Review Of Scheduling Methodologies Of Virtual Machines (Vms) In Heterogeneous Cloud Computing

S.Rekha, Dr.C.Kalaiselvi

Abstract: Cloud computing is a most excellent technology and users are in require of more service and higher effectiveness load balancing and scheduling plays the major significant role in cloud computing. In cloud efforts are reduced and also give service interaction to the user. By this users are able to help to pay for what they use. Lots of methods are introduced for describing the several scheduling tasks. Because Cloud computing handles with many user's suitable decisions are needed for every scheduled job. In this paper, we have reviewed about the details of different algorithms proposed to resolve the issue of task scheduling in cloud computing. This paper provides detailed review of various studies on different algorithms with the purpose of is explained to overcome the common issues recognized in different scheduling tasks. If the scheduling tasks are performed efficiently then it results to balance the load in cloud based on the resources and workloads. Since the variety of scheduling algorithms are used by load balancers to determine which backend server to send a request to VM. It is also the responsibility of the provider to dynamically reallocate or migrate the VM across physical machines for workload consolidation and to avoid over utilization or under utilization of resources.

Keywords: Cloud Computing system, Virtual Machine scheduling, scheduling algorithms, heterogeneous workload, and Quality of Service (QoS), Virtual Machine (VM).

1. INTRODUCTION

Cloud Computing is an emerging technique. Recently it is found that researchers are interested in using cloud for performing scientific applications and even the big organizations are on the verse of switching over to hybrid cloud. Many complex applications require parallel processing to execute the jobs effectively. Due to the communication and synchronization among parallel processes there is a decrease in utilization of Central Processing Unit (CPU) resources. It is necessary for a data center to achieve the utilization of nodes while maintaining the level of responsiveness of parallel jobs. The cloud computing is attracting an increased number of applications to run in the remote data centers. Many complex applications require parallel processing capabilities. Some of the parallel applications show a decrease in utilization of CPU resources whenever there is an increase in parallelism if the jobs are not schedule correctly then it reduces the computer performance.

- *S.Rekha is currently pursuing Ph.D. Department of computer science, Tirupur kumaran college for women. Also working as an Assistant Professor, Department of IT, Dr.N.G.P. Arts and Science college. E-mail: rekhaishanth@gmail.com*
- *Dr.C. Kalaiselvi, Head and Associate professor, Dept of Computer Applications, Tirupur kumaran college for women, tiruppur. E-mail: kalaic29@gmail.com*

Clouds provide a very large number of resources, including platforms for computation, data centers, storages, Networks, firewalls and software in form of services. At the same time it also provides the ways of managing these resources such that users of cloud can access them without facing any kind of performance related problems. Several algorithms & protocols are proposed in the regarding of scheduling mechanism [1] in cloud computing. But very few algorithms are proposed to detect the scheduling mechanism in cloud computing. Many of them consider a regular monitoring region in their protocol, which is not a real life scenario. Practically the monitoring region with improved performance is always irregular as the clouds are randomly deployed [2-3]. Virtual Machine (VM) is an execution unit that acts as a foundation for cloud computing technology. Virtualization consists of creation, execution, and management of a hosting environment for various applications and resources. The VMs in the cloud computing environment share resources like processing cores, system bus, and so forth. The computing resources available for each VM are constrained by total processing power. In this model of environment the job arrival pattern is unpredictable and also the capabilities of each virtual machine vary from one another. Hence, load balancing becomes a critical task leading to a poor system performance and maintaining stability. Thus, it becomes imperative to develop an algorithm which can improve the system performance by scheduling the task to handle the demand among virtual machines. There are various load balancing algorithms available, such as round robin, weighted round robin, dynamic load balancing, Equally Spread Current Execution (ESCE) Algorithm, First Come First Serve, Ant Colony algorithm, and Throttled algorithm. The most frequently used scheduling techniques for a non preemptive system are First In First Out (FIFO) and Weighted Round Robin (WRR) [4]. CloudSim-3.0.3 is the simulation environment for the cloud computing research works. It supports both system and behaviour modelling of cloud system components such as data centres, hosts, VMs, and resource provisioning policies. It supports

modelling and simulation of cloud computing environments consisting of both single and internetworked clouds. It exposes custom interfaces for implementing scheduling and load balancing policies of jobs into VMs and provisioning techniques for allocation of VMs under internetworked cloud computing scenarios. It can leverage virtualized services even on the fly based on requirements varying with time. This paper focuses on the review of several scheduling algorithms in order to optimize the VM scheduling in cloud system, aiming at reducing the delay performance of each and every one of the jobs over time. This paper review of scheduling algorithms is studied in three different categories: normal scheduling methods, Hybrid Scheduling methods and scheduling with delay constraint. The organization of this work is given as like follows: In this section, short introduction about the focus of this research is discussed. In section 2 discusses about various research methodologies. In section 3, the findings of this overall research work is concluded shortly.

2. LITERATURE REVIEW

This paper review about the details of various scheduling algorithms with three major categories: normal scheduling methods, Hybrid Scheduling methods and scheduling with machine learning algorithms. At section 2.1 discuss about the details of usual scheduling methods with optimization, section 2.2 discuss about the details of hybrid optimization methods which are used for scheduling issue in cloud computing, and finally section 2.3 discuss about the details of optimization methods with delay which are used for scheduling issue in cloud computing. After the discussion of all of these algorithms, then inference found from the review work is discussed in the tabular form.

A. NORMAL SCHEDULING

Atiewi et al [5] studied a review of several scheduling algorithms. The aim of cloud task scheduling is to obtain high system throughput and to assign several computing resources to applications. The complexness of scheduling problem increases with the size of the task and becomes extremely issue to handle successfully. Min-Min algorithm is used towards reduce the make span of tasks with assuming the task length. Maintenance this in mind, cloud providers must attain user satisfaction. Agarwal and Jain [6] presented a new Generalized Priority algorithm for proficient execution of task and comparison with First Come, First Served (FCFS) and Round Robin (RR) Scheduling. Algorithm must be tested in cloud Sim toolkit and result demonstrate that it gives improved results when compared to other existing scheduling algorithm. Zuo et al [7] presented a resource cost model with the purpose of describes the demand of tasks on resources with more details. This model returns the association among the user's resource costs and the budget costs. A multi-objective optimization scheduling method has been introduced depending on this resource cost model. This multi-objective optimization algorithm includes the makespan and the user's budget costs as metrics of the optimization issue, obtaining multi-objective optimization of both performance and cost. An improved Ant Colony Optimization (ACO) Algorithm has been introduced to handle this problem. Two constraint functions were used to measure and give feedback regarding the results and budget cost. These two constraint functions complete the algorithms adjust the quality of the solution in an appropriate way depending on feedback in order to obtain the optimal results.Shu-Ching et al., [8] introduced Load Balance Min-Min (LBMM) method uses Min-Min Scheduling algorithm as its base. It uses a three level hierarchical framework. Request manager which is in the first level of the architecture is responsible for receiving the task and assigning it to one service manager in the second level of LBMM. After receiving the request, service manager divides it into subtasks to speed up the processing. Then service manager assigns the subtask to a service node for execution based on CPU space, remaining memory and the transmission rate. This algorithm improves the load unbalance of Min-Min and minimizes the execution time of each node but does not specify how to select a node for a complicated task requiring large-scale computation.Zhan and Huo[9] introduced the improved Particle Swarm Optimization(PSO) algorithm in resources scheduling strategy of the cloud computing. Through experiments, the results demonstrate that this improved PSO method is able to decrease the task average running time, and increases the rate accessibility of resource.Basker et al [10] proposed an enhanced scheduling in Weighted Round Robin (WRR) for the cloud infrastructure services, which considers job length and resource capabilities. This type of algorithm minimizes the response time of the jobs by optimally utilizing the participating VMs using static and dynamic scheduling by identifying the length of the jobs and resource capabilities and effectively predicting the underutilized VMs and avoiding the overload on any of the VMs. The multilevel interdependent tasks have been considered. Load balancing in the heavily loaded scenarios for the task migrations has not been considered.Sandeep et al [11] this concept is used to evaluate this with the up rise of fourth paradigm, that is invention of science over a long interval of time, scientific workflows commence to amend their status amongst in numerous science subject areas indulging physics, astronomy, biology, chemistry, earthquake science and many more. The scheduling algorithms are brushing off the individual dependent and independent tasks. Max-Min algorithm is implemented for scheduling of workflow tasks that is focalized on the consideration of dependent and independent tasks and process independent tasks in parallel that directly gives profit in minimizing computation time.

B. HYBRID SCHEDULING

Elmougy et al [12] proposed a new hybrid Shortest job first and round Robin with Dynamic variable Quantum time (SRDQ) task scheduling algorithm integrating Shortest-Job-First (SJF) and Round Robin (RR) schedulers using a dynamic variable task quantum. The proposed methods majorly depends on two basic keys the first having a dynamic task quantum towards balance waiting time among short and long tasks at the same time as the second considers splitting the ready queue into two sub-queues, Q1 designed for the short tasks and the other for the long ones. Experimentations results indicated that the control of the proposed algorithm over the state of art in reducing waiting time, response time and partially the hunger of long tasks.Ren et al [13] proposed a new elastic service ability of cloud computing platform, several applications are moved

here, which makes able load balancing into an issue. Focusing on the single features of long-connectivity applications which are gradually more popular nowadays, an improved algorithm is introduced depending on the weighted least connection algorithm. In the weighted least connection algorithm, load and processing power are measure, and solitary exponential smoothing forecasting mechanism is used. At last, the results proves by experiments with the purpose of the new algorithm be able to decrease the server load tilt, and increase client service quality successfully.Paul et al [14] described the tasks are partitioned into various groups and they are replicated to local middleware of the system. It makes the system fault tolerant and load balancing improves response time and resource utilization. Lexi search method is employed here to schedule the tasks to various resources along with reducing the cost. The task is assigned based on a probabilistic measurement which is calculated based on the availability of the resource and execution time of the task. Load balancing reduces the overhead created at the scheduler in each resource.Huankai et al [15] an improved load balanced algorithm is introduced on the ground of Min-Min algorithm in order to reduce the makespan and increase the resource utilization (LBIMM). At the same time, Cloud providers offer computer resources to users on a pay-per-use base. In order to accommodate the demands of different users, they may offer different levels of quality for services. Then the cost per resource unit depends on the services selected by the user. In return, the user receives guarantees regarding the provided resources. To observe the promised guarantees, user-priority was considered in our proposed PA-LBIMM so that user's demand could be satisfied more completely.Guo et al [16] described the delay-optimal Virtual Machine (VM) scheduling problem in cloud computing systems, which have a constant amount of infrastructure resources such as CPU, memory and storage in the resource pool. First adopt a queuing model for the heterogeneous and dynamic workloads. Then, we formulate the VM scheduling in such a queuing cloud computing system as a decision-making process, where the decision variable is the vector of VM configurations and the optimization objective is the delay performance in terms of average job completion time. A low-complexity online scheme that combines the shortest-job-first (SJF) buffering and min-min best fit (MMBF) scheduling algorithms, i.e., SJF-MMBF, is proposed to determine the solutions. Another scheme that combines the SJF buffering and Reinforcement Learning (RL)-based scheduling algorithms, i.e., SJF-RL, is further proposed to avoid the potential of job starvation in SJF-MMBF.Al-Arasi and Saif [17] described and implemented an independent task scheduling algorithm with the intention of assigns the users' tasks towards numerous computing resources. The proposed hybrid algorithm designed for task scheduling in CC depending on a Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). The algorithm is implemented and simulated using CloudSim simulator. The results demonstrate that the proposed hybrid algorithm performs better than the other GA and PSO algorithms by reducing the makespan and improving the resource usage.

C. SCHEDULING WITH CLASSIFICATION

Wu et al [18] proposed a market-oriented hierarchical scheduling strategy in cloud workflow systems. The author explains about service level scheduling and task-level scheduling where service-level scheduling deals the task to service assignment. In task-level scheduling deals with the task to VM assignment in the local data center to minimize total execution cost. The scheduling strategies presented random scheduling along with meta-heuristic scheduling algorithm. They use a meta-heuristic algorithm such as GAs, ACO and PSO. For each algorithm Quality of Service (QoS) metrics, makespan, cost and CPU time compared, results show Ant Colony Optimization (ACO) is better than the other. Wu et al [19] proposed a task scheduling algorithm based on QoS-driven for cloud computing. Firstly, in order to reflect the precedence relation of tasks, the proposed algorithm computes the priority of tasks according to the special attributes of tasks, and then sorts tasks by priority. Secondly, the algorithm evaluates the completion time of each task on different services, and schedules each task onto a service which can complete the task as soon as possible according to the sorted task queue. Tan et al [20] evaluated trust service oriented task workflow scheduling algorithm to find the optimal solution for execution time with deadline constraint. To meet user constraints the scheduling algorithm combines direct trust and recommendation includes QoS metrics time, cost. The trust oriented workflow scheduling allocates task suitable for cloud service, with fuzzy and multi-objective functions. Trust is calculated by weight corresponds to the similarity between the active user and each of the other users. Calculations are compared with Minimum Critical Path and greedy cost model. The trust workflow scheduling has statically predetermined schedule but in a dynamic runtime environment, some adaptive method is needed to get optimum result.Zhu et al [21] proposed rolling horizon architecture for real-time aperiodic independent task scheduling and energy saves. Author focus on virtualization technique applied for task through rolling horizon. This rolling horizon architecture sorts the incoming task in order by its deadline. Energy-aware scheduling also handled through virtual machine by calculating start time and execution time. If task can be allocated, then it selects the VM yielding minimal energy consumption to execute the task, otherwise, the algorithm rejects task. The energy-aware scheduling is tested in simulator CloudSim but since this is real time task if it is applied in real cloud environment will improve energy efficiency.Sun et al [22] presented a review study on delay-optimal scheduling of replication in together centralized and distributed multi-server systems. Low-complexity scheduling policies are introduced and are established towards be delay-optimal or near delay-optimal in stochastic ordering between each and every one causal and non-preemptive policies. Results are measured for common system settings and delay metrics with the purpose of allow for subjective arrival processes, random job sizes, random due times, and heterogeneous servers by means of data position restrictions. New sample-path tools are introduced to prove these results. Zhan et al [23] proposed a new optimal scheduling method for cloud storage resources depending on two-threshold load balancing control. The storage resources are optimized by using grid region collocation method, and the problem of

cloud storage resource distribution information flow is created by nonlinear time series reorganization method. The threshold value of resource scheduling is computed by adaptive optimization by piecewise interpolation. The partition scheduling of cloud storage resources is experimented related towards the priority list. The results demonstrate that the proposed algorithm has efficient load balance, high throughput, and it has good application value in resource optimization.

D. ANALYSIS

The performance analysis of this work is done to identify the merits and demerits of these methodologies. The analysis of this work is given in the following table 1.

Table 1. Analysis of the discussed methodologies

S.No	Title	Author	Merits	Demerits
1	A review energy-efficient task scheduling algorithms in cloud computing	Atiewi et al [5]	1. Cloud providers must attain user satisfaction 2. Better resource utilization.	Processing time of each request is not considered.
2	Efficient optimal algorithm of task scheduling in cloud computing environment	Agarwal and Jain [6]	Proficient execution of task. Gives improved results	Doesn't specify how to select a node for a complicated task requiring large-scale computation.
3	A multi-objective optimization scheduling method based on the ant colony algorithm in cloud computing	Zuo et al [7]	Give feedback regarding the results and budget cost Optimization of both performance and cost	Has takes higher execution time
4	Towards a Load Balancing in a Three-level Cloud Computing Network	Shu-Ching et al., [8]	Improves the load unbalance of Min-Min and minimizes the execution time of each node.	Does not specify how to select a node for a complicated task requiring large-scale computation.
5	Improved PSO-based task scheduling algorithm in cloud computing	Zhan and Huo [9]	Decrease the task average running time, and increases the rate accessibility of resource	Task processing time is not considered.
6	An enhanced scheduling in weighted round robin for the cloud infrastructure services	Basker et al., [10]	1. Less complexity and load is balanced more fairly. 2. Equal distribution of work load.	1. Pre-emption is required. 2. Job processing time is not considered.
7	Optimizing workflow scheduling using max-min algorithm in cloud environment	Brar and Rao [11]	1.Works as the Min-Min algorithm. But jobs having large execution time are executed first. 2. Reduces the makespan.	Smaller jobs have to wait for long time.
8	A novel hybrid of Shortest job first and round Robin with dynamic variable quantum time task scheduling technique	Elmougy et al [12]	Reducing waiting time, response time and partially the hunger of long tasks	1. Complexity and long time consumption. 2. Does not consider the existing load on a resource.
9	A Dynamic Load Balancing Strategy for Cloud Computing Platform Based on Exponential Smoothing Forecast	Ren et al [13]	1. Dynamic task assignment strategy proposed, task heterogeneity is considered. 2. Capabilities of each node are considered.	1. Considering only load balancing feature. 2. Complex calculations are involved.
10	Dynamic job scheduling in cloud computing based on horizontal load balancing	Paul et al [14]	Energy is reduced meeting the deadline of tasks	Makespan and cost are less considered here
11	User-priority guided Min-Min scheduling algorithm for load balancing in cloud computing,	Huankai et al[15]	Produces a schedule which improves load balancing and also reduces the overall completion time.	Does not consider priority of a job while scheduling.
12	Optimal Scheduling of VMs in Queueing Cloud Computing Systems With a Heterogeneous Workload.	Guo et al., [16]	1. CPU is allocated to the process with least CPU burst time. 2. Prioritized is given to users improving load balancing and without increasing total completion time.	1. Difficult to understand and code. 2. Rescheduling of tasks to perform load balancing will increase the complexity and time
13	HTSCC: A Hybrid Task Scheduling Algorithm in Cloud Computing Environment	Al-Arasi and Saif [17]	Reducing the makespan and increasing the resource usage.	Applied for specific application only

14	A market-oriented hierarchical scheduling strategy in cloud workflow systems	Wu et al [18]	Matchmaker is to prioritize cloud resources more suitable for the Virtual machine.	Based on highest rank capacity, data stores are allocated, over committing capacity will result in negative values.
15	A task scheduling algorithm based on QoS-driven in cloud computing	Wu et al [19]	The sorted task queue is completed as soon as possible	Machine failure, communication overheads and dynamic workloads are not considered
16	A trust service-oriented scheduling model for workflow applications in cloud computing	Tan et al [20]	CRON (command Run scheduler on notice) is most suitable for scheduling repetitive tasks	Cron is not suited for complex, event-driven tasks, the type of jobs that automate tasks across the enterprise.
17	Real-time tasks oriented energy-aware scheduling in virtualized clouds	Zhu et al [21]	Uses cluster sharing, capacity scheduler partition the resources into pools in the form of multilevel queue, achieve maximum resource utilization and throughput	Only abstraction is queues which are setup by the administrator that reflects cost of the shared cluster, if short job comes after long one it will have to wait until the long one finish.
18	On Delay-Optimal Scheduling in Queueing Systems with Replications	Sun et al [22]	delay metrics is introduced to permit for random arrival processes, random job sizes, random due times, and heterogeneous servers	It is not easily for complex, event-driven tasks.
19	Optimal Scheduling Model of Cloud Storage Resources Based on Double Threshold Load Balancing Control.	Zhan et al [23]	efficient load balance, high throughput, and it has good application value	Higher running time.

From this analysis table, we can conclude that the every methodologies proposed previously consists of various merits and demerits in their way of application. All the merits and demerits involved in these works are considered for the review from which new methodology can be proposed by combining the merits of all the methodologies.

3. INFERENCES FROM THE REVIEW WORK

The delay performance based a great deal on the various scheduling decisions related to how many VM instances is able to be served in parallel and which instances to run first. On the other hand, the heterogeneous and dynamic workload feature of CC environments makes the design of a delay optimal scheduling scheme becomes a major important issue. In a CC environment, users normally request many jobs in terms of several types of VMs and several job lengths, where a category of VM denotes a specified resource set, consists of the Central Processing Unit(CPU), memory and storage [23]. The job length denotes the durations in which a job is expected to run. This exacting feature make use of VM scheduling challenges in a CC system considerably varied from usual scheduling issues, where a job needs only a single dimensional resource, i.e., a task scheduling issue regarding the CPU resource. To solve all of these issue future work will focuses on the develop and implementation of several online algorithms towards optimize the VM scheduling in such a queuing cloud system, aiming at reducing the delay performance of each and every one jobs over time.

4. RESULTS AND DISCUSSION

In this section, simulations are used to measure the results of SJF-RL, SJF-MMBF, SRDQ , LBMM and Min-Min Scheduling. This section compares the delays as well as resource utilization rate performance of SJF-MMBF, SJF-RL, SRDQ , LBMM and Min-Min Scheduling was action with the purpose of maximizes the reward to make best use of the guarantee rate. The guarantee rate is described as the amount of the no. of delay- assured jobs to the number of finished jobs; the greatest guarantee rate denotes the highest no. of delay-assured jobs.

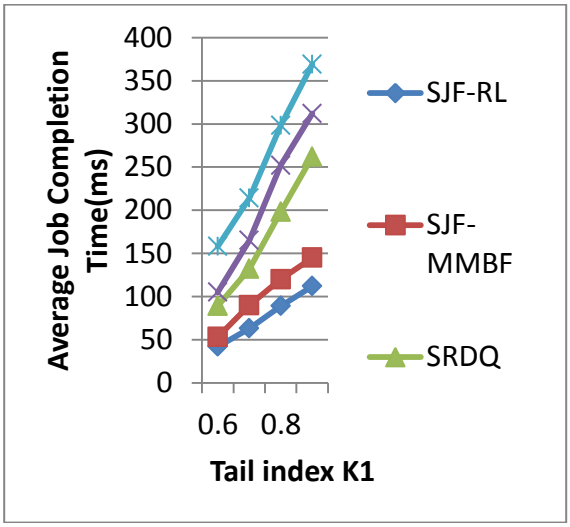

Figure 1. *Delay performance vs. tail index of job lengths (type-1 VM)*

As shown in Figure 1, SJF-RL performs better than the other methods such as SJF-MMBF, SRDQ, LBMM and Min-Min Scheduling by giving the lowest average job completion time in various tail indexes of type-1 jobs. SJF-RL

algorithm performs best average job completion time of 112 ms for tail index of 0.9, whereas other methods such as SRDQ, LBMM and Min-Min Scheduling has needs higher job completion time of 145 ms, 262 ms, 312 ms and 369 ms respectively.

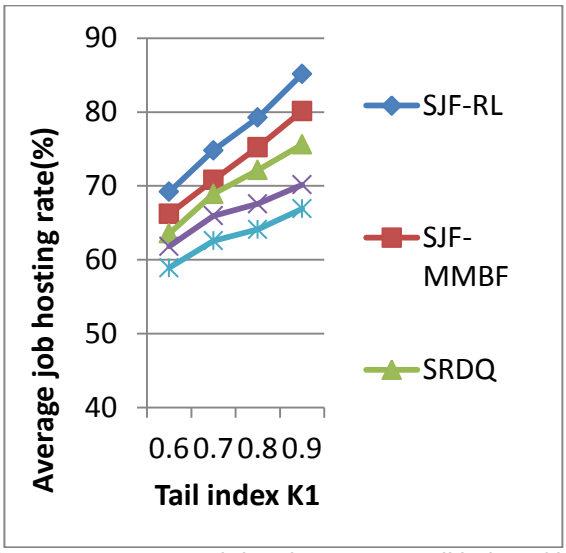

Figure 2. *Average Job hosting rate vs. tail index of job lengths (type-1 VM)*

Figure 2 shows the performance comparison results of average job hosting rate with respect to several scheduling methods such as SJF-RL, SJF-MMBF, SRDQ, LBMM and Min-Min Scheduling in various tail indexes of type-1 jobs. From the results it concludes that the proposed SJF-RL algorithm has highest average hosting rate of 85.15% for tail index of 0.9, whereas other methods such as SRDQ, LBMM and Min-Min Scheduling has needs higher job completion time of 80.15%, 75.62%, 70.15% and 66.92% respectively. The average job hosting rate provides higher rate since the proposed work decision-making is performed by using reinforcement learning.

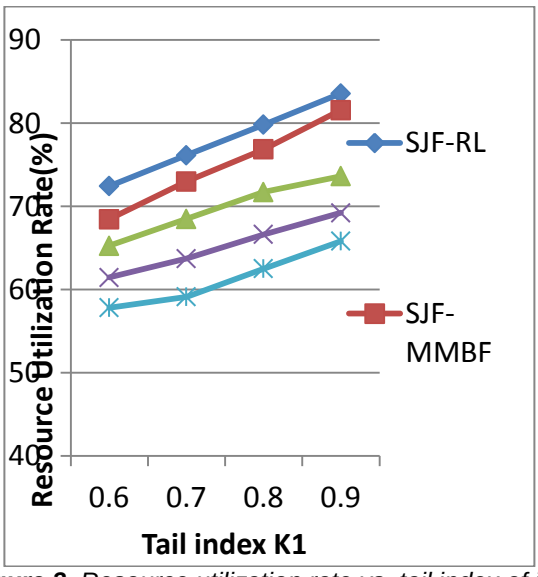

Figure 3. *Resource utilization rate vs. tail index of job lengths (type-1 VM)*

The resource utilization rate of the scheduling methods such as SJF-RL, SJF-MMBF, SRDQ, LBMM and Min-Min Scheduling in various tail indexes of type-1 jobs are shown in figure 3. From the figure 3 it concludes that the resource utilization rate of proposed SJF-RL algorithm is 83.51% for tail index of 0.9, whereas the methods like SJF-MMBF, SRDQ, LBMM and Min-Min Scheduling has reduced resource utilization rate of 81.56%, 73.62%, 69.21% and 65.81% respectively.

5. CONCLUSION AND FUTURE WORK

In this review work, multiple algorithms for load balancing and scheduling for the task in Cloud Computing are analysed with their issues. Review work conclude that the every scheduling algorithms proposed previously includes of various merits and demerits in their way of application. The delay performance based a great deal on the various scheduling decisions related to how many VM instances is able to be served in parallel and which instances to run first. On the other hand, the heterogeneous and dynamic workload feature of CC environments makes the design of a delay optimal scheduling scheme becomes a major important issue. To solve all of these issue future work will focuses on the develop and implementation of several online algorithms towards optimize the VM scheduling in such a queuing cloud system, aiming at reducing the delay performance of each and every one jobs over time. Future work will focus on the efficient utilization of resources; increase overall throughput and decrease the job completion time. All these will reduce the operational cost and will attract more users towards cloud computing.

REFERENCES

[1] Tripathy, L. and Patra, R.R., 2014, “Scheduling in cloud computing”, International Journal on Cloud Computing: Services and Architecture (IJCCSA), 4(5), pp.21-7.

[2] Wilson C., H. Ballani, T. Karagiannis, and A. Rowtron, “Better never than late: Meeting deadlines in datacenter networks,” SIGCOMM Comput. Commun. Rev., vol. 41, no. 4, pp. 50–61, 2011.

[3] Arapakis I., X. Bai, and B. B. Cambazoglu, “Impact of response latency on user behavior in web search,” in Proc. ACM SIGIR, pp. 103–112, 2014.

[4] Devi, D.C. and Uthariaraj, V.R., 2016, “Load balancing in cloud computing environment using improved weighted round robin algorithm for nonpreemptive dependent tasks”, The scientific world journal, 2016, vol.2016, no. 3896065, pp.1-14.

[5] Atiewi, S.; Yussof, S.; Ezanee, M.; Almiani, M, “A review energy-efficient task scheduling algorithms in cloud computing”, In Proceedings of the 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT), Farmingdale, NY, USA, 2016; pp. 1–6.

[6] Agarwal, D. and Jain, S., 2014, “Efficient optimal algorithm of task scheduling in cloud computing environment”, International Journal of Computer Trends and Technology (IJCTT), vol. 9, no. 7, pp. 344-349.

[7] Zuo, L., Shu, L., Dong, S., Zhu, C. and Hara, T., 2015, “A multi-objective optimization scheduling method based on the ant colony algorithm in cloud computing”, IEEE Access, 3, pp.2687-2699.

[8] Wang, S.C., Yan, K.Q., Liao, W.P. and Wang, S.S., "Towards a Load Balancing in a Three-level Cloud Computing Network", 3rd International Conference on Computer Science and Information Technology (ICCSIT), vol. 1, 2010, pp. 108-113.
[9] Zhan, S. and Huo, H., 2012, "Improved PSO-based task scheduling algorithm in cloud computing", Journal of Information & Computational Science, vol.9, no.13, pp.3821-3829.
[10] Basker R., V. Rhymend Uthariaraj, and D. Chitra Devi, "An enhanced scheduling in weighted round robin for the cloud infrastructure services," International Journal of Recent Advance in Engineering & Technology, vol. 2, no. 3, pp. 81–86, 2014.
[11] Brar, S.S. and Rao, S., 2015, "Optimizing workflow scheduling using max-min algorithm in cloud environment", International Journal of Computer Applications (0975 – 8887), vol.124 ,No.4,pp.44-49.
[12] Elmougy, S., Sarhan, S. and Joundy, M., 2017, "A novel hybrid of Shortest job first and round Robin with dynamic variable quantum time task scheduling technique", Journal of Cloud Computing, vol.6, no.1, pp.1-12.
[13] Ren, X., Lin, R. and Zou, H., 2011, "A Dynamic Load Balancing Strategy for Cloud Computing Platform Based on Exponential Smoothing Forecast", Proceedings of IEEE CCIS2011, pp. 220-224.
[14] Paul, M., Samanta, D. and Sanyal, G., 2011, "Dynamic job scheduling in cloud computing based on horizontal load balancing", International Journal of Computer Technology and Applications (IJCTA), vol.2, no.5, pp.1552-1556.
[15] Chen, H., Wang, F., Helian, N. and Akanmu, G., 2013, "User-priority guided Min-Min scheduling algorithm for load balancing in cloud computing", In National Conference on Parallel computing technologies (PARCOMPTECH) , pp. 1-8.
[16] Guo, M., Guan, Q. and Ke, W., 2018, "Optimal scheduling of VMs in queueing cloud computing systems with a heterogeneous workload", IEEE Access, 6, pp.15178-15191.
[17] Al-Arasi, R.A. and Saif, A., 2018, "HTSCC: A Hybrid Task Scheduling Algorithm in Cloud Computing Environment", International Journal of Computers & Technology, vol.17, no.2, pp.7236-7246.
[18] Wu Z, Liu X, Ni Z, Yuan D, Yang Y, "A market-oriented hierarchical scheduling strategy in cloud workflow systems", Journal of Supercomputing., vol.2013, no.16, pp.256–293, 2013.
[19] Wu, X., Deng, M., Zhang, R., Zeng, B. and Zhou, S., 2013, "A task scheduling algorithm based on QoS-driven in cloud computing", Procedia Computer Science, vol.17, pp.1162-1169.
[20] Tan W, Sun Y, Li LX, Lu G, Wang T, "A trust service-oriented scheduling model for workflow applications in cloud computing", IEEE Systems Journal, vol.8, no.3, pp.868–878, 2014.
[21] Zhu, X., Yang, L.T., Chen, H., Wang, J., Yin, S. and Liu, X., 2014, "Real-time tasks oriented energy-aware scheduling in virtualized clouds", IEEE Transactions on Cloud Computing, vol.2, no.2, pp.168-180.
[22] Sun Y., C. E. Koksal, and N. B. Shroff, "On delay-optimal scheduling in queueing systems with replications," CoRR, vol. abs/1603.07322, 2016.
[23] Tchernykh, A., Schwiegelsohn, U., Alexandrov, V. and Talbi, E.G., 2015, "Towards understanding uncertainty in cloud computing resource provisioning", Procedia Computer Science, vol.51, pp.1772-1781.