Decision Support Systems 78 (2015) 80–92

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Pricing strategy for cloud computing: A damaged services perspective

Jianhui Huang [a], Robert J. Kauffman [b], Dan Ma [b,*]

[a] The Corporate Executive Board, Singapore
[b] Singapore Management University, Singapore

ARTICLE INFO

Available online 28 November 2014

Keywords:
Cloud computing
Damaged services
Pricing strategy
Services interruption
Spot prices

ABSTRACT

How effective is a hybrid pricing strategy for a cloud computing services vendor that mixes fixed-price reserved services with spot-price on-demand services? This research offers a decision support model to create the appropriate strategy for IT services based on prior research on information goods, electricity pricing, product versioning, and revenue yield management. The goal is to establish whether interruptible spot-price on-demand cloud computing services — which we view as *damaged services* — are valuable to the vendor. The results from the analysis of an economic model show that a hybrid strategy outperforms a one-service-only strategy in most cases, especially when clients are sensitive to services interruptions or when task values are highly differentiated. A more intriguing finding is that a vendor should permit the possibility of services interruptions even when clients are highly sensitive to their occurrence. The presence of interruptions serves as a quality differentiator between the on-demand services and reserved services, assuring the efficacy of the hybrid strategy. Moreover, a vendor may use capacity limit, in the hybrid strategy, as a tool to further improve its profit. To our knowledge, this research is the first to propose the damaged services perspective as an analogy for damaged goods in the cloud software market.

1. Introduction

Cloud services vendors deliver IT resources and software application services via the Internet. The services are scalable and accommodate fluctuations in client demand [2]. The cloud services market has grown rapidly over the past decade. According to Gartner [12], revenue in the global cloud services market was US$111 billion in 2012, a 21.4% increase from US$91.4 billion in 2011, and it is expected to reach US$206.6 billion by 2016.[1]

Cloud services suit clients with unpredictable demand for computing power and large batch processing tasks [2]. Flexible provision of services and usage-based pricing are key enablers [43]. As prices are driven lower [16], clients will rely on cloud computing for all IT-related services. Cloud services are more economical than in-house systems, especially for clients with data-intensive computing. The cost savings can be 95%, an indicator of the high value it provides [44,19].

Various pricing mechanisms have been adopted in the cloud market.[2] Since cloud services are consumed similar to utility services such as electricity or water, most vendors have applied usage-based pricing with services charged by the hour or minute, and client payments are tied to actual usage. Clients, however, have shown concern, since it is difficult to calculate total cost [21,22]. Innovative pricing schemes have been implemented recently. For example, Amazon introduced its Elastic Compute Cloud (EC2) in 2006, and charged hourly fees. In 2009, Amazon employed a new pricing method: clients could purchase a reserved-services contract by paying an upfront fee. Thereafter, they were allowed to access the IT resources as reserved-services instances. By 2009, Amazon was delivering its EC2 services as spot-price on-demand instances. Spot prices often are lower than reserved-services instance prices to encourage use of on-demand services. Amazon retains the right to interrupt running tasks and takes back the IT resources without notifying clients. The completion of spot-price on-demand services is not guaranteed.

Such interruptible spot-price services are an inferior version of interruption-free reserved services; clients will assign a lower value to

* Corresponding author at: 80 Stamford Road, School of Information Systems, Singapore Management University, Singapore 178902. Tel.: +65 68280926.
E-mail address: madan@smu.edu.sg (D. Ma).

[1] Three main types of cloud computing services initially characterized the market: infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), and software-as-a-service (SaaS). As the market matured, more categories emerged, such as data storage-as-a-service, hardware-as-a-service, desktop-as-a-service, business processes-as-a-service, and others [32].

[2] There are quite a few recent useful survey articles on the pricing of cloud computing services. The research covers: the economics of cloud computing and pricing [18]; the design of a pricing service for grid computing [6]; pricing models and their relationship to the quality of cloud services [1]; fixed and variable cost analysis for monopoly cloud services vendor who offers subscription and pay-per-use pricing [7] with two-part tariffs [8]; and strategic differentiation for cloud vendors with pricing approaches, including price clarity and transparency [20].

http://dx.doi.org/10.1016/j.dss.2014.11.001
0167-9236/

them compared to reserved services. The lower value is not associated with lower service costs on the vendor's side. Spot-price on-demand services share the same IT infrastructure as fixed-price reserved services, and require the same delivery effort by the vendor. Instead, the lower value comes from operational uncertainty: the vendor can interrupt a running task and reallocate the resources. It will deliberately impose interruption risks on clients and make some of its services less attractive as a result.

There is a strong analogy to traditional *damaged goods strategy* [10], and this is leverageable and useful for cloud services vendors. A vendor will downgrade or "damage" some features of a product to create a lower quality version to segment its market through price discrimination. This strategy is effective with software and information goods. But will a cloud vendor be effective by damaging its cloud services? This issue has not been studied.

We study a monopoly cloud services vendor with many potential clients. The vendor considers what to offer and how to price its services. The vendor can employ a *one-service-strategy* — either fixed-price reserved services that are interruption-free, or spot-price on-demand services with or without interruptions. The vendor can also use a *hybrid strategy*. This involves making fixed-price reserved services and spot-price on-demand services available simultaneously. Clients with different demand for IT services will choose what suits them the best. When is it beneficial for a vendor to use a hybrid strategy? What is the appropriate level of service interruption? And how are clients affected? These are the questions this work aims to answer.

This research offers rich insights. First, moving from offering fixed-price reserved services to spot-price on-demand services always improves vendor profit. When it only offers on-demand services, the vendor should not version service quality. Second, a hybrid strategy mostly outperforms a one-service-only strategy. It will lead to more market coverage, and higher profit and social welfare, but probably lower consumer surplus. Third, a hybrid strategy performs well when clients are sensitive to interruption or task values are highly differentiated. Fourth, with hybrid strategy, the vendor should version its spot-price services by creating services interruption risks. Their presence is desirable due to two functions. It works as the quality differentiator between reserved and on-demand services, so they will not compete intensely. It also provides the vendor with resource reallocation flexibility, so if demand outstrips supply in the reserved-services segment, the vendor can adjust to avoid serious economic and reputation loss. Finally, we also show that service provision capacity limits can be used as a tool to further improve vendor profitability with a hybrid strategy.

2. Literature review

This work is built on four streams of research: fixed and usage-based pricing for information goods, peak-load pricing, revenue management, and quality differentiation. Past research on pricing information goods has analyzed business models with fixed and usage-based fees. Sridhar et al. [33] modeled uncertain client usage demand with frequency and utility, and showed that a monopolist should employ usage-based pricing when transaction costs are low. With competition though, fixed pricing often outperforms usage-based pricing. This is shown by research on electronic goods sales with fixed-price subscriptions and pay-per-use fees [13,14]. Sundararajan [36] reported that a monopolist using both performed no worse in the presence of positive transaction costs, and sometimes better. Cachon and Feldman [5] reported that fixed subscription fees are effective for a monopoly services vendor when there is services congestion. In contrast, Wu and Banker [42] showed that a monopoly cloud services vendor does best to use two-part tariff pricing (usage-based pricing plus a fixed fee) with heterogeneous client demand when capacity is constrained.

We compare three types of pricing: fixed prices for reserved services, spot prices for on-demand services, and a mixture of them in a hybrid strategy. Our findings contribute to the debate about the fixed and usage-based pricing for information goods. This study has a key difference compared to all previous works. In our context, spot prices are associated with services interruption risk. So a client's task completion is not guaranteed when it uses spot-price services. The vendor may interrupt any running spot instances and limit the IT services offered. This creates the possibility that business functions run on cloud services are subject to unexpected service termination by vendors [17]. This has never been studied in past research.

The second line of work deals with *peak-load pricing* for non-storable products such as electric power, where price discrimination is appropriate, and linear [35] or convex [3] costs have been assumed. The relationship between marginal production cost and production capacity is critical [41]. Capacity planning was not considered though. We will analyze how capacity planning, together with hybrid pricing, can be effective for vendors.

The third line of research emphasizes *revenue yield management.* Pricing, along with inventory control and overbooking, is an important aspect [27]. Gallego and Van Ryzin [11] showed that, for airline revenue management, there is a connection between ticket prices and seat-allocation decisions. Li [24] proved that it is better to offer a small number of fare classes with different restrictions, such as no luggage. The airlines use spot prices with infrequent changes, and leverage high and low prices to ration capacity [9]. In contrast, Amazon.com's EC2 cloud services have used interruptible services to support resource allocation. We are interested in examining its impact on the vendor's revenue management.

Another stream focuses on *quality differentiation*, which enables vendors to segment the market and price discriminate. Large sunk costs and low variable costs for creating products of different quality make versioning suitable for value-based pricing and quality differentiation. Prior studies on information goods suggest that a limited number of versions of information goods with different quality levels should be offered. Varian [39] advocated different versions when clients have different preferences for quality, or their preferences are hard to observe. A rule of thumb is to offer only two versions: high and low quality. This works best when two groups of clients are different in their valuation of services [30,34]. Manufacturers also can improve their profits by damaging their goods and price discriminating [10]. The outcome depends on the value of the damaged and undamaged goods, rather than how client valuations are distributed [26]. We study service quality and price issues for cloud services vendors: the vendor can offer damaged and undamaged services, set interruption risk as a quality differentiator, and price to maximize profit.

3. The model

Consider a monopoly cloud services vendor that can offer two types of services. One is fixed-price reserved services. To use reserved services, clients must buy a contract. They pay a *reserved-services price* P_R in advance to reserve N_R units of computing resources. When the client requires a task to be done, it submits for reserved services without any additional payment. The vendor guarantees task completion since the resources required to execute the task have been reserved exclusively for this client (Appendix A has our modeling notation).

The second type of services the vendor can offer is spot-price on-demand services. Clients are not required to pay in advance. Whenever a task arrives, the client can submit it to the vendor. It knows the task will face service interruption risk: task completion is not guaranteed. Instead, there is a probability $r \in [0,1]$ that the vendor will interrupt the running task and retrieve the cloud resources. The interruption risk r is common knowledge in the marketplace.[3]

[3] Our model follows Amazon's practice of using publicly-available historical price information. Amazon EC2 spot price information is available at aws.amazon.com/console. It publishes 90-day spot price information to make it common knowledge. Customers are able to estimate services interruption risks based on this data if they use spot-price services.