Exploring Resource Provisioning Cost Models in Cloud Computing

P.Aradhya[#1], K.Shivaranjani[*2]

[#1]M.Tech, CSE, SR Engineering College, Warangal, Andhra Pradesh, India
[#]Assistant Professor, Department of CSE, SR Engineering College, Warangal, Andhra Pradesh, India

Abstract--Cloud computing services are being provided in pay per use fashion. The cloud consumers are given two ways in which they can obtain computing resources. They are on-demand and reservation plans. The reservation plan is relatively cheaper than that of on-demand. In case of reservation plan the cloud consumer has to pay money in advance. However, making reservations for computing resources in well advance has some problems to consumers and also service providers due to the optimization problem. Optimization of resource provisioning cost is essential to make is suitable for both cloud service provides and cloud service consumers. Recently Chaisiri et al. proposed an algorithm that will take care of optimized resource provisioning cost. In this paper we built a prototype application that practically implements the algorithm and various approaches for resource provisioning. The empirical results revealed that the algorithm is very useful.

Index Terms – Cloud computing, virtualization, resource provisioning

I. INTRODUCTION

Cloud computing is a new approach is sharing computing resources. In fact it is the technology that helps individuals and organizations to gain access to huge computing resources in pay as you use fashion. The cloud services are accessed through Internet. There are many cloud service providers such as Microsoft, IBM, Google, Amazon and so on. Cloud provides various kinds of services such as Platform as a Service (PaaS), Software as a Service (SaaS), and Infrastructure as a Service (IaaS). These service models are provided by all service providers. The cloud also has many deployment models. They include private cloud, public cloud, community cloud and hybrid cloud [1]. The computing resources provided by cloud service providers include network bandwidth, software, storage, processing power and so on. The Infrastructure as a Service is the model which is computing service model which is widely being used. The cloud computing is implemented on top of other technology known as virtualization. With cloud computing services, the consumers can specify the kind of services they need and utilize the services in pay per use fashion. Virtual Machines (VMs) are used in cloud computing to provide cloud services to in affordable prices. The pricing of services play an important role in the future of cloud services. The pricing influences the consumers in the usage of cloud services. The consumers can specify both hardware and software resources in pay per use fashion. The computing resources are actually maintained by service providers and therefore it is possible to try to reduce cost of ownership of resources to cloud consumers.

The consumers of cloud computing resources can have two plans namely reservation and on-demand. The reservation plan is cheaper than on-demand plan. Generally the on-demand plans are pertaining to reservation and on-demand plan is of short term in nature. For instance in the real world both the plans are supported by GoGrid [2] and also EC2 [3] with respect to IaaS services. The advantage with on-demand plan is that the consumers can divide the required resources on demand based on their needs. It is in pay per use fashion. However, the reservation plan is given with prior payment for a period of time. Reservation plans generally reduce pricing. For instance Amazon's EC2 reservation plans can save money up to 46 percent. Both resource provisioning plans cause problems to consumers and service providers when the pricing process is not optimized. For instance, when more resources are reserved in advance, there might be scarcity for on-demand users. Keeping this in mind in this paper, we implement an algorithm that ensures optimal resource provisioning pricing. We built a prototype which demonstrates the proof of concept. The empirical results reveal that the algorithm is effective in reducing cost of ownership to cloud consumers and at the same time improve profitability to cloud service providers. The remainder of this paper is structured as follows. Literature review is made in section II. Section III provides the proposed algorithm and implementation. Section IV provides prototype implementation details. Section V presents experimental results while section VI concludes the paper.

II. RELATED WORKS

Resource provisioning is essential in cloud computing as the consumers are supported avail cloud services in pay per use fashion. For this reason there must be plans for resource provisioning prices. Earlier many researchers focused on the resource provisioning in distributed systems as explored in [4], [5], [6], [7], and [8]. On-demand service provisioning and profile based approaches were discussed. Resource slot concept is presented in [6]. Experiments were made to maximize revenues to cloud service providers by increasing utilization of cloud services [7]. However, these researchers ignored the future uncertainty of consumer demands. In [8] it is explored to some extent including QoS with workloads under uncertainty. For predicting demand of resources K-nearest neighbors algorithm is used in [9]. Probability distributions concept is used in this paper for resource provisioning pricing. Virtualization technology made the cloud computing easier. However many researches came into existence on placement of VMs [10], [11], [12], [13], [14].

In [15] Optimal Virtual Machine Placement (OVMP) algorithm was proposed for optimal solution for both VM placement and also the resource provisioning. Motivated by this work, in [16] an algorithm is proposed for optimal resource provisioning plans for cloud users. This algorithm uses various strategies in order to implement a roust plan that can reduce cost of ownership to cloud consumers while improving utilization of cloud thus making it beneficiary to cloud service providers.

III. PROPOSED RESOURCE PROVISIONING MODEL

The proposed resource provisioning model assume cloud computing environment as shown in figure 1. The aim of the model is to optimize resource provisioning and pricing in terms of both reservation and on-demand models. The reservation model and on-demand resource provisioning models are integrated into a single framework that are governed by resource provisioning algorithm which ensures reduction of cost to consumers and increasing profits to cloud service providers.

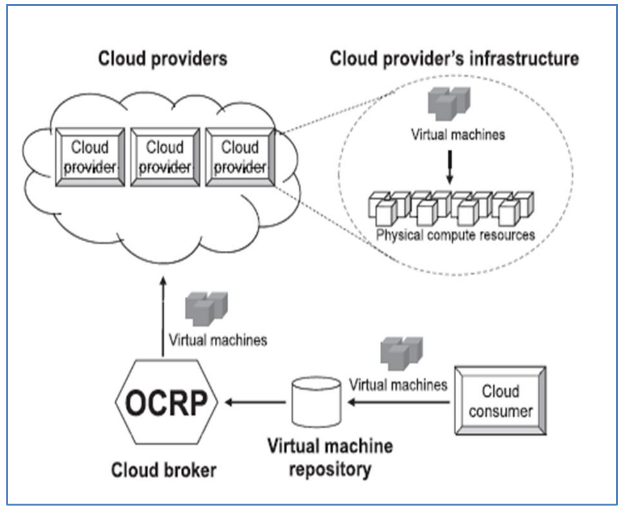

Fig. 1 – Overview of the system model

As can be seen in figure 1, the system model include cloud service providers, the cloud providers' infrastructure, the proposed algorithm running in virtual machines, VM repository and cloud consumers. The algorithm uses different set of virtual machines.

Provisioning Plans

There are two provisioning plans supported by cloud service providers. They are known as reservation and on-demand. The cloud broker has to determine the plan to be chosen. When plan is taken in advance, its provisioning cost gets reduced [3], [17]. The on-demand plan is considered to be on-demand plan while the reservation is considered medium to long term plan.

Provisioning Phases

These are the time intervals in which the cloud broker makes decisions regarding plans. These are known as provisioning phases. They are known as reservation, expending and on-demand phases. Reservation plan is done in advance. In the expending phase the resources are utilized as per the plan. When cloud broker pays additional money for resources then on-demand phase starts.

Provisioning Costs

The provisioning costs are of three types. They are known as reservation, expending and on-demand costs. The objective of the proposed algorithm is to reduce the cost of all kinds of provisioning. This will help consumers to get services in affordable prices while it brings more consumers to cloud providers thus making it profitable. As per the algorithm, the reservation cost is computed as follows.

$$c_{ijk}^{(\mathrm{R})} = \sum_{r \in \mathcal{R}} b_{ir} c_{jkr}^{(\mathrm{R})}.$$

Stochastic Integer Programming Model

In order to reduce the cost of provisioning of resource on the part of cloud provider, objective function is computed. The purpose of the objective function is to minimize provisioning cost. The objective function is computed as follows.

$$z = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} c_{ijk}^{(\mathrm{R})} x_{ijk}^{(\mathrm{R})} + \mathbb{E}_{\Omega}\left[\mathcal{Q}\left(x_{ijk}^{(\mathrm{R})}, \omega\right)\right],$$

subject to:

$$x_{ijk}^{(\mathrm{R})} \in \mathbb{N}_0, \quad \forall i \in \mathcal{I}, \forall j \in \mathcal{J}, \forall k \in \mathcal{K}.$$

Benders Decomposition

In order to break down the stochastic problem with respect to objective function of the algorithm, benders decomposition algorithm is used [11]. It will reduce the time required to bring about a solution. As per this algorithm the flow is as shown in figure 2.

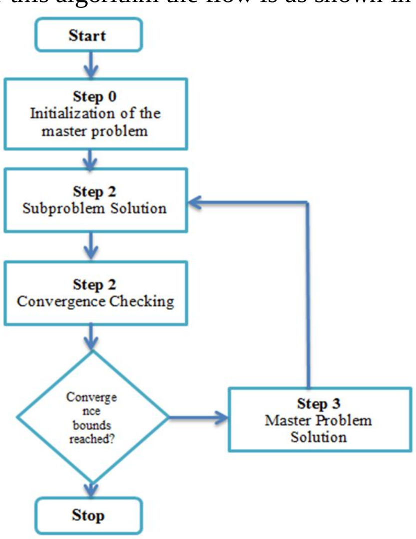

Fig. 2 – Flowchart of benders decomposition algorithm

As can be seen in figure 2, it is evident that the benders decomposition takes the main problem and decomposes it further to smaller problems to get fast convergence in terms of providing best solution. In this paper we used it to break down the resource provisioning algorithm for efficiency and in order to reduce the provisioning cost. In step1 multiple sub problem solutions are worked out and solved. Convergence checking is made in step 2. Then finally the master problem solution is found. More technical details of algorithm can be found in [29].

IV. PROTOTYPE IMPLEMENTATION

We built a prototype application to demonstrate the efficiency of the algorithm proposed. The platform used for the implementation is Microsoft .NET. The programming language used is C#. The application is to implement the proof of concept. The environment used for the development of application is a PC with 4 GB RAM, core 2 dual processor running Windows 7 operating system.

V. EXPERIMENTAL RESULTS

Experiments are made with our prototype implementation of resource provisioning algorithm. The experiments are made in terms of service provider, price per VM in reservation phase, unit price in expending phase, and unit price in on-demand phase. It also considers the processing time, storage, and network. The experimental results are shown in table 1.

Provider	Price per VM in reservation phase			Unit price in expending phase			Unit price in on-demand phase		
	3-M	6-M	1-Y	Processing time	Storage	Network	Processing time	Storage	Network
J_1	N/A	N/A	\$357	\$0	\$0	\$0.10	N/A	N/A	N/A
J_2	\$56.90	\$87.63	\$227.50	\$0.03	\$0.10	\$0.15	\$0.085	\$0.10	\$0.15
J_3	\$69.38	\$106.85	\$277.50	\$0.02	\$0.10	\$0.15	\$0.10	\$0.14	\$0.19
J_4	N/A	N/A	N/A	N/A	N/A	N/A	\$0.09	\$0.075	\$0.15

Table 1 – Experimental results

As can be seen in table 1, the experimental results are presented. The cost of on-demand phase is more than that of other phases. The following series of graphs visualize the results of experiments.

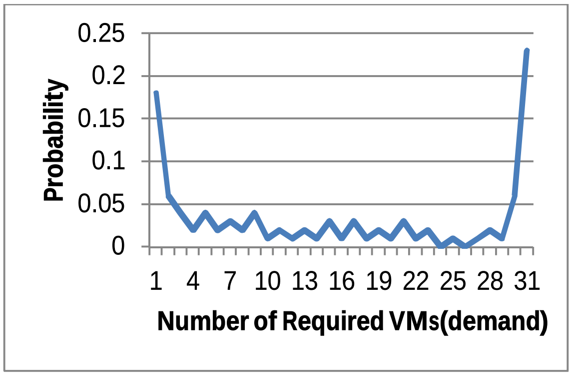

Fig 3 The probability distribution of the real data.

As shown in the above figure 3 represents the horizontal axis represents number of required VMs while vertical axis represents Probability.

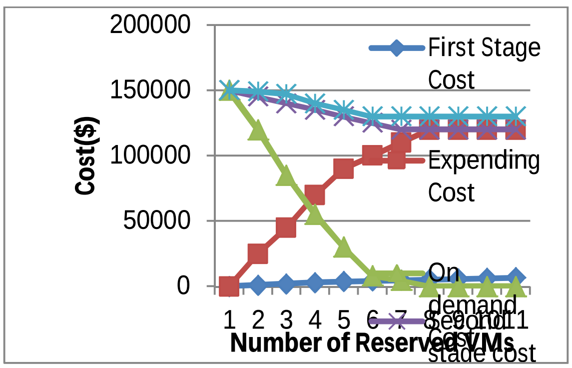

Fig. 4. The optimal solution in a simple cloud computing environment.

As shown in the above figure 4 represents the horizontal axis represents number of Reserved VMs while vertical axis represents cost.

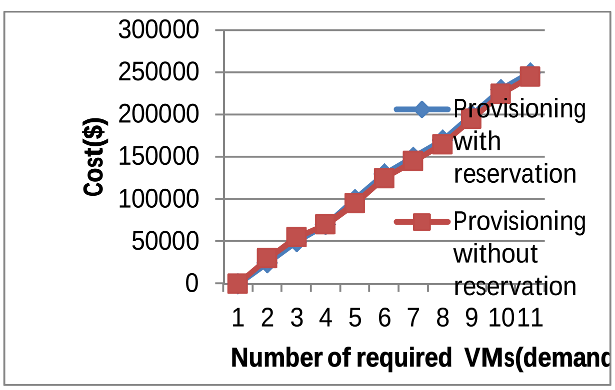

Fig. 5. Comparison between total costs of resource provision with and without reservation.

As shown in the above figure 5 represents the horizontal axis represents number of Reserved VMs while vertical axis represents cost.

VI. CONCLUSION

In this paper we implement an algorithm that is meant for resource provisioning to cloud consumers. The cloud consumers are provided two plans for resource provisioning namely reservation and on-demand. The resource provisioning is uncertain in both plans when the consumers are not certain about their future demands and the service providers can't give indication of future prices. For this reason optimization of provisioning of computing resources in cloud computing is the possible solution. Towards this end in this paper we implement an algorithm that makes use of various resource provisioning strategies to ensure that the resource provisioning is done optimally that helps both consumers and service providers. We built a prototype application that demonstrates the proof of concept. The empirical results revealed that the application is useful for optimal resource provisioning.

REFERENCES

[1] I. Foster, Y. Zhao, and S. Lu, "Cloud Computing and Grid Computing 360-Degree Compared," Proc. Grid Computing Environments Workshop (GCE '08), 2008.
[2] GoGrid, http://www.gogrid.com, 2012.
[3] Amazon EC2, http://aws.amazon.com/ec2, 2012.
[4] Z. Huang, C. He, and J. Wu, "On-Demand Service in Grid: Architecture Design, and Implementation," Proc. 11th Int'l Conf. Parallel and Distributed Systems (ICPADS '05), 2005.
[5] Y. Jie, Q. Jie, and L. Ying, "A Profile-Based Approach to Just-in- Time Scalability for Cloud Applications," Proc. IEEE Int'l Conf. Cloud Computing (CLOUD '09), 2009.
[6] Y. Kee and C. Kesselman, "Grid Resource Abstraction, Virtualization, and Provisioning for Time-Target Applications," Proc. IEEE Int'l Symp. Cluster Computing and the Grid, 2008.
[7] A. Filali, A.S. Hafid, and M. Gendreau, "Adaptive Resources Provisioning for Grid Applications and Services," Proc. IEEE Int'l Conf. Comm., 2008.
[8] D. Kusic and N. Kandasamy, "Risk-Aware Limited Lookahead Control for Dynamic Resource Provisioning in Enterprise Computing Systems," Proc. IEEE Int'l Conf. Autonomic Computing, 2006.
[9] J. Chen, G. Soundararajan, and C. Amza, "Autonomic Provisioning of Backend Databases in Dynamic Content Web Servers," Proc. IEEE Int'l Conf. Autonomic Computing, 2006.
[10] L. Grit, D. Irwin, A. Yumerefendi, and J. Chase, "Virtual Machine Hosting for Networked Clusters: Building the Foundations for Autonomic Orchestration," Proc. IEEE Int'l Workshop Virtualization Technology in Distributed Computing, 2006.
[11] H.N. Van, F.D. Tran, and J.-M. Menaud, "SLA-Aware Virtual Resource Management for Cloud Infrastructures," Proc. IEEE Ninth Int'l Conf. Computer and Information Technology, 2009.
[12] M. Cardosa, M.R. Korupolu, and A. Singh, "Shares and Utilities Based Power Consolidation in Virtualized Server Environments," Proc. IFIP/IEEE 11th Int'l Conf. Symp. Integrated Network Management (IM '09), 2009.
[13] F. Hermenier, X. Lorca, and J.-M. Menaud, "Entropy: A Consolidation Manager for Clusters," Proc. ACM SIGPLAN/ SIGOPS Int'l Conf. Virtual Execution Environments (VEE '09), 2009.

[14] N. Bobroff, A. Kochut, and K. Beaty, "Dynamic Placement of Virtual Machines for Managing SLA Violations," Proc. IFIP/IEEE Int'l Symp. Integrated Network Management (IM '07), pp. 119-128, May 2007.
[15] S. Chaisiri, B.S. Lee, and D. Niyato, "Optimal Virtual Machine Placement across Multiple Cloud Providers," Proc. IEEE Asia- Pacific Services Computing Conf. (APSCC), 2009.
[16] Sivadon Chaisiri, Bu-Sung Lee and Dusit Niyato, "Optimization of Resource Provisioning Cost in Cloud Computing". IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL. 5, NO. 2, APRIL-JUNE 2012.
[17] Amazon EC2 Reserved Instances, http://aws.amazon.com/ec2/reserved-instances, 2012.

AUTHORS

P.Aradhya, She is pursuing M.Tech (CSE) in SR Engineering College, Warangal, AP, INDIA. She has received B.Tech degree in Information technology. Her main research interest includes Cloud Computing .

K.Shivaranjani is working as an Assistant Professor in SR Engineering College, JNTUH, Warangal, Andhra Pradesh, India. She has completed M.Tech from JNTUH. His main research interest includes Cloud Computing .