Optimization of Resource Provisioning in Cloud

D.Seenivasan M.E.,
Assistant Professor,
K.S.Rangasamy College of Technology,
Tiruchengode.

P.Gayathri
IV-B.E (CSE),
K.S.Rangasamy College of Technology,
Tiruchengode.

R.Anitha
IV-B.E (CSE),
K.S.Rangasamy College of Technology,
Tiruchengode.

P.Dhivya
IV-B.E (CSE),
K.S.Rangasamy College of Technology,
Tiruchengode.

ABSTRACT
Cloud computing is an emerging technology which helps us to use the resources on the fly and pay as per the usage. In case of resource provisioning there are two ways viz. On-demand subscription and Reservation scheme. Although an upfront fee is required for reservation scheme, the reservation scheme is much cheaper than On-demand subscription. But the reservation scheme also suffers from two main issues. If the allocated resources are more than the actual requirement, it leads to over provisioning which causes waste of upfront fee whereas on the other hand if the allocated resources are less than the actual requirement, it leads to under provisioning of resources. If some effective predictions are done with uncertainties from users and providers and Virtual Machines are allocated based on those predictions then these two problems can be solved to a greater extent. Main objective is to implement a repository called Virtual Machine Repository (VMR) for cloud storage such that the problems of over and under provisioning can be solved to a greater extent. It also saves the cost and time of customers.

General Terms
Cloud computing, Storage, Security, Reservation, On demand.

Keywords
VMR, Store, Virtual machine, Provider, Customer, Cost.

1. INTRODUCTION
Cloud computing is emerging and widely used technology in which a pool of computing resources is available to users. The resources could be provided to cloud consumers through Internet based on their request. One of the core technologies in cloud is Virtualization. There are mainly three services where virtualization is employed. These services include IaaS(Infrastructure as a Service), PaaS (Platorm as a Service) and SaaS (Software as a Service). One of the widely used service is Infrastructure as a service. The virtualization include virtual machine (VM) placement. Resources are requested by the Cloud consumers. The resources may be software, operating system and applications which are integrated as virtual machine. The cloud providers provide the resources according to the VM requirement. Providers are responsible for checking the quality of service. Cloud computing requires a provisioning method for allocating resources to the customers. Resource provisioning consists of two provisioning plan for allocating resources in cloud. These are long term Reservation plan and short term On-demand plan. Reservation plan is long term plan and On-demand plan is a short term plan. On-demand plan is a scheme where the consumers can access resources at the time when they need. Reservation plan is a scheme where the resources could be reserved earlier. Hence the cloud providers could charge the resources before consumers could use it. The on-demand pricing is done as pay-per-use basis but in reservation plan pricing is charged by one-time fee. In Reservation plan computing resources could be utilized by consumers in a much cheaper amount than the on-demand scheme. 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. Some other problems with reservation plan is over provisioning of resources, where the reserved resources will be more than what actually needed by the customers. Hence the resources reserved will not be fully used and leads to over provisioning. The goal is to achieve an optimal solution for provisioning resource which is the most critical part. Here, Virtual Machine Repository (VMR) is used. The customers are allocated with a VMR with the space requested by the customer. The customer can use this space until it is filled. Once the storage exceeds the allotted space, then the files are stored directly in the Cloud. Thus the time and cost of the customers are reduced to a greater extent. The problems of Over and Under provisioning are also solved to a greater extent.

2. RELATED WORK
The OCRP algorithm was proposed in [1] this paper makes use of the OCRP algorithm to make an optimal decision based on which Virtual machines can be allocated to the consumers. In this paper, different approaches to obtain the solution of the OCRP algorithm are considered including deterministic equivalent formulation, a method called sample-average approximation, and a decompositions called Benders decomposition. Extensive numerical studies are performed in which the results clearly show that with the OCRP algorithm, consumers can minimize the total cost of resource provisioning in cloud computing environments. Capacity planning and Instant VM Provisioning was proposed in [2]. In this paper, research efforts on improving the service quality for the capacity planning and instant cloud resource

provisioning problem were proposed. First both of the two problems were formulated as a generic cost-sensitive prediction problem. Then, the highly dynamic environment of cloud is considered and an asymmetric and heterogeneous measure was proposed to quantify the prediction error .Finally, an ensemble prediction mechanism was designed by combining the prediction power of a setoff prediction techniques based on the proposed measure. Anchor, a general resource management architecture that uses the stable matching framework to decouple policies from mechanisms when mapping virtual machines to physical servers was proposed in [3]. Cost optimization of elasticity and cloud resource subscription policy was proposed in [4]. In this paper a two-phase algorithm is proposed for service operators to minimize their service provision cost. Various protocols and methods for Resource allocation in cloud was proposed in [5], [6], [7], [8]. In [8] RCRP algorithm was proposed for optimization of resource provisioning. In this paper an algorithm is proposed to minimize the under provisioning and over provisioning problems under four uncertainties. Demand, price resource utilization and consumers cost are the uncertainties considered. In particular, robust computing resource provisioning (RCRP) algorithm is used in which the demand, profit, resource utilization and cost uncertainty are considered to find a robust solution. Hence the effective prediction is obtained through considering demand uncertainty of cloud consumers side, profit from cloud provider side, and the idle time and wait time uncertainty. In order to make an optimal decision, the demand, price, idle-time and waiting-time uncertainties are taken into account to adjust the tradeoffs between on-demand and oversubscribed costs. A stochastic programming approach for supply chain network design under uncertainty was proposed in [10]. A crucial component of the planning activities of a manufacturing firm is the efficient design and operation of its supply chain. This strategic level supply chain planning involves deciding the configuration of the network, the number, location, capacity, and technology of the facilities are implied by the configuration. The tactical level planning of supply chain operations is deciding the aggregate quantities and material flows for purchasing the products, processing the products and distribution of products. The key factor which influences the efficient tactical operations is the strategic configuration of the supply chain, and hence has a long lasting impact on the firm. Moreover, the fact that the supply chain configuration involves the commitment of substantial capital resources over long periods of time makes the supply chain network design problem an extremely important one. This was discussed in this paper.

3. SYSTEM MODEL

3.1 VMR Model

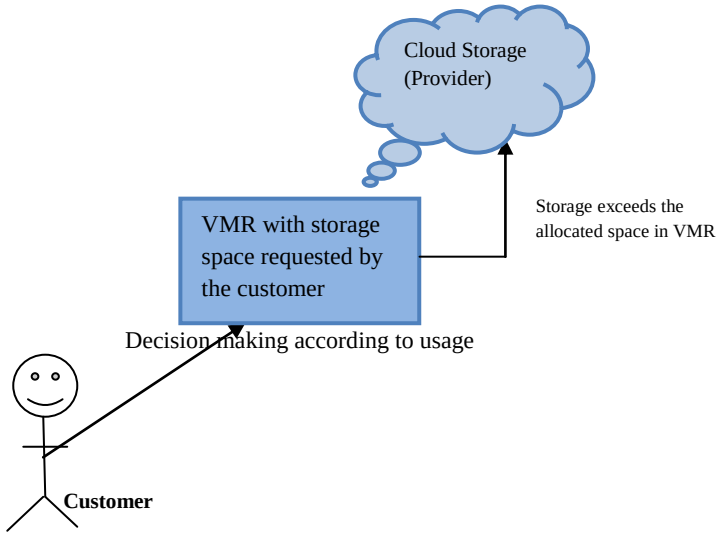

Fig. 1 System model of VMR

According to this model the customer initially requests the storage space to the provider. This request is handled by the Virtual Machine Repository (VMR) i.e. the files of the customer are stored in the VMR. The files stored by the customer are sent to the cloud once the space in the VMR gets over. Thus the cost and time consumption is avoided in the customer side. The problem of under and over provisioning is also handled in efficient manner. Effective load balancing is implemented with the provider to handle too much of requests from the customer.

3.1.1 Calculation of cost

Cost of Service C = d*cs
cs = no.of bytes*e

Penalty cost P(d)= { d*p, if d>current service capacity
{ 0, otherwise

where,
d= Current demand for storage calculated per hour
e= Cost per byte
cs= Cost of storage space
p= Penalty and p < cs

3.1.2 Solution for Over Provisioning

Since the reservation is based on the client's request, the client can determine the space required for him and for the safer side the client can reserve the space more than his requirement. This space is available for the client forever i.e., Over provisioning will not happen.

3.1.3 Solution for Under Provisioning

Under provisioning is also avoided up to 95%. The reason is, excess of space than the client's requirement is available in the VMR. Hence there is much less chances of occurrence of Under Provisioning. Even if under provisioning occurs, it can

be handled with minimum penalty which is acceptable to both provider and customer.

3.2 Advantages

- Solving cost-sensitive prediction
- Resolve over and under provisioning problems
- Improves the service quality of the cloud
- Reduce the time and cost of the customers and meets the increasing demand of service providers

4 Conclusion and Future Work

There are several promising directions for future works. The scope of this paper is confined to optimize the utility of existing resource based on the middle-term optimization strategy. The real-time cloud utility optimization is one of our future works. First, only a subset of the features of the trace records is used to build the prediction model in the current work. The service quality can be further improved by incorporating more features into the prediction model. Second, we can further improve the prediction accuracy by providing customer-oriented personalized prediction. Such a target can be fulfilled by leveraging the customer profile information. Third, the current solution for capacity planning is a kind of mid-term prediction. The irregularity of small granularity time series makes the prediction inherently difficult. To conquer this difficulty, leveraging the techniques in control theory is considered to conduct the short-term capacity planning. Finally, efficient formulation of calculating penalty costs can be done in future thus making it acceptable for both cloud resource provider and consumer in the situation of under provisioning of resources.

5 ACKNOWLEDGEMENTS

First and foremost, we dedicate our sincere gratitude to our honorable correspondent **Lion Dr. K. S. RANGASAMY, M.J.F** for providing immense facilities in our institution.

We proudly render our thanks to our principal **Dr. K. THYAGARAJAH, M.E., Ph.D., SMIEEE, MISTE** for the facilities and the encouragement given by him to the progress and completion of my work.

We proudly render our immense gratitude to the Head of the Department **Dr.B.G.GEETHA M.E., Ph.D.** for her effective leadership, encouragement and guidance in the work.

We are highly indebted to provide our heart full thanks to our guide **Mr.D.SEENIVASAN M.E., Assistant Professor** for his valuable ideas, encouragement and supportive guidance throughout the work.
Finally, we wish to thank all teaching and Non-teaching staff members of Department of Computer Science and Engineering and our friends for their support towards the successful completion of this work.

5 REFERENCES

[1] Bu-Sung Lee, Dusit Niyato and Sivadon Chaisiri (2012) "Optimization of resource provisioning cost in cloud computing", IEEE Transactions On Services Computing, Vol.5, No.2, pp.164-177.

[2] Chang-Shing Perng , Tao Li , Rong N. Chang and Yexi Jiang (2013) "Cloud Analytics for Capacity Planning and Instant VM Provisioning", IEEE Transactions On Network And Service Management, Vol.5, No.3 pp.170-185.

[3] Alexander Shapiro, Marc Goetschalckx, Shabbir Ahmed and Tjendera Santoso (2005) "A stochastic programming approach for supply chain network design under uncertainity," Elsevier On Computer Networks, pp.96-115.

[4] Chung-Nan Lee, Da-Jing Zhang-Jian, Ren-Hung Hwang, Yi-Ru Chen(2013), "Cost optimization of Elasticity and Resource Subscription Policy", IEEE Transactions On Services Computing, Manuscript Id 10.1109/TSC.2013.35, Vol.1, No.6, pp.968-974.

[5] Daniel Warneke and Odej Kao (2011) "Exploiting dynamic resource allocation for efficient parallel data processing in the cloud," IEEE Transactions On Parallel And Distributed Systems, Vol. 22, No. 6, pp.985-997.

[6] David Schanzenbach, Frederic Vivien, Henri Casanova and Mark Stillwell (2010) "Resource allocation algorithms for virtualized service hosting platforms", Elsevier On Parallel And Distributed Computing,pp.962-974.

[7] Fetahi Wuhib, Mike Spreitzer and Rolf Stadler (2012) "A gossip protocol for dynamic resource management in large cloud environments", IEEE Transactions On Network And Service Management, Vol. 9, No. 2, pp.213-225.

[8] Chaisiri, S. ;Bu-Sung Lee, Niyato, D. (2010) "Robust cloud resource provisioning for cloud computing environments", Service-Oriented Computing and Applications (SOCA), 2010 IEEE International Conference, Vol. 6, No. 3, pp.1-8.

[9] Fumitaka Higashitani, Kazuo Miyashita and Kazuyuki Masuda (2008) "Coordinated service allocation through flexible reservation," IEEE Transactions On Services Computing, Vol.1, No.2, pp.117-128.

[10] Alexander Shapiro, Marc Goetschalckx, Shabbir Ahmed and Tjendera Santoso (2005) "A stochastic programming approach for supply chain network design under uncertainity," Elsevier On Computer Networks, pp.96-115.

[11] Gagan Agrawal, and Qian Zhu, (2012) "Resource provisioning with budget constraints for adaptive applications in cloud environments", IEEE Transactions On Services Computing, Vol. 5, No. 4, pp.497-511.

[12] Gagan Agrawal, Qian Zhu and Smita Vijayakumar (2010) "Dynamic Resource Provisioning for Data Streaming Applications in a Cloud Environment", 2nd IEEE International Conference on Cloud Computing Technology and Science,Vol.6, No.2, pp.166-178.

[13] Kandasamy N. and Kusic D.(2007) "Risk aware limited lookahead control for dynamic resource provisioning in enterprise computing systems", Springer On Cloud Computing, pp.395-408.

[14] Kwang Mong Si and Seokho Son (2012) "A price-and-time slot negotiation mechanism for cloud service reservations", IEEE Transactions On Systems, Man And Cybernatics, Vol. 42, No. 3, pp.713-728.

[15] Torsten Eymann, and Werner Streitberger (2009) "A simulation of an economic, self-organizing resource allocation approach for application layer networks", Elsevier On Computer Networks, pp.1760-1770.

[16] Weisong Shi, Ying Song and Yuzhong Sun (2013) "A two-tiered on-demand resource allocation mechanism for VM-based data centers", IEEE Transactions On Services Computing, Vol. 6, No. 1, pp.116-129.