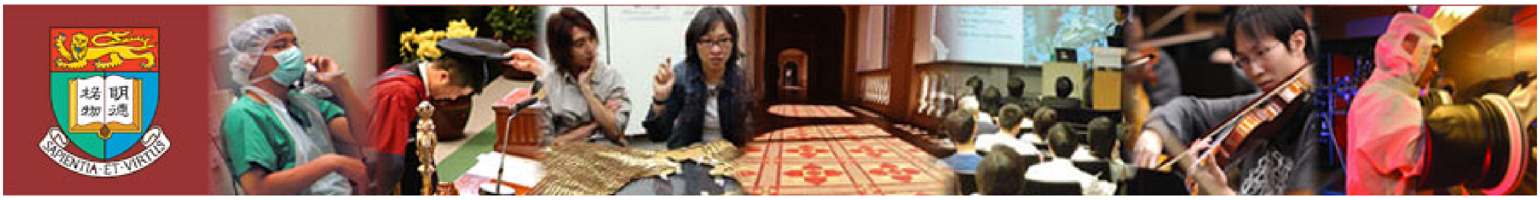

Competitive Cloud Resource Procurements via Cloud Brokerage

Xin Jin, Yu-Kwong Kwok, and Yong Yan
Department of Electrical and Electronic Engineering
The University of Hong Kong
Pokfulam, Hong Kong
Email: {tojinxin,ykwok}@eee.hku.hk, fuxin04@gmail.com

***Abstract*—In current IaaS cloud markets, tenant consumers non-cooperatively compete for cloud resources via demand quantities, and the service quality is offered in a best effort manner. To better exploit tenant demand correlation, cloud brokerage services provide cloud resource multiplexing so as to earn profits by receiving volume discounts from cloud providers. A fundamental but daunting problem facing a tenant consumer is competitive resource procurements via cloud brokerage. In this paper, we investigate this problem via non-cooperative game modeling. In the static game, to maximize the experienced surplus, tenants judiciously select optimal demand responses given pricing strategies of cloud brokers and complete information of the other tenants' demands. We also derive Nash equilibrium of the non-cooperative game for competitive resource procurements. Performance evaluation on Nash equilibrium reveals insightful observations for both theoretical analysis and practical cloud resource procurements scheme design.**

***Keywords*—Cloud computing, resource allocation, pricing scheme design, game theory, distributed learning.**

I. Introduction

The Infrastructure-as-a-Service (IaaS) view of cloud computing is widely adopted by several large cloud providers, which has fundamentally changed the operation of many industries [1]–[3]. Indeed, large cloud providers such as Amazon Web Services [4], Windows Azure [5], and Google App Engine [6] offer Internet-scale distributed computing facilities, where tenant users can dynamically reserve cloud resources including CPU, memory, and bandwidth so as to satisfy their own service requirements [7]. Tenants including application developers and small startups potentially reduce their investment risk and operating cost by renting computing and storage facilities from the cloud, while cloud providers benefit financially from multiplexing their data center networks [8]. In such a multi-tenant cloud computing environment, cloud brokers exploit demand correlation among tenants and obtain volume discounts from cloud providers via tenant demand aggregation. Therefore, tenants dynamically procure resources via cloud brokerage services due to lower offered price rates. In practice, tenants may be rejected of cloud services due to the inherent quantity competition among tenant consumers. In particular, tenant consumers judiciously decide optimal demand responses via tenant surplus maximization. Such demand competition, largely unexplored, fundamentally determines the tenant demand dynamics, which in turn affects the optimal pricing rules of both cloud brokers and the cloud provider in an interrelated market.

In this paper, we consider resource procurements from cloud brokers, and tackle the problem of tenant demand competition and competitive cloud resource procurements with a realistic broker pricing policy. In a practical cloud market, resource demands and prices will be cleared at an equilibrium level, where tenant consumers maximize their surplus and cloud brokers optimize the collected revenue given optimal tenant demand responses. Specifically, we propose a non-cooperative game to tractably investigate the competition among tenants for dynamic resource procurements and its impact on broker revenue and pricing scheme design. In this study of competitive cloud resource procurements, our specific contributions are three-fold.

In this paper, we build a general game model to realistically capture broker pricing scheme design. Tenant surplus (i.e., tenant utility minus dollar cost) is realistically formulated to model tenant rationality. We then analytically perform equilibrium analysis for competitive resource procurements under the assumption of perfect information. The remainder of this paper is organized as follows. In Section II, we present the system model of competitive cloud resource procurements via cloud brokerage. We propose a non-cooperative game and perform equilibrium analysis in Section III. Section IV validates our model with preliminary evaluation results on Nash equilibrium. In Section V, we articulate recent advances in cloud resource pricing. Finally, we present conclusions and future work in Section VI.

II. System Model

A. Cloud Brokers and Tenants

We consider a cloud system with multiple cloud brokers and a large number of tenant users. Cloud brokers share cloud resources such as CPU, memory, and bandwidth with tenant users as sellers. Tenant users as buyers dynamically procure cloud resources in units of virtual instances. A virtual instance is a resource bundle with one single resource type or a bundle of multiple resources. Throughout this paper, we study virtual instances as units of commodities sold in such a cloud system. Spot prices are provided by brokers so as to accommodate

978-0-7695-5095-4/13 $31.00 © 2013 IEEE
DOI 10.1109/CloudCom.2013.92

demand dynamics, which is widely used in realistic cloud markets, such as SpotCloud [9].

Denote by N the number of tenant users in the cloud system. The number of cloud brokers is M. The broker i sells the cloud resources at price rate p_i per virtual instance. Each broker provides services to multiple tenants, the demands of which depend on both the experienced service quality and the price charged by the cloud cloud brokers. The service quality of tenants is dependent on network delay (i.e., transmission delay due to request routing) and queueing delay (i.e., delay incurred by waiting for the service of the cloud broker). To this end, the queueing delay is implicitly considered in the pricing policy of each cloud broker, which will be elaborated in Section III. We explicitly consider network delay in the utility function of tenant users.

B. Tenant Competition in an Oligopoly Market

We approach the problem of dynamic cloud resource procurements via noncooperative game modeling of tenant users. In microeconomics, oligopoly describes a situation in which a small number of companies (i.e., oligopolists) dominate the entire market. This is what happens in current cloud markets, where several large companies such as Amazon and Microsoft dominantly occupy the market share with few cloud brokers. Under this market structure, the few cloud brokers own the control of cloud resource prices to tenant consumers. The decision making of each tenant user is to reserve cloud resources properly so as to maximize their individual benefits, which is quantified by tenant surplus in our study. Under the assumption of observable prices of different cloud brokers, the tenant users compete with each other and make demand requests in a noncooperative manner. The competition among tenant users is in terms of the resource demand, which is determined by the utility obtained from requested virtual instances and the price charged by cloud brokers. Each tenant user dynamically learns the equilibrium by adapting the amount of reserved cloud resources to the strategies of other tenant users. Indeed, tenants may also obtain cloud resources from cloud providers directly, but here we focus our discussions on cloud brokers due to the lower offered price rates [10].

III. Tenant Competition for Cloud Resource Procurements and Equilibrium Analysis

In this section, we build a game theoretic model for competitive resource procurements among tenant users. We first define the pricing scheme of cloud brokers and formulate our tenant surplus defintion. Based on this, we propose a game theoretic formulation to model the noncooperative competition among tenant users. However, this static game assumes that each tenant user possesses perfect information about strategies and surplus of all the other tenant users. To this end, we propose a dynamic game formulation, by relaxing the assumption of perfect information, so as to provide realistic resource procurement algorithms for tenant consumers.

A. Pricing Model of Data Centers and Tenant Surplus

The commodity sold in the cloud market is in the units of virtual instances. To model prices offered by cloud broker i, we consider a realistic pricing function:

$$p_i(\mathbf{d}_i) = \alpha + \beta \cdot \left(\sum_{j=1}^{N} d_{ij} \right)^{\tau}, \quad \forall i \in \{1, \cdots, M\}, \tag{1}$$

where d_{ij} is the amount of resources reserved by tenant j from cloud broker i, and $\mathbf{d}_i = [d_{i1}, \cdots, d_{ij}, \cdots, d_{iN}]^T$ is the vector of all resource demands at broker i. This practically reflects the situation that the price increases with the growth of aggregate demand at one cloud broker due to the limited amount of cloud resources reserved from cloud providers in the interrelated market. With the surge of resource prices, the demand will decrease accordingly. In this manner, the demand can be maintained at an equilibrium level so as to provide sufficient service quality as measured by the queueing delay.

Denote by l_{ij} the network delay due to tenant j's resource procurements from cloud broker i. L represents the maximum experienced network delay in the entire cloud system. Then, the utility of unit virtual instance can be modeled as

$$b_{ij} = \ln \left(1 + (L - l_{ij}) \right), \tag{2}$$

where $L \geq l_{ij}$ and L represents the maximum tolerated delay by tenant consumers. Then, the total utility obtained by tenant user j is $\sum_{i=1}^{M} b_{ij} \cdot d_{ij}$, with the financial cost of $\sum_{i=1}^{M} b_{ij} \cdot p_i(\mathbf{d}_i)$. Therefore the surplus of tenant j can be formulated as follows:

$$\begin{aligned} \pi_j(\mathbf{s}_j) &= \sum_{i=1}^{M} b_{ij} \cdot d_{ij} - \sum_{i=1}^{M} d_{ij} \cdot p_i(\mathbf{d}_i) \\ &= \sum_{i=1}^{M} b_{ij} \cdot d_{ij} - \sum_{i=1}^{M} d_{ij} \cdot \left(\alpha + \beta \cdot \left(\sum_{j=1}^{N} d_{ij} \right)^{\tau} \right) \end{aligned} \tag{3}$$

where $\mathbf{s}_j = [d_{1j}, \cdots, d_{ij}, \cdots, d_{Mj}]^T$ is a vector of tenant user j's demands from all the cloud brokers.

B. A Static Game and Nash Equilibrium

Based on the tenant surplus formulation in the above, we can formulate a noncooperative game among competing tenant users. In a static game, the most fundamental three elements are players, the strategy of each player, and the payoff of each player. The players in this game are all the tenant users. The strategy of each player (e.g., tenant user j) is the demand vector of resources reserved from different cloud brokers (i.e., $\mathbf{s}_j$ for tenant j). The payoff of each tenant user j is the surplus earned from the usage of cloud resources (i.e., $\pi_j(\mathbf{s}_j)$). We use Nash equilibrium to solve the game.

The Nash equilibrium of a game is a solution concept in which no player can increase his own payoff by unilaterally changing its own strategy. The Nash equilibrium can be obtained by solving the best response function, which is the optimal strategy of one player given the others' strategy choices. That is, the best response function of tenant j can be formulated as:

$$BR_j(\mathbf{S}_{-j}) = \arg\max_{\mathbf{s}_j} \pi_j(\mathbf{S}), \quad (4)$$

where $\mathbf{S} = [d_{ij}],\ \forall 1 \leq i \leq M,$ and $1 \leq j \leq N$ denotes the strategy matrix of all tenant users and $\mathbf{S}_{-j} = [d_{ik}]$ with $i \neq j$ represents the strategy matrix of all tenants except tenant j.

Denote by $\mathbf{S}^* = \{\mathbf{s}_1^*, \cdots, \mathbf{s}_j^*, \cdots, \mathbf{s}_N^*\}$ the Nash equilibrium of the noncooperative resource procurement game. Then, we have:

$$s_j^* = BR_i(\mathbf{S}_{-j}^*), \quad \forall j. \quad (5)$$

where the Nash equilibrium is given by the best response function. To this end, we can obtain the Nash equilibrium by solving the following equation array:

$$\begin{aligned} \frac{\partial \pi_j(\mathbf{s}_j)}{\partial d_{ij}} &= b_{ij} - \alpha - \beta \cdot \left(\sum_{j=1}^{N} d_{ij}\right)^{\tau} \\ &\quad - \beta \cdot \tau \cdot \sum_{i=1}^{M} d_{ij} \cdot \left(\sum_{j=1}^{N} d_{ij}\right)^{\tau-1} \\ &= 0. \end{aligned} \quad (6)$$

The solution $\mathbf{S}^*$ of the above equations is a Nash equilibrium. In practice, the tenant consumers set their optimal demand levels using the Nash equilibrium, given the pricing policies of the cloud broker. When all the strategies among all the tenant users are available in a centralized manner, the Nash equilibrium can be solved numerically by:

$$\min_{\mathbf{S} \succeq 0} \sum_{j=1}^{N} |\mathbf{s}_j - BR_j(\mathbf{S}_{-j})|, \quad (7)$$

where $|\mathbf{x}|$ is the norm of vector $\mathbf{x}$. That is, the Nash equilibrium can be solved by minimizing the sum of the differences between d_{ij} and the corresponding value obtained via best response functions. The closer to 0 the objective function is, the more accurate of the numerical solution.

In the following theorem, we investigate the analytical solution of Nash equilibrium for the special case of $M = 1$. That is, $b_{ij} = b_j$ and $d_{ij} = d_j$, $\forall i$.

THEOREM 1. For the special case of $M = 1$, there exists a unique Nash equilibrium given by

$$d_j^* = (\frac{b_j - \alpha}{\beta \cdot \tau \cdot Q^{\tau-1}} - \frac{Q}{\tau})^+, \forall 1 \leq j \leq M, \quad (8)$$

where $Q = \frac{\sum_{j=1}^{N} b_j - \alpha \cdot N}{\beta \cdot (N+\tau)}$ and $(x)^+ = \max(x, 0)$.

Proof. From Equation array 6, we get

$$\begin{aligned} \frac{\partial \pi_j(\mathbf{s}_j)}{\partial d_j} &= b_j - \alpha - \beta \cdot \left(\sum_{j=1}^{N} d_j\right)^{\tau} \\ &\quad - \beta \cdot \tau \cdot d_j \cdot \left(\sum_{j=1}^{N} d_j\right)^{\tau-1} \\ &= 0. \end{aligned} \quad (9)$$

Summing up the left side and the right side of the above equations, we have

$$\begin{aligned} &\sum_{j=1}^{N} b_j - \alpha \cdot N - \beta \cdot N \cdot \left(\sum_{j=1}^{N} d_j\right)^{\tau} \\ &- \beta \cdot \tau \cdot \left(\sum_{j=1}^{N} d_j\right)^{\tau} = 0. \end{aligned} \quad (10)$$

Suppose that $Q = \sum_{j=1}^{N} d_j$. We can readily get

$$Q = \left(\frac{\sum_{j=1}^{N} b_j - \alpha \cdot N}{\beta \cdot (N+\tau)}\right)^{1/\tau}. \quad (11)$$

Substitute Q into Equation 9, we obtain the unique Nash equilibrium:

$$d_j = \frac{b_j - \alpha}{\beta \cdot \tau \cdot Q^{\tau-1}} - \frac{Q}{\tau}. \quad (12)$$

However, this is on that condition that

$$d_j = \frac{b_j - \alpha}{\beta \cdot \tau \cdot Q^{\tau-1}} - \frac{Q}{\tau} \geq 0; \quad (13)$$

otherwise, the best response of tenant j is $d_j = 0$. To sum it up, we obtain the unique Nash equilibrium:

$$d_j^* = \max(\frac{b_j - \alpha}{\beta \cdot \tau \cdot Q^{\tau-1}} - \frac{Q}{\tau}, 0). \quad (14)$$

$\square$

IV. Performance Evaluation

In this section, we present our evaluation results of our proposed game model and learning algorithms in our dynamic game.

A. Setup

We consider a cloud system with one cloud broker and two tenant users procuring virtual instances from the broker (i.e., $M = 1$ and $N = 2$) so as to get clear insights about competitive cloud resource procurements. For the cloud pricing model, we use $\alpha = 0$ and $\beta = 1$.

B. Equilibrium Analysis

We first examine Nash equilibrium and the impact of network delay in Fig. 1 for the special case of two tenant users. In our game model, the best response of one tenant consumer is a linear function of the strategy of the other tenant user. The Nash equilibrium can be calculated by the intersection point of the best response functions of the two tenant users. Here, we investigate the impact of network delay on the equilibrium demand levels. With the decrease of network delay (i.e., better service quality), the corresponding tenant user would like to procure more resources from the cloud broker. On the other hand, the network delay of one tenant user affects the other's procurement of cloud resources. This clearly explains the impact of network delay and the interactions among tenants for resource procurements, when a large number of tenants coexist in the cloud system.

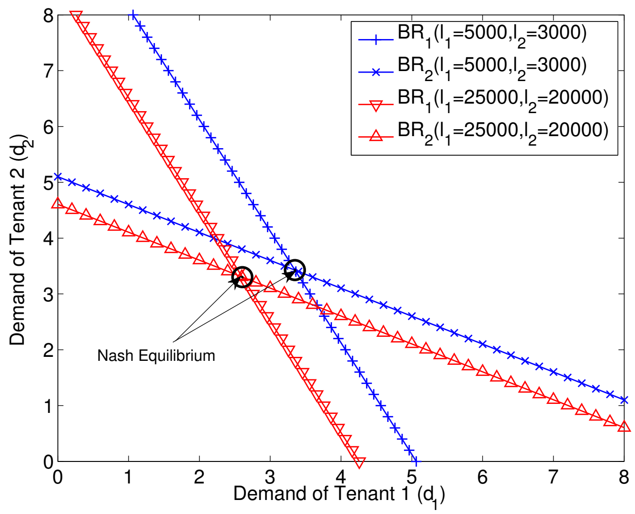

Fig. 1: Illustration of Nash equilibrium with two tenant users: best response functions.

V. Related Work

Pricing has been discussed for more than a decade by computer scientists for network resource allocation [11]. Recently, cloud resource pricing is widely adopted as the dominant resource allocation scheme in a cloud computing environment with multi-tenancy. Therefore, there already exist some studies on pricing scheme design and tenant resource procurements. Wang *et al.* [12] examine the importance of cloud resource pricing from the perspective of economics. Due to the co-existence of spot pricing and usage based pricing, Wang *et al.* [13] investigate optimal data center capacity segmentation between both pricing schemes with the objective of total cloud revenue maximization. Niu *et al.* [14], [15] propose a pricing scheme to better leverage the demand correlation among tenant consumers with VoD traffic and argue the necessity of brokers in a free cloud market. Most recently, Xu *et al.* [17] propose centralized schemes so as to maximize the revenue of the cloud provider. Wang *et al.* [10] investigate dynamic resource reservation via cloud brokers. Wang *et al.* further discuss optimal resource reservation with multiple purchasing options in IaaS clouds in [18]. While the above studies acknowledge the dominant role of the cloud provider and brokers in pricing, they ignore the competitive cloud resource procurements and its impact on broker revenue and pricing.

VI. Concluding Remarks

In this paper, we explore the problem of competitive cloud resource procurements in a cloud broker market. We realistically model the pricing scheme of the cloud broker and tenant surplus. We propose a noncooperative game to model such competitive resource procurements. We then conduct equilibrium analysis under the assumption of perfect information. To relax the assumption of perfect information, we propose the adoption of dynamic game to reach Nash equilibrium in a distributed manner by using local information only. The results revealed insightful observations for practical pricing scheme design. In the future, we would like to extend our model to the more general case of an interrelated market formulated by the cloud provider, brokers, and tenant consumers.

References

[1] C. Joe-Wong, S. Sen, T. Lan, and M. Chiang, “Multi-Resource Allocation: Fairness-Efficiency Tradeoffs in a Unifying Framework,” in *Proc. of INFOCOM*, March 2012.
[2] H. Ballani, P. Costa, T. Karagiannis, and A. Rowstron, “Towards Predictable Datacenter Networks,” in *Proc. of SIGCOMM*, August 2011.
[3] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. H. Andrew, “Greening Geographical Load Balancing,” in *Proc. of SIGMETRICS*, June 2011.
[4] Amazon EC2, 2013, http://aws.amazon.com/ec2/.
[5] Windows Azure Pricing Calculator, 2013, http://www.windowsazure.com/en-us/pricing/calculator/.
[6] Google App Engine, 2013, https://appengine.google.com/start.
[7] A. Ghodsi, M. Zaharia, B. Hindman, A. Konwinski, S. Shenker, and I. Stoica, “Dominant Resource Fairness: Fair Allocation of Multiple Resource Types,” in *Proc. of USENIX NSDI*, March 2011.
[8] B. C. Tak, B. Urgaonkar, and A. Sivasubramaniam, “Cloudy with a Chance of Cost Savings,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 6, pp. 1223–1233, June 2013.
[9] SpotCloud, 2013, http://spotcloud.com/.
[10] W. Wang, D. Niu, B. Li, and B. Liang, “Dynamic Cloud Resource Reservation via Cloud Brokerage,” in *Proc. of ICDCS*, July 2013.
[11] S. Shenker, D. Clark, D. Estrin, and S. Herzog, “Pricing in Computer Networks: Reshaping the Research Agenda,” *SIGCOMM Computer Communication Review*, vol. 26, no. 2, pp. 19–43, April 1996.
[12] H. Wang, Q. Jing, R. Chen, B. He, Z. Qian, and L. Zhou, “Distributed Systems Meet Economics: Pricing in the Cloud,” in *Proc. of USENIX HotCloud*, June 2010.
[13] W. Wang, B. Li, and B. Liang, “Towards Optimal Capacity Segmentation with Hybrid Cloud Pricing,” in *Proc. of ICDCS*, June 2012.
[14] D. Niu, C. Feng, and B. Li, “A Theory of Cloud Bandwidth Pricing for Video-on-Demand Providers,” in *Proc. of INFOCOM*, March 2012.
[15] ——, “Pricing Cloud Bandwidth Reservations under Demand Uncertainty,” in *Proc. of SIGMETRICS*, June 2012.
[16] H. Xu and B. Li, “A Study of Pricing for Cloud Resources,” *ACM SIGMETRICS Performance Evaluation Review, Special Issue on Cloud Computing*, March 2013.
[17] W. Wang, B. Li, and B. Liang, “To Reserve or Not to Reserve: Optimal Online Multi-Instance Acquisition in IaaS Clouds,” in *Proc. of ICAC*, June 2013.