A Survey Of Various Qos-Based Task Scheduling Algorithm In Cloud Computing Environment

Ronak Patel, Hiren Mer

Abstract: Cloud computing is a new terminology which is achieved by Distributed, Parallel and Grid computing. Cloud computing is a new design pattern for large, distributed data centers. Cloud computing offers end consumers an “a pay as go” model – a powerful shift for computing towards a utility model like the electrical system, the telephone system, or more recently the Internet. Cloud computing provides different types of resources like hardware and software as services via internet. Task scheduling is the key role in cloud computing. Tasks require minimum completion time, better performance, utilization of resources and quick response time for which cloud uses the concepts of the virtualization. For task allocation cloud provides virtual machine which is scalable but scheduling them is a major problem. In this paper we study different types of task scheduling algorithms with QoS facility and issues related to them in cloud computing. These scheduling algorithms focus on resource management, response time, load balancing and performance.

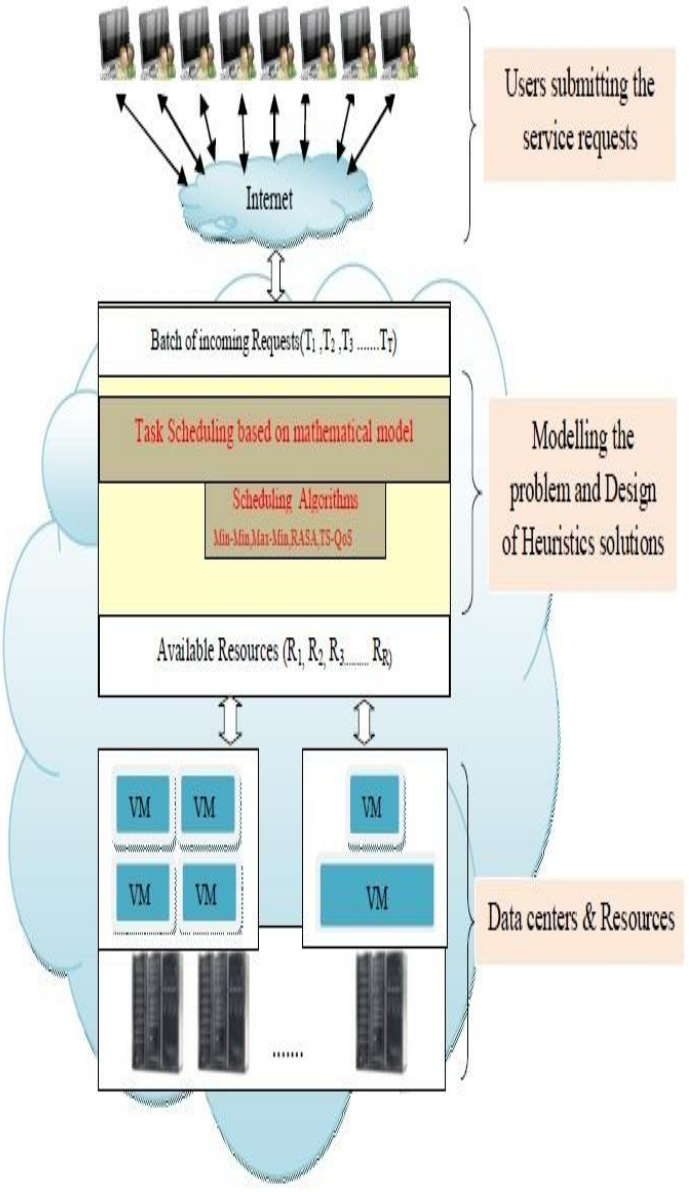
Index Terms: Cloud computing, Make Span, Priority, Quality of Service, Resource Allocation, Task Scheduling, Task Completion Time

1 INTRODUCTION

Cloud computing is a new technological trade which provide trustworthy services through cloud centers. Cloud centers hold the facility for virtualized computing and also storage technologies [1]. With cloud computing, user can increase the computing power and storage. Also variety of software services can be utilized from the wide range of available applications on the basis of pay-per-use, irrespective of the location. Cloud computing is a style of computing where dynamically scalable and virtualized resources are provided as services. 'Cloud' provides a flexible execution environment for resources which involve multiple cloud users and provide a superior service at multiple granularities for the specified level of quality of service (QoS). QoS is the collective effort of services performance, which determines the degree of the satisfaction of a user for the services [2]. QoS is expressed on the basis of qualitative measures like completion time, latency, execution price, packet loss rate, throughput and reliability. Task scheduling is an important issue which greatly influences the performance of cloud computing environment. The cloud service provider and clients have different objectives and requirements. To fulfill user’s requirement, cloud provides an extra level of virtualization in the whole task allocation business which come with the gain of being easily scalable, but scheduling them is a complicated problem. The key issue is how to dispatch efficiently and reasonably the tasks of users to different resources according to the Quality of Services (QoS) requirements of both cloud computing center and users, which belongs to task scheduling [10]. The paper is organized as follows: Section 2 gives a review of some of the task scheduling techniques used in the literature. In Section 3, Compression of QoS-based task scheduling technique. Section 4 metrics of task scheduling in cloud. Section 5 concludes the paper.

- *Ronak Patel is currently pursuing masters degree program in Computer science and engineering in Gujarat Technical University, India, PH-09409330900. E-mail: er.ronak@hotmail.com*
- *Hiren Mer is currently working as Assistant Professor in Computer Science & Engineering Department in Parul Institute of Technology,Gujarat,India. PH-09227885688. E-mail: prof.hirenmer@gmail.com*

Figure 1
conceptual modeling of the cloud computing environment for task scheduling



2 VARIOUS TASK SCHEDULING TECHNIQUES

A very beginning concern cloud is coming from distributed and grid, so mostly scheduling strategy is applied in cloud which is related to distribute and grid. Generally two category of the scheduling algorithm; Static Scheduling and Dynamic Scheduling. In Static Scheduling, Tasks are arrives simultaneously and available resource schedule updated after each task is schedule. In Dynamic Scheduling, task and machine set location and allocation is not going to fix. Dynamic strategy applied in two fashions: On-line mode heuristic scheduling and Batch mode heuristic scheduling. In on-line mode heuristic scheduling, tasks are scheduled when they arrived in the system. In Batch mode, tasks are queued and collected into set when they arrive in the system. The scheduling will start after a fixed period of time. In 2003, H.xiaoshan et al [3] suggested a QoS Guided Min-Min heuristic [Batch mode heuristic algorithm] was introduced in that some task require higher network bandwidth to exchange a large amount of data among processors, whereas some can be satisfied with the lower network bandwidth. In this algorithm the matching of the QoS request and services between the tasks and hosts based on conventional Min-Min. Firstly each task with the high QoS request in the Meta task, the algorithm finds the earliest completion time and the host that obtains it, in the entire QoS Qualified host. Secondly find the task with the minimum earliest completion time and assigns the task to the host that give the earliest completion time to task. In this algorithm they have addressed only one-dimension QoS issue, because they worked only bandwidth constraint. In 2006, F.Dong et al. [4] proposed a QoS priority grouping algorithm which considers deadline and acceptation rate of the task and the makespan as main factor of task scheduling in whole system. It achieves better acceptance rate and completion time for the submitted task then Min-Min and QoS Guided Min-Min. In 2008, C.Hsu et al [5] carried out two optimization schemes MOR (Makespan Optimization Rescheduling) and ROR (Resource Optimization Rescheduling). MOR focus on improving the makespan to pull off the better performance and in ROR focus on the re-dispatch tasks from the machine with the minimum number of tasks to other machine, which is helpful to reduce the resource need. Both this technique achieves low complexity, high effectiveness, good performance than QoS Guided scheduling algorithm and Min-Min algorithm. In 2008, M.Singh et al [6] proposed a QoS based predictive Max-Min, Min-Min switcher algorithm. In this algorithm, scheduling of the next job is based on appropriate selection among QoS based min-min or QoS max-min algorithm. The effect on the execution time grid jobs has been reduced due to non-dedicated resources. It normally uses the history information about the execution jobs to predict the performance of non-dedicated resources. This algorithm merges the efficiency of max-min along with min-min and also considers both QoS and non-dedicated property of grid resources. In 2009, S.Parsa et al [7] introduced a new task scheduling algorithm called RASA which has the advantage of both Min-Min and Max-Min algorithm. In this first estimate the completion time of the tasks on each resource and then applied both the algorithm. RASA use the Min-Min strategy to execute the small task first then long task and then applied Max-Min to avoid the delays in the execution of large task and support concurrency in the execution of the large and small tasks. It achieves the lower Makespan with good QoS. In 2010, Mrs.S.Selvarani et al [8] introduced an improved cost-based scheduling algorithm for making efficient mapping of tasks to available resources in cloud. The improvisation of traditional activity based costing is proposed by new task scheduling strategy for cloud environment where there may be no relation between the overhead application base and the way that different tasks cause overhead cost of resources in cloud. This scheduling algorithm divides all user tasks depending on priority of each task into three different lists. This scheduling algorithm measures both resource cost and computation performance, it also Improves the computation/communication ratio. In 2011, C.Zhao et al [9] proposed a Berger Model in Cloud computing in that algorithm scheduling process establish dual fairness constraint. First constraint is to classify user task by QoS preferences, and establish the general expectation function in accordance with the classification of tasks to restrain the fairness of the resources in the selection process. Second constraint is to define resource fairness justice function to judge the fairness of the resources allocation. According to constraint, the algorithm always assigns tasks on the optimal resources in order to satisfy the QoS requirement of user and it avoid to consider a long task for execution. Experiment result of this algorithm shows effective execution of the user tasks and manifest better performance. In 2013, X.Wu et al [10] introduce a task scheduling algorithm based on QoS-driven in cloud computing (TS-QoS). In this TS-QoS algorithm compute the priority of the task according to the special attributes of the tasks, and then sort tasks based on priority. Then the algorithm calculate the completion time of each task on different services, and schedule each task onto a service which can complete the task as soon as possible according to the sorted task queue. But in this process priority can change dynamically an increase continuously this can help to solve the “starvation” problem and follow FCFS principle. Experimental result achieves well performance and load balancing by QoS driving form both priority and completion time.

3 Comparison of QoS- based Task Scheduling Algorithm [11]

TABLE 1

EXISTING QOS-BASED TASK SCHEDULING ALGORITHM

Scheduling Algorithm	Scheduling Method	Scheduling Parameters	Scheduling Factors	Finding	Environment	Tools
QoS Guided Min-Min Heuristic [3]	Batch Mode	Quality of service, Make span	Bandwidth of tasks	1. Reduce the Makespan then Min-Min 2. Use only bandwidth parameter for QoS	Grid	GridSim
QoS Priority Grouping Algorithm[4]	Batch Mode	Acceptance rate, completion time	Grouped tasks	1. Deadline and acceptance rate of the tasks 2. Makespan	Grid	GridSim
Towards Improving QoS-Guided Scheduling[5]	Batch Mode	Makespan	Grouped tasks[Jobs]	1. Improving makespan to achieve better performance 2. Reduce the Resource Need by Rescheduling	Grid	GridSim
QoS based predictive Max-min, Min-min switcher[6]	Batch Mode	Makespan	Heuristic	Better performance with QoS	Grid	GridSim
RASA[7]	Batch Mode	Make span	Grouped tasks	Use to reduce the make span	Grid	GridSim
Improved Cost-Based Algorithm for Task Scheduling[8]	Batch Mode	Cost, Performance	Unscheduled task	1. Measure both resource cost and computation performance 2. Improve the computation /communication ratio	Cloud	CloudSim
Job[Task] Scheduling based on Berger Model[9]	Batch Mode	QoS, Fairness Constraint , Completion time	Bandwidth of tasks	1. Improving Task execution time and get better performance. 2. Better meet the user expectation	Cloud	CloudSim
Task scheduling algorithm based on QoS-driven in cloud computing (TS-QoS)[10]	Dynamic Batching mode	Priority of Task, Minimum Completion time	Grouped Tasks	1. Great performance 2. Load balancing	Cloud	CloudSim

TABLE 2
METRICS CONSIDERED BY THE QOS-BASED TASK SCHEDULING ALGORITHM

Techniques	Time	Cost	Scalability		Scheduling Success Rate	Make span	Speed	Quality of Service	Resource Utilization	Reliability	Availability
			Dynamic	Static							
T1	x	x	x	√	x	√	x	√	x	x	x
T2	√	√	x	√	x	√	√	√	√	√	x
T3	x	x	x	√	√	√	x	√	√	x	√
T4	x	x	x	√	x	√	√	√	x	x	x
T5	x	x	x	√	x	√	x	x	x	x	x
T6	x	√	x	√	x	x	x	x	x	x	x
T7	x	x	x	√	x	√	x	√	√	x	x
T8	x	x	√	x	x	√	√	√	√	√	x

4 METRICS FOR TASK SCHEDULING IN CLOUD

The Existing task scheduling algorithm considers various parameters like time, cost, and make span, speed, scalability, throughput, resource utilization, scheduling success rate, quality of service and so on. Also multiple task scheduling algorithm, reliability and availability can be considered. Table 2 gives the details about the different metrics considered for QoS-based task scheduling.

5 CONCLUSION AND FUTURE WORK

Scheduling is considered as the major factor for task execution in cloud environment. In this paper, we have surveyed the various scheduling algorithm and tabulated different parameters. The existing scheduling algorithm considered as topic of research can be used to introduce more efficient and better improved performance of algorithms based on arriving time of task, execution of task on resources and cost of the communication. So we need to improve availability and reliability of cloud computing for task allocation.

REFERENCES

[1] R. Buyya, C.S. Yeo, S.Venugopal, J. Broberg, I.Brandic . “Cloud computing and emerging IT platforms: vision ,hype and reality for delivering computing as the 5th utility ” Future Generation Computing System, 2009, 25(6), p.599-616.

[2] Syed Muhammad Ahsan. “A framework for QoS computation in web service and technology selection” computer standards & Interfaces.2006,28(6),p.714-720

[3] XiaoShan He,Xianhe Sun and Gergor von Laszewski. “QoS guided Min-Min heuristic for grid task scheduling”. Journal of Computer Science and Technology, 2003, 18(4), p.442-451.

[4] Dong. F, Luo. J, Gao. L and Ge. L, "A Grid Task Scheduling Algorithm Based on QoS Priority Grouping," In the Proceedings of the Fifth International Conference on Grid and Cooperative Computing (GCC’06), IEEE, 2006.

[5] Ching-Hsien Hsu,Zhan. J.,Wai-Chi Fang, et al. “Towards improving QoS-guided scheduling in grid” Third ChinaGrid Annual Conference(CHINAGRID). Dunhuang, Gansu, China, 2008, p.89-95.

[6] M.Singh and P.K.Suri; “QPSMax-Min<>Min-Min : A QoS Based Predictive Max-Min, Min-Min Switcher Algorithm for Job Scheduling in a Grid”, “International Technology Journal7(8)” :p.1176-1181,2008

[7] Saeed Parsa and Reza Entezari-Maleki,” RASA: A New Task Scheduling Algorithm in Grid Environment” in World Applied Sciences Journal 7 (Special Issue of Computer & IT): 152-160, 2009.

[8] Mrs. S.Selvarani1, Dr.G.Sudha Sadhasivam,” improved cost-based algorithm for task scheduling in Cloud computing “in IEEE 2010.

[9] Baomin Xu,Chunyan Zhao,Enzhao Hua,Bin Hu. “Job scheduling algorithm based on Berger Model in Cloud Environment” Advance in Engineering Software, 2011, 42(7), p.419-425.

[10] Xiaonian Wu, Mengqing Deng, Runlian Zhang, Bing Zeng, Shengyuan Zhou. “A task scheduling algorithm based on QoS-driven in Cloud Computing” Information Technology and Quantitative Management(ITQM2013).2013,p.1162-1169

[11] Anju Bala, Dr.Inderveer Chana. “A survey of various workflow scheduling algorithms in cloud environment”. International Journal of Computer Application (IJCA).2011,p.26-30.