International Journal of Scientific Engineering and Technology Research

www.ijsetr.com

ISSN 2319-8885
Vol.07,Issue.04,
April-2018,
Pages:0707-0710

Joint Pricing and Capacity Planning in the IaaS Cloud Market

B. SHIRESHA[1], DR. G. ANJAN BABU[2]

[1]PG Scholar, Dept of Computer Science, S.V.University, Tirupati, AP, India.

[2]Professor, Dept of Computer Science, S.V.University, Tirupati, AP, India.

Abstract: In the cloud context, pricing and capacity planning are two important factors to the profit of the Infrastructure-as-a-Service (IaaS) providers. This paper investigates the problem of joint pricing and capacity planning in the IaaS provider market with a set of Software-as-a-Service (SaaS) providers, where each SaaS provider leases the virtual machines (VMs) from the IaaS providers to provide cloud-based application services to its end-users. We study two market models, one with a monopoly IaaS provider market, the other with multiple-IaaS-provider market. For the monopoly IaaS provider market, we first study the SaaS providers' optimal decisions in terms of the amount of end-user requests to admit and the number of VMs to lease, given the resource price charged by the IaaS provider. Based on the best responses of the SaaS providers, we then derive the optimal solution to the problem of joint pricing and capacity planning to maximize the IaaS provider's profit. Next, for the market with multiple IaaS providers, we formulate the pricing and capacity planning competition among the IaaS providers as a three-stage Stackelberg game. We explore the existence and uniqueness of Nash equilibrium, and derive the conditions under which there exists a unique Nash equilibrium. Finally, we develop an iterative algorithm to achieve the Nash equilibrium.

Keywords: Pricing, Capacity Planning, Cloud Computing Market, Economics.

I. INTRODUCTION

Cloud computing provides an attractive paradigm for the dynamic provisioning of computing services in a "pay-as-you-go" manner. These services are typically classified into three categories: Infrastructureas-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS). With IaaS, such as Amazon EC2, each physical machine is virtualized into multiple virtual machines (VMs), and the computing resources are leased to cloud costumers in the form of VMs. With PaaS, such as Google App Engine, a computing platform is delivered on which cloud users can develop and run their applications. With SaaS, the applications can be accessed over the Internet by end-users without software related cost and effort. In general, SaaS providers utilize the internal resources of their own data centers or rent resources from a specific IaaS provider. Fig. 1 shows a typical three-tier cloud computing market. The end-users can get access to the applications provided by SaaS providers over the Internet. To serve their customers, SaaS providers lease computing resources from IaaS providers. The scenario we consider in this paper is that SaaS providers can not afford to establish their own data centers to server users' requests, and thus lease resources from IaaS providers without upfront investment in infrastructure and software. For example, Animoto is a company that creates videos out of images, music and video fragments submitted by end-users. It does not own a single server and bases its computing infrastructure entirely on Amazon Web Services. In this paper, we mainly focus on the IaaS layer and its interaction with SaaS layer.

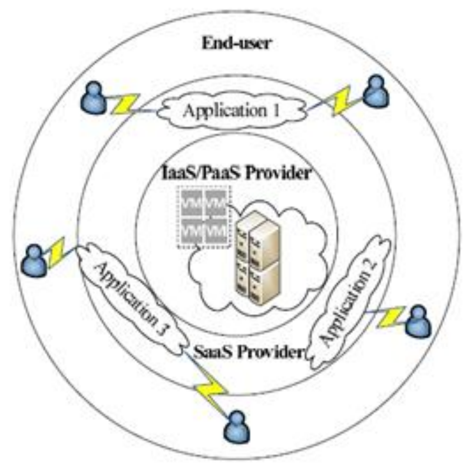

Fig.1. The three-tier cloud computing market.

In the cloud context, pricing is an important factor to the economics of the provider. Recently, lots of pricingbased scheme have been studied, including auction-style pricing, competition-based pricing game, etc. Most of these works focus on designing a better pricing scheme to maximize the cloud provider's profit with a fixed cloud capacity. It is worth noting that under-provision of the cloud capacity can cause resource shortage and revenue loss, while over-provision of the cloud capacity can result in idled resources and unnecessary energy cost. Notably, some works studied both the problem of optimal pricing and capacity right-sizing for the case with one IaaS provider, but the competition among multiple IaaS providers were not considered. Thus, we are motivated to study the problem of joint pricing and capacity planning in the IaaS cloud market with two market models: one with a monopoly IaaS provider, the other with multiple IaaS providers. In the market, both the SaaS providers and IaaS providers have their own interests. On one hand, the SaaS provider's objective is to maximize its revenue by serving end-users' requests and minimize the cost of rent resources while guaranteeing quality of service (QoS) (e.g. response time) specified in service level agreement (SLA). On the other hand, the IaaS provider aims to maximize its profit by optimizing the price and capacity. Since price has a great impact on SaaS providers' demand for resources, we first study the optimal decisions and behavior of SaaS providers. Then we investigate the IaaS provider's profit maximization problem in two market models.

We summarize below the main contributions of this paper:

- We formulate the profit maximization problem for each SaaS provider by taking SLA into consideration. Given the resource price, we derive the analytical expressions for its optimal decisions in terms of the amount of end-user requests to admit and the number of VMs to lease.
- For the monopoly IaaS provider market, we study the profit maximization problem for the IaaS provider by jointly optimizing the price and cloud capacity, and derive the optimal solutions.
- For the market with multiple IaaS providers, we formulate the pricing and capacity planning competition among the IaaS providers as a three-stage Stackelberg game. We derive the conditions where there exists a unique Nash equilibrium, and develop an iterative algorithm to reach the equilibrium.

II. RELATED WORK

We present a review of related works centered around pricing and capacity right-sizing in cloud environment, respectively.

A. Pricing in the Cloud Environment

In cloud computing, cloud resource pricing algorithm by employing financial option theory, where the cloud resources are treated as underlying assets to capture its realistic value. Menache proposed a usage-based pricing scheme to maximize the long-term social welfare, which comprises of the aggregate utility of users minus the operating cost of service provider. Kantere designed a novel demand-pricing model for cloud caching services and proposed a dynamic pricing scheme to maximize the cloud provider's profit. By exploiting the interplay between pricing and workload scheduling, Ren developed an online algorithm to maximize a wireless service provider's profit.

B. Capacity Right-sizing

Lin investigated the problem of dynamically right-sizing the data center by turning off servers, and proposed an online algorithm to achieve the energy saving. In cloud computing environment, Xu conducted price analysis with capacity right-sizing, and showed that the benefit of right-sizing critically depends on the unit cost of running the cloud. In the authors proposed an economic model that can be used to maximize the cloud provider's profit based on choosing the right-size in cloud data center. Most of previous work considered optimal pricing with fixed cloud capacity. Notably, in the authors studied both the optimal pricing and capacity right-sizing for the monopoly market. Our work differs from the previous papers in that we studied the joint pricing and capacity planning (right-sizing) problem for both the monopoly IaaS provider market and the market with multiple IaaS providers.

III. SYSTEM MODEL

We consider an IaaS provider market with a set of SaaS providers, denoted by $M = 1, 2, . . .,M$. The set of IaaS providers in the market is denoted by $N = 1, 2, . . .,N$. Each IaaS provider is modeled as a collection of homogeneous servers, each of which is virtualized to K VMs. Similar to the assumption in the existing works , we assume that the VMs are the same in terms of bandwith, CPU and storage, etc. The IaaS providers sell computational resources in the form of VMs. Each SaaS provider can lease the VMs from one IaaS provider to provide cloud-based application services to its endusers. For simplicity, we only consider a single application type. We consider a discrete time model, where the timeslot length matches the timescale at which the decisions of pricing and capacity planning are made, e.g., hourly. Without loss of generality, we model the IaaS provider's cloud capacity as the number of VM instances that the IaaS provider can support. Let x_j (t) denote the number of active servers in IaaS provider j at timeslot t, $\forall_J \in N$. Then the cloud capacity of IaaS provider j at timeslot t can be written as $c_j(t) = Kx_j(t)$. Note that the capacity model looks a little simplistic for the real world, but it could be a good start for the future analysis with the capacity model in practice. The IaaS providers adopt a pay-as-you-go charging model: SaaS provider i that requires $n_i(t)$ VMs from IaaS provider j incurs a monetary cost of $p_j(t)n_i(t)$ at timeslot t, where $p_j(t)$ denotes the price per VM set by IaaS provider j at timeslot t. Joint optimization of price and capacity periodically occurs at each timeslot. We therefore consider a single time slot, and drop the timeslot notation in what follows.

A. The SaaS Provider's Profit Model

For a given price p, SaaS provider i decides the enduser request rate $\lambda_i \in [\ 0, \lambda_i^m]$ to admit and the number of VMs n_i to rent, where λ_i^m denotes the maximum request arrival rate at the SaaS provider i. Note that λ_i^m can be predicted based on the prediction methods. Given the admitted request rate λi, SaaS provider i's utility by serving the requests can be written as vi $\log(1 + \lambda_i)$, where v_i denotes the utility level of SaaS provider i. The nonlinear utility function we adopted in the paper is the same as that in, which is based on the law of diminishing marginal utility in economics. SaaS provider i's profit is the difference between the utility obtained through serving the end-user requests and the cost sustained for using the VMs supplied by the IaaS provider, which can be written as

$$L_i(\lambda_i, n_i) = v_i \log(1+\lambda_i) - pn_i. \quad \textbf{(1)}$$

The end-user requests hosted in a VM is modeled as a M/G/1 queue. Let μ denote the maximum service rate on the VM. Assuming that the work load is evenly shared among the rent VMs, the average response time of SaaS provider i is given.

$$r(\lambda_i, n_i) = \frac{1}{\mu - \frac{\lambda_i}{n_i}} = \frac{n_i}{\mu n_i - \lambda_i}. \quad \textbf{(2)}$$

To guarantee the equilibrium conditions for theM/G/1 queues, the condition $\lambda i < \mu_{ni}$ must hold for a given μ, since otherwise the queue will go infinitely long and the system will not have a stationary distribution. Without loss of generality, the service rate of a VM is assumed to be 1 in what follows. We assume that the SLA between SaaS provider i and its users takes the form of an upper bound on the response time $r(\lambda_i, n_i)$, λ_i M. Thus, SaaS provider i should satisfy the constraint $r(\lambda i, ni) \leq di$, where di is the maximum average response time guaranteed by SaaS provider i.

B. The IaaS Provider's Profit Model

One common model of energy cost for a typical server is the affine function $e(\lambda') = e_0 + e_1 \lambda'$, where λ' is the end-user request arrival rate at this server, and e_0 models the fixed energy cost independent of workload, and e_1 is the energy cost per unit of end-user request rate to serve. For any given IaaS provider, let $\lambda'_1, \lambda'_2, \ldots, \lambda'_x$ denote the end-user request arrival rates at servers 1, 2, . . ., x, respectively. Then the energy cost of the IaaS provider can be written as

$$E(x, \lambda') = \sum_{j=1}^{x} (e_0 + e_1 \lambda'_j) = e_0 x + e_1 \sum_{j=1}^{x} \lambda'_j. \quad \textbf{(3)}$$

In the monopoly market with IaaS provider j, its profit is the difference between the revenue obtained by providing VMs to SaaS providers and the energy cost for maintaining the active servers, which is given by

$$P_j(p_j, x_j) = p_j \sum_{i=1}^{M} n_i - \left(e_0 x_j + e_1 \sum_{i=1}^{M} \lambda_i \right). \quad \textbf{(4)}$$

Note that the system model we considered is different from in our model, SaaS providers have direct interaction with end-users, and thus they are responsible for the delay of processing end-users' requests, i.e., each SaaS provider has a maximum response time guarantee for its users. Therefore, in this paper the delay cost is not considered in the IaaS provider's profit model.

IV. JOINT PRICING AND CAPACITY PLANNING IN THE MONOPOLY IAAS PROVIDER MARKET

Based on the optimal decisions of the SaaS providers, in this section we study the joint pricing and capacity planning for the IaaS provider to maximize its profit P(p, x) in the monopoly IaaS provider market. We assume that the SaaS providers send the information of guaranteed maximum average response time dis and the utility levels vis to the IaaS provider. In this section, we drop the subscript of p_j and x_j, and use p and x to denote the two decision variables of the IaaS provider in the monopoly market.

A.Pricing And Capacity Planning Game In The Multiple-Iaas-Provider Market

We consider the case where there are multiple IaaS providers in the cloud market. They compete with each other in terms of prices and available capacities in order to maximize their profit. We assume that both the IaaS and SaaS providers are selfish and rational. Observe that the interaction between the IaaS and SaaS providers is a typical leader-follower game that can be analyzed by using the Stackelberg game framework. Specifically, we cast this interaction as a three-stage Stackelberg game, where the IaaS and SaaS providers adapt their decisions dynamically to reach an equilibrium point. In the three-stage Stackelberg game, the IaaS providers first simultaneously determine their available capacities at Stage One, and then at Stage Two they simultaneously determine the VM prices to the SaaS providers. Finally, each SaaS provider chooses one IaaS provider to process its end-user requests at Stage Three.

V. CONCLUSION

This paper studied joint pricing and capacity planning in the IaaS provider market, where each SaaS provider leases VMs from the IaaS providers to provide cloudbased application services to its end-users. For the monopoly IaaS provider market, we first formulated the profit maximization problem for each SaaS provider, and derived its optimal decisions in terms of the amount of end-user requests to admit and the number of VMs to lease. Based on the best responses of the SaaS providers, we then derived the optimal solution to the problem of joint pricing and capacity planning to maximize the IaaS provider's profit. Next, for the market with multiple IaaS providers, we formulated the pricing and capacity planning competition among the IaaS providers as a threestage Stackelberg game. We characterized the conditions under which there exists a unique Nash equilibrium, and developed an iterative algorithm to achieve the equilibrium. For future work, it would be interesting to consider the heterogenous case where the VM instances provided by each IaaS provider are different. Also, it would be interesting to consider the case where the VM migration and consolidation are exploited to increase IaaS providers' profit further.

VI. REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. A view of cloud computing. Commun. ACM, 53(4):50–58, 2010.

[2] S. K. Garg, S. Versteeg, and R. Buyya. A framework for ranking of cloud computing services. Future Gneration Computer Systems, 29:1012–1023, 2013.

[3] L. Wu, S. Kumar Garg, and R. Buyya. Sla-based admission control for a software-as-a-service provider in cloud computing environments. J. Comput. Syst. Sci., 78(5):1280–1299, 2012.

[4] Y.C. Lee, C. Wang, A.Y. Zomaya, and B. Zhou. Profit-driven service request scheduling in clouds. In IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, 2010.
[5] R. Buyya, C. Vecchiola, and S.T. Selvi. Mastering cloud computing: foundations and applications programming. Elsevier, 2013.
[6] V. Abhishek, A. K. Ian, and P. Key. Fixed and market pricing for cloud services. In Proceedings of 7th Workshop on the Economics of Networks, Systems, and Computation, 2012.
[7] Q. Zhang, Q. Zhu, and R. Boutaba. Dynamic resource allocation for spot markets in cloud computing environments. In IEEE International Conference on Utility and Cloud Computing, 2011.
[8] H. Zhang, B. Li, H. Jiang, F. Liu, A. V. Vasilakos, and J. Liu. A framework for truthful online auctions in cloud computing with heterogeneous user demands. In Proceedings of International Conference on Computer Communications (INFOCOM), 2013.
[9] Q. Wang, K. Ren, and X. Meng. When cloud meets ebay: Towards effective pricing for cloud computing. In IEEE INFOCOM, 2012.
[10] H. Xu and B. Li. Dynamic cloud pricing for revenue maximization. IEEE Transactions on Cloud Computing, 1(2):158–171, July 2013.
[11] M. Macias and J. Guitart. A genetic model for pricing in cloud computing markets. In ACM SAC, 2011.
[12] T. Truong-Huu and C. Tham. A novel model for competition and cooperation among cloud providers. IEEE Transactions on Cloud Computing, PP(99), 2014.

Author's Profile:

B Shiresha received Bachelor of science (Computer science) degree from yogi vemana University, kadapa in the year of 2010-2013. Pursuing Master of Computer Applications from Sri Venkateswara University, Tirupati in the year of 2015-2018. Research interest in the field of Computer Science in the area of Cryptography-Network Security, Data Mining, Cloud Computing and Software Engineering.

dddddd

Dr. G Anjan Babu, working as an HOD & Professor in Dept. of Computer Science, Sri Venkateswara University College of Commerce Management and Computer Science, Tirupati (AP) India. Received Master of Computer Applications and Doctorate in Computer Science from S.V.University, Tirupati, Andhra Pradesh India and also attending in twenty three International and nineteen National conferences and also N.C.C: 1.C Certificate with A Grade 2.B Certificate with A Grade Two Nationals : 1.RDC, New Delhi 2.NIC, Balwade, Pune 3. diploma in Advanced Software Technology (eDAST) from CMC Ltd. Research interest in the field of Computer Science in the area of Data Mining, Neural network Graph databases and Big data.