Webology (ISSN: 1735-188X)
Volume 18, Number 5, 2021

Comparative Analysis Of Task Scheduling Algorithms In Cloud Environment In Term Of Their Future Prospective And Risk

Dr. Sunil Kumar[1] , Dr. Subhash Chander[2]

[1]Assistant Professor, Guru Nanak College for Girls, Sri Muktsar Sahib, INDIA,

[2]Assistant Professor, University College, Jaito, Faridkot, INDIA,

Abstract:

In the present scenario, due to covid19 pandemic dependents on the computer have increased at rapid speed, which can access from growth level gained by IT companies. For the processing of different applications, there are two types of environments Centralized or Distributed. Distributed Environments originate in reality in various avatars in the field of computing for example Grid Computing, Utility computing, Cloud Computing. Cloud computing is an evolving technology that is a virtual pool of resources and is provided to users on a pay-per-use basis without the burden of managing the infrastructure. For the best utilization of resources, efficient task scheduling is a major issue from the last eras, and amply task scheduling algorithms have been proposed for this purpose by well-known researchers. The problem of discovering an optimal feasible allocation is known to be NP-Hard and Heuristic techniques have found the optimal Solutions so this paper gives the extensive study of various heuristic task scheduling algorithms especially in terms of future perspectives and risks associated with them.

Keyword:

Cloud Computing, Task Scheduling, Distributive Environment, Resource Utilization

Introduction:

Cloud computing has grown in popularity in recent years as a result of its ability to share resources at a low cost and with high reliability. Various techniques, such as cluster, grid, and cloud computing, have emerged over the years. High Performance Distributed Computing (HPDC) is supported by cluster computing. Parallel or distributed computing processes are applied to computationally heavy applications over networks of computers in HPDC environments. Grid computing, on the other hand, can be conceived of as a distributed system with non-interactive jobs that work with a huge number of files. Grids are another type of distributed computing in which a

virtual network of supercomputers is made up of loosely connected computers that work together to complete long tasks. Cloud computing is a new technology that uses the internet to deliver a variety of services (R. Buyya et. al., 2009) such as infrastructure, software, and other applications. Cloud computing can alternatively be characterized as a distributed computing paradigm that consists of a collection of networked and virtualized computers that are supplied and shown dynamically (Shukur, H. et. al., 2020) as unified computing resources and are available as a pay-per-use service. All distributed computing aims to make better use of distributed resources, combining them to obtain increased throughput and the ability to solve large-scale computational problems. It is necessary to sequence the operations for this goal. Task scheduling and resource scheduling are two processes for managing resources and tasks, and they are both critical issues in these systems. There has been a lot of studies done on this topic, and it can be divided into two categories: independent and process scheduling.

Independent task scheduling refers to a situation in which all tasks are independent of one another and have no precedence relationships. As a result, tasks can be assigned to processors according to a priority list as they become available, and there is no need to analyses a project digraph before doing so. Many complicated and computationally intensive applications, on the other hand, can be described as workflows. These workflow applications are generally described as the sequence of tasks to be processed in well-defined order to accomplish a specific goal (Buyya RK et. al.,2008). Cloud workflow systems can be regarded as a platform service that facilitates the automation of distributed large-scale applications in cloud computing. Workflows are represented as directed acyclic graphs (DAG), which consists of nodes that are linked according to their flow dependencies: G = (V, E), which G represents the graph, V represents a vertex or a node that is an operation or a task, and E represents the edge that shows the relationship between two nodes (Barrionuevo JJD et. al., 2012).

As we further classified, it can be categorized into three categories i.e. Heuristic, Meta-Heuristic, and Hybrid task scheduling algorithms. Heuristic task scheduling algorithms are the practical methods that provide ease to schedule the task and provide the best possible solutions, but it doesn't assure that provided result is optimal but these methods can be used to speed up the process of finding satisfactory results.

The rest of the paper is organized as follows. Section 2 gives a detailed survey about the various scheduling algorithms based on the task Scheduling techniques. Section 3 provides the analysis of the selected algorithms and section 4 gives the conclusion about the complete study.

Task scheduling Algorithms:

In the present scenario, there are three types of scheduling algorithm mainly in existence, Heuristic and Meta-Heuristic and fusion of these algorithms to achieve the optimum results. The word heuristic comes from the Greek language in which its meaning is discovered, it is an approach to solve any problem with the nearest to an optimal result. It is an educated guess to solve any problem, where the optimal solution is impractical. In task scheduling our basic purpose is to reduce the execution time

of the system. For this purpose, various types of heuristic techniques have been proposed out of which some popular algorithms are named as Max-Min, Min-Min, Sufferage, HEFT, and TANH. All of these algorithms are basically used to get the nearest optimum result for task scheduling. Max-Min, Min-Min, and sufferage algorithms are the batch mode algorithms by nature, which means in these algorithms' jobs are collected at once and provide these jobs to the different processors to get the parallel execution and get the minimum makespan. On the other hand, Meta-heuristic means a technique which is independent from the problems and can be used for large set of data. means high level of problems. Various meta-heuristic algorithms have been proposed till date such as genetic algorithm, PSO, Artificial bee colony etc.to achieve the optimum results in the real-life application of cloud computing in term of different parameters.

The idea of the Min-Min algorithm is given by O. H. Ibarra and C. E. Kim in 1977 but it gets its real shape in 1999 by one of renowned researcher (Muthucumaru Maheswaran et. al. 2012), one of the faculty members of the University of Manitoba, Canada, and his friends. They have also proposed the concept of Max-Min and Suffrage Algorithms. A task having a minimum MCT value overall tasks is chosen to be scheduled first and given to the faster machine in the min-min algorithm, on the other hand, a task having maximum MCT is chosen to be scheduled to the faster machine in the max-min. The Max-Min will perform better than the Min-Min algorithm when a queue of jobs has shorter tasks than long tasks. Instead to these, Sufferage works on the idea that assigns a machine to a task that would suffer means who doesn't allocate the resources due to its expected completion time. All of these proposed algorithms performed better than conventional algorithms

(Soheil Anousha et. al.,2013) Soheil proposed an algorithm for task scheduling in Grid Environment with the name Improved Min-Min algorithm. This algorithm takes the vital features of min-min and removes the drawbacks. It improves the total completion time as compared to the conventional min-min algorithm by estimating the complete task on each resource and arranging all the tasks in ascending order. After that to select the task for scheduling, it computes the average completion time and Standard Deviation of existing tasks and if ACT is less than SD it will select the task from the front of the queue otherwise it will assign the rear of the queue to assign the next task. It outperforms in terms of makes pan and load balancing as compared to Min-Min.

Today, Cloud computing is getting more popularity, so a lot of algorithms have been proposed to date. (Gang et. al., 2012) authors have proposed an Improved Min-Min algorithm for Cloud computing. As compared to the conventional min-min algorithm, the proposed algorithm gives more resource utilization, executes the long tasks in a reasonable time and it fulfills the users' requirements. It maintains the three constraints quality of service, dynamic priority, and cost of service as per users' requirements. Its quality to manage the cost of service gives the trait to cloud to pay-per-use feature.

(Yingchi Mao et. al.,2014) authors has given an improved algorithm which is based on the conventional technology Max-Min but it gives better resource utilization and reduced response time for elastic clouds. It allocates the resource in an efficient way to ensure better user satisfaction. This algorithm provides the elasticity in cloud computing by maintaining the task status table to access the load on every resource and the expected completion time of tasks. By designing the table, the

appropriate tasks can be allocated to resources and load can be managed with a reduction in response time.

Researchers has materialized the task-based access control in the Min-Min algorithm to solve the resource leveling problem. The concept of TBAC was given by RK Thomas and R. Sandhu, to provide security to the resource by setting the access control to the resources. By emerging, the features of TBAC model with Min-Min algorithm authors given an improved algorithm with the capability to maintain the balanced load on resources with low cost and security of available resources. The use of TBAC gives the timeliness of the scheduling algorithm and min-min gives the rational allocation of resources (Lin Guoyuan and He sahn, 2011).

One of the renowned researchers in the field of cloud computing from India has introduce an algorithm named as M-Level Sufferage-based scheduling Algorithm (MSSA) to reduce the schedule length by minimizing the makes pan. In this algorithm, firstly, the earliest time is calculated using M-Level, and finally, the priority is assigned to the task to select the optimal machine for it. The proposed Algorithm always gives the optimal or near to optimal result. It gives better results in terms of reduced scheduling length and resource utilization as compared to MET, MCT, Min-Min, and Max-Min. Authors also have recommended adding the QoS constraint to the algorithm as future scope to get the realistic application (Sanjaya Kumar Panda et and Pabitra Mohan Khilar, 2013).

(Haluk Topcuoglu et. al., 2002) Authors have proposed two algorithms Heterogeneous Earliest Finish Time (HEFT) and Critical Path on a Processor (CPOP) to get high performance and fast scheduling in the heterogeneous environment. HEFT is based on the insertion-based technique, it selects the task with the highest rank value and assigns it to the processor in every step and it repeats the step till all the tasks are not allocated. By this process proposed algorithm gives the minimum earliest finish time. CPOP algorithm is summing up the upward and downward rank values to prioritizing the tasks. CPOP manages the critical task on a processor in such a way that it provides the minimum total execution time. In this work, a parametric graph was designed to generate WDAG with different characteristics. The results show that both of the algorithms outperform as compared to conventional approaches in terms of quality and cost of schedules.

In 2002, HEFT was proposed to manage Heterogeneous environments like Grid or cluster computing, with time Cloud computing was introduced in this environment with more complexity. (Nitish Chopra and Sarbjeet Singh, 2013) Researchers proposed a new algorithm based on HEFT for workflow scheduling for cost optimization within the deadline in a cloud environment mainly hybrid cloud, which is a collaboration of public and private resources to meet the requirement of users. The purpose of this algorithm is to provide the resources to users at minimum cost and manage the resources from private or take the resources from the public cloud on lease as per requirement and within the deadline. Experimental results show the betterment of HEFT as compared to the Greedy Method and Min-Min algorithm in terms of cost and fulfill the schedule within the deadline as per the Service Level agreement settled by the user and Cloud service provider.

(Rashmi Bajaj and Dharma P. Agrawal, 2004) Both of the researchers have given one of the renowned algorithms in 2004 for heterogeneous environments where parallel tasks are available with some precedence relationship. For the environment, they are given a task duplication-based

Author	Strong point	Weaknesses	Prospect	Risk
(Muthucumaru Maheswaran et. al. 2012)	1. It provides the minimum makes pan values as compared to the conventional algorithm. 2. It also gives the low average sharing penalty	Performances of algorithms are depending on the consistency property of the ETC Matrix and the arrival rate of tasks.	These are renowned algorithms (Max-Min, Min-Min, and Sufferage) that have been used to design more realistic applications, and further, there is more scope.	Selection of heuristic-based on various factors, choice of the wrong heuristic for a solution can lead to worse performance.
(Soheil Anousha et. al.,2013)	It gives the better makes pan as compared to Min-Min algorithm and load balancing	Makes pan and load balance are factors that are paradoxical to each other so it is impossible to get optimum results.	This algorithm outperforms so there is scope to consider some issues like deadlines on tasks and resources to make them more realistic applications.	It is a heuristic technique so it can give a better result but not the optimum result
(Gang et. al., 2012)	1. More resource Utilisation. 2. Capable to run long tasks at reasonable time.	It is defining the three different constraints but priority and cost constraint is part of QoS Constraint	Algorithm can be used with different QoS parameters which can make it rational.	It considered the priority constraint which can lead to more cost when the long task comes with higher priority
(Yingchi Mao et. al.,2014)	1. It gives more response time with better resource utilization. 2. It shortens the average pending time for activities	The algorithm works with a batch processing system but the due environment is more dynamic in nature	Algorithm works for elastic cloud so there is an option to get more optimum results by adding some new constraints.	It is not considering the QoS parameters so it can lead to unrealistic solutions.
(Lin Guoyuan and He Shan, 2011)	1. It assures the allocation of resources to balance the load. 2. Ensure the security of Grid resources.	It is not considering the cost, which is the primary element of every task scheduling.	Popular technologies min-min and TBAC have emerged for Grid, which can extend for cloud also.	It is used the Min-Min algorithm for minimum makes pan but it is better for small size grid,

				so it can inefficient for large size Gird.
(Sanjaya Kumar Panda et and Pabitra Mohan Khilar, 2013)	It gives maximum resource utilistaion with minimum scheduling length. It works better in even complex DAG problems	It is based on assumption that communication time and computation time is provided before execution with can lead to unrealistic schedule.	It can be more realistic if security, time constraints and duplication of task related parameters are considered in this algorithm.	Due to the Network problem there may be loose or delay of communication between task and resources, which can lead to inefficient schedule length.
(Zhao Tong et. al., 2021)	• It achieves the lowest average task response time while ensuring a high task completion rate. • It has good adaptability, robustness and scalability.	It doesn't consider the factor of energy consumption and economic benefits.	As it works on real-time applications like google, Alibaba, it can be more realistic with involvement of factor of energy consumption.	As it led to more complex application it can lead to inaccurate results.
(Nitish Chopra and Sarbjeet Singh, 2013)	It gives the schedule with the minimum cost within the deadline recommended by the user.	It is only considering the cost and deadline and it gives the same makes pan as compared to the existing algorithm	It surpasses the other algorithms in terms of cost, it can be better with some amendment.	In hybrid-cloud, we merge the private cloud with public which arise the issue of security, which is not considered in this research work.
(Rashmi Bajaj and Dharma P. Agrawal , 2004)	Algorithm optimize the waiting time, turnaround time, and response time.	Implemented on limited number of tasks even the real time applications work on a greater number of tasks.	With emergence of more meta-heuristic algorithms it can predict the best quantum values of tasks automatically.	It works on limited number of tasks so with increase of tasks can lead to reduction of its performance.

(Shivani Dubey et. al., 2013)	It gives the minimum makes pan by reducing the communication cost with an assignment of the dependent task to the same server.	It leads to duplication which may cause overhead and can lead to a higher overall cost.	It is considering the cost and Makes pan and there is later scope for improvement by considering various QoS parameters like budget, makes pan performance, deadline, penalty cost, file availability.	It is not considering the various parameters and the cloud is a large network so it may not be a cop to a realistic environment.
(Kashik olaei et. al.,2019)	It gives the best stability criteria, CPU time, load balancing, the percentage of efficiency, and makes pan as compared to other reputed algorithms.	Proposed algorithm form with a fusion of two algorithms can lead to the collision of their properties.	In today's age, if the fusion of different algorithms is performed in a better way than it can lead to better optimum results.	Limited population sets used are used so it can lead to non-optimum results even management of efficiency and load balancing are contradicted to each other.

scheduling algorithm named TANH, which gives the optimum results for applications that are represented through DAG by reducing the makes pan. This is an extensive improvement on the existing work on Task Duplication based scheduling (TDS). It also works better in the case of a homogenous network environment.

Author has proposed an innovative algorithm for a cloud computing environment based on the TANH algorithm. In the cloud environment, task management is more critical, where the environment is heterogeneous. In this work, researchers have proposed an algorithm based on cluster technology with emergence to task duplication-based scheduling algorithm for Heterogeneous System (TANH). In this algorithm, the same type of dependent task is grouped into one cluster so that the communication cost required between the clusters could be decreased and we can achieve the minimum completion time. It is compared with various scheduling algorithms and gives comparatively low schedule time in terms of minimum makes pan (Shivani Dubey et. al., 2013).

Comparative Analysis in terms of Prospective and Risk and Performance Parameters:

The analysis is a most renowned tool for audit and analysis of the overall capability of any element. Basically, it is considered that it is used only to judge the business or a venture but it gives the basic capability and weakness of elements in terms of internal and external environments in the term of Strongpoint, Weakness, Risk, and Prospect. So, we have chosen the analysis to assess the capability of

research work to date in the field of task scheduling in a Cloud Computing. As above described above a lot of work has been purposed by various researchers, so there is a need for analysis of these works. In the paper, we have given an extensive Analysis of Task Scheduling Algorithms in the cloud computing presented in Table 1 and Table 2. Where Table 1 shows the Strong points of every algorithms being proposed by researchers with weakness, future prospective and risk associated with every proposed algorithm by renowned researchers. Table 2 shows the complete analysis of various proposed algorithms in term of their performance parameters which are taken into consideration by the researchers.

Table 1: Analysis of Task scheduling Algorithms

Authors	**Algorithm Type**	**Platform Used**	**Performance Parameters**						
			Make Span	Load Balancing	Resource Utilization	Execution Time/ Response Time	Security	Cost	Deadline Constraint
(Muthucumaru Maheswaran et. al. 2012)	W	Grid Sim	√	×	×	×	×	×	×
(Soheil Anousha et. al.,2013)	I	Grid Sim	√	√	×	×	×	×	×
(Gang et. al., 2012)	I	Cloud Sim	×	×	√	√	×	×	×
(Yingchi Mao et. al.,2014)	I	Cloud Sim	×	√	√	×	×	×	×
(Lin Guoyuan and He Shan, 2011)	I	Grid Sim	×	√	×	×	√	×	×
(Sanjaya Kumar Panda et and Pabitra Mohan Khilar, 2013)	I	Grid Sim	√	×	√	×	×	×	×
(Zhao Tong et. al., 2021)	I	Python-Tesnorflow	×	×	×	√	×	×	×
(Nitish Chopra and Sarbjeet Singh, 2013)	W	Cloud Sim	×	×	×	×	×	√	√

(Rashmi Bajaj and Dharma P. Agrawal, 2004)	I	Cloud Sim	√	×	×	√	×	×	×
(Shivani Dubey et. al., 2013)	W	Cloud Sim	√	×	×	×	×	√	×
(Kashikolaei et. al.,2019)	W	Dot Net	√	√	√	√	×	×	×

Table 2: Comparative Analysis based on performance parameters used

Here **I** stand for Independent tasks and **W** stands for Workflow based Tasks

Conclusion:

After the extensive Analysis of task scheduling techniques, this has been come out that every technique has its own strength and weakness so only one of the techniques is not capable to give the optimum result in the task scheduling so there is a need to give the fusion of various techniques to get the optimum results and in the present era, various authors have given various meta-heuristic based techniques which can be used with the fusion of these heuristic techniques to give the optimum results. In the present era, data computational activities have been increased at rapid speed so scope of refinement in the task scheduling algorithms always remain. So, after complete review of eminent researcher works, as concern to future work there is need to proposed an algorithm with fusion of various techniques **to** achieve optimum result for real time application with multi-objective algorithm with concentration of QoS.

Acknowledgement:

This work is not sponsored by any funding organization but I am really thankful to Dr. Manpreet Singh for his Guidance and thankful to my Parents and My Wife Mrs. Pink Sethi and Son Master Akshit for their valuable Support.

References:

R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic. 2009. Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Generation computer systems, 25(6), 599-616.

Buyya RK, Kotagiri R, Yu J. 2008. Workflow scheduling algorithm for grid computing. In: Meta-heuristics for scheduling in distributed computing environment, 146. Berlin Heidelberg: Springer; 173–214.

Barrionuevo JJD, Fard HM, Prodan R, Fahringer T. 2012. A multi-objective approach for workflow scheduling in heterogeneous environment: cluster, cloud and grid computing. In: 12th IEEE International conference, 300–309.

Muthucumaru Maheswaran, Shoukat Ali, Howard Jay Siegel, Debra Hensgen_, and Richard F. Freund. 1999. Dynamic Matching and Scheduling of a Class of Independent Tasks onto Heterogeneous Computing Systems. IEEE, eighth Heterogeneous Computing Workshop. 30-44.

Soheil Anousha1, and Mahmoud Ahmadi. 2013. An Improved Min-Min Task Scheduling Algorithm in Grid Computing. Springer-Verlag Berlin Heidelberg, 103-113.

Gang Liu, Jing Li, and Jianchao Xu, "An Improved Min-Min Algorithm in Cloud Computing", Springer, Proceedings of the 2012 International Conference of MCSA, AISC 191, pp. 47–52.

Yingchi Mao, Xi Chen and Xiaofang Li. 2014. Max–Min Task Scheduling Algorithm for Load Balance in Cloud Computing. Springer, Proceedings of International Conference on Computer Science and Information Technology, Advances in Intelligent Systems and Computing, 257-265.

Lin Guoyuan, He Shan. 2011. The Application and Implementation of TBAC in Min - Min Scheduling Algorithm. IEEE 2nd International Conference on Software Engineering and Service Science (ICSESS), 213 – 215.

Sanjaya Kumar Panda and Pabitra Mohan Khilar. 2013. MSSA: A M-Level Sufferage-based Scheduling Algorithm in Grid Environment", Springer Berlin Heidelberg, Distributed Computing and Internet Technology, 410-419.

Zhao Tong, Feng Ye, Bilan Liu, Jinhui Cai, Jing Mei. 2021.DDQN-TS: A novel bi-objective intelligent scheduling algorithm in the cloud environment. Neurocomputing, 455, 419–430.

Nitish Chopra, Sarbjeet Singh. 2013. HEFT based Workflow Scheduling Algorithm for Cost Optimization within Deadline in Hybrid Clouds. IEEE, Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), 1-6.

Fahd Alhaidari and Taghreed Zayed Balharith. 2021. Enhanced Round-Robin Algorithm in the Cloud Computing Environment for Optimal Task Scheduling. Computers, 10(63), 1-27.

Shivani Dubey, Vismay Jain, Shailendra Shrivastava.2013. An Innovative Approach for Scheduling of Tasks in Cloud Environment. IEEE, Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), 1-8.

Kashikolaei, Seyedeh Monireh Ggasemnez had; Hosseinabadi, Ali Asghar Rahmani; Saemi, Behzad; Shareh, Morteza Babazadeh; Sangaiah, Arun Kumar; Bian, Gui-Bin .2019. An enhancement of task scheduling in cloud computing based on imperialist competitive algorithm and firefly algorithm. The Journal of Supercomputing 76(2).

Shukur, H., Zeebaree, S. R., Ahmed, A. J., Zebari, R. R., Ahmed, O., Tahir, B. S. A., et al. 2020. A State of Art Survey for Concurrent Computation and Clustering of Parallel Computing for Distributed Systems. Journal of Applied Science and Technology Trends, 1(4), 148-154.

H. Topcuoglu, S. Hariri , Min-You Wu. 2002. Performance-effective and low-complexity task scheduling for heterogeneous computing. IEEE Transactions on Parallel and Distributed Systems, 13(3), 260 – 274.

Rashmi Bajaj and Dharma P. Agrawal. 2004. Improving Scheduling of Tasks in a Heterogeneous Environment. IEEE Transactions On Parallel And Distributed Systems, 15,(2), 107-117.