Efficient Resource Provisioning in Cloud Environment in Terms of Profit using Hybrid Load Balancing Algorithm

P.Karthika, K.C.Palanisamy

M.E Computer Science and Engineering ,Tamilnadu College of Engineering,Coimbatore,Tamilnadu,India
Assistant Professor Computer Science and Engineering ,Tamilnadu College of Engineering,Coimbatore,Tamilnadu,India

Abstract - *Resource provisioning plays a major role in the cloud computing environment due to increased number of cloud users. The user satisfaction level can be improves in the considerable manner by provisioning the resources to the users which can process them with more profit. In this research work, resource provisioning is done with the consideration of the objective called profit. The profit of cloud service providers and as well as users are improved in the considerable manner through the method named as double renting scheme. This double renting method collects the amount for dealing the user request in several way depends on type resource access. The public cloud resources charges for specific amount as like secret cloud resources charged for particular amount. The proposed hybrid load balancing method is used to increase the maximal profit values more significantly rather than the existing system. In this scenario, it shows that the hybrid algorithm insert an important improvements on average response time and profit values. This research provides QoS assured resource selection for the user submitted task in the considerable manner. The experimental tests conducted was proves that the proposed research scenario provides better result than the existing research scenario in terms of increased profit.*

Keywords - *SLA parameter, Profit, Resource provisioning, dynamic pricing, Hybrid load balancing.*

I.INTRODUCTION

Cloud computing is described as a model for allowing everywhere, convenient and ongoing system access to a distributed database of configurable computing resources which is speedily provisioned along with small administration effort from the user as well as service provider communication. It is used to release the computing services over the internet. Cloud services permit individuals and productions to employ software and hardware which are organized by services providers in isolated places. The fulfilment level of cloud users are augmented through providing the resources to the cloud users along with fulfilment of their conditions.

Cloud computing process enhances association, liveliness, scaling and availability produces the potential for expenditure reduction by optimized as well as efficient computing.Cloud service providers are focused to satisfy the number of user requests more effectively. The major objective of service providers is to amplify the profit obtained through provisioning the efficient resources. Besides, when expanding the profit, they should concern cloud users profit also, so that the reputation level of the cloud service providers can be increased in the considerable manner.

B. Rochwerger et al [1] proposed extensible cloud architecture that provides intrinsic support for the business service management and the federation of clouds. The proposed work provided an open, service-based online economy with high-quality service. The proposed architecture implemented enables the providers of cloud infrastructure that partners with each other for the creation of pool of IT resources and also preserving their individual autonomy. This system leverages and also improved virtualization where the autonomous management was embedded in the infrastructure. This proposed approach also achieved the creation of foundation for next-generation enterprise-grade cloud computing. But the proposed system does not overcome the interoperability challenges.

Saurabh Kumar Garg et al [2] proposed three new techniques for scheduling parallel applications on the Utility Grids for the management and optimization of trade-off between time and cost constraints in the system. The proposed method minimized and managed the execution cost and makes span of user applications. It presented a cost metric to manage the trade-off between execution cost and time. The results were evaluated through extensive simulations of the proposed scheduling algorithms against existing ones. Results proved that the presented algorithm outperformed by reducing the time and cost. The proposed system does not provide the needs such as memory and network bandwidth.

Mahdi Ghamkhari et al [3] focused on the reduction of the data center's energy expenses. The practical service-level agreements(SLAs) were taken into account that was between data centers and their customers. The proposed scheme centered for two different cases, with and without behind-the-meter renewable generators and also the formulated optimization problems are convex programs in both the cases, hencethey are simple and suitable for practical implementation. Various experimental data and simulations were done to verify the performance of the proposed optimization-based profit maximization strategy. It showed that it significantly outstood two comparable energy and performance management algorithms in the literature. The main drawback in the paper is that cost elements are not included as the cost models are not generalized.

HamzehKhazaei et al [4] presented a new approximate analytical model for the purpose of performance evaluation of cloud server farms. And also solves it to get an accurate estimation of the complete probability distribution of the request response time and also the other important performance indicators of the farms. This proposed model allowed cloud operators for the determination of the relationship between the number of servers and the input buffer size. The performance indicators namely the mean number of tasks in the system, probability of a task's immediate servica and blocking probability were also allowed. The system did not support for burst arrival of requests.

Michele Mazzucco et al [5] presented an energy efficient server allocation problem in order to increase the profit of the cloud service providers. Most prominent feature of the cloud service providers was energy consumption that could be increased as there is an increase in the number of task execution. This issue caused the revenue deflation directly due to high power consumption. An energy aware resource allocation was presented here to increase the profit. Resources were selected for allocating the tasks based on the energy consumption by calculating the load level and processing capacity. A better selection of the resources with profit and reduced energy consumption are major advantages of the proposed system. But there is a trade off between frequency of the CPU and the number of running servers.

Junwei Cao et al [6] presented an optimal multi server configuration methodology in order to increase of the profit of cloud service providers where the configuration parameter values, energy consumption and the load level of the user were taken into account. Thus the proposed approach ensures efficient allocation of the task to the resources with reduced profit. The methodology called the M-M-1 queuing model was introduced to solve the multi server system problem. Two server speed values and the power consumption values are the input parameter values to the M-M-1 queuing model. The main limitation of this work was that the energy parameter value was not considered for process.

Jing Mei et al [7] proposed an Energy aware scheduling to schedule the tasks to the resources with less energy consumption. Server's energy level is measured by methodology called the dynamic voltage scaling approach. The approach reduced the energy consumption level in terms of handling of sporadic tasks in the time scalable manner and scaling the voltage of the processors in run time. Efficient way of handling the different configuration parameter values of resources were ensured by the proposed method. The results on this work provide the better result, but the proposed system is of greater time and energy consumption.

Hong Xu et al [8] proposed a dynamic pricing in order to schedule the tasks with the consideration of the increased revenue of the cloud service providers. This process is conducted in the Amazon cloud environment to evaluate the performance improvement of the resources in terms of increased revenue factor. This methodology adapts to the spot price fixing technique in terms of improved performance measure values. When the resources are executing the tasks present in the environment, based on the reputation level of cloud service providers and the level of the cloud users pricing would be fixed in the spot. This spot price fixing can lead to improved profit values of the cloud service providers.

Mahdi Ghamkhari et al [9] proposed a novel approach which focuses on the factors called profit maximization, Energy consumption reduction and the SLA agreement satisfaction. This methodology attempts to increase the satisfaction level of both cloud service providers and the cloud users in terms of increasing the profit value and assuring the SLA agreement. To do so, trade off between the different functional parameter values are considered in this work, which focus to achieve balanced output between all. Green data center is adapted in this research work for selecting and provisioning the resources which can execute the tasks with less energy consumption value. This approach leads to an optimal pricing strategy which is proved in the simulation environment.

Yi-Ju Chiang et al [10] introduced the finite capacity queuing model, which attempts to resolve the multi objective problem with the consideration of various performance measures. Some of the performance measures that are considered in this work are operational cost, waiting time, loss probability, system capacity control and final arrival rate. These performance measures are calculated and processed by giving them as input to the finite capacity queuing model. The result of this queuing model would lead to an efficient handling of the performance measure values in terms of different functional values. This approach mainly concentrates on the loses which can

be occurred in the environment like blocking loses, user abandoned loses and so on. The experimental tests conducted in this work proves that this approach can be provide better and flexible framework for provisioning the resources to the users in terms of improved profit and result.

The existing system used a profit maximization scheme with multi queue model which can execute the tasks with the satisfaction of SLA agreement, increased profit based on double renting scheme and the decreased power consumption by considering energy value while calculating profit. These mechanisms adapt the working procedure scenario of the M/M/m+D queuing model which can satisfy all the SLA parameters which are required by the users to run their task. This mechanism leads to efficient handling of the user submitted task in terms of constraints based on which efficient result can be retrieved. This approach adapts to a double renting scheme which is based on both static and dynamic pricing. This is implemented based on situation of cloud resources. The static pricing would be fixed for the available resources and the dynamic pricing is fixed for the extra resources. The methodology is proves to be better than all of the proposed mechanism in terms of the improved performance measure values.

II. MATERIALS AND METHODS

A. Cloud system model

In this module, the cloud consists of three parties such as infrastructure providers, service providers and users. In this set up, infrastructure provider is used to produce system facilities. A service provider rents sufficient resources from infrastructure producers and organize quality of services. There are two renting methods such as long term renting and short term renting. Generally, user gives a service request to a service provider and they returns more accurate services on time.

B. A multi server model

In this module, we assume the cloud service as a multi server system along with a service request queue. The cloud produces the sufficient resources for requested job in terms of virtual machine. The user submits the corresponding jobs to the cloud on the queuing system. The tasks are allocated to the various virtual machines through the job scheduler. Workload organization method produces job queuing services, scheduling policy, priority methods, resource checking and organization of resources. The multi server system is used to renting the user's requested services from the infrastructure provider along with service providers.

In this module, the multi server system includes m long term rented identical servers and it is improved through provisionally renting short term servers from the providers. There is a Poisson stream of service requests along with arrival rate R also identically disseminated and exponential random variables along with mean 1/R.A multi server system keeps a queue with unlimited capability. While the arriving service requests cannot be accessed instantly after they appear, they are initially located in the queue until they can be hold through any available server. The first-come-first-served queuing method is taken. The job finishing necessities are self-governing and identically distributed exponential random variables r with mean r'.

C. Revenue modelling

In this module, the pricing scheme is focused to discover the revenue model and server level agreement. The scenario considered usage based pricing scheme because cloud produces services to the users and charges them. The service level agreement is cooperation among service providers and users on the service quality and the cost. Since of the inadequate servers, the service requests which cannot be handled directly after incoming, the system should wait in the queue until any server is accessible. Though, to convince the quality-of-service necessities, the waiting time of every service demand must be restricted within a positive range which is discovered through the server level agreement.

D. Cost modelling

The cost of a service provider consists of two major parts, i.e., the rental cost of physical resources and the utility cost of energy consumption. Many existing research such as only consider the power consumption cost. As a major difference between their models and ours, the resource rental cost is considered in this paper as well, since it is a major part which affects the profit of service providers. The resources can be rented in two ways, long-term renting and short-term renting, and the rental price of long-term renting is much cheaper than that of short-term renting. This is reasonable and common in the real life. The cost of energy consumption is determined by the electricity price and the amount of energy consumption.

E. A quality guaranteed scheme

In this section, the method named as the double-quality guaranteed resource renting scheme which combines long term renting along with short-term renting. The main computing capability is presented through the long-term rented servers due to their low price. The short-term rented servers provide the extra capacity in peak period. The detail of the scheme is shown in Algorithm 1. For every service request incoming the system, it records its waiting time. The requests are allocated and implemented on the long-term rented servers in the order of coming times. Once

the waiting time of a request reaches D, a temporary server is rented from infrastructure providers to process the request.

1: A multi server system along with m servers is running and waiting for the events as follows

2: A queue Q is initialized as empty

3: Event – A service request arrives

4: Search if any server is available

5: if true then

6: Assign the service request to one available server

7: else

8: Put it at the end of queue Q and record its waiting time

9: end if

10: End Event

11: Event—A server becomes idle

12: Search if the queue Q is empty

13: if true then

14: Wait for a new service request

15: else

16: Take the first service request from queue Q and assign it

to the idle server

17: end if

18: End Event

19: Event – The deadline of a request is achieved

20: Rent a temporary server to execute the request and release

the temporary server when the request is completed

21: End Event

F. Hybrid load balancing method

The Hybrid algorithm is a load balancing algorithm applied through the data center to deal out the obtained jobs competently over the virtual machine under a normal work load via discovery the best VM among the group of VMs to allocate the load in mixed cloud computing background. The hybrid algorithm contains both random and greedy methods. The hybrid algorithm considers the current resource information and the CPU capacity factor. The hybrid algorithm will select k nodes (VM) randomly, and choose the current load for each VM selected. Then the hybrid algorithm will choose a VM that have least VM current loads and return the VM ID to Data Center.

Input: new request

1. Parses VM_List() to LoadBalancer:

2. For i← 0 to k //Select VM randomly

3. TempVMid← random(VM_List()).

4. VM_id← TempVMid

5. If vm_id Exist in Cl_Table(VM_id) then

6. currCount← Cl_Table(VM_id)

7. Else

8. currCount← 0

9. VMids() ← (VM_id, currCount).

10. End for

11. TempVMid← -1

12. currCount← 0

13. For i ← 0 to k

14. TempVMid ← i

15. currCount← VMids(TempVMid)

16. If currCount<minCount then

17. minCount= currCount

18. VM_id← TempVMid

19. End if

20. End for

21. Cl_Table(VM_id) ← Cl_Table(VM_id)+ 1

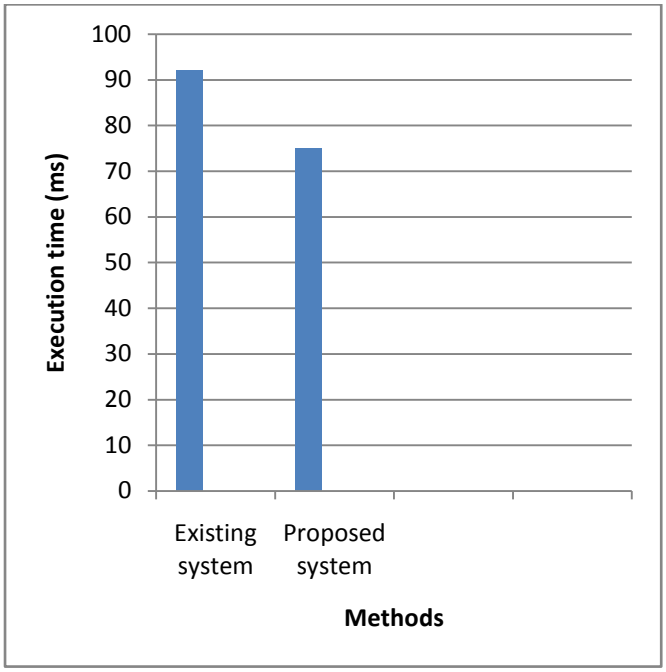
III.RESULT AND DISCUSSION

In this section, the performance metrics are evaluated using existing and proposed methodologies. The performance metrics are such as execution time and maximal profit. The existing double renting method is used to show the maximal profit values in the cloud setting. However the existing system has shown the lower performance in results. The proposed hybrid load balancing method has shown the higher

performance in profit results. The proposed hybrid load balancing algorithm provides optimal results. From the experimental result, we can conclude that the proposed system is better than the existing system in terms of higher performance.

A.Execution time

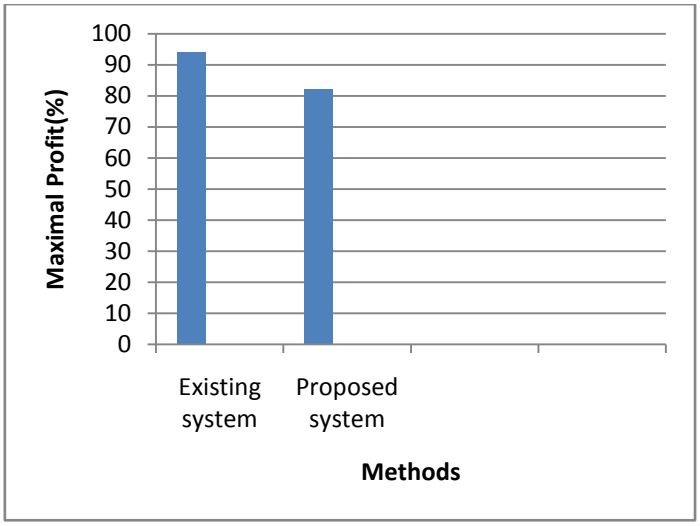
The network is better when the system takes minimum execution time for implementing the process.



From the above figure we can observe that the comparison of existing and proposed system in terms of execution metric. In x axis we plot the methods and in y axis we plot the execution values. In existing scenario, the execution value is higher by using double renting algorithm. The execution value of existing scenario is 92 sec. In proposed system, the execution value is reduced by using the proposed hybrid load balancing algorithm. The execution value of proposed scenario is 75 sec. Thus it shows that effective analysis is performed by using proposed hybrid load balancing algorithm. From the result, we conclude that proposed system is superior in performance.

B.Maximal Profit

The system is superior when the algorithm provides maximum of profit values.



From the above figure we can observe that the comparison of existing and proposed system in terms of maximal profit metric. In x axis we plot the methods and in y axis we plot the maximal profit values. In existing scenario, the maximal profit value is lower by using double renting algorithm. The maximal profit value of existing scenario is 94%. In proposed system, the maximal profit value is increased by using the proposed hybrid load balancing algorithm. The maximal profit of proposed scenario is 82% sec. Thus it shows that effective analysis is performed by using proposed hybrid load balancing algorithm. From the result, we conclude that proposed system is superior in performance.

IV.CONCLUSION

In this section, the conclusion decides that the proposed system is increased the maximal profit performance using hybrid load balancing algorithm. This cloud requests are implemented using various techniques and the methods are focused on the fast execution and maximal profit results. The existing double renting scheme is used to build model for multi server system along with varying system size. The proposed hybrid load balancing algorithm is used to improve the workload of requests, the server-level agreement, the rental cost of servers, and the cost of energy consumption. Thus, the experimental result proved that the proposed system is better than the existing system.

ACKNOWLEDGEMENT

We the authors assure you that, this is our own work and also assure you there is no conflict of interest.

REFERENCES

[1] Rochwerger, Benny, et al. "The reservoir model and architecture foropen federated cloud computing." IBM Journal of Research and Development 53.4 (2009): 4-1.

[2] Garg, Saurabh Kumar, RajkumarBuyya, and Howard Jay Siegel. "Time and cost trade-off management for scheduling parallel applications on utility grids." Future Generation Computer Systems 26.8 (2010): 1344-1355.

[3] Ghamkhari, Mahdi, and HamedMohsenian-Rad. "Energy and performance management of green data centers: A profit maximization approach."Smart Grid, IEEE Transactions on 4.2 (2013): 1017-1025.

[4] Khazaei, Hamzeh, JelenaMišić, and Vojislav B. Mišić "Performance analysis of cloud computing centers using m/g/m/m+ r queuing systems"Parallel and Distributed Systems, IEEE Transactions on 23.5 (2012): 936-943.

[5] Mazzucco, Michele, and DmytroDyachuk. "Optimizing cloud providers revenues via energy efficient server allocation." Sustainable Computing: Informatics and Systems 2.1 (2012): 1- 12.

[6] Cao, Junwei, et al. "Optimal multiserver configuration for profit maximization in cloud computing." Parallel and Distributed Systems, IEEE Transactions on 24.6 (2013): 1087-1096.

[7] Mei, Jing, et al. "Energy-aware preemptive scheduling algorithm for sporadic tasks on DVS platform." Microprocessors and Microsystems37.1 (2013): 99-112.

[8] Xu, Hong, and Baochun Li. "Dynamic cloud pricing for revenue maximization." Cloud Computing, IEEE Transactions on 1.2 (2013): 158-171.

[9] Ghamkhari, Mahdi, and HamedMohsenian-Rad. "Profit maximization and power management of green data centers supporting multiple slas" Computing, Networking and Communications (ICNC), 2013 International Conference on. IEEE, 2013.

[10] Chiang, Yi-Ju, and Yen-ChiehOuyang "Profit optimization in SLA-aware cloud services with a finite capacity queuing model" Mathematical Problems in Engineering 2014 (2014).