ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Profit Maximization with Guaranteed QOS in Cloud Computing

Ch Padmaja [1], V Divya[2], Sk Abidhussain[3]

MCA Student, Dept. of MCA, Narayana Engineering College, Nellore, A.P, India[1]

MCA Student, Dept. of MCA, Narayana Engineering College, Nellore, A.P, India[2]

Assistant Professor, Dept. of MCA, Narayana Engineering College, Nellore, A.P, India[3]

ABSTRACT: As an effective and efficient way to provide computing resources and services to customers on demand, cloud computing has become more and more popular. From cloud service providers' perspective, profit is one of the most important considerations, and it is mainly determined by the configuration of a cloud service platform under given market demand. However, a single long-term renting scheme is usually adopted to configure a cloud platform, which cannot guarantee the service quality but leads to serious resource waste. In this paper, a double resource renting scheme is designed firstly in which short-term renting and long-term renting are combined aiming at the existing issues. This double renting scheme can effectively guarantee the quality of service of all requests and reduce the resource waste greatly. Secondly, a service system is considered as an *M/M/m+D* queuing model and the performance indicators that affect the profit of our double renting scheme are analyzed, e.g., the average charge, the ratio of requests that need temporary servers, and so forth. Thirdly, a profit maximization problem is formulated for the double renting scheme and the optimized configuration of a cloud platform is obtained by solving the profit maximization problem. Finally, a series of calculations are conducted to compare the profit of our proposed scheme with that of the single renting scheme. The results show that our scheme can not only guarantee the service quality of all requests, but also obtain more profit than the latter.

KEYWORDS: Cloud computing, guaranteed service quality, multiserver system, profit maximization, queuing model, service-level agreement, waiting time.

I. INTRODUCTION

AS an effective and efficient way to consolidate computing resources and computing services, clouding computing has become more and more popular [1]. Cloud computing centralizes management of resources and services, and delivers hosted services over the Internet. The hard-ware, software, databases, information, and all resources are concentrated and provided to consumer's on-demand [2]. Cloud computing turn's information technology into ordinary commodities and utilities by the pay-per-use pricing model [3, 4, 5]. In a cloud computing environment, there are always three tiers, i.e., infrastructure providers, services providers, and customers (see Fig. 1 and its elaboration in Section 3.1). An infrastructure provider maintains the basic hardware and software facilities. A service provider rents resources from the infrastructure providers and provides services to customers. A customer submits its request to a service provider and pays for it based on the amount and the quality of the provided service [6]. In this paper, we aim at researching the multiserver configuration of a service provider such that its profit is maximized.

Like all business, the profit of a service provider in cloud computing is related to two parts, which are the cost and the revenue. For a service provider, the cost is the renting cost paid to the infrastructure providers plus the electricity cost caused by energy consumption, and the revenue is the service charge to customers. In general, a service provider rents a certain number of servers from the infrastructure providers and builds different multiserver systems for different application domains. Each multiserver system is to execute a special type of service requests and applications. Hence, the renting cost is proportional to the number of servers in a multiserver system [2]. The power consumption of a multiserver system is linearly proportional to the number of servers and the server utilization, and to the square of

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

execution speed [7, 8]. The revenue of a service provider is related to the amount of service and the quality of service. To summarize, the profit of a service provider is mainly determined by the configuration of its service platform.

To configure a cloud service platform, a service provider usually adopts a single renting scheme. That's to say, the servers in the service system are all long-term rented. Be-cause of the limited number of servers, some of the incoming service requests cannot be processed immediately. So they are first inserted into a queue until they can handle by any available server. However, the waiting time of the service requests cannot be too long. In order to satisfy quality-of-service requirements, the waiting time of each incoming service request should be limited within a certain range, which is determined by a service-level agreement (SLA). If the quality of service is guaranteed, the service is fully charged, otherwise, the service provider serves the request for free as a penalty of low quality. To obtain higher revenue, a service provider should rent more servers from the infrastructure providers or scale up the server execution speed to ensure that more service requests are processed with high service quality. However, doing this would lead to sharp increase of the renting cost or the electricity cost. Such increased cost may counterweight the gain from penalty reduction. In conclusion, the single renting scheme is not a good scheme for service providers. In this paper, we propose a novel renting scheme for service providers, which not only can satisfy quality-of-service requirements, but also can obtain more profit. Our contributions in this paper can be summarized as follows.

II. EXISTING SYSTEM

In general, a service provider rents a certain number of servers from the infrastructure providers and builds different multi-server systems for different application domains. Each multiserver system is to execute a special type of service requests and applications. Hence, the renting cost is proportional to the number of servers in a multiserver system. The power consumption of a multiserver system is linearly proportional to the number of servers and the server utilization, and to the square of execution speed. The revenue of a service provider is related to the amount of service and the quality of service. To summarize, the profit of a service provider is mainly determined by the configuration of its service platform.

To configure a cloud service platform, a service provider usually adopts a single renting scheme. That's to say, the servers in the service system are all long-term rented. Because of the limited number of servers, some of the incoming service requests cannot be processed immediately. So they are first inserted into a queue until they can handle by any available server.

III. PROPOSED SYSTEM

- In this paper, we propose a novel renting scheme for service providers, which not only can satisfy quality-of-service requirements, but also can obtain more profit.
- A novel double renting scheme is proposed for service providers. It combines long-term renting with short-term renting, which can not only satisfy quality-of-service requirements under the varying system workload, but also reduce the resource waste greatly.
- A multiserver system adopted in our paper is modeled as an $M/M/m+D$ queuing model and the performance indicators are analyzed such as the average service charge, the ratio of requests that need short term servers, and so forth.
- The optimal configuration problem of service providers for profit maximization is formulated and two kinds of optimal solutions, i.e., the ideal solutions and the actual solutions, are obtained respectively.
- A series of comparisons are given to verify the performance of our scheme. The results show that the proposed Double-Quality-Guaranteed (DQG) renting scheme can achieve more profit than the compared Single-Quality-Unguaranteed (SQU) renting scheme in the premise of guaranteeing the service quality completely.

ADVANTAGES OF PROPOSED SYSTEM:

- ✓ Since the requests with waiting time D are all assigned to temporary servers, it is apparent that all service requests can guarantee their deadline and are charged based on the workload according to the SLA. Hence, the revenue of the service provider increases.
- ✓ Increase in the quality of service requests and maximize the profit of service providers.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

This scheme combines short-term renting with long-term renting, which can reduce the resource waste greatly and adapt to the dynamical demand of computing capacity.

IV. SYSTEM MODELS

In this section, we first describe the three-tier cloud computing structure. Then, we introduce the related models used in this paper, including a multiserver system model, a revenue model, and a cost model.

A. A Cloud System Model

The cloud structure (see Fig. 1) consists of three typical parties, i.e., infrastructure providers, service providers and customers. This three-tier structure is used commonly in existing literatures [2, 6, 10].

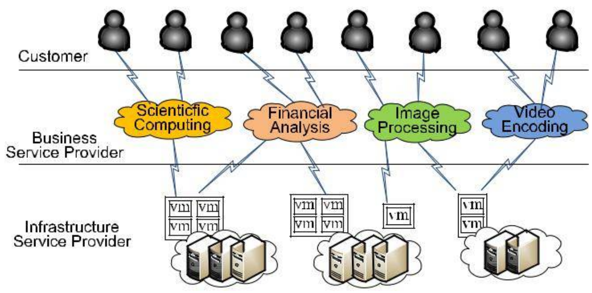


Fig. 1: The three-tier cloud structure.

In the three-tier structure, an infrastructure provider the basic hardware and software facilities. A service provider rents resources from infrastructure providers and prepares a set of services in the form of virtual machine (VM). Infrastructure providers provide two kinds of resource renting schemes, e.g., long-term renting and short-term renting. In general, the rental price of long-term renting is much cheaper than that of short-term renting. A customer submits a service request to a service provider which delivers services on demand. The customer receives the desired result from the service provider with certain service-level agreement, and pays for the service based on the amount of the service and the service quality. Service providers pay infrastructure providers for renting their physical resources, and charge customers for processing their service requests, which generates cost and revenue, respectively. The profit is generated from the gap between the revenue and the cost.

B. A Multiserver Model

In this paper, we consider the cloud service platform as a multiserver system with a service request queue. Fig. 2 gives the schematic diagram of cloud computing [32].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

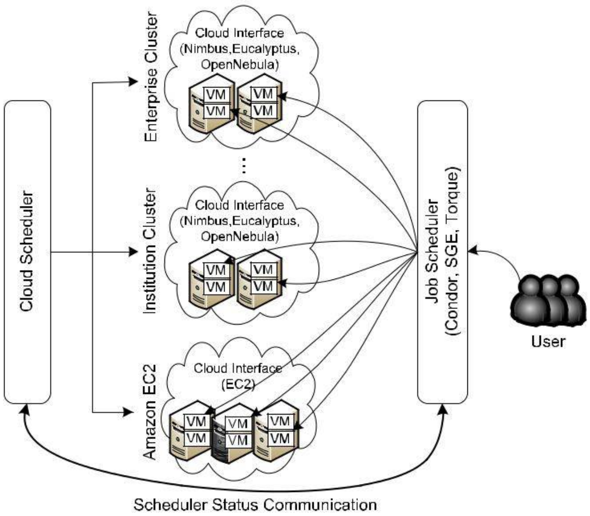


Fig. 2: The schematic diagram of cloud computing.

In an actual cloud computing platform such as Amazon EC2, IBM blue cloud, and private clouds, there are many work nodes managed by the cloud managers such as Eucalyptus, Open Nebula, and Nimbus. The clouds provide resources for jobs in the form of virtual machine (VM). In addition, the users submit their jobs to the cloud in which a job queuing system such as SGE, PBS, or Condor is used. All jobs are scheduled by the job scheduler and assigned to different VMs in a centralized way. Hence, we can consider it as a service request queue. For example, Condor is a specialized workload management system for compute-intensive jobs and it provides a job queuing mechanism, scheduling policy, priority scheme, resource monitoring, and resource management. Users submit their jobs to Con-dor, and Condor places them into a queue, chooses when and where to run them based upon a policy [33, 34]. Hence, it is reasonable to abstract a cloud service platform as a multiserver model with a service request queue, and the model is widely adopted in existing literature [2, 11, 35, 36, 37].

V. IMPLEMENTATION

1. MODULES DESCRIPTION

1.1. Service providers module:

A service provider rents resources from infrastructure providers and prepares a set of services in the form of virtual machine (VM). Infrastructure providers provide two kinds of resource renting schemes, e.g., long-term renting and short-term renting. In general, the rental price of long-term renting is much cheaper than that of short-term renting. A customer submits a service request to a service provider which delivers services on demand. The customer receives the desired result from the service provider with certain service-level agreement.

1.2. Infrastructure provider's module:

In an actual cloud computing platform such as Amazon EC2, IBM blue cloud, and private clouds, there are many work nodes managed by the cloud managers such as Eucalyptus, Open Nebula, and Nimbus. The clouds provide resources for jobs in the form of virtual machine (VM). In addition, the users submit their jobs to the cloud in which a job queuing system such as SGE, PBS, or Condor is used.
In the most basic cloud-service model - and according to the IETF (Internet Engineering Task Force) - providers of IaaS offer computers – physical or (more often) virtual machines – and other resources. IaaS refers to online services that abstract user from the detail of infrastructure like physical computing resources, location, data partitioning, scaling, security, backup etc.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

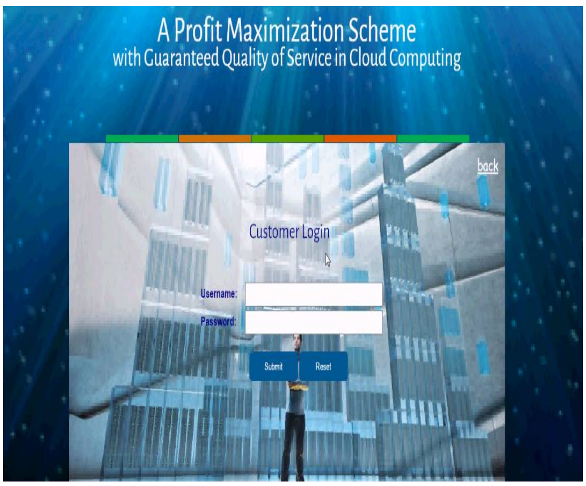
1.3. Customers module:
A customer submits a service request to a service provider which delivers services on demand. The customer receives the desired result from the service provider with certain service-level agreement. The customer rent the two types of renting scheme viz long term and short term renting. The revenue model is determined by the pricing strategy and the server-level agreement (SLA). In this paper, the usage-based pricing strategy is adopted, since cloud computing provides services to customers and charges them on demand. The SLA is a negotiation between service providers and customers on the service quality and the price. Because of the limited servers, the service requests that cannot be handled immediately after entering the system must wait in the queue until any server is available. However, to satisfy the quality-of-service requirements, the waiting time of each service request should be limited within a certain range which is determined by the SLA. The SLA is widely used by many types of businesses, and it adopts a price compensation mechanism to guarantee service quality and customer satisfaction.

1.4. Queing Model:
When the incoming service requests cannot be processed immediately after they arrive, they are firstly placed in the queue until they can be handled by any available server. The first-come-first-served (FCFS) queuing discipline is adopted. Because the fixed computing capacity of the service system is limited, some requests would wait for a long time before they are served. According to the queuing theory, we have the following theorem about the waiting time in an M/M/m queuing system.

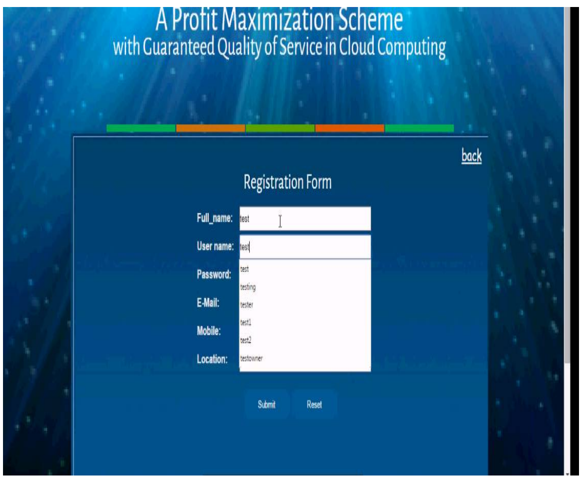
1.5. Double Renting Scheme:
It combines long-term renting with short-term renting, which can not only satisfy quality-of-service requirements under the varying system workload, but also reduce the resource waste greatly. The Double-Quality Guaranteed (DQG) resource renting scheme which combines long-term renting with short-term renting. The main computing capacity is provided by the long-term rented servers due to their low price. The short-term rented servers provide the extra capacity in peak period. The requests are assigned and executed on the long-term rented servers in the order of arrival times.

VI. RESULTS AND EXPERIMENTAL STUDIES

In this paper, to overcome the shortcomings mentioned above, a double renting scheme is designed to configure a cloud service platform, which can guarantee the service quality of all requests and reduce the resource waste greatly. Moreover, a profit maximization problem is formulated and solved to get the optimal multiserver configuration which can product more profit than the optimal configuration in [2].



1. Customer Login



2. Registration Form

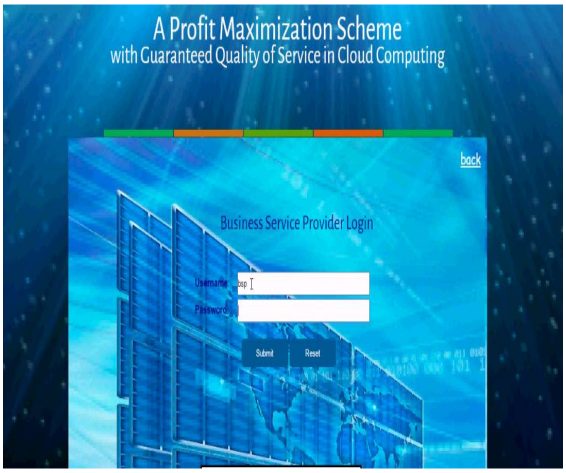
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

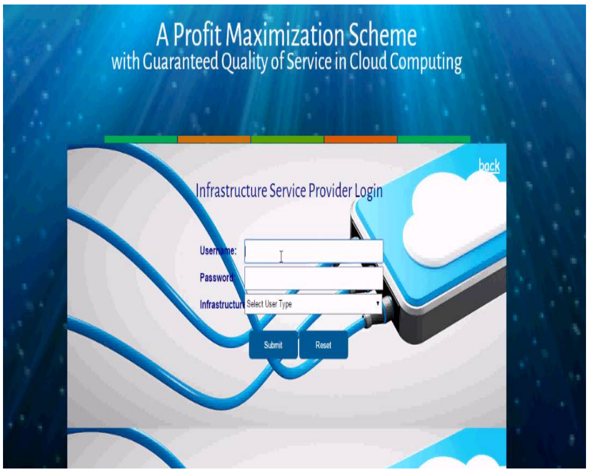
(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017



3. Business Service Provider Login



4. Infrastructure Service Provider Login

VII. Conclusion and Future Work

In order to guarantee the quality of service requests and maximize the profit of service providers, this paper has proposed a novel Double-Quality-Guaranteed (DQG) renting scheme for service providers. This scheme combines short-term renting with long-term renting, which can reduce the resource waste greatly and adapt to the dynamical demand of computing capacity. An *M/M/m+D* queuing model is build for our multiserver system with varying system size. And then, an optimal configuration problem of profit maximization is formulated in which many factors are taken into considerations, such as the market demand, the workload of requests, the server-level agreement, the rental cost of servers, the cost of energy consumption, and so forth. The optimal solutions are solved for two different situations, which are the ideal optimal solutions and the actual optimal solutions. In addition, a series of calculations are conducted to compare the profit obtained by the DQG renting scheme with the Single-Quality-Unguaranteed (SQU) renting scheme. The results show that our scheme outperforms the SQU scheme in terms of both of service quality and profit.

References

[1] K. Hwang, J. Dongarra, and G. C. Fox, *Distributed and Cloud Computing*. Elsevier/Morgan Kaufmann, 2012.
[2] J. Cao, K. Hwang, K. Li, and A. Y. Zomaya, "Optimal multiserver configuration for profit maximization in cloud computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 6, pp. 1087–1096, 2013.
[3] A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and I. Stoica, "Above the clouds: A berkeley view of cloud computing," *Dept. Electrical Eng. and Comput. Sciences*, vol. 28, 2009.
[4] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Gener. Comp. Sy.*, vol. 25, no. 6, pp. 599– 616, 2009.
[5] P. Mell and T. Grance, "The NIST definition of cloud com-puting. national institute of standards and technology," *Information Technology Laboratory*, vol. 15, p. 2009, 2009.
[6] J. Chen, C. Wang, B. B. Zhou, L. Sun, Y. C. Lee, and A. Y. Zomaya, "Tradeoffs between profit and customer satisfaction for service provisioning in the cloud," in *Proc. 20th Int'l Symp. High Performance Distributed Computing*. ACM, 2011, pp. 229–238.
[7] J. Mei, K. Li, J. Hu, S. Yin, and E. H.-M. Sha, "Energy-aware preemptive scheduling algorithm for sporadic tasks on dvs platform," *MICROPROCESS MICROSY.*, vol. 37, no. 1, pp. 99–112, 2013.
[8] P. de Langen and B. Juurlink, "Leakage-aware multipro-cessor scheduling," *J. Signal Process. Sys.*, vol. 57, no. 1, pp. 73–88, 2009.
[9] G. P. Cachon and P. Feldman, "Dynamic versus static pricing in the presence of strategic consumers," Tech. Rep., 2010.
[10] Y. C. Lee, C. Wang, A. Y. Zomaya, and B. B. Zhou, "Profit-driven scheduling for cloud services with data access awareness," *J. Parallel Distr. Com.*, vol. 72, no. 4, pp. 591– 602, 2012.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

[11] M. Ghamkhari and H. Mohsenian-Rad, "Energy and per-formance management of green data centers: a profit max-imization approach," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 1017–1025, 2013.
[12] A. Odlyzko, "Should flat-rate internet pricing continue," *IT Professional*, vol. 2, no. 5, pp. 48–51, 2000.
[13] G. Kesidis, A. Das, and G. de Veciana, "On flat-rate and usage-based pricing for tiered commodity internet services," in *42nd Annual Conf. Information Sciences and Systems*. IEEE, 2008, pp. 304–308.
[14] S. Shakkottai, R. Srikant, A. Ozdaglar, and D. Acemoglu, "The price of simplicity," *IEEE J. Selected Areas in Commu-nications*, vol. 26, no. 7, pp. 1269–1276, 2008.
[15] H. Xu and B. Li, "Dynamic cloud pricing for revenue maximization," *IEEE Trans. Cloud Computing*, vol. 1, no. 2, pp. 158–171, July 2013.
[16] M. Macıas´ and J. Guitart, "A genetic model for pricing in cloud computing markets," in *Proc. 2011 ACM Symp. Applied Computing*, 2011, pp. 113–118.
[17] "Amazon EC2," http://aws.amazon.com, 2014.
[18] "Amazon EC2 spot instances," http://aws.amazon.com/ ec2/spot-instances, 2014.
[19] R. L. Hall and C. J. Hitch, "Price theory and business behaviour," *Oxford economic papers*, no. 2, pp. 12–45, 1939.
1. D. Kahneman, J. L. Knetsch, and R. Thaler, "Fairness as a constraint on profit seeking: Entitlements in the market,"

BIOGRAPHY

Mr. Shaik Abid Hussain is an Assistant Professor of MCA at Narayana Engineering College, Nellore, A.P, India. He had done his Post Graduation at Audisankara College of Engineering and Technology, Gudur, A.P, India. He guided many projects for PG students. His research interests include Data mining, Computer Networks and Software Engineering etc.

Ms. CH. Padmaja is a student pursuing MCA at Narayana Engineering College, Nellore, A.P, India. During my final semester project work we prepared this paper for publishing it in an international journal.

Ms. V. Divya is a student pursuing MCA at Narayana Engineering College, Nellore, A.P, India. During my final semester project work we prepared this paper for publishing it in an international journal.