A Two-Phase Gaming Model for Resource Pricing in Elastic Cloud Environments

Xilong Qu, Hong He and Peng Xiao

Department of Computer and Communication, Hunan Institute of Engineering
xlq1978@126.com; hhong1970@126.com

Abstract

With the development of cloud computing, many organizations are deploying their IT infrastructures through cloud platforms so as to provide services for more and more users. Therefore, resource pricing mechanism becomes a key component in these open cloud platforms. In this paper, we propose a gaming theory based cloud resource pricing model, in which a cooperative gaming model is applied to optimize the resource benefits and non-cooperative gaming model is used to balance the user’s costs and provider’s benefits. Theoretical analysis is presented to validate the correctness of the proposed gaming model, and extensive experiments are conducted to investigate the effectiveness of the proposed resource pricing mechanism. The results indicate that our resource pricing mechanism outperforms many existing approaches in terms of resource profits and response time, especially when the cloud system is in presence of intensive workloads.

Keywords: Cloud Computing, Resource Pricing, Gaming Theory, Job Scheduling

1. Introduction

Cloud computing has emerged as an promising technology and it has been increasingly adopted in many areas including science, engineering, and commercial business, due to its inherent flexibility, scalability, cost-effectiveness and security [1, 2]. Clouds are primarily motivated by the conception of utility computing, in which users have to pay resource providers for executing their applications. While the pay-per-use pricing model is very appealing for both service providers and consumers, conflicting objectives between the two parties hinder its effective application [3]. In other words, the service provider aims to accommodate as many requests as possible with aiming to maximizing their profits, which inevitably conflict with consumer’s performance requirements.

In the past few years, there have been plenty of studies exploiting market pricing mechanism for distributed resource allocation, and the well-known distributed systems [4-7]. Beside this systems, many economic-based policies and scheduling algorithms have also be widely studied, include Auction Mechanism [8], First Price [9], First Profit [10], and Proportional-Share [11]. In typical distributed systems, economic-based model is has been proven to be effective for resource allocation, however, it also raise other problems that cannot be ignored. Firstly, economic models bring about extra communicational and computational overhead to applications [5-7, 12]; Secondly, when the system is in presence of high-end applications that require co-allocating multiple resources across sites, the price negotiation process is often low-efficient [8, 13].

Currently, many existing cloud systems adopted fixed pricing mechanism whose advantages are easy implementation and low maintain cost [14]. Unfortunately, fixed
pricing mechanism will lead to many negative effects on system performance with the increasing of system scale, such as low resource utilization [14-16], load unbalancing [12, 17, 18], undesirable QoS satisfaction [19, 20]. To address the above issues, in this work we present a cloud resource pricing model with aiming at overcoming the shortcomings of existing price mechanisms in terms of efficiency and fairness. In our pricing model, virtual resource configuration and provision are defined as a two-phrase gaming procedure, in which cooperative gaming model is applied to optimize the resource benefits and non-cooperative gaming model is used to balance the user’s costs and provider’s benefits.

The rest of this paper is organized as following: Section 2 presents the related work; In Section 3, the gaming models are presented with problem description; In Section 4 the solutions of game models are presented theoretically; In Section 5, experiments are conducted to examine the effectiveness of the proposed approach. Finally, Section 6 concludes the paper with a brief discussion of the future work.

2. Related Work

Researches on market-based resource allocation have been widely emerged in the past decades. In these studies, resource pricing mechanism plays an important role for solving some practical problems, in which the resource price can reveal the true needs of users who compete for shared resources and allocate resources more efficiently [6, 9-13, 21]. For example, in [9] the authors build a prototype cluster that provides a market for time-shared CPU usage for various jobs. In [11], a time-varying resource valuation approach is proposed for jobs submission in virtual clusters. The changing values were used for prioritizing and scheduling batch sequential and parallel jobs. In [21], the authors argued using the market-based method to address flash crowds and traffic spikes for clusters hosting Internet applications. The above studies mainly concentrate on how to using scheduling decision for influencing resource prices. In these studies, the role of resource users can almost be ignored.

As the suitable resource price is difficult to be defined considering the various of user’s requirements, many existing cloud systems adopted fixed pricing mechanism whose advantages are easy implementation and low maintain cost [14, 18]. Unfortunately, fixed pricing mechanism will lead to many negative effects on system performance with the increasing of system scale. To solve these problems, dynamic pricing mechanisms have been extensively proposed, i.e. Resource Auction [22], SLA Negotiation [16, 19], Hot-spot Pricing [23]. For example, in [14], the authors presented a dynamic pricing scheme which takes efforts on improving the efficiency of batch resource trading in federated cloud environments. In their scheme, the whole cloud system is considered as a uniformed resource market where resource supply and demand can be balanced by using macro-economic equivalence theory. Unfortunately, the scheme relies on market self to automatically obtaining equivalent price; therefore it is low-efficient comparing with the opening feature of cloud platform. In [22], the authors proposed a dynamic second-priced auction mechanism to solve the allocation problem of computation capacity in the environment of cloud computing. During the auction procedure, it assumes that resource pricing will be increased significantly when the system workload is in peak state. Such an assumption is validating for those systems whose resource quantity is constant during a long time interval. In [20], the authors proposed a hierarchical game model to analyze the decisions of resource providers when cloud resources are shared by both internal users and public users. The game model is composed of two interrelated cooperative games: (1) The low-level game model is to
describe the revenue sharing process between various cloud providers, and its game solution can be figured out by stochastic linear programming technique; (2) The upper-level game model formulate the coalitional process when a group of providers contribute their resources to a common pool, and its analytical solution is presented by Markov Chain technique.

3. Problem Description

The framework of resource management in elastic clouds is shown in Figure 1. In such a framework, single cloud providers can serve as independent cloud system when it can provide sufficient resource for user application; if the resource capability of single cloud provider is not enough to satisfy the requirements of some large-scale applications, it can federate together and provide service for users. As shown in Figure 1, each cloud provider is independent in some times meanwhile keeping connected when needed. This work model is very similar to the ‘Production Broker’ model in normal business area [24]. Motivated by this, we introduce the conception of ‘Virtual Resource Broker’ (VRB) to describe the working of individual cloud providers with aiming at analyzing the resource provision and configuration in cloud platform.

![Figure 1. Framework of Resource Management in Federated Elastic Clouds](image-url)

Let the set of cloud providers be as \{S_1, S_2, ..., S_n\}, their resource quantity is noted as \{v_1, v_2, ..., v_n\}. The set of VRB be note as \{B_1, B_2, ..., B_n\} and their resource quantity is noted as \{c_1, c_2, ..., c_n\}. The user application is consist of a set of tasks noting as \{t_1, t_2, ..., t_m\}, each being characterized as \(<r_i, d_i>\), where \(r_i\) is the resource requirement and \(d_i\) is the deadline constraint. The utility function of cloud provider \(S_i\) is noted as \(u_i^S = p \cdot v\), indicating the profits of \(S_i\) when it sells its resource with price \(\bar{p}\). The utility function of VRB \(B_i\) is noted as \(u_i^B = \mu_i \cdot p_i \cdot c_i - \bar{p} \cdot c_i\), where \(p_i\) is retail price decided by \(B_i\); \(\mu_i\) is the resource utilization of \(B_i\). Therefore, \(U_i^B\) indicating the profits of \(B_i\) when it sell its resource to users with price \(p_i\) after it obtained resources from cloud providers.
with price $\overline{p}$. The global utility function of the federated cloud system is noted as $U^* = \sum_{i=1}^{n} (v^i + u^i)$, which is summation profits of cloud providers and virtual resource brokers. The utility function of user application is noted as $U^T = \sum_{j=1}^{n} (r^j \cdot p^j)$, where $r^j_i$ is resource quantity that task $t_i$ obtained from $B_i$.

It is clear that the profits of each VRB will be different after a period of time, since they use different retail prices. Therefore, we categorize them into three set by their profits. The set of VRB with positive profits is noted as $\mathcal{X}^+ (\overline{p}) = \{ B_i \mid U^i > 0, i \in [1 \ldots n] \}$; that with negative profits is noted as $\mathcal{X}^- (\overline{p}) = \{ B_i \mid U^i < 0, i \in [1 \ldots n] \}$; that with zero profits is noted $\mathcal{X}^0 (\overline{p}) = \{ B_i \mid U^i = 0, i \in [1 \ldots n] \}$. From the perspective of resource configuration, the profits of any VRB is decided by $c_i$ and $\overline{p}$; from the perspective of resource provision, it is affected by retail price $p_i$ and its resource utilization $\mu_i$. Therefore, for any individual $B_i \in \mathcal{X}^g(\overline{p})$, we say it is in balance state under condition $<\overline{p}, c_i>$. If $B_i \in \mathcal{X}(\overline{p})$, then we say the whole resource trading system is in balance state under condition $<\overline{p}, c,c' >$.

Combing Figure 1 and the above definitions, we can see that there are three classes of participants: cloud providers, VRBs, and resource consumers. The cloud providers and VRBs cooperate with each other, since they both aim at maximizing resources utilization and resource profits. On the other hand, the relationship between the VRBs and the resource consumers is non-cooperative, as the clients hope to minimize their costs, which would inevitably lower down the benefits of cloud providers. In the following sections, we will present the validity and solution of this framework in theory.

4. Game Models and Solutions

4.1. Cooperative Gaming Model

As mentioned in Section 3, the gaming model between cloud providers and VRBs is cooperative, and the former needs to decide an optimal resource price $\overline{p}$, while the latter needs to decide the optimal resource configuration noted as $\{ c_1, c_2, \ldots, c_n \}$. Therefore, the solution of this cooperative game can be noted as $<\overline{p}, c, c', \ldots, c_n>$. Since all the resource owned by cloud providers are brokered by VRBs, it satisfies $\sum_{i=1}^{n} v_i = \sum_{i=1}^{n} u_i^g$. When the whole system is not in balancing state, we can have $U^G = \sum_{i=1}^{n} (p_i \cdot \overline{p} \cdot c_i)$. When the system is in balancing state, according to the definitions in Section 3, we know that $\sum_{i=1}^{n} u^1_i = 0$, therefore $U^G = \sum_{i=1}^{n} v^1_i = \overline{p} \cdot \sum_{i=1}^{n} r^1_i$. That is saying, when the whole resource trading system is in balance state, the global profits $U^G$ is independent with $\overline{p}$; otherwise, $U^G$ is decided only by $\overline{p}$.

Assuming that the system is in balancing state and the current condition is $<\overline{p}, c, c', \ldots, c_n>$, we note the VRBs’ profits as $<U^1_i, U^2_i, \ldots, U^n_i>$. When the resource trading system is in balancing state, $\forall B_i$ satisfies $u^1_i = \mu_i \cdot p_i \cdot c_i - \overline{p} \cdot c_i = 0$, that is $\mu_i = \overline{p} = p_i = 0$. Assume that the VRBs’ resource configurations are changed as $<c_1, \Delta c_1, \ldots, c_n + \Delta c_n>$, and their profits are noted as $<U^1_i, U^2_i, \ldots, U^n_i>$ under condition $<\overline{p} + \Delta \overline{p}, c_1 + \Delta c_1, \ldots, c_n + \Delta c_n>$. Combing the definitions of utility functions, we have $U^1_i = (c_i + \Delta c_i) \cdot \mu_i \cdot p_i = (\overline{p} + \Delta \overline{p}) \cdot \Delta \overline{p}$. As the total resource quantity is
constant, we have \( \sum_{i=1}^{n} (c_i + \Delta c_i) = \sum_{i=1}^{n} c_i = \sum_{i=1}^{n} \tilde{c}_i \). Since the system is in balancing state, we know that \( U^G = \sum_{i=1}^{n} (U^G_i + U^G_0) = \sum_{i=1}^{n} \tilde{U}_i = \tilde{p} \sum_{i=1}^{n} \tilde{c}_i \). Combining the above equations, under condition \(< \tilde{p}, <c_1, \ldots, c_n>>\), the global profit \( U^G \) satisfies

\[
U^G = \sum_{i=1}^{n} (U^G_i + U^G_0) = (\tilde{p} + \Delta \tilde{p}) \cdot \sum_{i=1}^{n} \tilde{c}_i = \tilde{p} \cdot \sum_{i=1}^{n} \tilde{c}_i = U^G.
\]

By Equation (1), we know that if the resource trading system is in balancing state under condition \(< \tilde{p}, <c_1, \ldots, c_n>>\), any change of \(< \tilde{p}, <c_1, \ldots, c_n>>\) will not change the global profits \( U^G \). As the pricing policy of VRBs and cloud providers will not improve the global benefits \( U^G \) when the resource trading system is in balancing state. Therefore, the Nash equivalent solution of cooperative gaming is the condition \(< \tilde{p}, <c_1, \ldots, c_n>>\) which can make the resource trading system in balancing state.

When the system is not in balancing state under condition \(< \tilde{p}, <c_1, \ldots, c_n>>\), we need to find a feasible approach to improve \( U^G \). It can be done by the following steps: (S1) We pick out any \( B_k \) that belong to \( \chi^*(\tilde{p}) \) or \( \chi^*(\tilde{p}) \). Change the resource price \( \tilde{p} \) as \( \tilde{p} = u_i \cdot p_i \), where \( i \in \chi^*(\tilde{p}) \). Meanwhile, we keep \( c_i = c_i \). Therefore, under condition \(< \tilde{p}', <c_1, \ldots, c_n>>\), we have the following conclusions: a) The system is still not in balancing state; b) \( U^G \) is unchanged. c) \( B_k \) is in balancing state. (S3) According to the second conclusion in Step 2, we know that \( \exists B_k \) satisfying \( B_k = \chi^*(\tilde{p}') \). Let the resource configuration of \( B_k \) being adjusted from \( c_k \) to \( c_k + \Delta c_k \), and the resource configuration of \( B_k \) being changed from \( c_k \) to \( c_k - \Delta c_k \), where \( \Delta c_k > 0 \). We have known that the above changing of \( B_k \)'s resource configuration will increase \( U^G \). At the same time, we also know that the above changing of \( B_k \)'s resource configuration will not affect \( U^G \), because \( B_k \) is now in balancing state. Therefore, under condition \(< \tilde{p}', \{c_1, \ldots, c_n\}>>\), we increase the global profit \( U^G \).

### 4.2. Non-Cooperative Gaming Model

According to the definitions of Section 3, we use \(< \tilde{p}, <p_1, \ldots, p_n>>\) to describe the gaming solution between VRBs and user applications, where \( \tilde{p} \) is the resource trading matrix. According to definition 2, the profits of \( B_i \) is affected by \( p_i \), \( \mu_i \), \( \tilde{p} \), and \( c_i \). During the cooperative gaming procedure, we have decided \( \tilde{p} \) and \( c_i \). At the same time, \( \mu_i \) is not adjustable since we can only obtain it by statistics. Therefore, the pricing policy can be noted as \( p_i(\mu_i) \), which means that VRBs adjusts their retail price \( p_i \) by observing its resource utilization \( \mu_i \). Therefore, the key point of solving the non-cooperative gaming model is to figure out the pricing function of VRBs.

Let \( p_i (\mu_i) \) be the pricing function of \( B_i \). According to definition 2, we have \( U^B_i = c_i \cdot p_i (\mu_i) - \tilde{p} \cdot c_i \). Let \( dU^B_i/d\mu_i = 0 \), we have the equation \( \mu_i \cdot d\tilde{p}_i(\mu) = d\mu_i + p_i (\mu_i) = 0 \), and \( U^B_i \) is maximized if this equation is solvable. By solve this equation, we can have following general solution as \( p_i (\mu_i) = K_{i,1} + K_{i,2} \cdot \mu_i \), where \( K_{i,1} \) and \( K_{i,2} \) are positive constant.

Clearly, if all \( \{B_1, B_2, \ldots, B_n\} \) are independent and their price polices are only related with their resource utilization, the condition of obtaining maximized profits is that the pricing function is inversely proportional to resource utilization. We can have multiple pricing function since \( K_{i,1} \) and \( K_{i,2} \) can be any positive constant. For example, assuming...
that the pricing bounds of \( B_i \) is \( \{ p_i^{\text{min}}, p_i^{\text{max}} \} \), then its pricing function can be defined as \( p_i(\mu_i) = p_i^{\text{max}} - (p_i^{\text{max}} - p_i^{\text{min}}) \cdot \mu_i \). We can easily know that \( B_i \) have maximal profits when \( \mu_i = \frac{p_i^{\text{max}}}{2(p_i^{\text{max}} - p_i^{\text{min}})} \). If \( p_i^{\text{max}} = 2 \overline{p} \) and \( p_i^{\text{min}} = 0.5 \overline{p} \), then when \( \mu_i = \frac{2}{3} \) the \( B_i \) have maximal profits. Therefore, once \( B_i \) decides its pricing function, it can obtain its optimal resource utilization by \( \mu_i^* = \frac{1}{K} \cdot \mu_i \). During the running time, if it observed that \( \mu_i < \mu_i^* \) it can low down its retail price, otherwise increases its price. If \( \mu_i = \mu_i^* \), then the current \( p_i \) is optimal for maximizing profits.

5. Experiments and Performance Comparison

In the experiments, we use CloudSim [25] to construct a simulative cloud platform, which consists of twelve high-performance clusters as underlying resources. The topology and setting of individual clusters are deprived from the grid test-bed DAS-2. The experimental workload (tasks stream) is generated by using Lublin-Feitelson model, which is derived from the workload logs of real supercomputers. In the experiment, we mainly concentrate on the effects of resource trading on application’s execution time. To analyze the performance, our hybrid gaming model (HGM) is compared with other four resource trading models, including Commodity Market Model (CMM) [26], Double Auction Model (DAM) [22], Vickery Auction Model (VAM) [26], and Batch Auction Model (BAM) [27]. To examine the efforts of resource requirements on performance, we enlarge the workload’s resource requirement \( \beta \) times. The experiments are conducted four times, with increasing \( \beta \) from 1.0 to 2.5. The results are shown in Figure 2.
Figure 2. Comparison of Responsive Time
The experimental results show that resource requirements affect not only the resource negotiation time but also the execution time. When the resource requirements is in low level ($\beta=1.0$), the performance of VAM and DAM is significantly higher that other policies, and their negotiation time is about 10.2% and 8.7% of the total completing time. As to HGM and BAM, the negotiation time is about 2.9% and 3.1% of the total completing time. With the increasing of $\beta$, the negotiation times of both VAM and DAM increase as well as their proportions to the total completing time. For example, when $\beta = 2.5$, negotiation time of VAM is about 3.82 times of the case $\beta = 1.0$, and its proportion is increased to 21.3%. BAM is a batch resource trading model, and it is very effective to allocate multiple resources to application. However, it has to take many rounds to complete the auction procedure, especially when the resource requirements are very large. Therefore, when $\beta = 1.0$ its performance is almost equal to HGM; however, with increasing of $\beta$, the time costs on auction procedure become dominator, which makes BAM’s performance reduced.

Comparing with auction model, CMM is effective to reduce communication-related costs. However, our experimental results indicate that its negotiation time increases significantly when $\beta$ increases to high level. By examining the logs, we found that re-negotiation occurs more frequently than before, that is, the CMM’s pricing policy can not efficiently finish the trading for all resource requirements. For example, when $\beta = 2.5$ about 43% tasks need to negotiation 2 times to obtain the required resources, and about 7% tasks need to do it at the third negotiation.

By the description of Section 3, we can see that HGM’s negotiation costs mainly come from the selection of suitable VRBs. As all VRBs decide their retail price independently according to their resource utilizations, therefore, it can avoids workload concentration which is very important for reducing the costs of re-negotiation. Before HGM is in balancing state, those VRB with positive profits will increase their resource configuration. By this mechanism, they can improve the capability of serving multi-resource requirements. Therefore, we can consider HGM as the combination of BAM and CMM. Based on the above experimental results, our conclusion is that HGM is effective to reduce the negotiation time, which in turn reduce the application execution time.

Secondly, we all investigate the HGM’s performance under constraints to costs and deadline. Because of the deadline constraint, we can not compare the policies directly. As a result, we select four typical resource matching algorithms to integrate those pricing mechanisms, including Round-Robin (RR), Capability-based Random (CR), Optimal Miss Rate (OMR) and Hierarchical Gaming Selection (HGS). The experiments conducted four times, each with different $\lambda$ parameter which is used to define the arrival interval of tasks. The results are shown in Figure 3.

![Figure 3. Resource Costs with Different Policies and $\lambda$ Parameter](image-url)
As shown in Figure 3, the performance of RR is the best if we only consider the resource costs. However, RR will result in high deadline miss rate when $\lambda$ increases. By our result logs, the deadline miss rate is about 61.22% when $\lambda=10$. By our experiment setting, if deadline occurs the user will not pay any cost for resource providers. That is the reason that the resource cost of RR is the lowest. Among the left policies, HGS can obtain lowest resource costs and its changes are very stable for different $\lambda$ parameters. We notice that the rejection rate of HGS is very high, which is significantly different from HGM. For example, when $\lambda=10$ the rejection rate of HGM is only about 6.17%. By the gaming policy, we know that the pricing function in HGM is inversely proportional to the resource utilization. Because of this pricing policy, HGM is capably of maintain a low rejection rate. As shown in the experimental results, when using the same pricing mechanism, CR and OMR perform very similar in terms of resource costs and rejection rate. However, their deadline miss rates are very different. In a whole, OMR is more effective to provide deadline guarantee than other policies especially when $\lambda=0.05$ and $\lambda=0.10$. However, when we increasing $\lambda$ from 0.1 to 1.0, the rejection rate of OMR+BAM increases about 2 times, and the deadline miss rate increases about 2.5 times. Such a result happens on OMR+CMM. That is, OMR is effective to reduce deadline miss rate when there is no cost constraint; when cost constraint should be considered, OMR is only suitable for those applications with uniform workloads. Based on this experiment, we can see that HGM is effective to maintain low rejection rate and obtain better tradeoffs between resource costs and cloud provider’s profits.

6. Conclusion

To address the issue of resource pricing mechanism in cloud environments, a hybrid gaming based pricing model is proposed to overcome the demerits of existing price mechanisms in terms of efficiency and fairness. In the proposed pricing model, virtual resource configuration and provision are described as a two-phrase gaming procedure, in which cooperative gaming model is applied to optimize the resource benefits and non-cooperative gaming model is used to balance the user’s costs and provider’s benefits. The validity and solution of the proposed price model are presented theoretically, and the experimental results indicate that the hybrid gaming model can significantly improve the price negotiation efficiency when a bundle of resources are negotiated concurrently, which in turn reduce the application execution latency that caused by conventional price negotiation mechanisms. In addition, it also outperforms other pricing mechanism in terms of user’s QoS requirements, such as resource cost and deadline guarantee, especially when the system workload is very intensive. In the future, we plan to incorporate resource reservation mechanism into our HGM framework, and design some elastic reservation mechanism to improve the QoS performance.

Acknowledgements

This work is supported by the Hunan Provincial Natural Science Foundation of China (No. 13JJ9022). Also, it is the Project Supported by Scientific Research Fund of Hunan Provincial Education Department of China (No. 13B015).

References


Authors

Xilong Qu, he received his M.S. and Ph.D. degrees in computer science from the Southwest Jiaotong University, China, in 2003 and 2006 respectively. He is currently an Associated Professor in the School of Computer and Communication of Hunan Institute of Engineering. His research interests are cloud computing, virtualization technology, network optimization and etc.

Hong He, he received his B.S. degree at Wuhan University of Technology in 1996 and M.S. degree at Xiangtan University in 2006. Currently, he works in Hunan Institute of Engineering as an associate professor. His research interesting is grid computing, cloud computing, distributed resource management.

Xiao Peng, he received the Ph. D degree in computer science in CSU at 2010. He is currently an associate professor in the Hunan Institute of Engineering. His research interests include cloud computing, distributed resource management. He is a member of ACM and IEEE.