INTERNATIONAL JOURNAL OF COMMUNICATIONS
DOI: 10.46300/9107.2020.14.2 Volume 14, 2020

Contemporary Perception Of Task Scheduling Techniques In Cloud: A Review

Samah Alshathri[1]
[1]*Information technology, College of Computer and Information Sciences*
Princess Nourah Bint Abdulrahman University
Riyadh, Saudi Arabia
sealshathry@pnu.edu.sa

***Abstract*— Typically, the constant changes in computers and communications technology led to the need of on-demand network access to a shared computing resources to reduce cost and time and this is known as Cloud computing, which delivers computing services to users as a pay-as-you-go manner by emerging several distributed and high performance computing concepts. The cloud makes reaching any information or source possible from anywhere eliminating the setup and instillation step such that the user and the hardware may co-exist in different places. This comes beneficial for the users or the small companies that cannot effort to pay for the hardware, storage or resources as the big companies. Many of the studies on cloud computing was dedicated to the performance efficiency of task scheduling. Scheduling is a wide concept and it is one of the most important issues that generally work on mapping tasks to appropriate resources efficiently and effectively using one or more strategy. This paper have reviewed and classified the most recent scheduling algorithms in cloud computing and gave examples on each.**

Keywords*— cloud, scheduling techniques, green computing, energy efficient, optimization, Algorithms

I. INTRODUCTION

The use of cloud computing is tremendously growing and it is becoming the driver for innovation in all companies to serve their customers around the world. The main base of cloud computing environments are data centers, which are used to house computers and their associated components. Moreover, applications run on servers in the data center that consume considerable energy to host cloud applications. Data center energy use has received much attention recently neglecting the energy consumption of the transmission and switching between networks that are the key to connect users to the cloud, which needs to be the focus of all the information and communications technology sectors aiming to cloud.

One of the key elements of cloud platform is the scheduler, which connect the users' application with the computing resource. scheduling technique is one of the most effective approaches that affects the performance of the cloud computing data centers but it is quite a challenge in cloud environment and it is done based on the user requirements.

Section II of this paper defines the concept of scheduling in cloud computing. In section III, gives an overview on the scheduling literature. Section IV, present the scheduling algorithm measurements and metrics. Section V, gives an overview on the most recent existing strategies for scheduling approaches in the cloud. Finally, the last section concludes and highlights an emerging research challenge to address.

II. SCHEDULING ALGORITHMS

One of the most current and main challenges of the data centers is to optimize the total operating cost while maintaining the desired Quality-Of-Service (QoS) standards. Using different scheduling algorithms can have a major effect on the energy consumption in the data center, and the overall performance of the cloud. Scheduling algorithm provide an alternative mechanism to optimize resource utilization and reduce power consumption while balancing between Quality of Service (QoS) and fairness among the jobs. In computer science, scheduling is the method by which access the process have to the system resources [1]. Scheduling algorithm is usually constructed based on one or more strategy and time, cost, energy, quality of service (QoS), and fault tolerance are the most important strategies used [2].

III. LITERATURE REVIEW

Task allocation and scheduling complications where comprehensively studied and efficient algorithms have been proposed. The task scheduling in a cloud environment aim to minimize the energy consumption, make span, cost, and deliver the best quality of service (QoS). Almezeini et al [2] propose a novel task-scheduling algorithm based on Lion Optimization Algorithm (LOA) for cloud computing. The inspiration of this scheduler was the behavior of lions and their characteristics and the results shows high performance compared to the GA and PSO scheduling techniques. Razaque et al. [3] proposed an efficient task-scheduling algorithm based on the availability of network bandwidth. A nonlinear programming model was used to assign tasks to each VM. Amjad et al. [4] proposed a greedy and a genetic algorithm with an adaptive selection of suitable crossover and mutation operations (named as AGA) to allocate and schedule real-time

tasks with precedence constraint on heterogamous virtual machines. N.Moganarangan et al. [5] presented a new Hybrid algorithm is proposed for reduction of energy consumption and make span by merging the advantages of ACO and cuckoo search algorithm. Deepika et al [6] proposed an algorithm that Categorize tasks based on their deadline and cost restrictions and assign them to different priority queue, and regarding the resource selection the scheduler select VM with the lowest turnaround time for each individual task. D.I Esa et al [7] proposed a new job scheduling mechanism using Firefly Algorithm to minimize the execution time of jobs. The proposed mechanism outperformed the FCFS algorithm. Komarasamy et al [8] proposed a new scheduling technique called Content-based Federated Job Scheduling (CFJS) algorithm in the cloud computing, to execute the deadline and non-deadline based jobs concurrently in a VM. The scheduler minimizes the waiting time of tasks; maximize the resource utilization and throughput. Paper [9] proposed an algorithm that assigns priority to tasks based on task size where the highest size task has highest priority, while the VMs selection is based on their MIPS values where the one having highest MIPS is given the top priority to be selected.

IV. Performance Metrics of scheduling Algorithm

There are many measurements that could be used to define the scheduling algorithm performance, and have to be considered by in the development process of the scheduling algorithm such as [10] [11]:

- Execution Time (ET)
  It is the time from submitting a task to the cloud until it executed.
- Response Time (RT)
  The time that the system start responding to the submitted task.
  RT=Service Time-Wait Time
- Execution Cost
  It is the over-all cost of the resources used at the task execution.
- Make span (M-s)

It is the total time to finish all the tasks scheduling.
M-s= Completion Time - Start

- Reliability

It means that the user should receive a continue service without any kind of failures.

- Scalability

The system power to develop itself due to the growing demand or the increasing of the data.

- Fairness

It means the tasks equality in sharing the CPU time.

- Job Rejection Ratio

It is the ration of the overall rejected tasks to the total submitted number of tasks.

- User Satisfaction Level

It is the satisfaction of the user on the resources like storage and computation.

- Budget constrains

The cost limitation for processing all tasks.

V. Contemporary Scheduling Techniques

Task scheduling is one of the crucial stages in the system that plays a significant role in the overall performance. The goal of scheduling algorithms is distributing the load of the system on processors to maximize the utilization and minimize the total tasks execution time. Scheduling techniques could be classified based on six criteria: Urgency (immediate vs batch), priority (pre-emptive vs non-pre-emptive), distribution (centralized vs decentralized), cooperation (independent vs workflow), prior knowledge (heuristic vs meta-heuristic), and flexibility (static vs dynamic) [12] [13] [14]:

A. Immediate (online) vs. Batch(offline) Scheduling

The immediate (online) scheduling used in real-time services. In event-triggered systems where the tasks parameters are not fully known, an online scheduler is invoked to take decisions based on pre-defined rules. The tasks are scheduled as soon as they arrive with no waiting time [15] [16] and there are several examples on this kind of scheduling as in Table 2.

Batch (offline) Scheduling is used in time triggered (TT) systems and it is applied when all tasks arrive at the same time. The offline scheduling usually uses a scheduling table, which lists tasks and their processing times. It has a specific time interval of execution where tasks are hold until their processing time is in [15] [16]. Table.1 show some of the offline scheduling.

TABLE 1. ONLINE AND OFFLINE SCHEDULING TECHNIQUES EXAMPLES

Scheduling Technique type	Scheduling technique	Description
Online Scheduling	First fit (FF)	This algorithm assigns the task to any available server it finds while a task is assigned to a new server once there is no enough resources in the available servers
	Best fit (BF)	The scheduler here chooses the server where the task best fits in with the least remaining resource after the task is assigned to it
	Worst fit (WF)	This algorithm attempts to choose the server that has the largest amount of remaining resources.
	Random fit (RF)	The scheduler assigns the task randomly to any available server with enough resources
Offline Scheduling	First fit decreasing (FFD)	This algorithm sorts all tasks in descending order based on metric stated by the algorithm
	Best fit decreasing (BFD)	This algorithm sorts all tasks in descending order based on the least amount of remaining resource

B. Preemptive vs. Non preemptive scheduling

Pre-emptive scheduling is based on priority and allow task processing to be interrupted and moved to another resource such as Shortest remaining time first (SRTF) or Pre-emptive priority based job scheduling algorithm in green cloud (PPJSGC). In the case of Shortest remaining time first, the task with the lowest remaining time to finish processing is selected to start processing [17], while for Pre-emptive priority based job-scheduling algorithm in green cloud, tasks are assigned based on the best fit as per their energy requirements and server frequency availability, which is executed by the DVFS Controller [18].

Non pre-emptive scheduling is unalike the pre-emptive one, where in this technique once tasks start processing it won't stop until it finish completely its execution. A couple of techniques that are typical examples on this scheduling are improved shortest job first scheduling algorithm to decrease starvation and Priority. The first technique is an enhancement Shortest Job First scheduling where the scheduler assigns the processes in order of task arrival as shortest hob first with a possibility to change the priority to diminish the waiting time of task. The second technique is Priority scheduling, where this scheduling assigns each task with a priority level and allocate it based on this level, so that tasks with higher priority are processed first [19].

C. Centralized vs.(Distributed)Decentralized Scheduling

In centralized scheduling, tasks in the system use a central scheduler, which make the process of monitoring the resources easy and efficient but complex and fault tolerance. The Priority Task Scheduling Strategy for Heterogeneous Multi-Datacenters is one of the examples on the centralized scheduling where in this task scheduling the strategy is based on three parameters: task deadline, task length and task age, where those parameters are used to improve the priority of task scheduling with regards to the deadlines [20].

The decentralized scheduling have no central control where the local schedulers keep the state of scheduling, which makes the resource scheduling more efficient but less control over scheduling process. The techniques in Table. 2 are examples on this scheduling.

TABLE 2. DECENTRALIZES SCHEDULING EXAMPLES

Scheduling Technique Type	Scheduling Technique	Description
Decentralized scheduling	Load balancing task scheduling algorithm based on feedback mechanism	The algorithm uses the weighted random strategy, overload assessment and feedback to submit tasks first to the resources with the greatest performance but insure not to overload it [21].
	De-centralized dynamic task scheduling using hill climbing algorithm	This dynamic scheduling algorithm uses hill climbing algorithm aiming to reduce the completion time of tasks and improve the throughput and resources utilization [22].

D. Independent vs. Workflow scheduling

The independent scheduling executes tasks independently and assigns tasks to the processors in the order given by the priority list once they become free.

Workflow scheduling in cloud becomes an important research topic and it is one of the prominent issues in this domain. In this type of scheduling, tasks are dependent on each other where a task can start its execution until all its preceding tasks are already finished. Workflow scheduling is described by a Directed Acyclic Graph (DAG), in which each task is represented by a node and the flow by edges. The main advantage of this scheduling is to reduce the make span and expand the utilization of resources [23].
The following techniques in Table 4 are examples on both of these scheduling techniques.

TABLE 3. INDEPENDENT AND WORKFLOW SCHEDULING EXAMPLES

Scheduling Technique Type	Scheduling Technique	Description
Independent Scheduling	Period ACO_based scheduling Algorithm (PACO)	PACO use ant colony optimization algorithm with the first proposed scheduling period strategy and the improvement of pheromone intensity update strategy, which optimize the make span and load balance [24].
	Self-adaptive ant colony optimization (SAACO)	This algorithm use the swarm optimization (PSO) to change the parameters of Ant colony optimization (ACO) to be self-adaptive. This algorithm improve the total make span and the load balance [25].
Workflow Scheduling	Improved particle swarm optimization (IPSO)	The IPSO is used to minimize the total cost of assigning tasks on available resources. Total cost values are obtained by varying the communication cost between the resources, task dependency cost values, and the execution cost of compute resources [26].
	PRS scheduling algorithm	This Uncertainty-Aware Real-Time Workflow Scheduling algorithm combines the proactive and the reactive scheduling methods, to achieve cheaper computational overhead [27].
	Cost effective genetic algorithm for workflow scheduling in cloud under deadline constraint (CEGA)	It is a novel scheduler based on genetic algorithm (GA) for encoding, population initialization, crossover, and mutation operators of the Genetic Algorithm [28].

E. Heuristic vs. Meta-heuristick scheduling

Heuristic scheduling algorithms are extracted from intuitions and usually get the easy and quick solution but not the best one. The performance of heuristic-based algorithms profoundly relied on the success of the heuristics [29] [30].
Unlike heuristic meta-heuristic scheduling methods, use a guided-random-search-based process for solution searching. Metaheuristic methods always have much higher computational cost than heuristic but can obtain better performance in terms of schedule quality [31][11]. Table 4 shows different examples on both techniques.

TABLE 4. HEURISTIC AND META-HEURISTICK TEQNIQUES EXAMPLES

Scheduling Technique type	Scheduling technique	Description
Heuristic Scheduling	The heterogeneous earliest finish time (HEFT)	This algorithm assigns priorities based on the earliest start time of each task and it minimizes the task's start time by allocating a task to the processor.
	Max- Min scheduling	Max-Min algorithm selects a task from the tasks list that have the maximum completion time on a resource that can execute it within a shorter period of time
	Enhanced Max-Min	It is an enhanced Max-Min algorithm where the scheduler assigns a task from the tasks list to resource based on the average time of job execution instead of highest completion time.
Meta-Heuristic Scheduling	Genetic scheduling algorithms(GA)	In GA, a chromosome is used to represent each possible solution and a random population is taken and used as an initial data.
	Ant colony optimization scheduling algorithms(ACO)	This algorithm imitates the ant's life style where a number of artificial ants exchange information through a communication scheme to help in creating solution for optimization problems
	Particle swarm optimization scheduling algorithms(PSO)	It is an intelligent algorithm based on the social behavior of animals such as a flock of birds searching for a food source or a school of fish defending themselves from a predator. It is practical for function optimization problems. İn PSO, the solutions are named particles and each particle will have a fitness value that will be defined by a fitness function to have trajectory based on its best position and the position of the best particle of the whole population.
	League championship algorithms(LCA)	It is inspired by the contests of sport teams in the league sport. A league schedule is designed periodically for individuals to play in pairs and the win or loss result depends on the fitness value of a team.

F. Static Scheduling vs. Dynamic Scheduling

In Static scheduling, all the existing resources and the data of the tasks are available in advance by the time of task is scheduled and there is no resource failures, and as an example on this technique the Enhanced load balanced Min-Min (ELBMM) for Static Meta Task Scheduling in Cloud Computing.
In dynamic Scheduling, tasks are scheduled dynamically over time and it is more flexible than static scheduling but more overhead. The following scheduling techniques are examples on this scheduling:

TABLE 5. STATISC AND DYNAMIC TECHNIQUES EXAMPLES

Scheduling Technique Type	Scheduling Technique	Description
Static Scheduling	Enhanced load balanced Min-Min (ELBMM)	This scheduling is based on Min-Min approach where it selects the task with maximum completion time and assigns it to appropriate resource, which effectively utilizes resource, and deliver better make span [32].
Dynamic Scheduling	Dynamic resource scheduling method based on fuzzy control theory	This algorithm predict user resource requirement using the Second moving average method then the relationships between resource availability and the resource requirements are determined[33].
	A delay-based dynamic scheduling algorithm for bag-of-task workflows with stochastic task execution times	This algorithm calculates the actual task execution times using the summation of task execution time expectation and standard deviation. It Minimize the cloud resource renting cost by the deployment of both a bag-based delay scheduling strategy and a single-type based virtual machine interval renting method [34].
	Self-adaptive layered sleep-based method for security dynamic scheduling	This algorithm improves the resource utilization, accuracy and efficiency of security resources scheduling by combining three different models, which are decision-making tree, top-down analytical and self-adaptive filtering [35].

CONCLUSION

In cloud computing environment, different resources are provided by the virtual machines which are scheduled by scheduling algorithm. Many of the studies on cloud computing was dedicated to the performance efficiency of task scheduling. Scheduling is a wide concept and it is one of the most important issues that generally work on mapping tasks to appropriate resources efficiently and effectively using one or more strategy. A good scheduling algorithm can be implemented using more number of parameters, which will deliver a good performance, and outputs that can be conveyed for deployment in a cloud environment in future. In this paper,

in the studied literature, most of the authors have focused on reduction of make span and execution cost whereas others have given significance to response time, throughput, flowtime and average resource utilization. We also presented scheduling metrics which can be coupled to formulate a framework for recourse scheduling in cloud computing. After that we have reviewed and classified several contemporary scheduling techniques and algorithms in cloud computing.

REFERENCES

[1] Shivani, H. Kaur, "Task Scheduling for Utilization of Resources using Cloud Computing," International Journal of Computer Applications (0975 – 8887), Vol. 174 – No.4, September 2017.

[2] N. Almezeini, A. Hafez, "Task Scheduling in Cloud Computing using Lion Optimization Algorithm" (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 8, No. 11, 2017

[3] A. Razaque, N. R. Vennapusa, N. Soni, G. S. Janapati, k. R. Vangala "Task Scheduling in Cloud Computing" IEEE Long Island Systems, Applications and Technology Conference (LISAT), April 2016; pp. 1–5.

[4] A. Mahmood, S. A. Khan, "Hard Real-Time Task Scheduling in Cloud Computing Using an Adaptive Genetic Algorithm" Computers 2017.

[5] N.Moganarangan, R.G.Babukarthik, S.Bhuvaneswari, M.S.Saleem Basha, P.Dhavachelvan, "A novel algorithm for reducing energy-consumption in cloud computing environment: Web service computing approach." Journal of King Saud University - Computer and Information Sciences, Vol.28, No.1, Jan 2016, PP 55-67.

[6] Deepika Saxena, R.K. Chauhan And Ramesh Kait "Dynamic Fair Priority Optimization Task Scheduling Algorithm In Cloud Computing: Concepts And Implementations", I. J. Computer Network and Information Security (IJCNIS), February 2016.

[7] D. I. Esa, A. Yousif, "Scheduling Jobs on Cloud Computing using Firefly Algorithm" International Journal of Grid and Distributed Computing, Vol. 9, No. 7 (2016), pp.149-158.

[8] D. Komarasamy, V. Muthuswamy, "Content-Based Federated Job Scheduling Algorithm in Cloud Computing", Australian Journal of Basic and Applied Sciences, Vol. 10(2), PP: 52-59, 2016.

[9] M. Lawanya Shri, M.B.Benjula Anbumalar, K. Santhi And Deepa.M "Task Scheduling Based On Efficient Optimal Algorithm In Cloud Computing Environment" International Conference on "Recent Research Development in Science, Engineering and Management (ICRRDSEM), May 2016.

[10] H. Goel, N. Chamoli, "Job Scheduling Algorithms in Cloud Computing: A Survey" International Journal of Computer Applications, Vol. 95, No.23, June 2014.

[11] N. Kaur, A. Chhabra, "Comparative Analysis of Job Scheduling Algorithms in Parallel and Distributed Computing Environments", International Journal of Advanced Research in Computer Science, Vol. 8, No. 3, April 2017.

[12] N. Panwar, M. S. Rauthan, "Analysis of various task scheduling algorithms in cloud environment: Review" 2017 7th International Conference on Cloud Computing, Data Science & Engineering – Confluence,Jan 2017.

[13] Santhosh B, Manjaiah D H, Dr.L.Padma Suresh, "A Survey of Various Scheduling Algorithms in Cloud Environment", 2016 International Conference on Emerging Technological Trends [ICETT],

[14] H. Vashishth, K. Prakash, "The Art Of Scheduling In Cloud Computing" International Journal of Advance Research, Ideas and Innovations in Technology, Vol.2, no.4,2016.

[15] N. Sharma, S. Tyagi, "A Survey on Heuristic Approach for Task Scheduling in Cloud Computing" International Journal of Advanced Research in Computer Science, Vol. 8, No. 3, March – Apr 2017.

[16] A. K. Mishra, Y. Mohapatra, A. K. Mishra, "A fundamental study of scheduling algorithm for Cloud Computing Job Selection" Journal of Environmental Science, Computer Science and Engineering & Technology, Vol.5. No.4, PP 282-293, Nov 2016.

[17] C. S. Bindu, A. Y. Reddy, P. D. K. Reddy, "Intelligent SRTF: A New Approach to Reduce the Number of Context Switches in SRTF," Springer Science+Business Media Singapore 2017,2017.

[18] A Preemptive Priority Based Job Scheduling Algorithm in Green Cloud Computing," 2016 6th International Conference - Cloud System and Big Data Engineering, Jan 2016.

[19] S. K. Nager, N. S Gill"An Improved Shortest Job First Scheduling Algorithm to Decrease Starvation in Cloud Computing," International Journal of Computer Science and Mobile Computing (IJCSMC), Vol. 5, Issue. 8, PP.155 – 161, Aug 2016.

[20] N. Er-raji, F. Benabbou, "Priority Task Scheduling Strategy for Heterogeneous Multi-Datacenters in Cloud Computing," (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 2, 2017.

[21] Z. Qian, G. Yufei, L. Hong, S. Jin "A Load Balancing Task Scheduling Algorithm based on Feedback Mechanism for Cloud Computing," International Journal of Grid and Distributed Computing Vol. 9, No. 4, PP.41-52, 2016 .

[22] M. Mohammadi, A. M. Rahmani "De-centralised dynamic task scheduling using hill climbing algorithm in cloud computing environments," International Journal of Cloud Computing (IJCC), Vol. 6, No. 1, 2017.

[23] M. Gautam , R. Kumar, "A Survey on Workflow Scheduling in Cloud Computing Environment", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 5, May 2016.

[24] M. Patel, R. Kadian, "A Review on ACO based Scheduling Algorithm in Cloud Computing," International Journal of Computer Science and Mobile Computing (IJCSMC), Vol. 5, Issue. 5, PP.489 – 493, May 2016.

[25] W. Sun, Z. Ji, J. Sun, N. Zhang, Y. Hu, "SAACO: A Self Adaptive Ant Colony Optimization in Cloud Computing," 2015 IEEE Fifth International Conference on Big Data and Cloud Computing, Aug 2015.

[26] N. Sadhasivam, P. Thangaraj , "Design of an improved PSO algorithm for workflow scheduling in cloud computing environment" Intelligent Automation & Soft Computing, Volume 23, Issue 3, PP.493-500, Aug 2016.

[27] H. Chen, X. Zhu, D. Qiu, L. Liu, "Uncertainty-Aware Real-Time Workflow Scheduling in the Cloud" The Journal of Supercomputing, 2017.

[28] J. Meena, M. Kumar, M. Vardhan, "Cost Effective Genetic Algorithm for Workflow Scheduling in Cloud Under Deadline Constraint," IEEE Access, vol. 4, pp. 5065–5082, 2016.

[29] P. Banga, S. Rana, "Heuristic based Independent Task Scheduling Techniques in Cloud Computing: A Review", International Journal of Computer Applications (0975 – 8887),Vol. 166, No.1, May 2017.

[30] M.Jaeyalakshmi, P.Kumar, "Task Scheduling Using Meta-Heuristic Optimization Techniques in Cloud Environment" International Journal Of Engineering And Computer Science, Vol. 5, NO. 11, Nov 2016, PP. 19050-19053.

[31] J. Jiao, W. Yu , L. Guo, "Research on Batch Scheduling in Cloud Computing" TELKOMNIKA, Vol.14, No.4, Dec 2016, pp. 1454-1461.

[32] G. Patel, R. Mehta, U. Bhoi, "Enhanced Load Balanced Min-min Algorithm for Static Meta Task Scheduling in Cloud Computing" Procedia Computer Science, Vol. 57, PP 545-553, 2015.

[33] Z. Chen, Y. Zhu, Y. Di, S. Feng,"A Dynamic Resource Scheduling Method Based on Fuzzy Control Theory in Cloud Environment" Journal of Control Science and Engineering, Jun 2015.

[34] Z. Cai, X. Li, R.Ruiz, Qianmu Li,"A delay-based dynamic scheduling algorithm for bag-of-task workflows with stochastic task execution times in clouds" Future Generation Computer Systems, Vol. 71, Issue C, PP. 57-72 ,June 2017.

[35] X. Liu, Y. Zhou, "A Self-Adaptive Layered Sleep-Based Method for Security Dynamic Scheduling in Cloud Storage" 4th International Conference on Information Science and Control Engineering (ICISCE) (2017), PP. 99-103,2017.