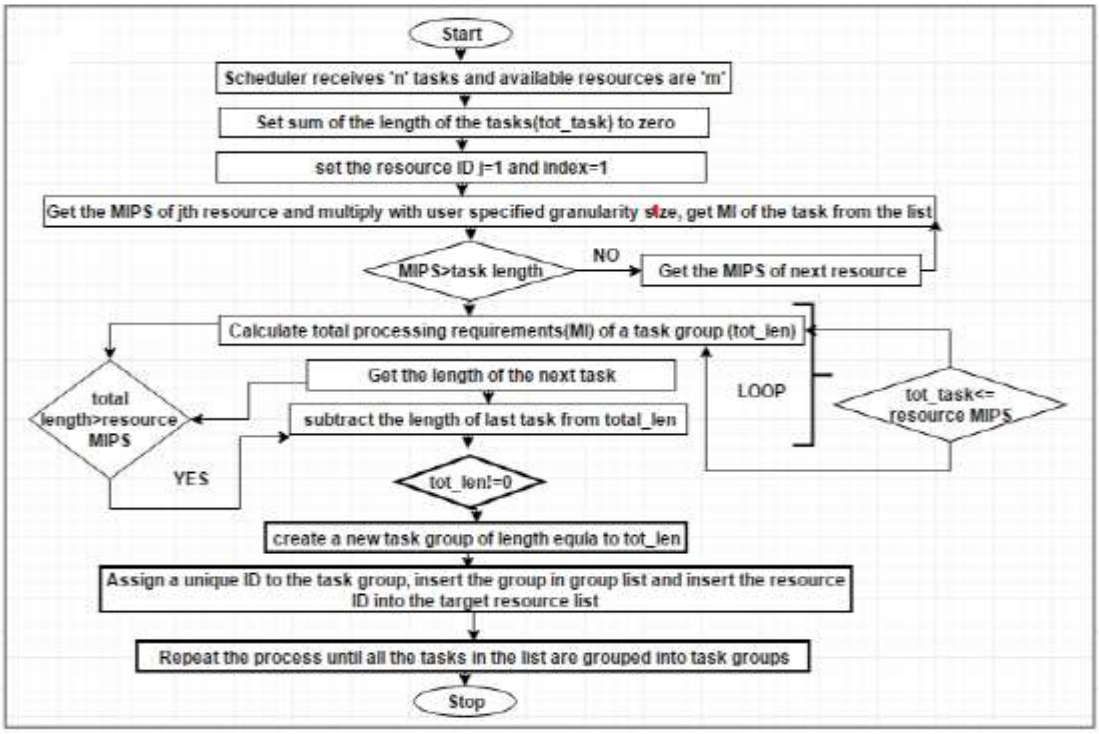

Fig 3: Cloud computing based on task schedule

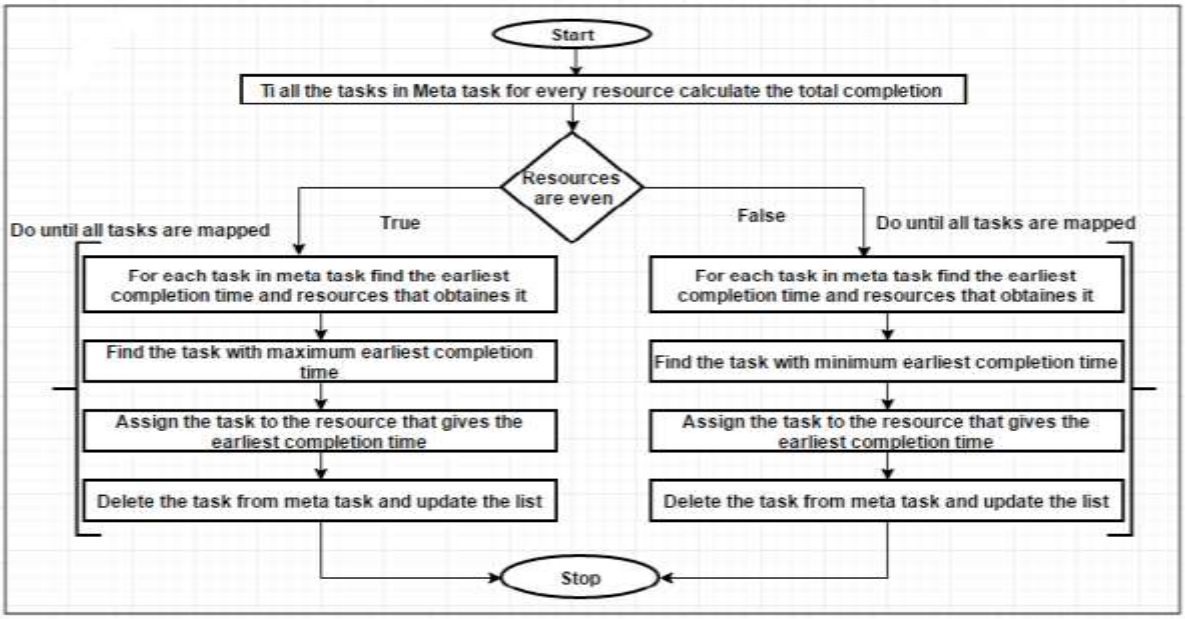

Fig 4: RASA

III. Multitasking Objective Scheduling Algorithms

The purpose of these algorithms is to balance conflicting QoS requirements, e.g, maximizing performance/minimizing cost. The methods A pareto optimisation is used in order to find a set of trade-offs that result in a set of optimal scheduling solutions. Task scheduling using QoS is a requirement in order to make service delivery in cloud computing quality and reliable. Cloud systems may dynamically allocate resources to meet user expectations by considering multiple QoS parameters with the help of complex algorithms: heuristic, metaheuristic, and machine learning approaches. As cloud infrastructures evolve, AI-based multi-objective approaches are the increasingly important future approach to scheduling solutions, to enable fulfilment of complex requirements in real time.

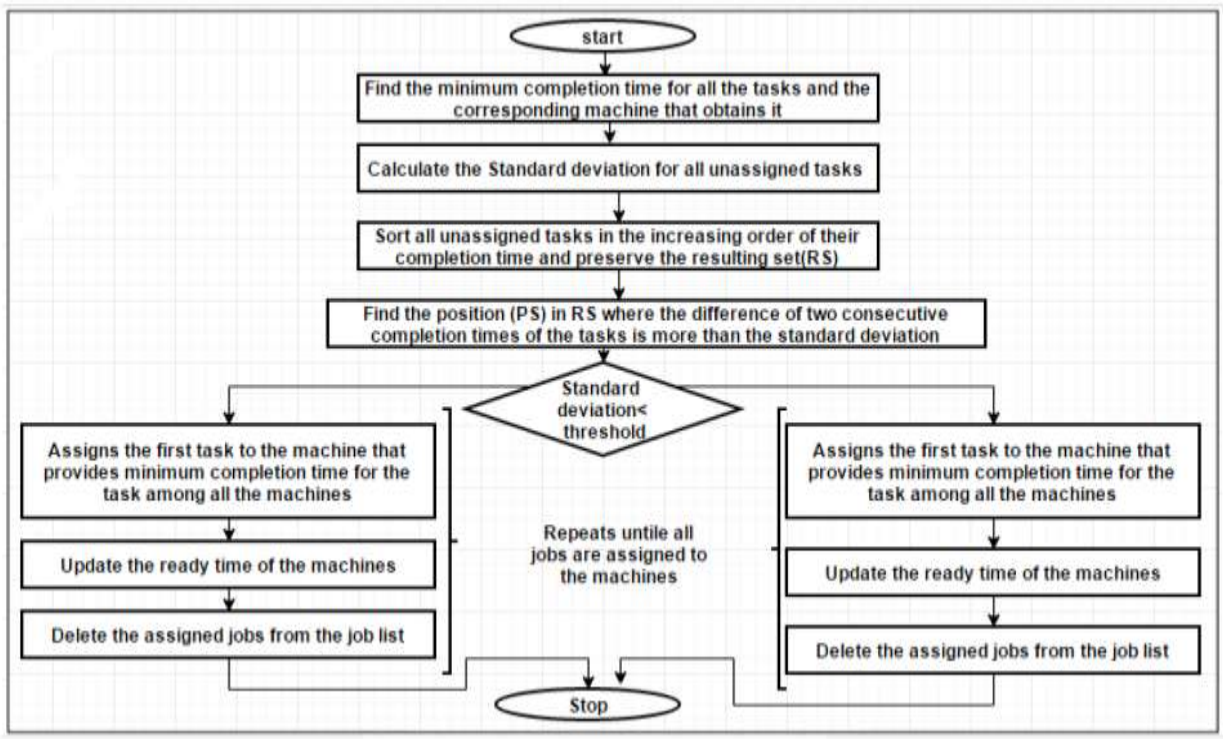

Fig 5: A QoS based Predictive scheduling task

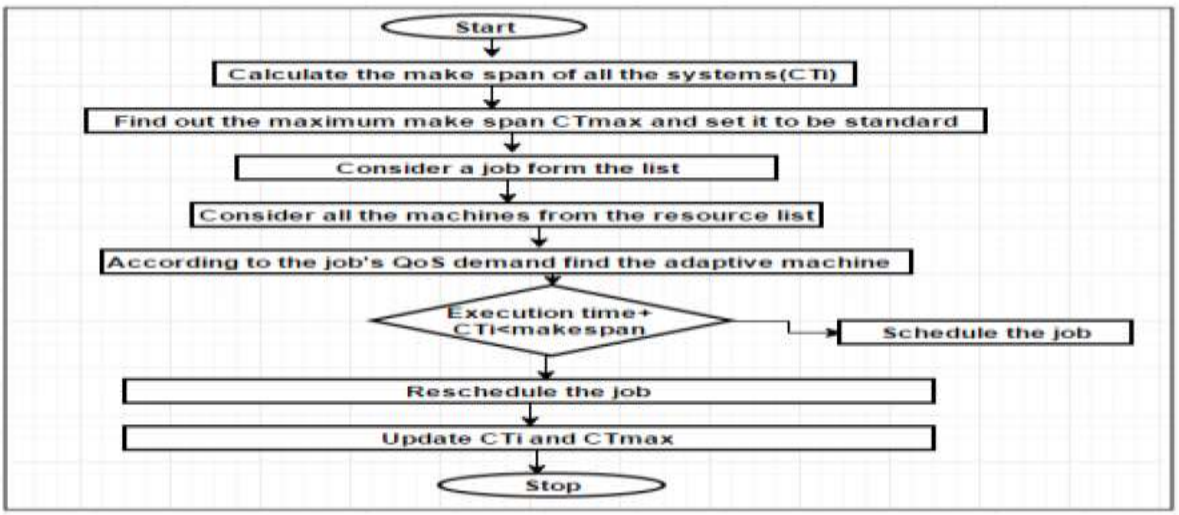

Fig 6: QoS-guided scheduling in grids

IV. Conclusion

In conclusion, the service-based task scheduling of cloud computing should include the algorithms that are necessary to ensure the efficient utilization of resources, a reduced execution time, and the Quality of Service (QoS). The frequent changes and complexities in cloud environment continuously deny the conventional scheduling methods the ability to meet the diverse requirements of users. Examples of advanced algorithms include heuristic, meta-heuristic, and machine learning based scheduling algorithms that are more intelligent and adaptive and can address several quality-of-service requirements simultaneously. It is through these algorithms that it becomes possible to assign tasks more effectively, expenses are saved, and it becomes time efficient in rendering services. Future advances in both AI and hybrid scheduling models play a key role in making cloud infrastructures and next generation computing services since they provide greater benefits in scalability, reliability, and swift decision-making.

References

[1] A. Ajith, M Madhavan, Murugesan and Ashok Kumar, "Research survey in cloud computing and big data", IEEE Trans. Chennai vol 4(7), pp 389-402.

[2] S Sekhar Kammula, VV Vinyaka and Mouli Chandra, "Survey on IoT and bigdata in cloud computing", Springer lecture notes conference, Noida, vol 4(8), pp 762-772, 2002.

[3] S. Naik kumar, Patel Yadav and Sohail Md, "Architecture on cloud computing on Amazon services", " *Journal of Cloud Computing and Services Science,* vol/issue: 5(4), pp. 142-152, 2020.

[4] R. Buyya, *et al.*, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Generation Computer Systems*, vol. 25, pp. 599-616, 2009.

[5] F Max Savio, M Sasi Kumar. "An Effective Control Technique for an Impedance Source Inverter Based Wind Energy System". 2012 IEEE International Conference on Emerging Trends in Electrical Engineering and Energy Management (ICETEEEM-2012)

[6] Sasikumar M and Chenthur Pandian S. "Characteristics Study of ZSI For PMSG Based Wind Energy Conversion Systems". Journal of Electrical Engineering (JEE). ISSN: 1582-4594.