A Behavioral Study of Task Scheduling Algorithms in Cloud Computing

Mohammad Riyaz Belgaum[1], Shahrulniza Musa*[2], M. S. Mazliham[3], Muhammad Alam[4]

Malaysian Institute of Information Technology, Universiti Kuala Lumpur, Kuala Lumpur, Malaysia[1, 2, 3, 4]

Institute of Business Management, Karachi, Pakistan[4]

Abstract—**All the services offered by cloud computing are bundled into one service know as IT as a Service (ITaaS). The user's processes are executed using these services. The scheduling techniques used in the cloud computing environment execute the tasks at different datacenters considering the needs of the consumers. As the requirements vary from one to one, and so the priorities also change. The jobs are executed either in a preemptive or non-preemptive way. The tasks in cloud computing also migrate from one datacenter to another considering load balancing. This research mainly focused on the study of how the Round Robin (RR) and Throttled (TR) scheduling techniques function subject to different tasks given for processing. An analysis is carried out to measure the performance based on the metrics like response time and service time at different userbases and data centers. The consumers have the option to select the server broker policy as they are the ultimate users and payers.**

Keywords—*Cloud computing; load balancing; service broker policy; scheduling*

I. INTRODUCTION

Distributed computing innovation has risen as another data innovation framework for the quick creating IT industry. In distributed computing, data is for all time put away in vast scale server farms on the Internet everywhere throughout the world. And is available to the customers, furthermore, including desktops and convenient PCs, sensors, and so forth. With the "cloud" as an allegory for the internet [1], distributed computing guarantees to convey exceedingly adaptable IT-empowered information, programming, and equipment capacities as support of outside customers with the internet.

Furthermore, the profoundly versatile calculation capacity of the cloud server farms can additionally help and quicken most computation intensive administrations and works viable. Distributed computing is imagined as the key innovation to accomplish economies of scale and in the arrangement and operation of IT. Various types of data are stored in the form of text, voice, images, videos; and through the internet, they access from any corner of the world. Moreover, the way how they are stored and are available is not the concern of the user and is all taken care of by the IT administrators through an interface with Cloud Computing. In combination, it can be called as "IT as a Service," or ITaaS [2], packaged to the end clients as a virtual server farm as shown below in Fig. 1. The cloud administrators are responsible for managing the relationship between the client and the service provider based on Service level agreement (SLA) [3]. Based on the various services provided by the service providers, they also monitor

the performance. With Software as a Service (SaaS), they ensure that customer satisfaction is guaranteed. Platform as a Service (PaaS) provides and supports the implementation of processes with provisioning, testing, and deployments. Infrastructure as a Service (IaaS) helps in operational management and control over the resources provided for the service. The collaboration between the PaaS and IaaS helps to reduce the IT capital expenditure and operating expenditure by providing virtual infrastructure, security requirements, and other essential requirements.

As to signs of progress in portable correspondence innovations, it is triggering another flood of the client request for prosperous, versatile administration. Versatile clients dependably expect broadband Internet get to wherever they go, communicate with each other employing informal organizations while moving; besides, they are looking for omnipresent access to an abundance of media-based substance and administrations. Since cell phones are resource limited naturally, it is necessary for the cloud to give computational help to numerous media-rich applications with authentication [4]. The mix of versatile media and distributed computing very emerges various specialized difficulties, and the central pressure between asset hungry interactive media streams and power-constrained cell phones exists. The exertion for giving a general rich-media encounter over any screen is ordinarily ruined by the heterogeneity among consistently developing cell phones, as showed in their unique physical shape factors, middleware stages, and natural capacities. Besides the improvements of creative inescapable portable administrations, e.g., versatile video spilling, rich media spread, observation, gaming, e-social insurance, and so on, can be enormously encouraged by versatile distributed computing stages utilizing rose and rising advances.

Cloud stages are empowering new, elaborate plans of action and organizing more internationally based incorporation arranges in coming years than numerous investigator and admonitory firms anticipated. Joined with cloud services appropriation expanding in the mid-level and little and medium organizations, driving specialist, including Forrester, are changing their gauges upward. The various cloud service models listed in Table I, shows the services they offer along with the type of flexibility with examples.

According to the prediction by IDC, as shown in Fig. 2, the need for public cloud is going to increase each year, and hence, the tasks scheduled at each data center needs to be managed. The functions in the cloud are to process the user's requirements like providing a platform, infrastructure, or

---

*Corresponding Author

software as a service. Also, being simple to use, most of the customers are moving their tasks to the cloud. The payment for the use is made based on the policy pay-as-you-use. So, there is no chance of being charged more for not using the service.
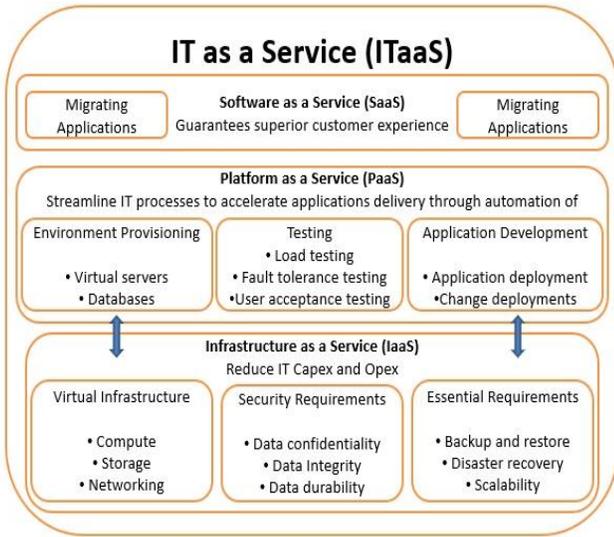

Fig. 1.   IT as a Service (ITaaS).

TABLE. I.   CLOUD SERVICE MODELS, SERVICES, EXAMPLES [5]

| Cloud Service Model	Services	Frequency	Examples
IaaS	Compute, storage, and network service	High	Amazon Elastic Compute Cloud (EC2), Defense Information Systems Agency's DISA milCloud, Google Compute Engine, Microsoft Azure
PaaS	Application program interfaces (API) and services	Medium	Amazon Elastic MapReduce, MathWorks Cloud, Red Hat OpenShift
SaaS	Full-fledged applications	Low	Google Gmail, Microsoft Office 365, Facebook
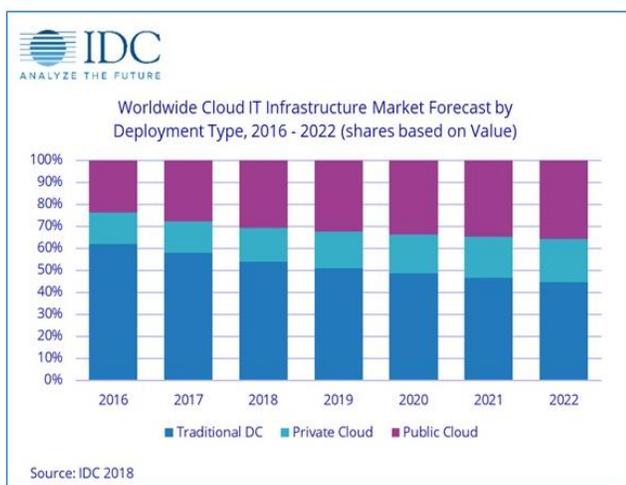
Fig. 2.   Worldwide Cloud IT Infrastructure Market Forecast[6].

The data centers located in remote places are responsible for providing the necessary service. The resources are allotted to the processes at the data center using the policy to complete the tasks efficiently. The performance of different task scheduling algorithms varies based on the policy.

In this study, the researcher plans to study the various task scheduling algorithms in cloud computing and analyze the behavior of those scheduling algorithms subject to different requirements of the users. The mentioned study shows the effects and impact of different scheduling algorithms, which allows the user to decide on choosing a specific scheduling algorithm with a better QoS [7]. This research focuses on the study of the performance of various task scheduling algorithms in the cloud, considering the service broker policies. The metric used to measure the performance of each of the task scheduling algorithm is response time. The following questions are to be addressed to do so.

1) What are the characteristics of different task scheduling techniques?

2) Match the user requirements with the task scheduling technique.

3) Analysis of these scheduling techniques to have better throughput and less waiting time based on the service broker policy chosen.

The researcher concentrates on the standard scheduling algorithms used in cloud computing. However, these algorithms can be customized and modified with the changing requirements. This study will be a considerable contribution in the area of Cloud Computing to select the scheduling policy from the available different scheduling policies while meeting the needs of the users. It shows the researchers how the processing, storage, platforms, software are provided to the user by optimizing the response time and minimizing the waiting time. This study can help other researchers to enhance different scheduling algorithms based on their behavior to improve the QoS. It can also be used by the enterprises to decide on choosing the appropriate service broker policy matching their requirements to utilize the services offered by the cloud service providers.

The following is the conceptual framework for this research. The study of the effective strategy of different scheduling algorithms to process the tasks is studied. There are various ways how the scheduling algorithms behave when they execute at the nearest data centers, and when they migrate to the datacenters without the consent of the user. A study of different scheduling algorithms helps in knowing the behavior of such algorithms in this research. Comparative outputs between Round-Robin and throttled policies subject to separate service broker policy selected. Therefore, it helps to choose for the specific scheduling algorithm taking into consideration the requirements of the users to have a safe and secure transaction.

The organization of the rest of the paper is as follows. Section 2 presents a literature review, with the methodology used to analyze the results provided in Section 3. Section 4 contains discussion on analysis of results, Section 5 contains a conclusion, and finally, Section 6 presents future work directions of the study.

II. LITERATURE REVIEW

This section reviews the works done by different researchers in the area of cloud computing from the perspective of balancing the load. The authors in [8] have conducted a study on the load balancing algorithms. They implemented load balancing in cloud computing using checkpoints. Ranks are calculated by considering the requirements from the user and keeping the objective to maintain QoS so that the customers know on which cloud services can be selected.

The smart devices which are now prevailing much are also used to access the services of the cloud. The method of quickly obtaining cloud computing applications with rapid and fast-moving communication media is known as mobile cloud computing [9]. It examines on reducing the energy consumption by dynamically scheduling the tasks and proposes an algorithm considering the time, voltage, and processor constraints of the cyber-physical system. Further, in [10], the authors have shown how HTML5 is used to implement the applications and services of the cloud efficiently. Still, it shows the gaps between traditional cloud computing and mobile cloud computing.

From the systematic review conducted in [11], it shows very clearly how the resources will be allocated. In the process of resource management, though there are some challenges about resources like allocation, adaption, brokering, discovery, mapping, modeling, provisioning, scheduling, the distribution of resource to a task is critical. The parameters like throughput, time, response time, speed, availability, and so forth, were used to compare different policies. The study also addresses the problems of green computing by minimizing energy consumption.

The authors in [12] have presented a scheduling strategy based on genetic algorithm for task scheduling considering energy requirements as it is given more considerable attention than before. The results showed that it achieved the best solution with least or no migration. While in real cloud computing where there are dynamic changes in the virtual machines and the computing cost increases with the unpredicted load, it is concluded that the cloud data center always has an optimal energy-efficiency ratio and it can be obtained by efficient resource allocation.

In [13], the authors have proposed a load balancing algorithm named Firefly algorithm with neighborhood attraction (NaFA), where the tasks are allocated to such a virtual machine which is richly equipped with the resources and simulates from the social behavior of the fireflies. Just as the brighter one leads the other fireflies; many are attracted in the population. As more tasks are allocated to the same virtual machine, the time complexity is high. The balancing of the load at all the virtual machines is paid less attention.

To allocate the virtual machines online in a distributed cloud environment, the cloud service provider allocates the resources without the knowledge of considering the tasks are joining the pool in the future. The authors in [14] proposed algorithms that serve the functions present on different cloud architectures. While with the new emerging virtualized applications which are geographically distributed the complexity still increases if the data centers are increased.

In [15], the authors focused on minimizing the total weighted job response time. To reduce the job response time, they proposed a model wherein which the jobs generated from the users are deployed to the servers with upload and download delays. They have used OnDisc by setting the weight for each job based on job latency. The results showed that the total response time is reduced when compared with the heuristic algorithms.

The authors in [16] considered the dynamic resizing of virtual machines, as the size of the virtual machines shrink and expand when the resources are added and removed from the pool. This feature of cloud computing affects the performance as the cloud infrastructure functions in prescribed limits because of the scarcity of resource availability. The adverse effects of the tasks which are scheduled at one virtual machine have to be migrated to another as the resource is not available due to its elasticity feature were to be paid much importance.

Cloud computing in coordination with the Internet of Things (IoT) has put forth many challenges to be addressed. In connection with building the smart homes, a framework needed to bind the applications and implementations of such with the gaps to enable such implementations were discussed in [17]. The authors integrated the technologies like IoT and cloud to have an efficient cloud-centric IoT based solution as the information.

An online auction-based mechanism was proposed in [18] to allocate the resources to the users by the cloud service providers. The users intended to utilize the resources like processor, memory, storage which are nothing, but the virtual machines are allocated based on the quoted price by the users. Moreover, the cloud service providers cite their services, which can be provided to the users matching the incentives. This policy is utilized when all the tasks are stable, but for dynamic works, the online auction mechanism fails.

In [19], the authors considered different criteria to allocate the task to a particular datacenter. The resources at the data center are assigned to the tasks to complete the execution. Resources being the costliest components are to be effectively utilized without overloading them with the tasks and without keeping them idle too. The authors used à CloudSim simulator and simulated the results to show that their proposed algorithm performed better over the existing one in terms of throughput. However, they did not pay much attention to other criteria to measure the efficiency of the algorithm.

Migration of the tasks from one virtual machine (VM) to another is a part of balancing the load at the data centers. The authors in [20] introduced collaborative agents to migrate the tasks considering various requirements like hardware diversity, dynamic user requirements, wearable resources, imbalanced load, and energy usage. These agents proved to be efficient in performing the intended tasks while they did not consider a significant constraint of trust.

The authors in [21] considered the bandwidth requirements for task scheduling in cloud computing. They have proposed a decentralized belief propagation-based method where the agents and the tasks continuously change. Also, the authors made a comparison of the proposed plan with two other methods prevailing in task scheduling. The proposed way out ruled the different techniques in terms of shorter problem-

solving time and lesser communication requirements. While the focus was on task allocation, by decomposing the network, the security issues with such were also to be considered.

A balanced scheduler [22] is used to balance the tasks by the cloud service provider and the applications. The authors proposed a Balanced and file Reuse-Replication Scheduling (BaRRS) by using the replication and data reuse techniques where a task is split into subtasks and was run parallel to improve the system utilization but the fact that if one subtask delays then the complete job will also be delayed has been overlooked. However, the results showed that it performs well in optimistic situations.

From all the above review, it is evident that though the priority is to balance the load to attain a better throughput, it lacks an essential point on guiding the customer to choose a policy at the time of signing the service level agreement (SLA).

III. METHODOLOGY

The main aim of this research is to study the behavior of the scheduling algorithms, which can be either preemptive or non-preemptive [23] subject to the user requirements and the geographic location of the data center. The study aims to consider the number of virtual machines accessing with the user requirements like either to select the closest data center or to optimize the response time or to reconfigure dynamically with load balancing.

The study addresses the following questions:

1) What are the various user requirements?
2) What are the various scheduling algorithms available to have better throughput?
3) After knowing the requirements and the scheduling algorithms, which scheduling algorithm has to be chosen to match the user requirements and can a generalized framework be proposed for better performance?

The study mainly uses the following load balancing policies:

- Round Robin (RR)

- Throttled

Factors considered are as follows:

- Response time

- Data Center request service time

The researcher adopted an analytical research methodology in conducting the study. Open source available cloud simulation software called à CloudSim is used to get the results. The various load balancing algorithms have their methods to execute the tasks at different data centers. The results of these algorithms are used to analyze the performance and propose a framework for the consumers to adopt a respective policy if the option of selection is given to them in SLA[24]. Both the algorithms are measured in terms of response time and other metrics too and put forth the opportunity to the consumer to select based on the requirements and the amount they bid.

The response time can be defined as the time from which the request has arrived at the data center to the time at which the request starts processing. Data center request service time is defined as the time from which the request comes at the data center to the time the request completes processing.

The research is organized as:

- Study the various load balancing algorithms.

- Analyze the performance of the algorithms mentioned earlier in terms of metrics

Propose a solution by analyzing all the conditions for the consumers as they go with the policy of pay-as-you-use for the services they are using. It will help them to select an appropriate one considering the complexity of the task to be allocated to the datacenter.

IV. ANALYSIS AND RESULTS

The simulator à CloudSim models and simulates various services offered by cloud computing, and it is an open source tool, which is widely used in academics and research. This tool allows the researchers to simulate the algorithms developed to meet the requirements of the users. From the various offered cloud computing services, Infrastructure as a Service (IaaS), is one of them, where the location of the data centers, servers, and the clients are widely scattered on a broad geographical area but still, there is uninterrupted service. The availability of the resources is a significant concern for any of the cloud service provider to serve the users without deadlock [25]. From the study of different algorithms implemented in à CloudSim, namely Round Robin (RR), Equally spread Execution load (EE), Throttled (TR), it is observed that these are used in two different types of scheduling like preemptive and non-preemptive scheduling. Preemptive scheduling is a type of schedule where the resources are allocated to the task either for a quantum of time or based on priority. While non-preemptive scheduling is adopted in a static environment where the resources of the task are determined initially so that the available resources are equally given to all the tasks based on the size and need.

The design of à CloudSim simulator covers the whole globe and is divided into five regions, with each region establishing the data centers and as many as user bases to be added manually providing an excellent graphical user interface to the users to configure the network.

Each region has its specified boundaries, and, in each area, there can be datacenters and the userbases. The user can configure the simulation, define the internet characteristics, and when everything is fixed, can run the simulation. So first, the data center is set to add many userbases by selecting the service broker policy. There are three types of service broker policies as closest datacenter, optimal response time, and reconfigure dynamically with the load which is to be attached to the application and is deployed at that data center.

In the main configuration after all the entities are set, then such an arrangement can be saved to perform the simulation. Once the datacenters and userbases with the server broker policy are set later, the advanced option to select the load

balancing system to be used in the data center is to be decided. By using the advanced option, the user tags the load balancing policy. If the tasks are to be executed in a preemptive environment, then RR and TR are to be used. However, the task to be implemented in a non-preemptive climate uses EE policy. Once the network is configured with the datacenters and userbases by selecting the policy to use as the closest data center for the service policy, Round Robin for the load balancing and then the simulator is run to get the results. The obtained results are summarized as follows.

The data centers are located in different regions like Region-0, Region-2, Region-5, while the userbases are found in regions 1, 3, 4, respectively. As the closest data center service policy is selected, it is observed that the data center (DC3) have no user bases allocated, and it is idle. So, the problem with such a policy is that there is no efficient use of resources and sometimes the nearest datacenters might be overloaded with the tasks to complete thereby increasing the response time and also the delay time. In addition to all these, the concern is to serve all the requests maintaining the QoS and avoiding the deadlock states. Sometimes some tasks might also migrate from one data center to another when such is overloaded. So, a list of all such tasks is to be considered for scheduling either before the resources are allocated or at the time of execution too.

Fig. 3 and Fig. 4 are the graphs derived by plotting the response times of different user bases using the RR policy and Throttled policy for three different types of service broker policies namely Closest datacenter, Optimal response time and reconfigure dynamically with loading respectively. The results show that, the response time for the userbase3 is very low as the load is considered dynamically, and the tasks are executed. As both UB2 and UB3 are assigned to DC2, it may result in high response time for UB3 so, according to re-configure dynamically policy, during runtime, the tasks at UB3 get shifted to DC3 and the response time is very less when compared with others.

Fig. 5 and Fig. 6 are the graphs derived by plotting the service times of different data centers using the RR policy and Throttled policy for three different types of service broker policies namely Closest datacenter, Optimal response time and reconfigure dynamically with loading respectively. The results show that the service time at the DC3 is more when compared with all the other data centers. It is because the load is considered here, and the network is reconfigured accordingly.


Response time using RR Scheduling

	CDC	OR	RD
UB1	201.35	199.8	201.72
UB2	299.39	299.99	300.28
UB3	301.08	300.67	53.14
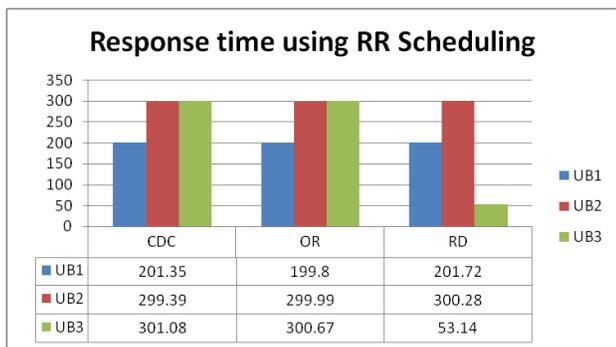
Fig. 3.   Response Times of different user bases using RR Policy.

Response time using Throttled Scheduling

	CDC	OR	RD
UB1	201.27	199.77	201.59
UB2	299.52	300.13	300.2
UB3	301.08	300.89	301.73
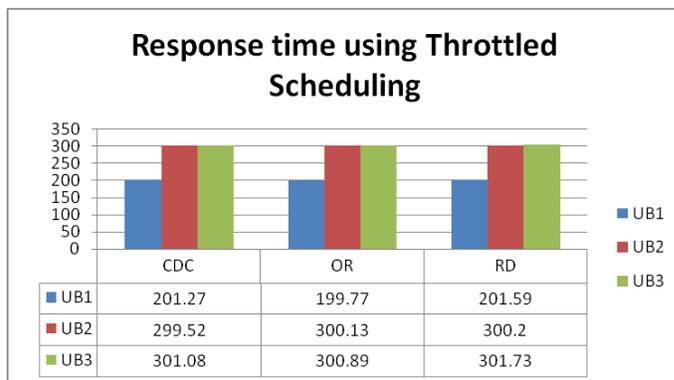
Fig. 4.   Response Times of different user bases using Throttled Policy.

Data Center Request Service time using RR Scheduling

	CDC	OR	RD
DC1	0.1	0.1	0.5
DC2	0.37	0.37	1.19
DC3	0	0	3.35
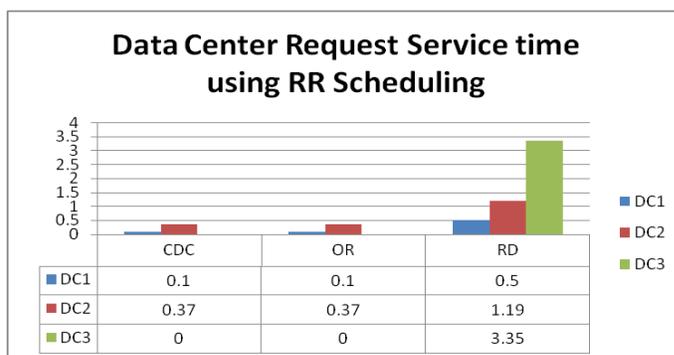
Fig. 5.   Data Center Service Times using RR Policy.

Data Center Request Service time using Throttled Scheduling

|	CDC	OR	RD
DC1	0.1	0.1	0.4
DC2	0.37	0.37	1.04
DC3	0	0	0
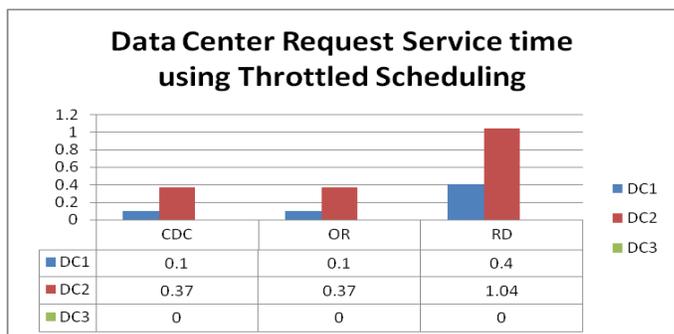
Fig. 6.   Data Center Service Times using Throttled Policy.

Therefore, for the consumers, it must be decided by them on which service they need as their requirements are different. Some consumers might want the response time to be less as they wish their tasks to get processed faster irrespective of the payment charged while there might be other groups of consumers, who are not bothered about the speed of the execution but are concerned much with the amount.

V.   CONCLUSIONS

Task scheduling is done under two strategies one being preemptive and the other being non-preemptive. This study analyzes the characteristics of different scheduling techniques used in both environments. Before the resources are allocated, the availability and the accessibility to those resources are estimated. One of the promising challenges is load balancing in cloud computing. Addressing this challenge will reduce the burden at the data center, and this time can be better utilized in

processing. The behavior of two of the resource scheduling policies, Round Robin and Throttled are compared under different service broker policies. The metrics, like the response time at each user base and processing time at each data center, are used. The simulative results show that the response time for the userbase3 under RR is very low as the load is considered dynamically and the tasks are executed. The data center service times are high in both the scheduling policies using reconfigure dynamically as the service broker policy, while with the other two are the same. The metrics used help us to conclude that if the tasks are to be executed faster and then the load is to be reconfigured and is to be allocated to the free datacenter, wherein it increases a little overhead. This information may guide the consumers to take appropriate decision in signing the SLA.

VI. Future Work

The authors have studied the behavior of two of task scheduling algorithms and compared them under different service broker policies. But still there is a scope to modify these algorithms according to the requirements and analyze its behavior. A new hybrid method can also be proposed to better satisfy the customers and improve the performance metrics. The overhead incurred while performing load balancing can be reduced further by using either traditional techniques or by using machine learning techniques, which is carried as future work.

References

[1] M. Shah and A. S. Shah, "Appraisal of the most prominent attacks due to vulnerabilities in cloud computing," Int. J. Grid Distrib. Comput., vol. 9, no. 7, 2016.

[2] S. Sarkar and S. Venkateswaran, "Best-Fit Containerization as a Brokered Service," in 2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing and Communications, Big Data and Cloud Computing, Social Computing and Networking, Sustainable Computing and Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom), pp. 940–947, 2018.

[3] K. Akiyama, K. Iwano, A. Ohkado, and T. Tsumura, "Dynamic service level agreement for cloud computing services." Google Patents, 15-Jul-2014.

[4] M. Shah, A. S. Shah, and I. Ijaz, "Implementation of User Authentication as a Service for Cloud Network," Int. J. Grid Distrib. Comput., vol. 9, no. 10, pp. 197–210, 2016.

[5] N. A. Schear, A. B. Yerukhimovich, V. N. Gadepally, T. M. Moyer, P. I. I. T. Cable, and R. K. Cunningham, "Secure and resilient cloud computing for the department of defense," MIT Lincoln Laboratory Lexington United States, 2015.

[6] IDC, "Worldwide Cloud IT Infrastructure Revenues Continue to Grow by Double Digits in the First Quarter of 2018 as Public Cloud Expands, According to IDC," 2018.

[7] M. H. Ghahramani, M. Zhou, and C. T. Hon, "Toward cloud computing QoS architecture: Analysis of cloud systems and cloud services," IEEE/CAA J. Autom. Sin., vol. 4, no. 1, pp. 6–18, 2017.

[8] M. R. Belgaum, S. Soomro, Z. Alansari, and M. Alam, "Cloud service ranking using checkpoint-based load balancing in real-time scheduling of cloud computing," in Progress in Advanced Computing and Intelligent Engineering, Springer, pp. 667–676, 2018.

[9] Y. Li, M. Chen, W. Dai, and M. Qiu, "Energy optimization with dynamic task scheduling mobile cloud computing," IEEE Syst. J. vol. 11, no. 1, pp. 96–105, 2017.

[10] M. Ali and M. H. Miraz, "Recent Advances in Cloud Computing Applications and Services," Recent Adv. Cloud Comput. Appl. Serv. Int. J. Cloud, Comput. (IJCC), ISSN, vol. 9445, pp. 1–12, 2014.

[11] S. H. H. Madni, M. S. A. Latiff, and Y. Coulibaly, "Recent advancements in resource allocation techniques for cloud computing environment: a systematic review," Cluster Comput., vol. 20, no. 3, pp. 2489–2533, 2017.

[12] Y. Shen, Z. Bao, X. Qin, and J. Shen, "Adaptive task scheduling strategy in cloud: when energy consumption meets performance guarantee," World Wide Web, vol. 20, no. 2, pp. 155–173, 2017.

[13] H. Wang et al., "Firefly algorithm with neighborhood attraction," Inf. Sci. (NY)., vol. 382, pp. 374–387, 2017.

[14] F. Hao, M. Kodialam, T. V Lakshman, and S. Mukherjee, "Online allocation of virtual machines in a distributed cloud," IEEE/ACM Trans. Netw., vol. 25, no. 1, pp. 238–249, 2017.

[15] H. Tan, Z. Han, X.-Y. Li, and F. C. M. Lau, "Online job dispatching and scheduling in edge-clouds," in IEEE INFOCOM 2017-IEEE Conference on Computer Communications, pp. 1–9, 2017.

[16] D. Breitgand et al., "Dynamic virtual machine resizing in a cloud computing infrastructure." Google Patents, 2018.

[17] B. L. R. Stojkoska and K. V Trivodaliev, "A review of Internet of Things for smart home: Challenges and solutions," J. Clean. Prod., vol. 140, pp. 1454–1464, 2017.

[18] L. Mashayekhy, M. M. Nejad, D. Grosu, and A. V Vasilakos, "An online mechanism for resource allocation and pricing in clouds," IEEE Trans. Comput., vol. 65, no. 4, pp. 1172–1184, 2016.

[19] A. V. Lakra and D. K. Yadav, "Multi-objective tasks scheduling algorithm for cloud computing throughput optimization," Procedia Comput. Sci., vol. 48, pp. 107–113, 2015.

[20] J. O. Gutierrez-Garcia and A. Ramirez-Nafarrate, "Collaborative agents for distributed load management in cloud data centers using live migration of virtual machines," IEEE Trans. Serv. Comput., vol. 8, no. 6, pp. 916–929, 2015.

[21] Y. Kong, M. Zhang, and D. Ye, "A belief propagation-based method for task allocation in open and dynamic cloud environments," Knowledge-Based Syst., vol. 115, pp. 123–132, 2017.

[22] I. Casas, J. Taheri, R. Ranjan, L. Wang, and A. Y. Zomaya, "A balanced scheduler with data reuse and replication for scientific workflows in cloud computing systems," Futur. Gener. Comput. Syst., vol. 74, pp. 168–178, 2017.

[23] M. Masdari, S. ValiKardan, Z. Shahi, and S. I. Azar, "Towards workflow scheduling in cloud computing: a comprehensive analysis," J. Netw. Comput. Appl., vol. 66, pp. 64–82, 2016.

[24] Z. Zhou et al., "Minimizing SLA violation and power consumption in Cloud data centers using adaptive energy-aware algorithms," Futur. Gener. Comput. Syst., vol. 86, pp. 836–850, 2018.

[25] F. Tian, R. Zhang, J. Lewandowski, K.-M. Chao, L. Li, and B. Dong, "Deadlock-free migration for virtual machine consolidation using Chicken Swarm Optimization algorithm," J. Intell. Fuzzy Syst., vol. 32, no. 2, pp. 1389–1400, 2017.