SLA BASED RESCHEDULING OF TASK FOR OPTIMUM ALLOCATION OF RESOURCES IN CLOUD COMPUTING ENVIRONMENT

Ms. Nikky Ahuja, Dr. Priyesh Kanungo, Dr. Sumant Katiyal

Abstract- The dynamic and highly-competitive business environment has made it mandatory for the resource providers to fulfill the business resource demand within stipulated time and cost. Because non-meeting of deadline may not only result in the loss of money but can also hamper the goodwill of the company. But, meeting these performance requirements within specified budget and deadline is a challenge for the cloud providers. This calls for an urgent need for reducing the number of request rejections due to missing deadlines and SLA violation. Present paper is an effort in the direction of reducing the number of request rejection due to missing deadlines, and rescheduling of tasks to ensure maximum task efficiency and optimum utilization of allocated resources. In this paper, we have presented algorithm for reducing penalty cost incurred by delay in task accomplishment, rescheduling of task and strategies focusing on optimum allocation of cloud resources to guarantee the quality constraints defined in SLA.

Keywords- Cloud computing, Service Level Agreement (SLA), Quality of Service (QoS), Admission control, Strategies, Penalty Delay and Rescheduling.

I. INTRODUCTION

The way in which the applications are developed in a commercial and scientific environment has witnessed a remarkable change because of the growth and development of networking technology and infrastructure. Cloud computing technique is a revolutionary change in the direction of providing infrastructural facilities to consumer for transfer of information over the internet with a very little or no effect on the system's performance. Cloud offers on demand resources and storage capacity make cloud an attractive computing platform [1, 2]. Its dynamic behavior enable customer to change their business needs as per their requirement at anytime. Virtual machines, bandwidth of the network, storage space or memory available, processors' speed etc. acts as the resources in cloud. In the initial stages, users were supposed to pay rent for the entire time slot hired for the task, whether they have used the resources or not. But, present day technology is focusing on the pay as much as resources used strategy for solving the problem of high cost of resource renting. This flexibility is created by the renting of virtual machines arranging request in parallel, resulting in guaranteed budget and deadlines. The huge number of virtual machines is made available to

serve consumer request, this helps in providing choices to consumers. At any point of time a consumer can select the number and the type f VMs meeting her demand. Also, if a rented VM is idle for a certain period of time it can be used by any other task or request.

II. LITERATURE SURVEY

To improve the performance of Cloud Computing, useful research has already been carried out as described in following paragraphs.

Various strategies for allocation of resources are required for gaining customer satisfaction and profit maximization. Study conducted by V.Vinothina *et al.* [7] has described classification of resource allocation strategies and its impact on cloud. Linlin Wu *et al.* [8, 9] have dealt with the challenge of efficient and feasible allocation of resources as per SLA norms. They have formulated four strategies and algorithms for satisfying the resource demand and meeting the expectations of the user. Goudarzi and Pedram [3] have considered the problem of SLA based resource allocation system and have presented a distributed solution for it.

According to Noha Hamdy *et al.* Cloud computing is an attractive processing technology that allows efficient computing for optimum management of cloud system resources [4]. In their work, they have defined different resource allocation strategies and those strategies affects various parameter of system like throughput, utilization, latency and response time in a cloud. D. Mahendranv *et al.*, in their research have presented techniques for effective management of resources and also define metrics associated with the resource management techniques [6]. As per Swarupa and Shahu [5] the main aim of service provider is to satisfy customer without violation the quality constraint. An extensive evaluation study was conducted by authors to find best solution to increase profit by minimizing cost and improving customer satisfaction level. Paper presented various scheduling strategies and algorithm to provide efficiently resource allocation and increase profit for SaaS provider.

III. ADMISSION CONTROL POLICY

Cloud service providers are supposed to have clear admission control policies with a vigorous mechanism for task admission to create a balance between client requirements and the resources in cloud's system. Policies of admission control are defined as per the expectations of the client and the resource requirement in the cloud's system. Present research works on the assumption that at its starting phase

each incoming request from client have accessibility to server pool and it will be queued in the resource allocation process. By using the most suited mechanism, the VMs will be assigned to these requested tasks with continuous monitoring of system's performance. Although, some of the requests not meeting SLA constraints or the systems' resource requirement may gets rejected by the control mechanism. In the current research we have tried to frame some strategies for reducing the number of request rejections and maximize profit.

IV. ALGORITHM FOR PENALTY DELAY AND TASK RESCHEDULLING

Present research focuses on the maximization of profit by best possible allocation of clouds' resources within SLA terms and conditions resulting in no or vey less penalty cost. Resource allocation is done in such a way so as to reduce penalty cost by effective utilization of derived strategies for controlling cost. Following algorithm fulfills our objective of reduction penalty by meeting deadlines:

Penalty delay ()

Input: *Multiple Task, Meta information of task –mit, vm matching=0, request (r) with QoS parameter;*

Output: *whole No of vm match task performs, result;*

1. Start
2. If (already a user)
3. login();
4. else (create an account)
5. then login();
6. loading SLA_packageinfo( );
7. put(meta info of task)
8. get request on the basis of package selected
9. For each package selected (check for the availability of server ready to process the request r)
10. Check for VM match as requested by customer c
11. If (VM found) then
12. {
13. For each initiated VM matching customer request r
14. Assign VMr to the request r
15. Check for enough space then
16. Check request for SLA terms

17. If VMr has enough space and SLA doesn't gets violated then
18. Put VMr into VM_list
19. Sort VM on the basis of space availability
20. Perform the scheduling on the basis of maximum space availability and minimum or no penalty cost
21. Else if(c is ready to wait)
22. {
23. Put(r into wl)
24. Queue(r)
25. Else
26. Schedule request r on the basis of priority and minimum penalty cost
27. }
28. }
29. Else
30. {
31. Search for new VM
32. If (VM found)
33. Initiate new VM with type S and assign the product type as requested by customer c
34. }
35. Else while(s+j<=S)
36. {
37. If(there is next initiated VM of type s+j, where type s+j matches the VM requirement of customer c)
38. Repeat step 14 to 26
39. J++
40. }
41. Else
42. {
43. Check for the availability of new server ready to process the request r
44. Repeat step 10 to 39
45. }
46. If(server found available)
47. Repeat 10 to 39
48. Else
49. {
50. Reject the request
51. }

Stated below is an algorithm for rescheduling of clients request to fulfill client resource demand for reducing the number of request rejection and earning maximum profit with no penalty delay. Requests with high penalty costs are rejected here.

Rescheduling ()

1. {

2. For each initiated VM matching customer request r
3. Check for available space
4. If enough space is available then
5. Perform scheduling on the basis of maximum profit and minimum penalty cost
6. Else
7. Initiate new VM
8. If cannot initiate new VM then
9. reject request
10. }

V. ADMISSION CONTOL AND RESOURCE ALLOCATION STRATEGIES

The strategies for improving cloud system's performance as per the quality (QoS) demands of customer and to maximize provider's profit are:

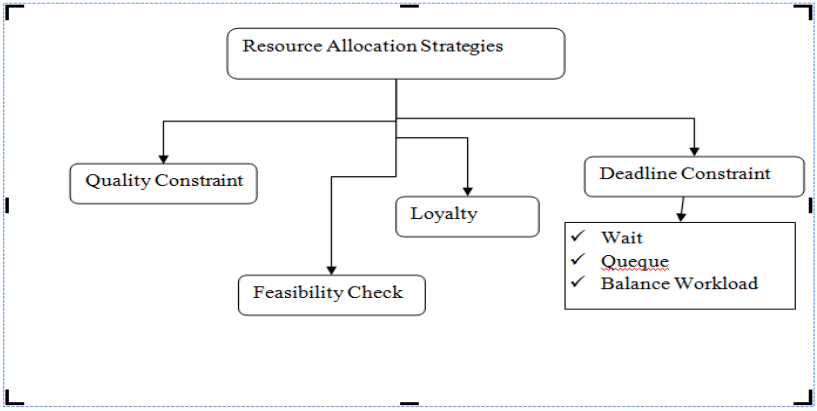


Figure 1.Admission Control Resource Allocation Strategies

1. Focusing on fulfilling the quality requirements of clients as defined in SLA will result in no or less SLA violation. This will reduce the number of errors and hence, the penalty delay cost incurred due to the violation.

2. Before accepting any contract if a check for contract's feasibility in terms of SLA is made then, this will guarantee reduction in SLA violation and help in controlling cost of resources.

3. Scheduling and rescheduling of requests must be performed to get the task finished on or before deadline. The tasks must be queued as per priority and the one with lesser priority should be made to wait. Also, prime focus must be on proper balancing of workload to get the task finished on time.

4. Following above strategies with surprise element of loyalty point scheme for customers entering into multiple contracts with provider or placing consecutive requests will help provider earn customer loyalty.

SLA is vital element for the success of above stated strategies that will help in reducing penalty cost for any work delayed due to violation of SLA terms.

VI. EXPERIMENTATION IMPLEMENTATION

In this work, CloudSim version 3.0.3 has been used as a simulator for cloud computing environment. Algorithm developed was tested in this simulated environment from both users' and providers' perspective. Windows provide the platform and other software used are:

- Net Beans 8.0.2
- JDK 1.8.0_40
- MySQL Server 5.1
- SQLyog Server ultimate v9.0.2.0
- JRE 1.8.0_40

Resources used for experimental setup:

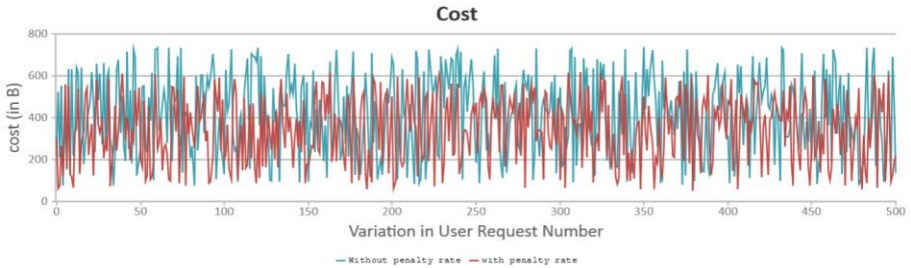
Our system offers four types of packages to the customers: basic, silver, gold and diamond. From these four packages three are constant and only one package titled diamond can be customized. The configuration of the resources is as: basic pack: - 2 GB RAM, CPU core processor and Windows as operating system, silver pack: - 4

GB RAM, CPU core processor and Windows as OS, gold: - 8 GB RAM, CPU intel core processor and windows as OS and for diamond configuration is tailored on the basis of request of the user (in diamond pack local storage space is variable but processor and RAM is constant).

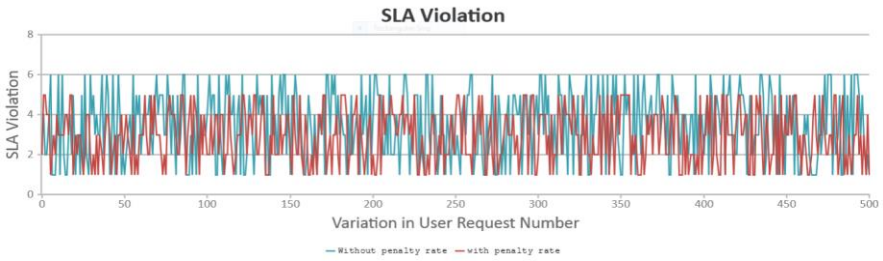
VII. ANALYSIS OF RESULT

For optimum scheduling of resources above algorithms and strategies are designed and they help in finding a best suited server meeting customer request.

Testing of the proposed algorithm and strategies was done with the help of data, which was synthesized randomly with the help of a statistical distribution method. Results are shown using following graphs:



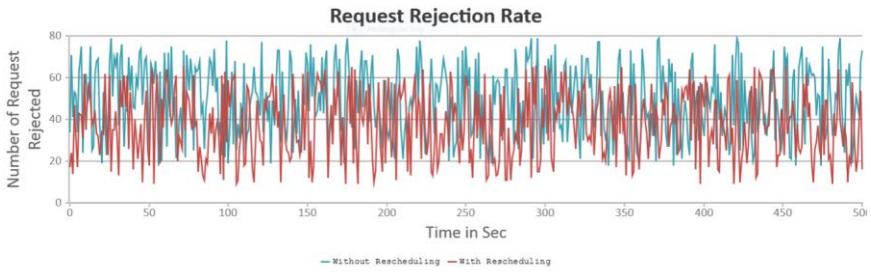
Graph 1 Variation in Cost



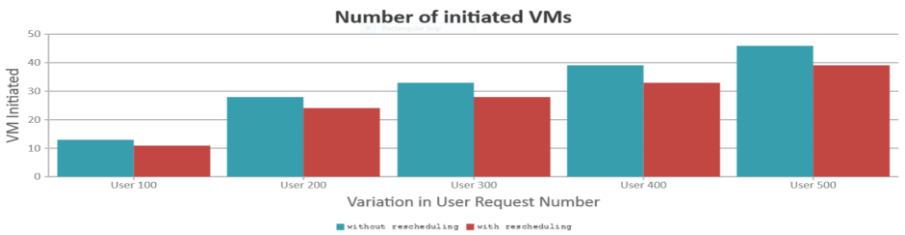
Graph 2 Variation in SLA violation

Graph 3 Variation in profit

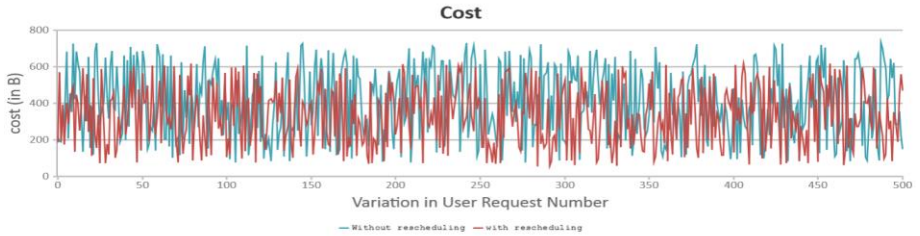
Graph 1, 2 and 3 shows the impact of penalty delay on cloud system. If the task is finished within stipulated deadline then SLA violations gets reduced and hence, the cost of infrastructural resources and penalty costs gets decreased resulting in increased profit.



Graph 4 Request Rejection Rate



Graph 5 Number of VM initiated



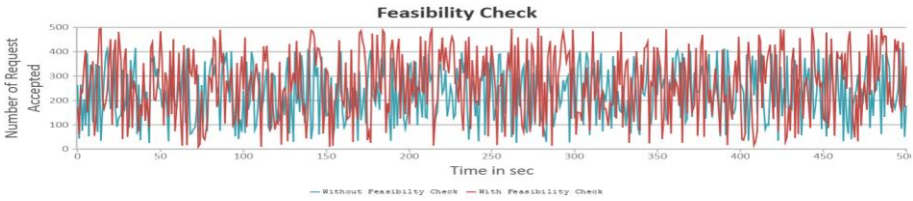
Graph 6 Variation in cost

Graph 7 Variation in profit

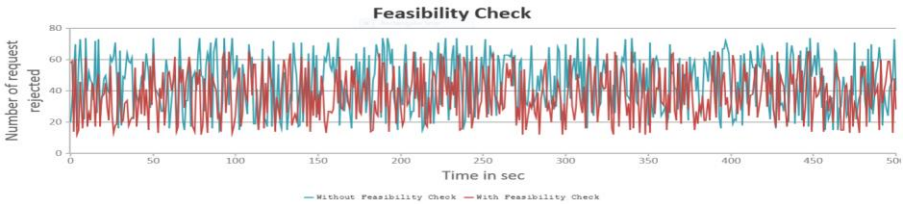
Graph 4, 5, 6 and 7 shows the impact of rescheduling of tasks on cloud performance. Rescheduling results in lesser number of request rejections and less number of VMs initiated. Thus, the cost of resources (i.e. VMs initiated) gets reduced resulting in increased profit.

Graph 8 Average Profit for meeting Quality Constraint as strategy

Graph 8 show an increase in the net profit of the service provider per second, which is achieved by reducing the number of SLA violation. The lesser the number of SLA violation the lesser will be the penalty charged and hence, helps in improving the overall profit of the system.



Graph 9 Request Acceptance rate

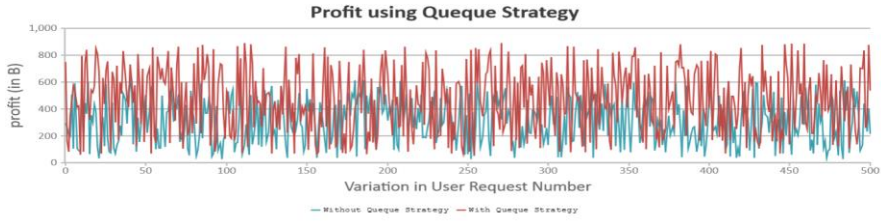


Graph 10 Request Rejection rate

Graph 9 and Graph 10 shows the average number of requests accepted and the average number of requests rejected after doing the feasibility check. It is obvious that with the proposed algorithm the average number of request acceptance gets increased and that of rejections gets decreased.

Graph 11 Average Profit using wait Strategy

It is evident from graph 11, that the wait strategy helps in enhancing the profit of the service provider. Here, the request with no or minimum penalty is accepted and rest rejected.



Graph 12 Average Profit using Queque Strategy

Graphs 12 shows that queuing strategy helps in enhancing the profit of the service provider by optimum utilization of available resources.

Graph 13 Average Profit using balance workload Strategy

It is evident from graph 13 that balanced workload strategy helps in enhancing the profit of the service provider by efficient distribution of workload.

Graph 14 Variation in Customer request

Loyalty strategy helps in retaining of customer and hence aids in profit of the provider. Graph 14 shows the increase in the providers profit due to increased customer request per time interval by using loyalty strategy.

Graph 15 Comparative Average profit earned using strategy

Graph 16 Comparative Average cost minimization using strategy

Graph 15 and Graph 16 shows the impact of resource allocation strategy on the cloud system. It clearly shows that the overall profit of the system increases and cost decreases after the scheduling of resources on the basis of strategies devised.

It could be inferred from above results that there is a positive impact of resource allocation strategies on the cloud system as the profit of the system increases and cost gets reduced along with the reduction in the number of request rejections rate.

VIII. CONCLUSION AND FUTURE WORK

Because of the convenience and scalability of system cloud computing technology is being used increasingly by business enterprises. For achieving the objectives of cloud an efficient and effective resource allocation strategy is required. It plays a pivotal role in the cloud environment. This helps in earning client satisfaction and maximizing the system's profit. Present research work is an attempt to present some resource allocation strategies and algorithms for optimum utilization of cloud systems' resources. The focus is to reduce request rejection rate, minimizing request response time, scalability and flexibility of the system and optimizing cost and profit. In the present work, the strategy for feasibility check before the

acceptance of request and the strategy of request/ task rescheduling will help in the optimum allocation of available resources with the reduction in the number of SLA violations. This will help in reducing penalty cost and timely completion of task allocated. Next the strategy for giving loyalty points will help in gaining customer satisfaction and retention. In this way, present study hopes to contribute towards providing a structure for strengthening the cloud computing model.

IX. REFERENCES

[1] Ahuja, N., Kanungo, P. and Katiyal, S.(2018), "SLA in Cloud Computing Environment: A New Breakthrough" 5th International Conference on "Computing for Sustainable Global Development" INDIACom-2018 organised by BVICAM, New Delhi, India, pp. 622-625.

[2] Ahuja, N., Kanungo, P., Katiyal, S. and Parandkar, P.(2015), "Optimum Management of Resources in Cloud Computing Environment: A Challenge" International Conference on "Computing for Sustainable Global Development" INDIACom-2015organised by BVICAM, New Delhi, pp.4.30-4.33.

[3] Goudarzi, H. and Pedram, M.(2011), "Maximizing Profit in Cloud Computing System via Resource Allocation," in ICDCSW' 11: 31st International Conference on Distributed Computing Systems Workshops, Minneapolis: IEEE, pp. 1 – 6.

[4] Hamdy, N., Aboutabl, A. E., ElHaggar, N. and Mostafa, M. (2017), "Resource Allocation Strategies in Cloud Computing: Overview", International Journal of Computer Applications (0975 – 8887) Volume 177 – No.4, pp. 18-22.

[5] Irugurala, S. and Chatrapati, K. S. (2013), "Various Scheduling Algorithms for Resource Allocation In Cloud Computing", The International Journal Of Engineering And Science (IJES), Vol. 2, Issue 5, pp. 16-24.

[6] Mahendran, D., Gopi, M., Priyadharshini, S. and Karthick, R.(2013), "A Study on Cloud Resource Management Techniques", International Journal of Innovative Research in Computer and Communication Engineering Vol. 1, Issue 1, pp.58-61.

[7] Vinothina, V., Sridaran, R. and Ganapathi, P.(2012), " A Survey on Resource Allocation Strategies in Cloud Computing", International Journal of Advanced Computer Science and Applications, Vol. 3, No.6, pp. 97-104.

[8] Wu, L., Garg, S. K. and Buyya, R. (2011), "SLA-based Resource Allocation for Software as a Service Provider (SaaS) in Cloud Computing Environments," 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), Newport Beach, CA, ISBN: 978-1-4577-0129-0, pp. 195 – 204.

[9] Wu, L., Garg, S. K. and Buyya, R.(2012), "SLA-based Admission Control for a Software-as-a-Service provider in Cloud Computing Environments," Journal of Computer and System Sciences, Elsevier, pp. 1280–1299.

AUTHOR'S PROFILE

Nikky Ahuja is a Research Scholar in School of Computer Science and IT, Devi Ahilya University, India. Her area of Research is Cloud Computing. She is having four years of teaching experience. She has presented and published more than 10 research papers.

Priyesh Kanungo is working as a professor of Computer Science in School of Computer Science and IT, Devi Ahilya University, India. His area of Research is Distributed and Cloud Computing, Operating systems, Artificial Intelligence etc. He is having 27+ years of teaching experience and has published more than 50 research papers.

Sumant Katiyal is working as a professor of Electronics in School of Electronics, Devi Ahilya University, India. His area of expertise is VLSI, Embedded System, IT and Electronics etc. He is having 28+ years of teaching Experience and has published more than 100 research papers.