Send Orders for Reprints to reprints@benthamscience.ae

The Open Cybernetics & Systemics Journal, **2015**, *9*, 639-647

Open Access

The Virtual Machine Resource Allocation based on Service Features in Cloud Computing Environment

Liu Yan* and Guo Chenxian

College of Information Engineering, Zhongzhou University, Zhengzhou, 450044, P.R. China

Abstract: This paper presents the virtual machine resource allocation algorithm based on business characteristics under a cloud computing environment. This algorithm describes the user service characteristic time delay factor and price factor and other parameters, resource scheduler computing data center according to the user on the business characteristics of the description and the status of cloud calculated the cost of the business index. According to the cost index distribution of cloud computing resources, firstly the system model of the algorithm is introduced, and then the model is established according to the corresponding problem of the system model, and then the corresponding resource allocation scheme is proposed. Finally, the performance of the algorithm is verified by simulation. The simulation results show that the proposed algorithm in cloud computing platform improves the utilization rate of resources and at the same time significantly reduces the computational characteristics of different business users of cloud use cost, and improves the cloud user service experience.

Keywords: Business characteristics, Cloud computing, Resource allocation, Virtual machine.

1. INTRODUCTION

With the development of cloud computing, more and more users will do business migration to the cloud service providers of cloud computing flat station, a virtual machine running in the cloud computing platform, and become the cloud users. Cloud service providers perform according to the specific needs of the cloud user business distribution of its corresponding virtual machine configuration and quantity.

Using the current resource allocation method can improve the rate of cloud resources better, and increase the cloud service provider revenue. However, from the cloud user's point of view, these methods can not bring much benefit to the cloud user, or even reduce the user experience, from a long-term point of view; the long-term income of a cloud service provider will also have a negative impact. The business characteristics of different clouds of different users vary from the virtual machine resource price, delay and resource deployment and other requirements are also not identical. Therefore, in the process of virtual machine resource allocation, allocation is necessary for virtual machine resources according to the characteristics of the cloud user, thereby reducing the cost of low cloud users, and improving the cloud user service experience, and then realizing maximum cloud service providers of long-term income [1]. In this paper, we from the cloud user perspective proposed the virtual machine allocation algorithm based on a cloud computing business characteristics, in order to improve the utilization rate of the platform in cloud computing resources at the same time reducing the cloud user cost, and improving the cloud user service experience.

2. THE SYSTEM MODEL

Considering a computing platform is distributed in a plurality of different geographic location data centers to form clouds. Capacity of each data center, delay performance and virtual machine resource prices are not the same. Cloud service providers perform according to the user's requirements for the allocation of cloud resources corresponding virtual machine and operation of cloud users of virtual machine in the data center [2]. In order to understand, following definitions are given:

Definition one: a resource scheduler, running state of computing data center according to the user's requirements and the cloud virtual machine resource allocation corresponding to the cloud user, a data center resource scheduler itself runs in cloud computing platform.

Definition two: data center status monitoring process, running on each data center, running state real-time monitoring data center, the amount of residual resources including data center and virtual machine resource price information, status information and to the resource scheduler sends location data center.

Definition three: data center status table, data storage center state information for resource scheduler module, according to the state information center data process state monitoring reports are updated in real time, including the data center of the residual amount of resources, time delay and resource price information.

According to the above definition, the system architecture can be represented as in Fig. (**1**).

*Address correspondence to this author at the College of Information Engineering, Zhongzhou University, Zhengzhou, Henan, 450044, P.R. China; Tel: 13803770071; E-mail: liuyan0407@126.com

1874-110X/15 2015 Bentham Open

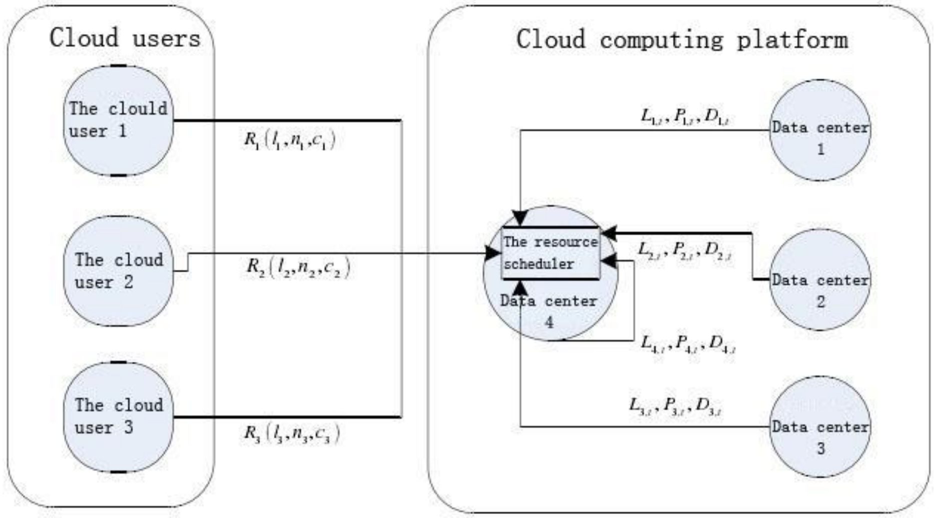

Fig. (1). Virtual machine system framework of distribution maps based on business characteristics.

In Fig. (**1**), $R_i(l_i,n_i,c_i)$ represents the cloud user submitted to the resource scheduler resource request, whereas $l_i = l_i(f,m,h,b)$ represents a cloud user i for virtual machine configuration, including the processor f, memory m, external memory bandwidth h and b. To represent a number of cloud users, *i* is needed for virtual machine; c_i represents business characteristics of the cloud user *i*; $L_{n,t}$, $P_{n,t}$, $D_{n,t}$ respectively represent the amount of remaining resources data center and *n* reports data center status monitoring process to a resource scheduler, resource price and time delay estimation information, whereas, the resource price $P_{n,t} = P(f_{n,t}, m_{n,t}, h_{n,t}, b_{n,t})$, including processor computing resource price $f_{n,t}$, resource in memory $m_{n,t}$, external storage resource $h_{n,t}$ and bandwidth resource $b_{n,t}$. In order to simplify the system model, we assume a cloud user with only one type of business [3].

3. MODEL CONSTRUCTION AND PROBLEM DESCRIPTION

3.1. Description of Service Features

For the specific business, such as massively multiplayer online game business, the business of large amount of internal information interaction, virtual machine deployment for the use of centralized deployment mode, and highly sensitive to the delay, resource price has little influence on it. As micro-blog social business, with strong regional characteristics, the smaller amount of internal business information interaction is more suitable for virtual machine deployment adopting distributed deployment mode, and at the same time, delay in this kind of business generates low sensitivity of price. The difference of business of cloud users leads to its variable business characteristics, and according to the virtual machine deployment requirements, time delay factor and price factor can be used to describe the business characteristics of cloud user i.

$$c_i = (s_i, a_i, b_i) \tag{1}$$

3.2. Status Updates of Data Center

On the pricing model of resource price data center and cloud provider, in the system model, taking into account the equalization of multiple data center load demand, we used dynamic pricing model based on the residual amount of resources. Specifically, data center resources price $P_{k,t}$ was calculated by the data center status monitoring process according to the following formula.

$$P_{k,t} = g_1 P_{k,o} + g_2 f(L_{k,t}) \tag{2}$$

In the formula, k represents the data centre number, t represent the current time, g_i as weighting factor, $\sum \gamma_i = 1$, $P_{k,o}$ is a reference price of resources data center, k is constant, $f(L_{k,t})$ is the function associated with the residual amount of resources.

Time delay estimation of data center renewal process: resource scheduler periodically Xiang Yun computing data center sending delay test signal, and record the response time delay $d_{k,t}$ the test signal, sending periodic testing signal for t_d, and whenever a resource scheduler receives the state information from the data center will also send to the data center of an experiment the test signal. Delay of data center is obtained by the formula:

$$D_{k,t} = \sum_{i=1}^{I} \frac{d_{k,t}}{I} \tag{3}$$

where I represents the most recent I signal, namely the mean response delay time delay I recently a test signal to the data center estimation.

3.2. Description of the Problem

In order to influence the intuitive description of data center delay and resource price on the business, we put forward the concept of business cost index, resource scheduler according to the cloud user service cost index for the allocation of resources to the corresponding business. Cost index can be calculated by the following formula:

$$T_{i,k,t} = a_i D_{k,t} + b_i P_{i,k,t} \tag{4}$$

Among them, $T_{i,k,t}$ represents the cloud user i arrived at t moment of its business with respect to data center k cost index, $P_{i,k,t} = P_{k,t} l_i$ represents the user i in the data center, k is deployed on a virtual machine that configured the requirements of cost, the same cloud users of its business delay factor α and price factor β are constant. From the type set, on the same cloud user service, different data centers and different time, its cost index are not the same.

According to business characteristics of cloud users assign virtual machine resource, minimum cloud user cost, cost index minimization problem is equivalent to solving a cloud user, therefore, our objective function can be described as follows:

$$\begin{aligned} &\min T_{i,t} = \sum_{k=1}^{K} T_{i,k,t} x_{i,k,t} \\ &s.t. \sum_{k=1}^{K} x_{i,k,t} = n_i \\ &x_{i,k,t} l_i \le L_{k,t}, \forall k = 1,2,...,K \end{aligned} \tag{5}$$

Among them, $T_{i,t}$ represents the total cost of the cloud user i, t represents the cloud user i service request arrival time, $x_{i,k,t}$ represents the virtual machine allocation index, where t users of i virtual machine in k data center are the upper department number. The first restriction represents a number of virtual machine resources allocated to the virtual machine scheduler total user i should be equal to the user i required; second limiting conditions indicates that the data center on k allocated to the virtual machine resources required of user i should be less than the total amount of resources available to the data center k.

Known by the first limiting conditions, the objective function is equivalent to:

$$\min T_{i,t} = n_i \sum_{k=1}^{K} \frac{x_{i,k,t}}{n_i} T_{i,k,t} \tag{6}$$

In the formula, for a particular user service, the number of virtual machine needs n_i is determined, and the cost of index $T_{i,k,t}$ user service relative to a particular data center is the fixed value. Therefore, we can solve the water filling algorithm through the optimization problem [4]. The steps are as follows:

1) According to the cost index to sort the data center, remember the ranked set for the D_T virtual machine, the initialization allocation index: $x_{i,k,t} = 0$;

2) Select the data center, cost index $T_{i,k,t}$ minimum D_T if $x_{i,k^*,t} l_i \le L_{k^*,t}$, then data center for the user service agency, a virtual machine, $x_{i,k^*,t} = x_{i,k^*,t} + 1$, and more new $n_i = n_i - 1$ and $L_{k^*,t} = L_{k^*,t} - x_{i,k^*,t} l_i$, if $x_{i,k^*,t} l_i > L_{k*,t}$, then $D_T = D_T - k^*$, execute step 1);

3) $D_T = \emptyset$, indicates that the current cloud computing platform does not have sufficient resources to provide services, for the user allocation failure, end; if $x_{i,k^*,t} = n_i$ and the distribution end, otherwise execute step 2);

Through many times of iterative water filling algorithm, we can find the best virtual machine user I deployment strategy.

4. RESOURCE ALLOCATION ALGORITHM BASED ON BUSINESS CHARACTERISTICS

According to the system model and the analysis of the problem, we propose a virtual machine resource allocation algorithm based on business characteristics. The specific process is shown in Fig. (**2**).

The concrete steps of the proposed virtual machine allocation algorithm are as follows:

Step one, the user to submit a request for a resource allocating resources $R_i(l_i, n_i, c_i)$;

Step two, if the user's virtual machine deployment strategy of $s_i = 1$, execute step three; if $s_i = v, (v > 1)$, then

execute step four;

Step three, centralized way for user i to the virtual machine resource allocation, the specific method for:

1) Resource scheduler queries the data center k^* status table, elected to meet the needs of the user i resources data center $R_i(l_i, n_i)$, must satisfy $L_{k^*,t} \ge n_i l_i$, and constitutes the candidate data center set G_i of user i.

2) If the candidate user i set data center G_i is null, it suggests that the current cloud computing platform has not enough resources allocated to the user I, resource scheduler refused to user i request, execute step five; otherwise, the calculation of candidate data center cost index $T_{i,k,t}, k \in G_i$ relative cloud users I collection of G_i in each data center;

3) Select the candidate data center G_i collection cost index and the lowest in $T_{i,k,t}$ data center k^* for user i department cloud virtual machine;

4) Remaining resources state of $L_{k,t}$ on the selected data center are updated, execute step five;

Step four, adopting distributed deployment of virtual machine as the user i, the specific method for:

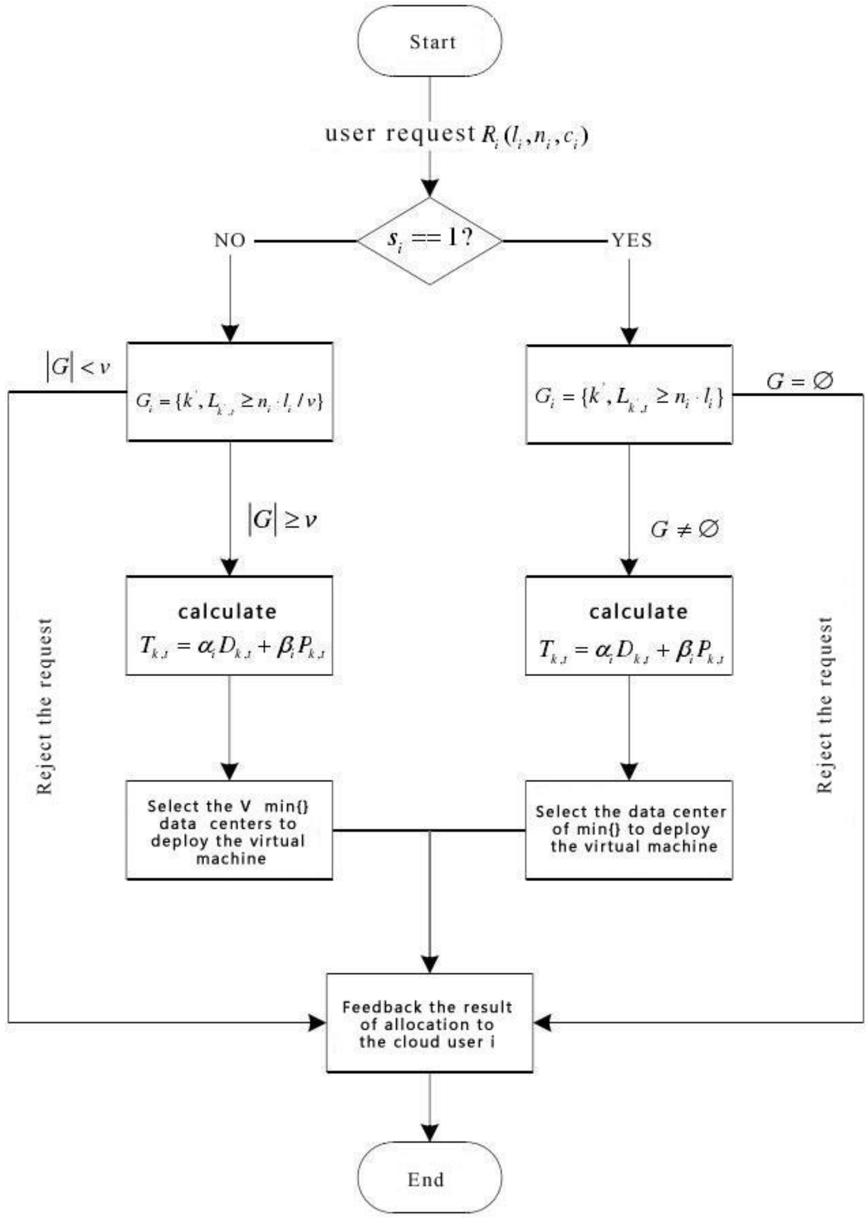

Fig. (2). Flow chart of algorithm.

1) Resource scheduler queries the data center status table, select the resource requirement $R_i(l_i, n_i)$to meet the user i data center k^*, must satisfy $L_{k^*,t} \geq n_i l$, wherein, $L_{k^*,t}$ is the surplus resources, n_i represents the number of user $i$needed for virtual machine, l_i represents the cloud user desired virtual machine configuration, v represents the user requirements of virtual machine which are deployed in v data center, the data center of a candidate data center user i in the set G_i;

2) If the candidate data center G_i collection of data center $|G_i| < v$ number, it represents that the current data are not adequate to meet the cloud center user i resource requirements, resource scheduler refused to user i requests for resources, execute step five; otherwise, the calculation of the candidate set of each data center in G_i phase to the cloud user $i$cost of index $T_{i,k,t}, k \mid G_i$ is performed;

3) In the candidate set G_i consisting of a collection of data center cost index $T_{i,k,t}, k$ minimum v data center is selected in G_i^* in the G_i^* collection of the data center are user i virtual machine deployment.

4) The remaining resource state $L_{k,t}$ on the selected data center are updated, execute step five;

Step five, as the virtual machine assignment completes, the resource scheduler to the cloud user I feedback results.

Compared with the water filling algorithm standards, the above algorithm considering the strategy requires the user for the virtual machine deployment with some performance loss, but it still significantly reduces the user's cloud computing cost, and especially requires the use of centralized virtual machine deployment strategy users for better performance [5].

5. SIMULATION AND PERFORMANCE ANALYSIS

5.1. Establishment of the Simulation Model

In practice, the data center with delay and price resource utilization rate, therefore, we first define the utilization data center resources η, t moment data center k resources utilization rate calculation formula is as follows:

$$\eta_{k,t} = \frac{\mu_1 f_{k,t} + \mu_2 m_{k,t} + \mu_3 h_{k,t} + \mu_4 b_{k,t}}{\mu_1 f_k + \mu_2 m_k + \mu_3 h_k + \mu_4 b_k} \tag{7}$$

In the formula, μ_i represents the corresponding to the data center resources influence factor, delay of $0 < \mu_i < 1, \sum_i \mu_i = 1$; the denominator indicates the total amount of resources data center k, the denominator represents the use of a corresponding amount of resources in the t time.

According to the utilization rate of data center resources, we define the relationship between delay and resource utilization rate in the simulation model as follows:

$$D_{k,t} = D_{k,0} \exp\left[\frac{1}{1-\eta_{k,t}} - e\right] \tag{8}$$

In the formula, $D_{k,0}$ reference delay data center k, according to the formula, when the resource utilization rate is less than 70%, the resource utilization rate increases slowly; when the resource utilization rate is higher than 70%, the increase of resource utilization rate increases rapidly [6].

Similarly, we define the rate relationship by the price of resource and resource data center in the simulation model as follows:

$$P_{k,t} = \frac{1}{1-\eta_{k,t}} P_{k,o} + P_{k,o} \tag{9}$$

Among them, $P_{k,o}$ represent the benchmark price of resources data center k, which can be seen from the above equation; with the increase of resource utilization, resource price data center also gradually increases.

5.2. The Results of Simulation Analysis

Consider a distributed cloud computing platform consisting of five different geographic location data formed in the center of each data center, all having the same capacity, but its reference resources price and reference delay are not the same. In each time slot, the business needs a number of virtual machines and the corresponding business characteristics randomly generated and executed in a data center business leaving a certain probability. The specific simulation parameters are shown in Table **1**.

Table 1: The simulation parameters.

Parameters	Value
The number of data center	5
Data center computing ability	8×3000MIPS
Data center memory capacity	16GByte
The bandwidth of data center	100Mbps
Leave the probability data center operations	0.02
Business average required number of virtual machine	3
The number of simulation time slot	288

Figs. (**3**) and (**4**), respectively, indicate the resource price five data center (Fig. **3**) and delay (Fig. **4**) in the 100th time slot to slot changes among the 200th. In the simulation model, the load is proportional to the price of resources and time delay and data centers, *i.e.* the larger the load, the higher the price of resources, and the delay is large. Five data centers' benchmark price and the reference delay are not the same. From the data center I to V, the benchmark price lowers and the reference delay is increased.

Fig. (**5**) shows the five data centers in the 100th to 200th time slot resource utilization curves, as can be seen from the graph. Resource allocation algorithm, five data center load difference, and the price model algorithm with balanced load effect are proposed in this paper. According to the simulation model, the average load of five data center is 60%, while the load in Fig. (**5**) data center fluctuates between 40%-70%, demonstrating that the proposed algorithm can achieve better load balancing data center between the target [7].

Fig. (**6**) shows the relation graph of rate and delay, resource price data center utilization III resources, as can be seen from the graph of time delay and resource prices and resource utilization rate is proportional to, and has a high correlation consistent with the simulation model theory, illustrating the correctness of simulation [8].

We concluded factor α more than 0.7 business as delay sensitive traffic, and the price factor β more than 0.7 business as price sensitive business. Both were statistically different virtual machine deployment strategies of delay sensitive traffic delay and resource cost and price sensitive business accumulated value, Fig. (**7**). Among them, the first acts as distributed virtual machine deployment delay and price different characteristics under the business strategy. It can be observed from the figure that for delay sensitive traffic, the cumulative delay value is far less than the price sensitive business Fig. (**7a**). This is because of the business office delay sensitive business requirements of smaller physical and time delay, and price sensitive business to business processing delay is not a strict requirement; on the contrary, the time delay sensitive service resource cost accumulation is greater than the price sensitive service, Fig. (**7b**). This is because the price sensitive business requirements of resource price are as low as possible, while the price of delay sensitive business to resource is not strict; Fig. (**7c**) and (**7d**) show centralized virtual machine deployment delay and

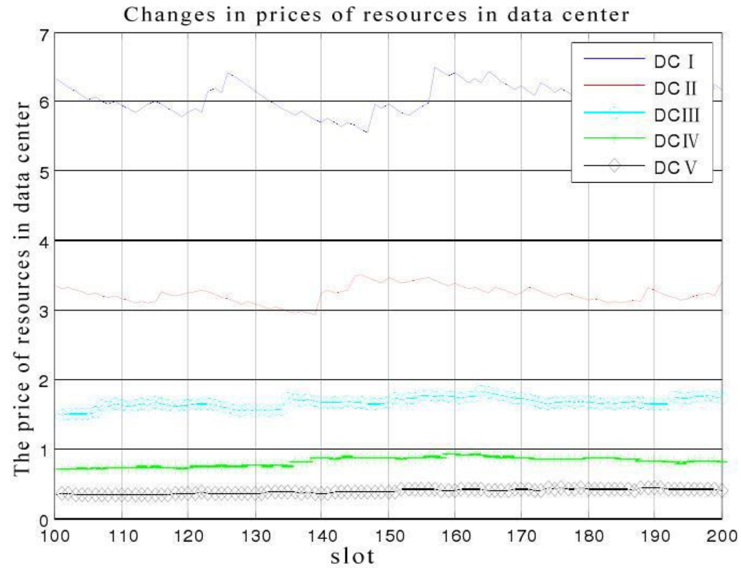

Fig. (3). Resource price curves of five data center.

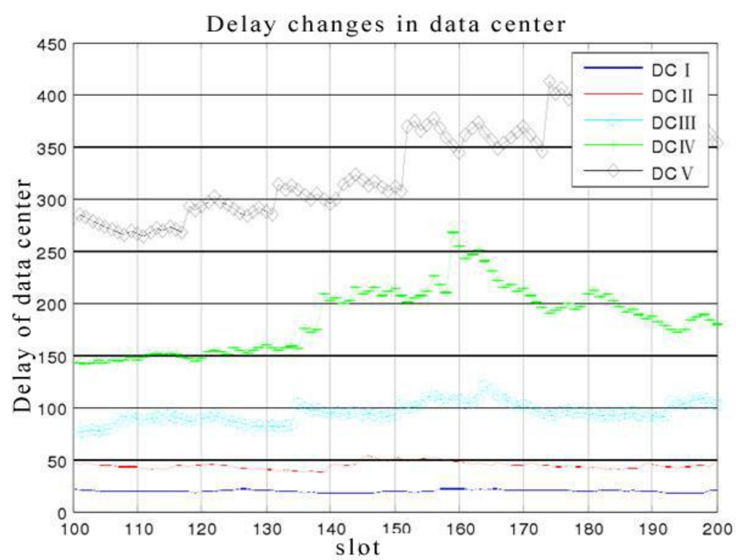

Fig. (4). Time delay curve of five data center.

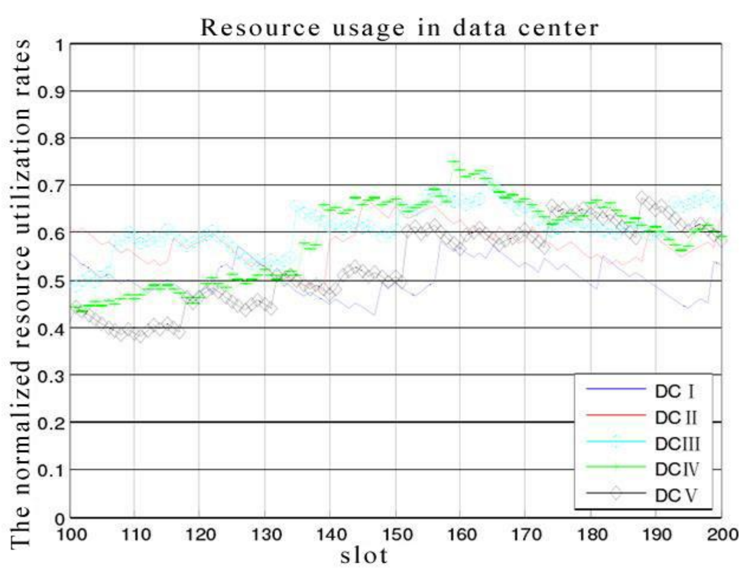

Fig. (5). The utilization rate of data center resources.

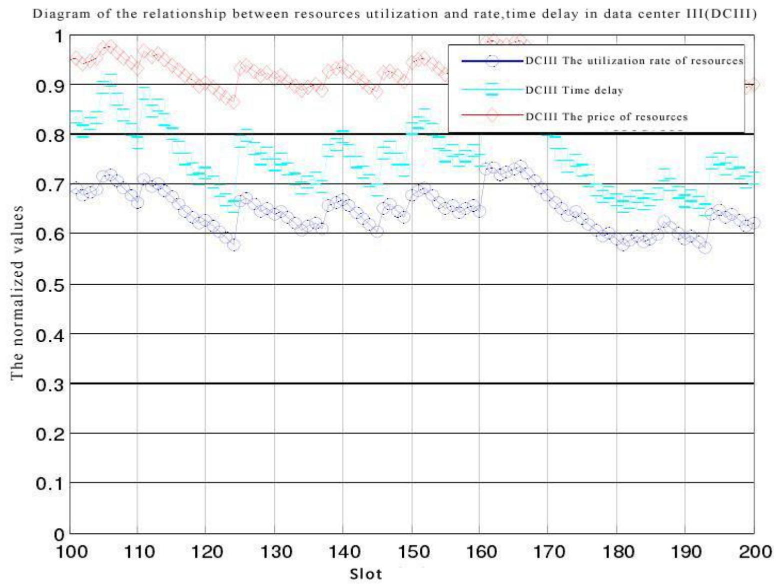

Fig. (6). Diagram of the relationship between resources utilization and rate, time delay in data center III(DCIII).

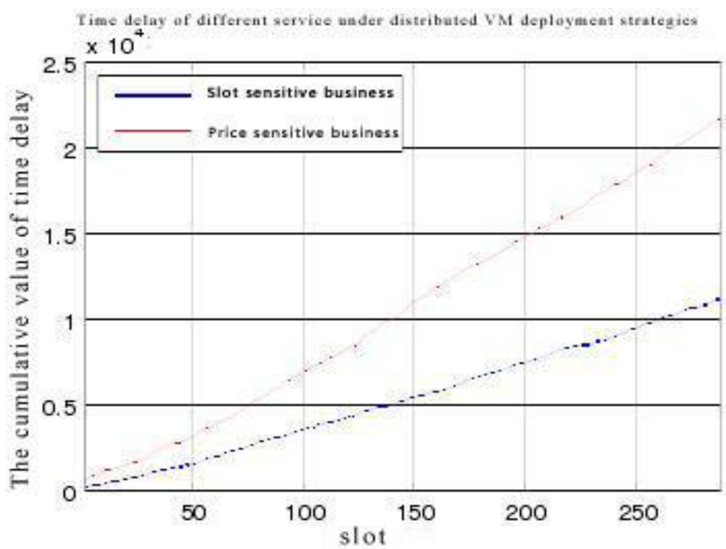

Fig. (7a). Time delay of different service under distributed VM deployment strategies.

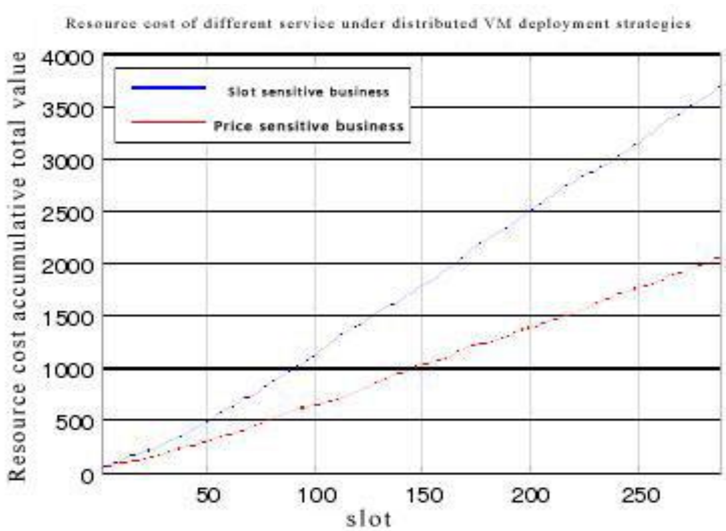

Fig. (7b). Resource cost of different service under distributed VM deployment strategies.

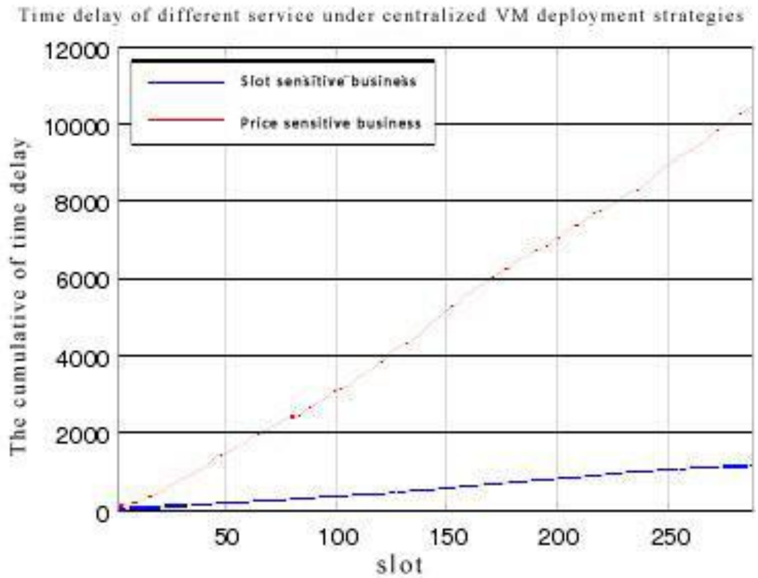

Fig. (7c). Time delay of different service under centralized VM deployment strategies.

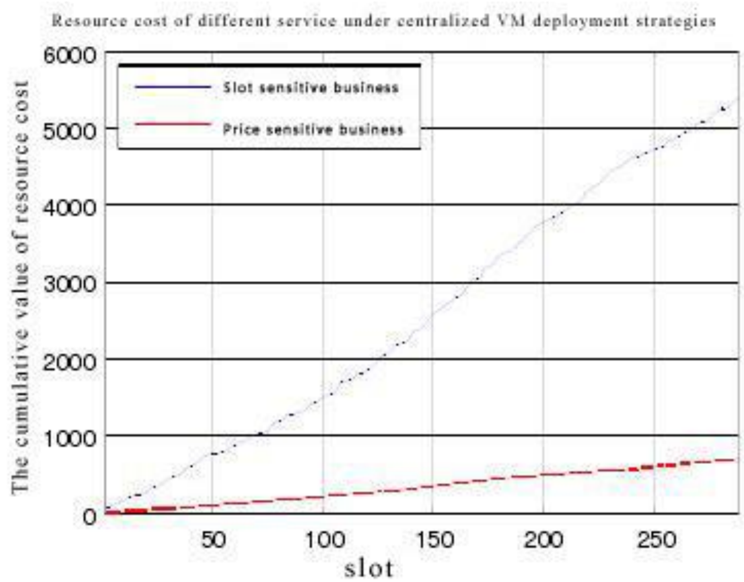

Fig. (7d). Resource cost of different service under centralized VM deployment strategies.

Fig. (7). Cost and time delay curve of the different characteristics business under different virtual machine deployment strategies.

resource price business strategies under the different characteristics of the cumulative curve; the result is similar to Figs. (**7a**) and (**7b**), which shows that no matter what the user virtual machine deployment strategy is, the proposed algorithm can be calculated using the corresponding cost reduce users of cloud. The centralized virtual machine deployment strategy users, differences between different types of traffic delays and resource prices are more obvious, the results show that the proposed algorithm is better to adopt for centralized virtual machine deployment strategy users return. In general, Fig. (**7**) shows that the virtual machine resource allocation algorithm is based on business characteristics that can allocate resources for users according to business characteristics of users, thus reducing the cost of the user [9, 10].

CONCLUSION

From the above-mentioned content, resource allocation algorithm based on business characteristics mainly introduces a cloud computing environment. This algorithm can perform according to the characteristics of user service to allocate appropriate resources for users, thus reducing the user cost, load balance and can achieve a cloud computing platform. This paper first introduces the system model of the algorithm, later the model is established according to the corresponding problem of the system model, and then the corresponding resource allocation scheme is proposed. Finally, the performance of the algorithm is verified by simulation. The simulation results show that the proposed algorithm can perform according to the characteristics of user service for a particular resource allocation to users, significantly reducing the cost to the user, and implementing a distributed cloud data center computing platform under different load balancing.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work was financially supported by the Liaoning Doctor Star-up Foundation (20111041).

REFERENCES

[1] L. Xu W. Chen and Z. Wang, "A strategy of dynamic resource scheduling in cloud data center," In: *Proc. of IEEE Cluster Workshops*, pp. 120-127, Sep. 2012.

[2] V. Jalaparti, and G. D. Nguyen, "*Cloud resource allocation games*," Dept of Computer Science, Tech. Rep, Dec. 2010.

[3] Kuilin Chen, C-RAN White Paper, version 2.5, China Mobile Research Institute, Oct. 2011.

[4] L. Guangjie, Z. Senjie and Y. Xuebin, "Architecture of GPP based, scalable, large-scale C-RAN BBU pool," In: *Proc. of IEEE GC Workshops*, Dec. 2012, pp. 267-272.

[5] Z. B. Zhu, P. Gupta, and Q. Wang, "Virtual base station pool: towards a wireless network cloud for radio access networks," In: *Proc. of the 8th ACM International Conference on Computing Frontiers* April 2011, p. 34.

[6] G. Bhanage, I. Seskar, and R. Mahindra, "Virtual basestation: architecture for an open shared WiMAX framework," In: *Proc. of the 2nd ACM SIGCOMM Workshop on Virtualized Infrastructure Systems and Architectures*, Sep. 2010, pp. 1-8.

[7] S. Namba, T. Warabin, and, S. Kaneko, "BBU-RRH switching schemes for centralized RAN," In: *Proc. of IEEE CHINACOM*, Aug. 2012, pp. 762-766.

[8] S. Bhaumik, S. P. Chandrabose, and M. K. Jataprolu, "CloudIQ: a framework for processing base stations in a data center," In: *Proc. of the 18th Annual International Conference on Mobile Computing and Networking*, Aug. 2012, pp. 125-136.

[9] X. F. Tao, Y. Z. Hou, and K. D. Wang, "GPP-based soft base station designing and optimization," *Journal of Computer Science and Technology*, vol. 28, no. 3, pp. 420-428, 2013.

[10] Y. Chen, X. Li and F. Chen, "Overview and analysis of cloud computing research and application," In: *Proc. of IEEE ICEE*, May 2011, pp. 1-4.

Received: April 02, 2015 Revised: May 23, 2015 Accepted: June 06, 2015