Scheduling Virtual Machines in Cloud Computing For Enhancing Income and Resource Utilization

Sanaz Yousefian[1] and Ahmad Habibi Zadnavin[1*]

[1]Department of Computer Engineering, Engineering Faculty, Azad University of Tabriz, Tabriz, Iran

*Corresponding Author: Ahmad Habibi Zadnavin

Received: Nov. 2014 & Published: Jan. 2015

Abstract: With the advent of information technology (IT), doing computational tasks has become an indispensable aspect of modern life. People should be able to perform heavy computing tasks even if they do not possess costly hardware and software; indeed, they can do such tasks through service provides. Cloud computing is regarded as the latest development in IT which can accommodate these needs. The rationales for using cloud computing are maximum efficiency and minimum cost. Among cloud computing challenges, server cost in relation to virtual machine can be considered as a significant issue. In this paper, server cost has been examined as the research problem; more precisely, scheduling virtual machines in cloud computing has been investigated. The advantages of using common scheduling methods are that significant parameters such as energy consumption optimization, migration time minimization, response time reduction, resource utilization enhancement and system efficiency improvement have been addressed and enhanced. These parameters have been extensively studied and optimized in the literature related to cloud computing. However, it should be pointed out that host machine cost in relation to service providers is another important parameter which has been mainly ignored and under-researched. To address this under-researched issue in this paper, the researchers have proposed a novel scheduling algorithm for cloud computing setting which is based on server cost. The merits of scheduling algorithm proposed here are to enhance income and resource utilization. The results of implementation revealed that the proposed scheduling algorithm improved income and resource utilization for 8% and 3 %, respectively, when compared with the basic method. Nevertheless, it should be pointed out that minor reduction of resource utilization in comparison with the rotational shift method is the partial drawback of the proposed method which should be addressed in future studies.

Keywords: cloud, scheduling, virtual machines, cost, income, data center, resource utilization.

1. Introduction

Cloud computing is a model aimed at providing easy access to a set of clouding resources based on user needs though network. The provided clouding services such as networks, servers, saving spaces, application programs and other services are variable and configurable. The access to these services is provided rapidly with the minimum need for resource management (Kloster, 2007). Cloud computing refers to providing hardware and software services in internet work context where users only need to pay for the resources and application programs (Sen Su, 2013). Cloud computing consists of three infrastructure layers, namely Service, Platform, and Software (Jorje Bernal Bernabe, 2012). The advantages and merits of cloud computing are so obvious and extensive that the year 2010 has been labeled as the cloud computing year. The significant reduction in hardware and software costs, efficient utilization of resources, reliability, sharing documents and group cooperation, global access to documents, unlimited saving capacity and ease of use are among the many advantages of cloud computing (Skandariyan, 2012). Nevertheless, it should be argued that the requirement for the permanent connection to high-speed and secure internet is a chief disadvantage of cloud computing.

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-1/MAGNT.41)

Researchers within the domain of cloud computing have tried to address various issues and challenges so as to provide feasible solutions to them. Some of the main issues in cloud computing are security, cost, vulnerability against economic records, access transparency, scheduling, consumption power improvement, load balance, resource utilization enhancement, efficiency improvement, response time reduction, migration time reduction, etc.

The present study has focused on scheduling virtual machines in cloud computing. Indeed, the purpose of this study is to optimize the parameters of cost and resource utilization. The rationales for using cloud computing are to maximize efficiency and minimize cost. In fact, server cost with respect to virtual machines is one of the significant challenges which faces cloud computing researchers. For example, in case there are some low-cost servers and some high-cost servers, the critical decision which machine should make is that which use and function should be allocated to which server. This is an issue which has not been taken into consideration in available algorithms. The rest of the paper is as follows: section two is concerned with the examination and discussion of the available methods on scheduling virtual machines in cloud computing. In section three, the novel method of the present study will be proposed. Section four will report the simulation conducted in the study and the evaluation of the proposed method. Section five will conclude the study and recommend further research and directions for future research.

2. Review of the Related Works

The scheduling and allocation of resources should be considered at different levels for supporting the deployment of programs in cloud. One of the significant issues which has attracted researchers' attention is scheduling in cloud computing infrastructure layer; the purpose of this type of scheduling is the optimal allocation of virtual machines on host servers. For doing this, different parameters should be taken into account. For instance, energy management, load balance, locating the best hosts for placing in virtual machine, reduction of management overloads, etc. have been investigated in cloud computing research. Each of these issues examine one aspect of cloud computing. In this section of the study, some of the latest studies on cloud in this research domain will be overviewed.

Raj and Setia (2012) proposed an efficient algorithm for enhancing the execution speed of the allocated tasks in which an effective communication is maintained between the employer and the virtual machine; hence, time and cost factors are optimized by the algorithm proposed by them. In this algorithm, the virtual machine is selected based on high priority and low-cost; the total cost is calculated based on the number of hops, network delay, bandwidth and security cost. Furthermore, the two policies of entrance sequence and rotating shift were simulated in the cloudsim in the above-mentioned algorithm; as a result, the service cost for the entrance sequence will be higher than the rotating shift and a rotating shift algorithm will be proposed (Gaurav Raj, 2012).

In order to maintain load balance, Gu et al, (2012) proposed a genetic-based algorithm; they conducted many experiments and investigations which revealed that their proposed algorithm is able to sort out the issue of load balance and high migration. The problem with this algorithm is that the cost of virtual machine is not considered in this algorithm (Gu, 2012).

A new method was proposed in 2013 which aimed at maintaining load balance in cloud computing. In that method, the proposed algorithm produces a C matrix in which *Cij* stands for completion of the task at time *Ti* and in the resource *Rj*. If the number of available resources is odd, for proceeding work, minimization algorithm will be carried out; otherwise, maximum minimum algorithm will be implemented. This algorithm leads to efficiency enhancement in cloud systems; however, the cost of resource utilization in parallel with efficiency enhancement has not been considered (Mohana priya, 2013).

With respect to the unique features of connection programs, a developed ESBWLC algorithm was proposed by Kapgate (2013); this algorithm optimizes the number of connections and directly reduces the delay time on the part of the service receiver. The purpose of the algorithm proposed by Kapgate (2013) which was simulated in the cloud setting was to achieve load balance.

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-1/MAGNT.41)

Hong Xu et al, (2013) optimized the income through dynamic pricing. Having examined spot prices, they came to the conclusion that it is improbable that the spot sample price would be compatible with the sample on demand. They also investigated income management and tried to formulate the maximum income. Indeed, the merit of dynamic pricing is that service providers can ensure about the service efficiency for the customers (Hong Xu, 2013).

Subscription pricing is a simple method for refunding money in which pricing is done at the time unit rather than the consumption unit. In contrast, peak-level pricing is based supervising and recording peak consumption of users. User-based pricing was analyzed in two 5-day periods. It was found that pricing in the first period was 50 percent faster than that of the second period which is due to the fact that in the second period, all users entered the system. Ticket-based pricing is similar to buying a film ticket where high-class services are given to high-price tickets. In Integral pricing, break-even analysis is used to obtain service-provision price for one month(Denne, 2007).

An investigation and analysis of related previous studies reveal that parameters such as power improvement, migration time reduction, response time reduction, resource utilization enhancement and system efficiency improvement have been some of the outstanding factors which have been considered and examined in previous works. It should be noted that each researcher has focused on one particular parameter since examining all parameters at the same time within one study can be highly complicated and requires confusing computational complexities. Hence, each researcher has been concerned with one parameter at a time. In the following table, a synopsis of previous works and recent studies is given. However, it should be noted that one significant parameter has been ignored by researchers; that is, host machine cost on the part of service provide has not been examined by relevant researchers. In cloud computing challenges, server cost related to the available virtual machine is of high significance. For instance, in a situation where there are some low-priced servers and some high-priced servers, a significant issue is which user will be allocated to which server by the machine.

Table1. The comparison of previous scheduling studies

Different available methods	Merits	Demerits
BVCF algorithm, executed with RCFs and RR scheduling policies	Execution speed enhancement, time and cost optimization	Increased power consumption
Genetic-based algorithm	Load balance, solving the problems of high migration	Genetic-based algorithm
RASA algorithm	Performance improvement, efficiency improvement in cloud systems	The cost of utilizing resources was not taken into consideration in parallel with increase in performance
ESBWLC algorithm	Optimization of the number of connections and reduction of delay time on the part of service receiver	Neglecting Cost
PALB algorithm	Energy optimization, load balance	Increase in response time

It should be noticed that host server costs on the part of service provider can increase as the resource utilization increases; however, this significant issue with respect to virtual machine cost has been neglected in the majority of previous studies. The present study intends to fill this research gap; hence, this parameter will be examined in this paper.

3. Designing a New Method

In this paper, the purpose for proposing a novel method was to enhance the income of service providers. The proposed method was designed and implemented which is described in this section. Since in the cloud computing-based data center, the trend is to move towards resource combination and integration, hence, costs cannot be managed manually according to the

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-1/MAGNT.41)

categorization of servers and determination of different pricing models based on different situations; rather, there is a tendency towards autonomous management of data centers. Indeed, the issue of resource management can be solved with regard to server cost through the following two methods:

- Traditional method: in this method, different resources are categorized and priced based on the traditional models of pricing.
- Automating pricing model and scheduling virtual machines with respect to server costs.

In this study, practical procedures have been proposed according to the second method of resource management. That is, at any moment, the user is notified of the average price of resources; however, at the scheduling time, low-cost servers have the highest priority. Since the sample can be allocated and given to different physical servers, the computation of the final cost of a computational sample can be variable. It is obvious that when all the data center resources are 100 percent in use, considering the variable prices is not important since the total price will be regarded as average. Nevertheless, since workloads are usually variable, taking this issue into consideration can increase profits and reduce losses.

In the proposed method, the *break-even* analysis was used to calculate the cost for providing service. Also, since the useful life of using resources is about 3 years, the value of *U* which denotes the time span for providing service in terms of hour can be obtained through equation 1 below:

$$U=3*365*24=26280 \quad (1)$$

O the other hand, if we assume that R % of resources is used and N % profit is obtained, the cost of each processing unit will be obtained through equation 2 below:

$$P=C/U*R*N \quad (2)$$

In this cost calculation method, the total cost of processing resources will be calculated for each request for the servers which have free capacity; then, after calculating head to head cost, the request cost will be measured in proportion with the optional parameters of the service provider. Next, a low-cost server which has a vacancy will be selected and request scheduling will implemented on it. Ultimately, the final cost for this request will measured through *break-even* analysis. For a better understanding of the procedure, it has been illustrated in the figure1.

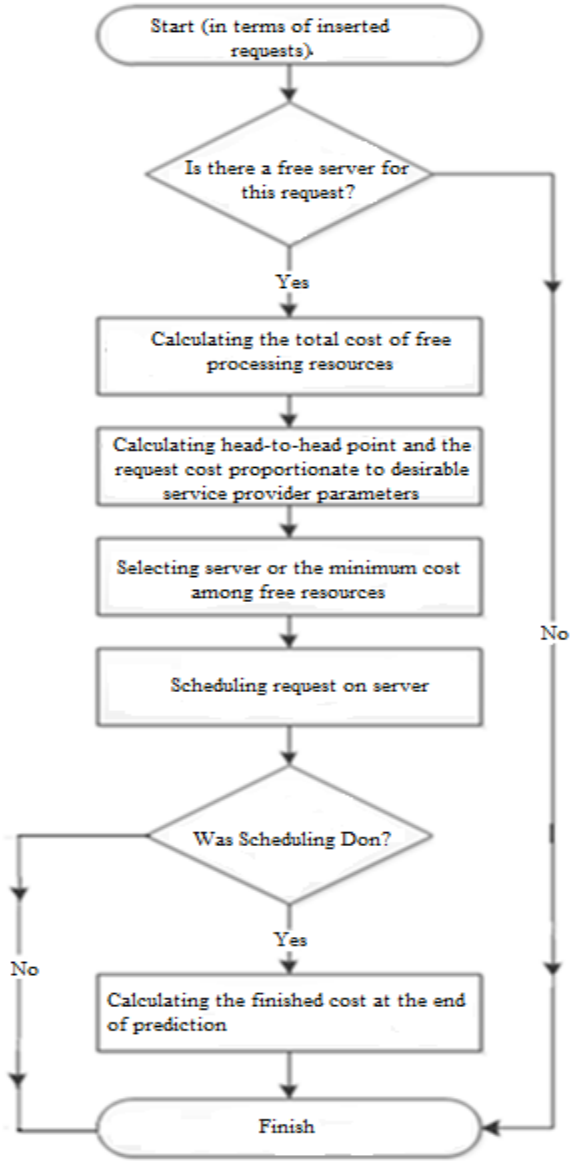

Figure 1. Stages of calculating cost by the method of low cost priority

Figure 2 illustrates the proposed algorithm.

Client Side Pricing Algorithm (CSPA):

1- *Calculate provisioning ratio R = 2*
2- *Calculate net ratio N = 1.2*
3- *For each new request r*

- *C = Calculate the cost of resources for r*
- *Calculate Marginal Price M = C / 3 / 365 / 24*
- *Calculate Instance type r Price P = M * R * N*

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-1/MAGNT.41)

- *Return P as price of r*

Cost based Resource Scheduling Algorithm (CRSA):

1- *r = new Instance to be schedule*
2- *FreeList = The hosts with free PEs for r*
3- *MVH = The host with the min value from FreeList*
4- *Schedule Instance r in MVH*

Figure 2. Pseudo code of the proposed algorithm

4. Simulating And Evaluating the Proposed Model

For simulating the proposed model, the researchers used the *Cloudsim* software. Conducting simulation in Cloudsim requires defining the following factors: defining the availabilities of the data center, defining physical resources in the data center, defining scheduler (equivalent to Cloudsim broker), defining virtual machine allocation policies used by the scheduler, defining users' request including virtual machine and the task which should be done on virtual machine, i.e. cloudlet. Since the cost of each physical machine should be considered in the proposed method, hence, it is imperative that the attributes and features required for implementing Cloudsim should be defined and given. Then, the virtual machine allocation policies should be implemented in the scheduling model. In this paper, three methods were implemented, i.e. basic method, rotational shifting and the proposed method.

In order to evaluate the model, the researchers produced 20 sets of random data. For each set of data, the simulation was replicated with different methods. Table 2 illustrates the summary of the simulation results for the following parameters: cost, resource utilization (the ratio of used resources in the simulation to the requested resources) and the number of requests which were not executed due to the shortage of resources.

Table2. Comparison of simulation results for different methods on different parameters

Test#	Proposed			RR			FirstFit		
	Cost	Drop	Utilization	Cost	Drop	Utilization	Cost	Drop	Utilization
1	14.27	4	76	12.75	4	76	12.82	4	76
2	10.31	8	52	10.35	6	63	10.35	6	63
3	11.39	5	73	10.21	5	73	10.21	5	73
4	10.06	6	63	8.75	6	63	9.46	5	66
5	11.87	1	91	9.53	1	91	8.41	2	81
6	13.2	4	70	11.4	4	76	10.11	4	68
7	12.78	2	79	11.94	2	79	11.94	2	79
8	11.11	4	63	9.9	4	63	9.9	4	63
9	13.72	4	75	11.84	4	75	11.84	4	75
10	13.77	4	69	12.21	4	69	10.01	6	53
11	12	7	58	12.21	5	63	9.84	6	58
12	11.59	5	68	9.29	5	68	9.29	5	65
13	12.55	4	69	10.96	4	69	11.47	4	72
14	14.68	4	83	13.43	4	80	12.21	5	71
15	10.67	4	77	8.72	3	81	8.72	3	81
16	13.92	3	73	11.53	3	73	10.85	4	68
17	12.71	4	71	10.01	5	64	10.68	4	71
18	14.43	4	74	12.04	4	74	11.36	4	72
19	11.33	5	64	9.5	5	64	9.23	5	61
20	13.18	3	76	10.24	4	71	10.65	4	71
Average	**12.477**	**4.25**	**71**	**10.8405**	**4.1**	**72**	**10.4675**	**4.3**	**69**
Max	**14.68**	**8**	**90.63**	**13.43**	**6**	**90.63**	**12.82**	**6**	**81.25**
Min	**10.06**	**1**	**52.08**	**8.72**	**1**	**62.50**	**8.41**	**2**	**53.33**

The values obtained for the comparison of cost parameter indicate that the data center has gained profits which are the result of applying different resource allocation policies in the scheduling algorithm. Figure 3 depicts these results. As shown in this figure, the proposed algorithm was able to obtained the highest profit among all other algorithms.

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-1/MAGNT.41)

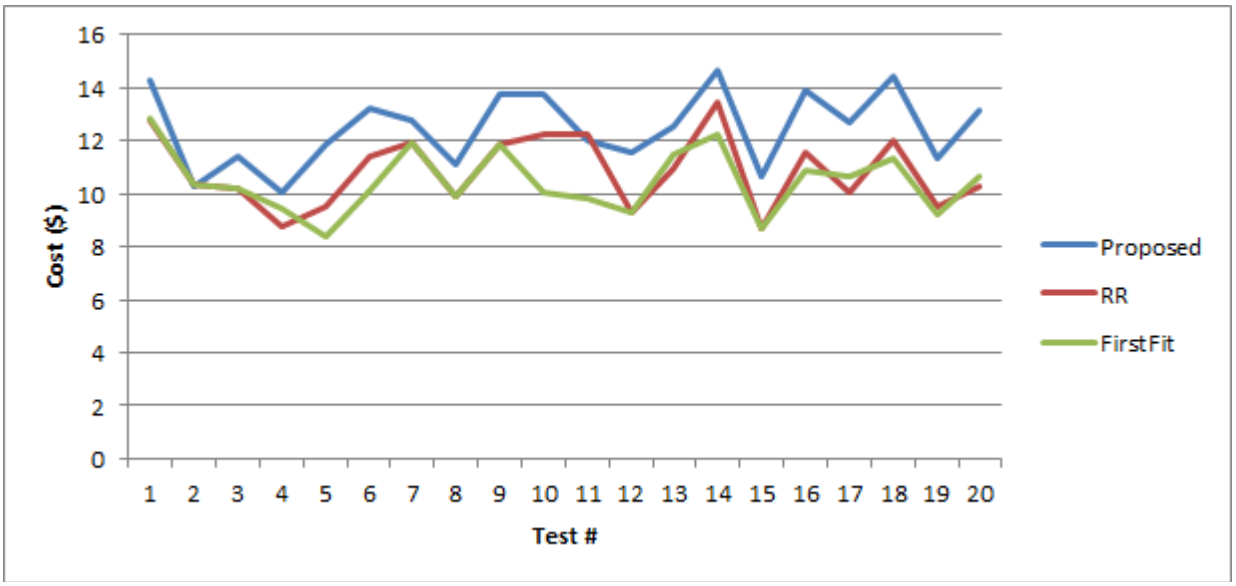

Figure 3. Comparing the cost of requests (data center profit) in 20 different simulations for the three examined methods

Furthermore, figure 4 below illustrates the average profit for the three investigated algorithms in different simulations.

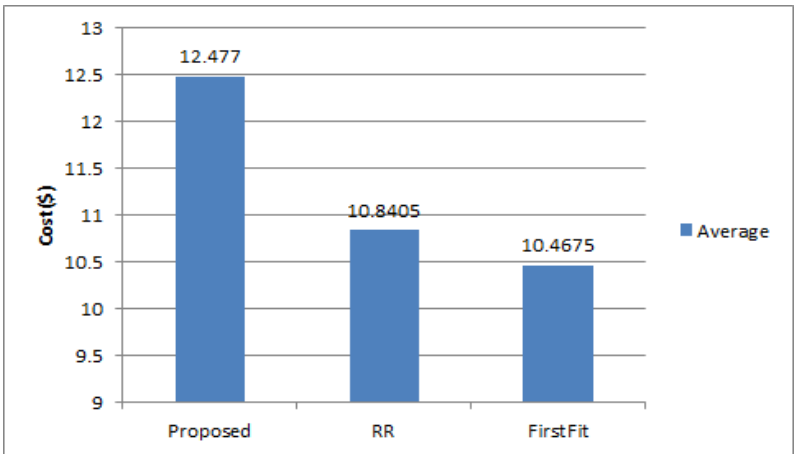

Figure4. Comparison of the average cost of requests (data center profit) in all conducted simulations

Moreover, figure 5 demonstrates the ratio of using data center resources in different simulations. This figure reveals that resource utilization levels in different methods are highly close to each other.

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-1/MAGNT.41)

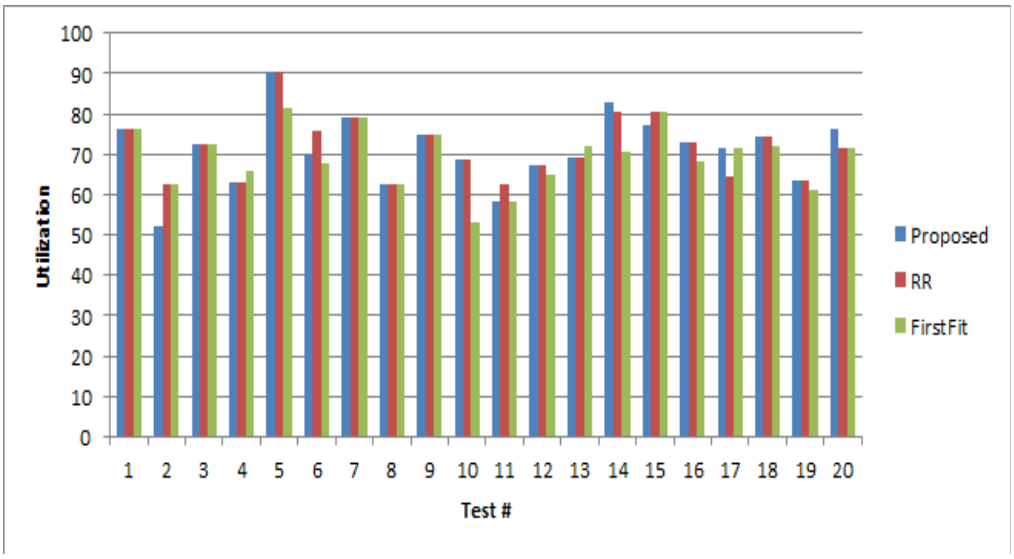

Figure5. Comparison of the ratio of resource utilization in data center'

The average results for more accurate comparison of the three methods are given in figure 6.

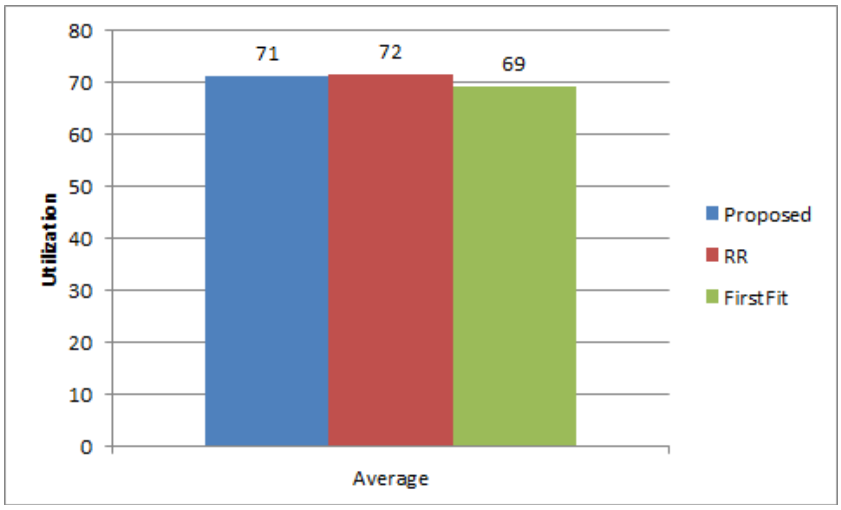

Figure6. Comparing average resource utilization in data center for different experiments

This figure indicates that average resource utilizations are very similar; thus, it can be mentioned that the proposed method can obtain further profits than the other two methods as well as maintaining high service provision level.

5. Conclusion

The issue and research problem which was examined in this paper was cloud scheduling. In the method proposed in this study, the virtual machines were scheduled optimally on host servers with respect to the cost. That is, at any moment, the user is notified of the average price of the available resources. However, it should be noted that in the scheduling time, the low-cost servers were the highest priority. In this research study, Cloudsim simulation was used. The proposed method was compared with the basic method and rotational shifting method in terms two important parameters, namely optimal management of cost and enhancement of

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-1/MAGNT.41)

resource utilization. The analysis of simulation results revealed that applying the proposed method leads to the increase in the data center profits. In other words, the results of simulation indicated that, in the proposed method, the parameters of income evaluation and resource utilization were optimized for 8 % and 3 %, respectively in comparison with the other two methods. Hence, the proposed method improved these two parameters significantly better than basic method and the rotational shift method. However, the drawback of the proposed method is a minor reduction in resource utilization of the rotational shift method. As a direction for future research, interested researchers can optimize the proposed method in different respects. The following issues can be considered in related future studied:

- Proposing a method for enhancing the income of service provide through optimizing consumption energy
- Reducing the response time in the method proposed in this study.

References

1. B.Mondala,, K.Dasguptaa, P.Duttab. (2012) . Load Balancing in Cloud Computing using Stochastic Hill Climbing-A Soft Computing Approach, Procedia Technology 4,pp. 783 – 789.
2. Bernal Bernabe.J,b, Juan M. Marin Perez, Jose M. Alcaraz Calero b, Felix J. Garcia Clementec, Gregorio Martinez Perez a, Antonio F. Gomez Skarmetaa(2012). Semantic-aware multi-tenancy authorization system for cloud architectures. Future Generation Computer Systems. Volume 32, Pages 154-167.
3. D.Kapgate, M.B.Narnaware.(2013). Predictive Load Balancing Strategy for reduction of Latency in Mobile cloud Computing. International Journal of Computer & Communication Engineering Research (IJCCER) Volume 1 - Issue 1
4. Galloway, Jeffrey M., Karl L. Smith, and Susan S. Vrbsky(2011). Power Aware Load Balancing for Cloud Computing. Proceedings of the World Congress on Engineering and Computer Science. Vol. 1.
5. Gaurav.R, Sonika Setia (2012). Effective Cost Mechanism for Cloudlet Retransmission and Prioritized VM Scheduling Mechanism over Broker Virtual Machine Communication Framework. International Journal on Cloud Computing: Services and Architecture(IJCCSA),Vol.2, No.3.
6. Gaurav.R, Sonika Setia2(2012). Effective Cost Mechanism for Cloudlet Retransmission and Prioritized VM Scheduling Mechanism over Broker Virtual Machine Communication Framework. International Journal on Cloud Computing: Services and Architecture(IJCCSA),Vol.2, No.3.
7. Gu.J, Hu.J, zhao.T, Sun.G.(2012). A New Resource Scheduling Strategy Based on Genetic Algorithm in Cloud Computing Environment. Journal Of Computers, VOL. 7, NO. 1.
8. Gu.J, Hu.J, zhao.T, Sun.G.(2012). A New Resource Scheduling Strategy Based on Genetic Algorithm in Cloud Computing Environment. JOURNAL OF COMPUTERS, VOL. 7, NO. 1.
9. H.Xu, B.Li.(2013).Dynamic Cloud Pricing for Revenue Maximization, IEEE Transactions on Cloud Computing, Vol. 1, No. 2.
10. Kloster, J. F. & Kristensen, J. & Mejlholm, A. (2007) A Comparison of Hardware Virtual Machines Versus Native Performance in Xen.
11. Mark Denne,M. (2007). Pricing Utility computing services. International Journal of Web Services Research, Accenture, USA , Volume 4, Issue 2.pp.1-14
12. Mohana Priya.S, Subramani.B.(2013). A NEW APPROACH FOR LOAD BALANCING IN CLOUD COMPUTING. International Journal Of Engineering And Computer Science ISSN:2319-7242, Volume 2, Issue 5. Page No. 1636-1640.
13. Mohana Priya.S, Subramani.B.(2013). A New Approch For Load Balancing In Cloud Computing. International Journal Of Engineering And Computer Science ISSN:2319-7242, Volume 2, Issue 5. Page No. 1636-1640.

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-1/MAGNT.41)

14. Mr. Kapgate.D, Mr. M. B. Narnaware.(2013). Predictive Load Balancing Strategy for reduction of Latency in Mobile Cloud Computing. International Journal of Computer & Communication Engineering Research (IJCCER),Volume 1, Issue 1.
15. Skandarian, P., J. Karimpour. (2012). Strategies for Enhancing Cloud Computing Service Efficiency through Virtulization. The First National Cloud Computing Workshop in Iran.
16. Su, S., Li, J., Huang, Q., Huang, X., Shuang, K., & Wang, J. (2013). Cost-efficient task scheduling for executing large programs. Parallel Computing, 188-177.
17. T.Sharma, Dr.Vi.Kumar .Banga,(2013.). "Proposed Efficient and Enhanced Algorithm in Cloud Computing", International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 2,

(DOI: dx.doi.org/14.9831/1444-8939.2015/3-1/MAGNT.41)