Cost Optimization in Cloud Services

N Latha
Assistant Professor
SSS Shasun Jain College for Women
T Nagar,Chennai-17.

S T Deepa. Ph.D.
Head ,Department of Computer Science
SSS Shasun Jain College for Women
T Nagar,Chennai-17

ABSTRACT
Cloud computing is a large distributed computing environment in which collection of resources is available to user through internet. It helps the consumers to reduce the cost of information management as they are not required to own their servers and can use capacity leased from cloud service providers. At present, cloud service providers provide two provisioning plans for computing resources namely reservation and on-demand plans. The main goal of this research work is to minimize the total cost of resource provisioning by reducing the on-demand cost and oversubscribed cost of under provisioning and over provisioning The proposed solution for minimizing the total cost of resource provisioning is implemented in a cloud that provides storage and encryption services

Keywords
Encryption, Cloud Computing, On-demand plans, Resource ,Service Provider.

1. INTRODUCTION
Cloud computing is a revolutionary technology that changed the way as computing and storage resources are acquired. If a company needs resources to deploy their applications, it is no longer necessary to make a great investment in infrastructure and resources management, because there are significant number of cloud providers that can offer a solution for specific requirements.

Cloud computing is reshaping the computing and Internet landscape. With breakthroughs made in relevant service and business models, cloud computing will inevitably expand its role as a backbone for IT services.

Cloud services are popular because they can reduce the cost and complexity of owning and operating computers and networks. Since cloud users do not have to invest in information technology infrastructure, purchase hardware, or buy software licenses, the benefits are low up-front costs, rapid return on investment, rapid deployment, customization, flexible use, and solutions that can make use of new innovations.

2. RESOURCE PROVISIONING
Cloud computing requires a provisioning method for allocating resources to cloud consumers. Cloud computing consists of two provisioning plan for allocating resources in cloud. They are: Reservation Plan and On-demand Plan.

Reservation plan is long term plan and On-demand plan is a short term plan. In On-demand plan the consumers can access resources at the time when they need. In Reservation plan the resources could be reserved earlier. Hence the cloud providers could charge the resources before consumers could use it. For on-demand pricing is done as pay-per-use basis but in reservation plan pricing is charged by one-time fee. With Reservation plan consumers could utilize the computing resources in a much cheaper amount than on-demand plan.

Even though with the reservation plan the cloud consumer could use the resources in advance some problems could occur with it. One is the under provisioning problem in which the consumers could not fully meet the required resources due to uncertainty of allocating resources. Other problem with reservation plan is over provisioning of resources, where the reserved resources will be more than what actually needed. Hence the resources reserved will not be fully used. The goal is to achieve an optimal solution for provisioning resource which is the most critical part.

3. RELATED WORKS
Dynamic resource allocation in cloud computing has attracted attention of the research community in the last few years. It is one of the most challenging problems in the resource management problems. Many researchers around the world have come up with new ways of facing this challenge.

The OCRP algorithm can provision computing resources for being used in multiple provisioning stages. The demand and price uncertainty is considered in OCRP and the different approaches to obtain the solution of the OCRP algorithm are considered including deterministic equivalent formulation, sample-average approximation and Benders decomposition [1].

The resource provisioning problem is posed as one of sequential decision making under uncertainty and solved using a limited look ahead control scheme [2]. Cloud platforms host several independent applications on a shared resource pool with the ability to allocate computing power to applications on a per-demand basis. The use of server virtualization techniques for such platforms provides great flexibility to consolidate several virtual machines on the same physical server, to resize a virtual machine capacity and to migrate virtual machine across physical servers. A key challenge for cloud providers is to automate the management of virtual servers while taking into account both high-level Quality of Service (QoS) requirements of hosted applications and resource management costs [3].

Dynamic consolidation is an approach that migrates tasks within a cluster as their computational requirements change, both to reduce the number of nodes that need to be active and to eliminate temporary overload situations. It allows Entropy, a consolidation manager for clusters to find mappings of tasks to nodes that are better than those found by heuristics based on local optimizations, and that are frequently globally optimal in the number of nodes. Because migration overhead is taken into account, Entropy chooses migrations that implemented efficiently by low performance overhead [4].

The objective was the performance of Resource Manager on the basis of resource utilization and cost in hybrid cloud environment [5]. Resource provisioning is important issue in cloud computing and in the environment of heterogeneous clouds.

4. METHODOLOGY

4.1 OCRP Algorithm*:*

The OCRP algorithm provides optimal solution for resource provisioning. It uses only two uncertainties namely demand and price. The OCRP algorithm is a derivative of stochastic programming model. Using in the multiple provisioning states the OCRP algorithm can furnish the computing resources.

The general form of stochastic integer program of the OCRP algorithm is formulated in (1) and (2). The objective function (1) is to minimize the cloud consumer’s total provisioning cost.

Subject to:

Where,

$$z = \sum_{i\in I}\sum_{j\in J}\sum_{k\in} c^{(R)}_{ijk}\, x^{(R)}_{ij} + IE_{\Omega}\,[\Omega(x^{(R)}_{ij},\omega)]$$

I - Set of virtual machine (VM) classes while iϵ I denotes the VM class index.

J - Set of cloud providers while j ϵJ denotes the cloud provider index.

K- Set of reservation contracts while k ϵ K denotes the reservation contract index.

Ω - Set of scenarios while ω ϵ Ω denotes the scenario index.

c(R) - Reservation cost subscribed to reservation contract k charged by cloud provider j to cloud consumer's VM class i in the reservation stage.

x(R)- Decision variable representing the number of VMs in class i provisioned in reservation phase subscribed to reservation contract k offered by cloud provider j in the reservation stage.

The aim of Equation (1) is to minimize the resource provisioning costs include: on demand cost, reservation cost, expected cost. In this calculation the stochastic two stage integer recourse is developed for solving complexity of resource cost optimization problems in uncertainty. In this formulation, representation IEΩ represents the decreased expected cost of resource provisioning. In Equation (2) the reservation cost represent to number of cloud provider, number of virtual machines, and set of provisioning stages.

4.2 Benders Decomposition Algorithm:

The Benders decomposition algorithm is applied to solve the stochastic programming problem. The goal of this algorithm is to break down the optimization problem into multiple smaller problems which can be solved independently and simultaneously. The Benders decomposition algorithm can decompose integer programming problems with complicating variables into two major problems as master problem and sub problem.

The algorithm contain of steps which are operated iteratively. At each iteration, the master problem signified by the complicating variables and sub problems constituted by the other decision irregulars are solved, then lower and upper bounds are estimated. The algorithm breaks when optimal solution converges, i.e., the lower and upper bounds are acceptably close to each other.

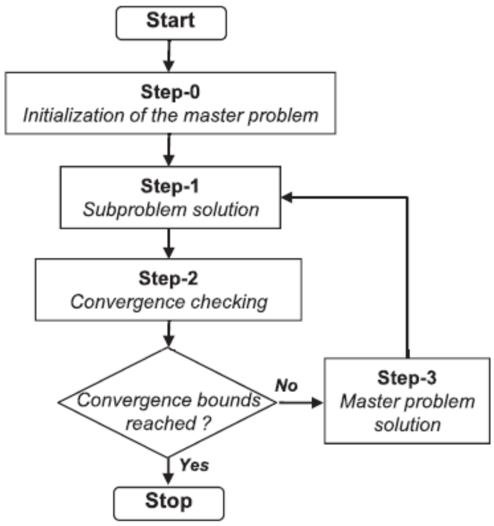

Figure 1 Flowchart showing steps in Benders Decomposition Algorithm

Step-0: Initialization of the master problem. In Step-0, the step is the initialization of the master problem. This Step-0 is performed only once, while Step-1 to Step-3 are repeatable in the algorithm

The objective function is directly derived from that represents variable in iteration of master problem, while variable provides the minimum cost given on-demand costs. This will be improved in consequent iterations. Initially, it can be fixed by constant. This can be estimated from an economic analysis or historical data of prior solutions. After solving the master problem, the algorithm proceeds to the Step-1.

Step-1: Sub problem solution. In Step-1, multiple sub problems are formulated and solved. Let’s assign the solution obtained from the master problem to variables. Given the fixed solution, sub problems, namely, S1 and S2 can be solved concurrently. The sub problem S1 is presented in which the objective function is to minimize the reservation cost. It denotes the optimal solution of the dual problem of S1 in iteration associated with constraint. The solution will be used in Step-3.

Step-2: Convergence checking. In Step-2, the convergence of lower and upper bounds of the solutions obtained from master problem and sub problems is checked. Both bounds are adjusted in each iteration. The lower bound in iteration can be obtained from the objective function of the master problem.

Step-3: Master problem solution. In addition, the solution of the master problem will adjust the cost and also the expending cost the Benders cuts are constructed from the optimal costs obtained from master problem and sub problems in the prior iterations. After solving this master problem, Step-1 is repeated and the same iterative process continues.

4.3 Proposed Work

Overview: At present, cloud service providers provide two provisioning plans for computing resources namely reservation and on-demand plans. With the reservation plan, the cloud consumers will previously reserve the resources in advance. There may be, occur under provisioning problem when the reserved resources are unable to fully meet the demand due to its uncertainty. Although this problem can be solved by ordering extra resources by the help of on-demand plan to fit the extra demand. But the cost incurred will be more due to expensive price of resource provision with on-demand plan. At the same time, the over provisioning problem can occur when the reserved resources are more than the real demand in which part of a resource pool will be under utilized.

The main goal of this research work is to minimize the total cost of resource provisioning by reducing the on-demand cost and oversubscribed cost of under provisioning and over provisioning.

Static pricing cannot guarantee cloud profit maximization. The cloud makes profit from selling its services at a price that is higher than the actual cost. Optimal pricing is achieved based on dynamic pricing scheme that adapts to time changes. This research work solves the problem of demand and price uncertainty in cloud through Optimal Cloud Resource Provisioning algorithm.

The proposed solution for minimizing the total cost of resource provisioning is implemented in a cloud that provides storage and encryption services. The resource allocation cost is dynamically fixed based on the number of users registered for the service. When more number of users registers for the service then the cost is automatically increased. At the same time if less number of users register the cost is decreased. The encryption service also considers the time factor and hence consumers can pay only for the exact duration for which they have utilized the service

5. APPLICATION ARCHITECTURE

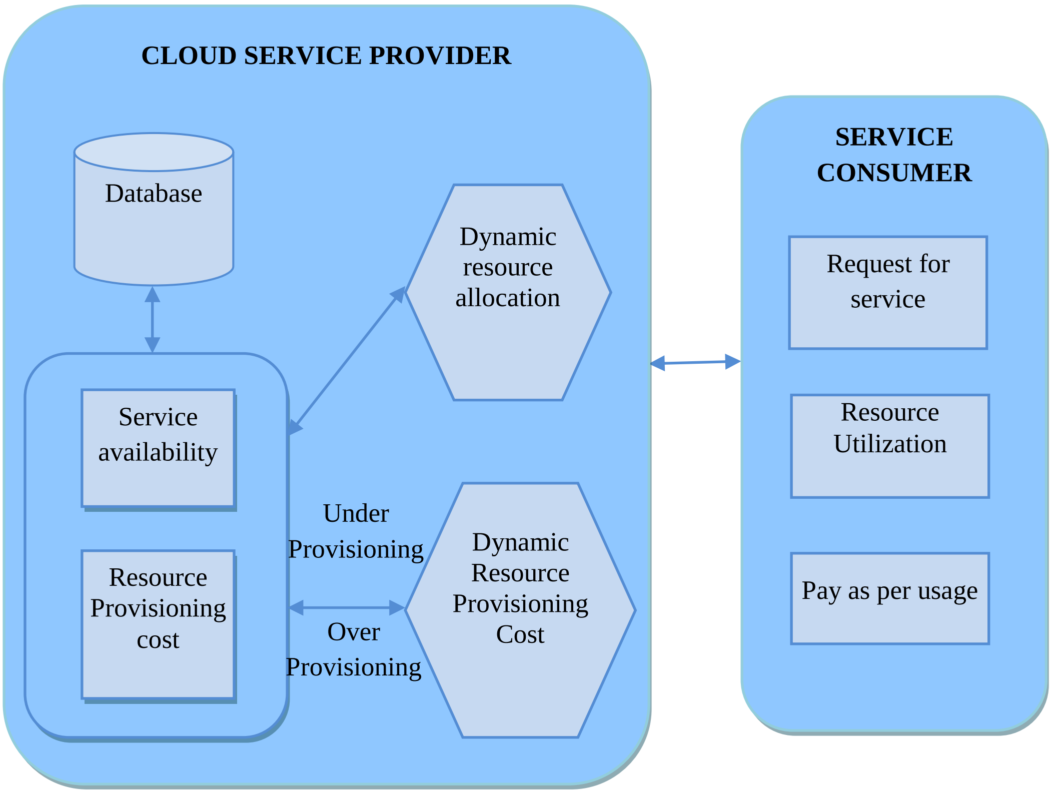

Figure 2 Application Architecture

5.1 Application Description:

In order to implement the proposed solution of minimizing the resource provisioning cost, a web application simulating the operations of cloud is developed in ASP.Net 4.0. SQL Server 2008 is used as back end for storing data. The cloud provides two types of services namely Data Storage and Encryption/Decryption Service. The home page of the application is as given below in Figure 3.

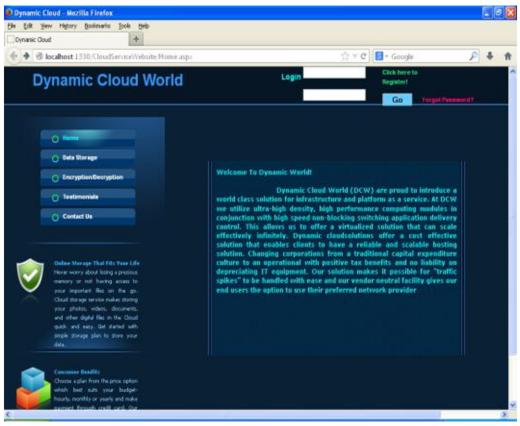

Figure 3 Home Page

The application is managed by the administrator who updates the charges for data storage, encryption and decryption charges. Users wishing to utilize the data storage and encryption/decryption service should register their personal details and then they can login to utilize the services.

In order to utilize the service users should first buy it. For Data Storage Service, user should select required storage space ranging from 1 GB to 10 GB and make payment. The validity period can be selected by the user. The storage space can be utilized by the user for a particular period which may be for 1 day or 1 year depending on the requirement.

Users can encrypt or decrypt their file use Encryption/Decryption Service. In order to utilize encryption and decryption service, users should first buy the service. Users should select required service namely Encryption or Decryption and make payment. The validity period can be selected by the user and they can utilize the service for that period.

The main aim of this research work is to overcome the problem of under provisioning and over provisioning. Hence, the cost of data storage and encryption/decryption services is not the same for all users. If less number of users have reserved the service then under provisioning occurs and in that case the cost charged for the service is the minimum. Whereas if more number of users have reserved the service then over provisioning occurs and the cost is increased which is proportionate to the number of users.

6. RESULTS AND DISCUSSION

The application is tested with number of users simultaneously buying and accessing Data Storage and Encryption/Decryption Services. The resource provisioning cost is dynamically fixed which depends on the number of users request for the service. Thus the chart given in Figure 4 shows the variation in resource provisioning cost for data storage of various sizes viz. 1 GB, 3GB, 5GB and number of users utilizing the service ranging from 2 to 27.

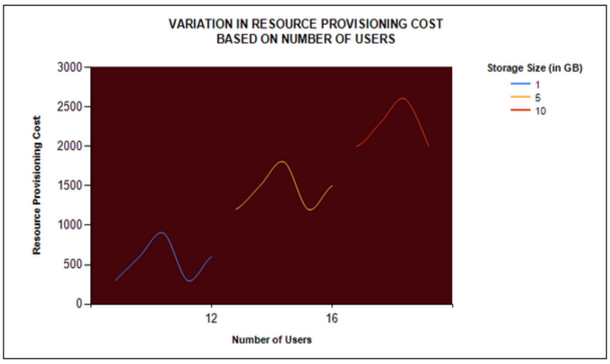

Figure 4 Resource Provisioning Chart

In this application, as more number of users request for the service the resource provisioning cost is increased proportionately and at the same time if less number of users request for the service the cost is decreased. In this way this system is quite profitable for the cloud service provider.

This application also contains the feature of auto repayment for the consumers based on their usage. For example, if a consumer has bought 5 GB for data storage for 1 day but utilized only 3 GB then the amount related to unused 2GB is refunded to the consumer's account by the end of 1 day. Thus consumers have to pay as per their usage. Thus this system is also profitable for the consumer also.

7. CONCLUSION

This research provides means of reducing cost of resource provisioning in cloud computing. As a practical implementation, a web application simulating the operations of cloud is developed in ASP.Net. The cloud provides storage and encryption services. The encryption services are done through web service. The resource allocation cost is dynamically fixed based on the number of users registered for the service. When more number of users registers for the service then the cost is automatically increased. At the same time if less number of users register the cost is decreased. The encryption service also considers the time factor and hence consumers can pay only for the exact duration for which they have utilized the service.

The success of this thesis lies in handling the problem of under provisioning and over provisioning and providing optimal resource provisioning cost. The application is beneficial to the cloud service provider as the cost is dynamically increased or decreased based on the number of users and at the same time it is beneficial to the consumer as they paying as per their usage.

8. REFERENCES

[1] SivadonChaisiri and Bu-Sung Lee, "Optimization of Resource Provisioning Cost in Cloud Computing", IEEE Transactions on Services Computing, VOL. 5, NO.2, APRIL-JUNE 2012

[2] D. Kusic and N. Kandasamy, "Risk-Aware Limited Lookahead Control for Dynamic Resource Provisioning in Enterprise Computing Systems", Proc. IEEE International Conference Autonomic Computing, 2006.

[3] H.N. Van, F.D. Tran, and J.-M. Menaud, "SLAAware Virtual Resource Management for Cloud Infrastructures", Proc. IEEE Ninth International Conference Computer and Information Technology, 2009.

[4] F. Hermenier, X. Lorca, and J.-M. Menaud, "Entropy: A Consolidation Manager for Clusters", Proc. ACM SIGPLAN/ SIGOPS International Conference Virtual Execution Environments (VEE '09), 2009

[5] RajkamalKaurGrewal&Pushpendra Kumar Pateriya, "A Rule-based Approach for Effective Resource Provisioning in Hybrid Cloud Environment", International Journal of Computer Science and Informatics ISSN (PRINT): 2231 –5292, Vol-1, Iss-4, 2012.