http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (22): 6882-6886, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.6882.6886

Utility-based Price Proportion in Cloud Resource Allocation

Zexiang Mao, Yanlei Shang, Chuanchang Liu and Junliang Chen
State Key Lab of Networking and Switching Technology,
Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract: As cloud computing is a new emerging distributed computing paradigm driven by economies of scale it is urgently to find better solutions for cloud resource allocation problem for economy reasons. Although there are lots of research efforts for cloud resource allocation, most of them introduce auction model to analyze the competition among cloud consumers in the condition of resource provisioned by means of indivisible VMs. For divisible resources, previous works are focus on minimizing the cost without considering features of diverse cloud consumers. In this study, utility-based price proportion approach for divisible resource is proposed. In the presented approach, utility is used to specify the features of diverse cloud consumers. With utility introduced, each cloud consumer is focus on the profit but not the cost, where the profit is the utility consumer gained from using resources minus the cost. Furthermore, experimental results show that the approach gains 21 and 15% more profit than common equal resource and cost minimization approaches separately with the same cost.

Key words: Cloud, price proportion, utility, resource allocation

INTRODUCTION

Cloud computing, a new term for the vision of computing as a utility like electricity and water, enables convenient network access to a centralized pool of configurable computing resources. As ‘the cloud is driven by economies of scale (Foster *et al*., 2008), cloud computing is different from other distributed computing paradigms, although it shares similar vision with grid computing. Grid computing forces on executing large calculations effectively by making use of networked and loosely coupled computers while cloud computing inclines to support Internet services of supplement, consumption and delivery through resource virtualization and allocation. The differences makes it is crucial reconsider the resource allocation algorithms in cloud.

Since cloud computing integrates dynamic, scalable, distributed resources available to enterprises and individuals (Vaquero *et al*., 2009) and cloud services involve frequent buying, selling, trading and exchanging behaviors between providers and consumers who are both financially rational, market mechanism turns out to be an appropriate approach for resource allocation in the complex and heterogeneous environment (Tan and Gurd, 2007). Currently, most cloud providers make use of both fixed price and auction models (Amazon, 2009, 2010). Lots of researchers study and improve auction models in cloud (Prasad *et al*., 2012; Zaman and Grosu, 2013; Wang *et al*., 2012), where the resources are indivisible provided by means of virtual machines. However, heterogeneous resource should be allocated suitable based on different demands, where these resources are divisible, e.g., computing, memory and network resources. Maheswaran and Basar (2003) presented a method for divisible resource allocation, named as price proportion method. Participants bid for one unit resource, where each achieves the resource equals to the proportion of his bid to the sum of all bids. With the use of price proportion method, the price of resource adjusts to the market and every participant gain appropriate resource with the same cost of one unit resource. Based on the research result above, other works (Bredin *et al*., 2003; Teng and Magoules, 2010) study the price proportion method both in Mobile-Agent Systems and cloud, where they assumed that every user's job has the same set of subtasks aiming at minimum cost. As cloud live in an open word, the scenarios in above works considered are idealized. Furthermore, different jobs completion gain various benefit for divers users. Thus, a utility-based price proportion approach for cloud divisible resources is proposed in this study.

UTILITY-BASED PRICE PROPORTION APPROACH

In this study, we consider multiple cloud consumers rent one cloud provider’s resources to execute their jobs.

Corresponding Author: Zexiang Mao, State Key Lab of Networking and Switching Technology,
Beijing University of Posts and Telecommunications, Beijing 100876, China

As mentioned above, cloud provider and consumers both have strategies to maximize own benefits. Cloud consumers gain benefits from completing jobs by using of the cloud provider's resources while cloud provider gain benefit by renting resources. Thus, the benefit cloud provider gained depends on the competition among cloud consumers. Every consumer chooses the appropriate bid to maximize his profit which is the benefit gained by job completion minus the cost of renting resource.

Linear utility function: To the various financial capacities of consumers, jobs completion gains diverse benefit based on SLA, where we use utility function model the benefit gained of different consumers. We use the linear utility function introduced from (Van *et al*., 2009; Ardagna *et al*., 2011, 2012) which is shown follow:

$$V_k(t_k) = \begin{cases} v_k & t_k \le t_k^0 \\ v_k + m_k(t_k - t_k^0) & t_k > t_k^0 \end{cases} \quad (1)$$

The linear utility function above specifies the revenue (or penalty) V_k incurred when the completion time is t_k. Linear utility functions are a flexible mechanism to rank different jobs (e.g., assigning higher slopes $|m_k|$ to more important jobs) and allow also to implement soft constraints on completion times.

Price proportion: Now we look at a general case as an example where cloud provider virtualizes one resource with fixed finite capacity C. Without loss of generality, we assume that each consumer owns one job simply. We clarify the optimization object of each service is the maximum profit and the cloud provider gain revenue from renting resource.

It is assumed that there are N consumers competing resource. The size of the job of ith consumer is q_i, who bid for resource at price b_i. The total bids for the resource is $\theta = \Sigma_{NbiN} b_i$ which is also the price of total resource while $\theta_{-i} = \Sigma_{j \ne i}^{N} b_i$ is given as the sum of other bids except ith consumer's bid b_i.

For ith consumer, the time t_i and expense c_i taken are defined as:

$$t_i = \frac{q_i}{\left(\frac{b_i}{\theta} \cdot C\right)} = \frac{q_i \cdot (b_i + \theta_{-i})}{b_i \cdot C}$$
$$c_i = b_i t_i = \frac{q_i \cdot (b_i + \theta_{-i})}{C} \quad (2)$$

As the optimization object of each consumer is the maximum profit, particularly for ith consumer, who bid at the price b_i to maximize profit u_i as follows:

$$\max_{b_i} \{u_i(t_i)\} = \max_{b_i} \{V_i(t_i) - b_i t_i\} \quad (3)$$

By using Equation (1) and (2), the expansion of Equation (3) is:

$$u_i(b_i) = \begin{cases} v_i - \frac{q_i \cdot (b_i + \theta_{-i})}{C} & \frac{q_i\theta}{b_i C} \le t_i^0 \\ v_i + m_i\left(\frac{q_i\theta}{b_i \cdot C} - t_i^0\right) - \frac{q_i\theta}{C} & \frac{q_i\theta}{b_i C} \le t_i^0 \end{cases} \quad (4)$$

To illustrate how the profit u_i changes with bid b_i, Eq. 5 is needed and the corresponding profit curve of ith player is shown in Fig. 1.

$$u_i'(b_i) = \begin{cases} -\frac{q_i}{C} < 0 & \frac{q_i\theta}{b_i C} \le t_i^0 \\ -\frac{m_i \cdot q_i \cdot \theta_{-i}}{b_i^2 \cdot C} - \frac{q_i}{C} & \frac{q_i\theta}{b_i C} > t_i^0 \end{cases}$$
$$u_i''(b_i) = \begin{cases} 0 & \frac{q_i\theta}{b_i C} \le t_i^0 \\ 2\frac{m_i \cdot q_i \cdot \theta_{-i}}{b_i^3 \cdot C} < 0 & \frac{q_i\theta}{b_i C} > t_i^0 \end{cases} \quad (5)$$

For every rational consumer making his own profit u_i maximum, he should bid at optimal price b_i to reach the vertex of u_i curve in Fig. 1. It is obvious that the curve of u_i is a piecewise curve which makes the bid function piecewise too. Thus, based on the analysis above, the *i*th consumer will bid at price:

$$b_i(\theta) = \begin{cases} \frac{q_i \cdot \theta}{C \cdot t_i^0} & \frac{q_i\theta}{b_i C} \le t_i^0 \\ \frac{m_i}{2} + \sqrt{\frac{m_i^2}{4} - m_i\theta} & \frac{q_i\theta}{b_i C} > t_i^0 \end{cases} \quad (6)$$

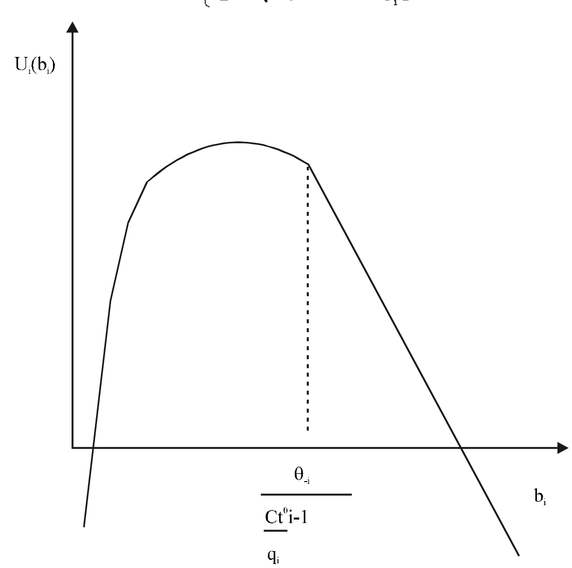

Fig. 1: Profit with different bid

For simply, we use the symbol:

$$h\left(b_i\right) = -\frac{m_i \cdot q_i \cdot \theta_{-i}}{b_i^2 \cdot C} - \frac{q_i}{C}$$

From Fig. 1 and Eq. 6 above it is obvious that:

$$b_i\left(\theta\right) = \frac{q_i \cdot \theta}{C \cdot t_i^0}$$

if and only if it satisfy the inequations:

$$b_i = \frac{q_i \cdot \theta}{C \cdot t_i^0} < \theta$$

and

$$h\left(\frac{q_i \theta}{C t_i^0}\right) \geq 0$$

otherwise:

$$b_i\left(\theta\right) = \frac{m_i}{2} + \sqrt{\frac{m_i^2}{4} - m_i \theta}$$

In order to facility the analysis, the symbol:

$$y_i = \frac{q_i}{t_i^0}$$

is introduced which stands for the minimum resource ith player needs to achieve the maximum utility v_i. Solving:

$$\frac{q_i \cdot \theta}{C \cdot t_i^0} < \theta$$

and

$$h\left(\frac{q_i \cdot \theta}{C \cdot t_i^0}\right) \geq$$

in the case of è>0 it is obtained that:

$$\theta \leq -\frac{m_i\left(C - y_i\right)}{y_i^2} \cdot C$$

That is, the bid function of *i*th consumer is:

$$b_i\left(\theta\right) = \begin{cases} \frac{y_i}{C} \cdot \theta & \theta \leq -\frac{m_i\left(C - y_i\right)}{y_i^2} C \\ \frac{m_i}{2} + \sqrt{\frac{m_i^2}{4} - m_i \theta} & \theta > -\frac{m_i\left(C - y_i\right)}{y_i^2} C \end{cases} \quad (7)$$

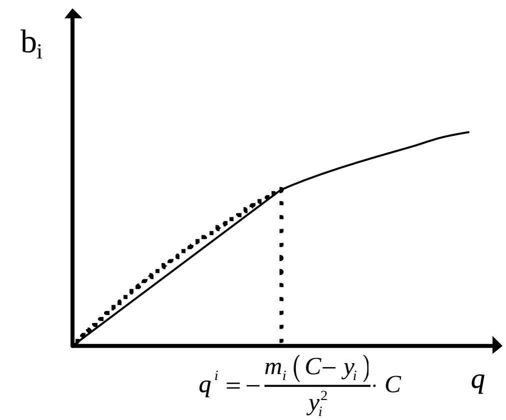

Fig. 2: Optimal bid of *i*th consumer

The corresponding curve of the ith consumer's bid b_i with resource price θ variation is shown in Fig. 2.

The curve in Fig. 2 indicate that *i*th consumer has two strategy based on different θ, where the critical price of changing strategy is:

$$\theta^i = -\frac{m_i\left(C - y_i\right)}{y_i^2} \cdot C$$

When $\theta < \theta^i$ it means that there are adequate resource with low price, the objective of ith player is achieving the maximum utility v_i with lowest expense. Otherwise, the demand is higher in the case of more expensive resource price. The player has to looking for the trade-off between utility and expense considering the slop m_i, for the excessive cost to gain the maximum profit. Moreover, the amount of resource C influences the supply condition of resource market, that is it increases the critical point price θ^i by provisioning more resource. Also, the strategy of ith player is restricted to his private information which is y_i and m_i. y_i and m_i reflect the minimum resource consumer needed to gain the maximum utility v_i and how execution time jitter impact utility V_i, respectively. Thus, the more resource y_i needed the lower critical price θ^i is. The lower value of m_i means that resource fluctuation has higher influence on utility V_i, in which case the value of θ^i is larger.

PERFORMANCE EVALUATION

In this section, another two common resource allocation approaches are introduced to verify the efficiency of the proposed allocation approach. As above it is assumed that there are N agents which stands for consumers. Furthermore, the allocation result among agents is denoted as $r = (r_1, r_2, \ldots, r_N)$.

The first introduced approach is the Equal Resource Mechanism, named as ERM for short. Resource provisions among agents are equal. Thus, the allocation result is $r^e = (r_1^e, r_2^e, \ldots, r_N^e)$, where $r_i^e = C/N$. The symbol $\ddot{o}_e$ denotes the price that unit resource cost in unit time. Thus, the expense of ith agent is $\varphi_e q_i$, the total expense is $c^e = \varphi_e \Sigma_N q_i$.

The other introduced approach is the Cost Minimization Mechanism, named as CMM for short, whose cost contains both service completion time and expense of renting resource. This mechanism is used widely in cloud resource allocation system (Mao and Humphrey, 2012). The fixed price of unit resource in unit time is φ_m and allocation result is $r^m = (r_1^m, r_2^m, \ldots, r_N^m)$. To minimize the cost which contains resource expense and completion time, the optimization function of each agent is shown below:

$$\begin{aligned} &\min_{r_i^m}\left\{w_e \cdot \varphi_m r_i^m \frac{q_i}{r_i^m} + w_t \cdot \frac{q_i}{r_i^m}\right\} \\ &= \min_{r_i^m}\left\{w_e \cdot \varphi_m q_i + w_t \cdot \frac{q_i}{r_i^m}\right\} \\ &\text{s.t.} \quad \sum_{i=1}^{N} r_i^m \le C \end{aligned} \tag{8}$$

where, w_e and w_t denote the weights of expense and completion time, respectively.

With KKT condition using, the function above is easily solved for each agent. The optimal resource allocation result should satisfy:

$$\frac{r_i^m}{r_j^m} = \frac{\sqrt{q_i}}{\sqrt{q_j}}$$

where i and j∈[1, N]. Thus, the resource provisions for ith agent is:

$$r_i^m = \frac{\sqrt{q_i}}{\sum_N \sqrt{q_j}} C$$

As mentioned above, the total expense of this approach is $c^e = \varphi_m \Sigma_N q_i$.

The approach proposed here is Price Proportion Mechanism, named as PPM for short. As expense of each agent shown in Eq. 2, the total expense of PPM is$^p = (\theta/C)\ \Sigma_n q_i$.

For the cloud provider's utility equals the sum of agents' expense, we assume that the utility gained by provider in the three approaches are the same to compare the profits brought. That is $\varphi_e = \varphi_m = \theta/C$.

For the scenario of one resource virtualized, memory in the simulation, there are 100 agents competing memory resource, where each agent holds one job. The size of each job q_i which stands for the memory needed to complete it in unit time, is chosen between 500M and 2500M with random probability. The completion time of maximum utility t_i^0 and slop of utility function m $_i$are generated randomly from [1, 100] and [-20, -1], separately. The amount of memory provision is C=8000M.

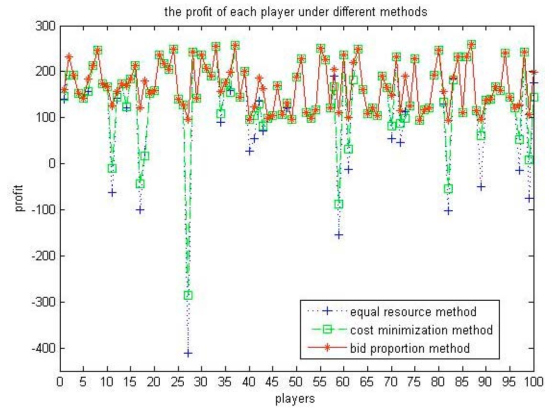

Fig. 3: Agents' profit in three approaches

The profit of each agent is shown in Fig. 3 with the assumption of maximum utility v_i=300. Without considering any features of services, the ERM must leads to over-provisioning or under-provisioning for a agent's job and brings the lowest profit to each agent. Although CMM more effective than the former it is not established efficiency allocation scheme because that utility functions of services are diverse. The utility-based price proportion approach guarantee positive profit of each agent and highest total profit. The total profits of three approaches are 13924 for ERM, 14647 for CMM and 16875 for PPM, respectively. In other words, the total profit gained by PPM is 21% more than ERM and 15% more than CMM bring.

CONCLUSION

Cloud computing is driven by economies of scale which is different from former distributed computing paradigms. Using auction, the resource price fluctuates according to the condition of supply and demand. Different from former works minimizing cost and response time, we use linear utility function characterize the profit brought. Thus, the utility-based price proportion approach is proposed. Furthermore, the approach proposed gains 21\ and 15\% more profit than general equal resource and cost minimization approaches with the same cost.

In the approach proposed, the total resource price θ reflects the scale of consumers. From the value of unit resource price θ/C fluctuates, cloud provider will understand the supply and demand of resource. Observing both total resource and unit resource price, the provider know how the scale of participants changes.

ACKNOWLEDGMENT

The work presented in this study is supported by the National Grand Fundamental Research 973 Program of China (2011CB302506), the National Key Technology Research and Development Program of China "Research on the mobile community cultural service aggregation supporting technology" (2012BAH94F02) and the Novel Mobile Service Control Network Architecture and Key Technologies (2010ZX03004-001-01).

REFERENCES

Amazon, 2009. Amazon elastic compute cloud (Amazon EC2). Amazon Web Services. http://aws. amazon. com/ec2/

Amazon, 2010. Amazon EC2 spot instances. Amazon Web Services. http://aws.amazon.com/ec2/spot-instances/.

Ardagna, D., B. Panicucci and M. Passacantando, 2011. A game theoretic formulation of the service provisioning problem in cloud systems. Proceedings of the 20th International Conference on World Wide Web, March 28-April 1, 2011, Hyderabad, pp: 177-186.

Ardagna, D., B. Panicucci, M. Trubian and L. Zhang, 2012. Energy-aware autonomic resource allocation in multitier virtualized environments. IEEE Trans. Services Comput., 5: 2-19.

Bredin, J., D. Kotz, D. Rus, R.T. Maheswaran, C. Imer and T. Basar, 2003. Computational markets to regulate mobile-agent systems. Autonomous Agents Multi-Agent Syst., 6: 235-263.

Foster, I., Y. Zhao, I. Raicu and S. Lu, 2008. Cloud computing and grid computing 360-degree compared. Proceedings of the Grid Computing Environments Workshop, November 12-16, 2008, Austin, Texas, USA., pp: 1-10.

Maheswaran, R.T. and T. Basar, 2003. Nash equilibrium and decentralized negotiation in auctioning divisible resources. Group Decis. Negotiation, 12: 361-395.

Mao, M. and M. Humphrey, 2011. Auto-scaling to minimize cost and meet application deadlines in cloud workflows. Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, November 12-18, 2011, Seatle, WA., USA., pp: 1-12.

Prasad, G.V., S. Rao and A.S. Prasad, 2012. A combinatorial auction mechanism for multiple resource procurement in cloud computing. Proceedings of the 12th International Conference on Intelligent Systems Design and Applications, November 27-29, 2012, Kochi, India, pp: 337-344.

Tan, Z. and J.R. Gurd, 2007. Market-based grid resource allocation using a stable continuous double auction. Proceedings of the 8th IEEE/ACM International Conference on Grid Computing, September 19-21, 2007, Austin, Texas, pp: 283-290.

Steng, F. and F. Magoules, 2010. Resource pricing and equilibrium allocation policy in cloud computing. Proceedings of the IEEE 10th International Conference on Computer and Information Technology, June 29-July 1, 2010, Bradford, UK., pp: 195-202.

Van, H.N., F.D. Tran and J.M. Menaud, 2009. SLA-aware virtual resource management for cloud infrastructures. Proceedings of the 9th IEEE International Conference on Computer and Information Technology, October 11-14, 2009, Xiamen, China, pp: 357-362.

Vaquero, L.M., L. Rodero-Merino, J. Caceres and M. Lindner, 2009. A break in the clouds: Towards a cloud definition. ACM SIGCOMM Comput. Commun. Rev., 39: 50-55.

Wang, Q., K. Ren and X. Meng, 2012. When cloud meets eBay: Towards effective pricing for cloud computing. Proceedings of the IEEE INFOCOM, March 25-30, 2012, Orlando, FL., USA., pp: 936-944.

Zaman, S. and D. Grosu, 2013. Combinatorial auction-based allocation of virtual machine instances in clouds. J. Parallel Distrib. Comput., 73: 495-508.