RSCAS 2013/35
Robert Schuman Centre for Advanced Studies
Florence School of Regulation

Migration to the Public Cloud

Tong Wang

European University Institute
Robert Schuman Centre for Advanced Studies
Florence School of Regulation

Migration to the Public Cloud

Tong Wang

EUI Working Paper **RSCAS** 2013/35

ISSN 1028-3625

Printed in Italy, June 2013
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu/RSCAS/Publications/
www.eui.eu
cadmus.eui.eu

Robert Schuman Centre for Advanced Studies

The Robert Schuman Centre for Advanced Studies (RSCAS), created in 1992 and directed by Stefano Bartolini since September 2006, aims to develop inter-disciplinary and comparative research and to promote work on the major issues facing the process of integration and European society.

The Centre is home to a large post-doctoral programme and hosts major research programmes and projects, and a range of working groups and *ad hoc* initiatives. The research agenda is organised around a set of core themes and is continuously evolving, reflecting the changing agenda of European integration and the expanding membership of the European Union.

Details of the research of the Centre can be found on:
http://www.eui.eu/RSCAS/Research/

Research publications take the form of Working Papers, Policy Papers, Distinguished Lectures and books. Most of these are also available on the RSCAS website:
http://www.eui.eu/RSCAS/Publications/

Florence School of Regulation

The Florence School of Regulation (FSR) is a partnership between the Robert Schuman Centre for Advanced Studies (RSCAS) at the European University Institute (EUI), the Council of the European Energy Regulators (CEER) and the Independent Regulators Group (IRG). Moreover, as part of the EUI, the FSR works closely with the European Commission.

The objectives of the FSR are to promote informed discussions on key policy issues, through workshops and seminars, to provide state-of-the-art training for practitioners (from European Commission, National Regulators and private companies), to produce analytical and empirical researches about regulated sectors, to network, and to exchange documents and ideas.

At present, its scope is focused on the regulation of Energy (electricity and gas markets), of Communications & Media, and of Transport.

This series of working papers aims at disseminating the work of scholars and practitioners on current regulatory issues.

For further information
Florence School of Regulation
Robert Schuman Centre for Advanced Studies
European University Institute
Via Boccaccio, 151
I-50133 Firenze
Tel.: +39 055 4685 751
Fax: +39 055 4685 755
E-mail: fsr@eui.eu
http://www.eui.eu/RSCAS/ProfessionalDevelopment/FSR/

Abstract

Along with the development of cloud computing technology, website owners begin to consider migrating their website from private in-house server to public cloud servers. In this paper, we use a principal-agent model to analyze the underlying economic trade-offs of such migration and then extend it into a dynamic environment. Our results indicate that the trade-off between market information precision and rent extraction affects the decision choice between private server and public cloud in the short run; in the long run, the rent extraction effect diminishes and the demand for public cloud increases. In a long run equilibrium, private servers exist but are constrained to a comparatively low level.

Keywords

Cloud Computing, Agency Cost, Dynamic Migration

1 Introduction*

This article is to analyze factors that affect firms' decision about adopting a public cloud infrastructure or a private one. Although cloud computing is a rather new concept that emerges in recent years, the basic idea was similar to so-called client-server mode; in both cases, there is a data centre that provides storage and/or computing services via local network or Internet to an amount of terminals that may geographically distributed far from each other. From the user perspective, there are quite similar, while from the business perspective, cloud computing bring "pay-as-you-go" pattern and auto-scalability properties, which seems more promising especially for SMEs (small and medium enterprises).

Undoubtedly, cloud computing shows us an alternative way of managing and delivering computing service. Some commentators even think that cloud computing represents the future of the development of IT infrastructure and it will subsequently change the nature of computing(Fox and Griffith 2009), however, other observers argue that cloud computing is simply another kind of outsourcing. Despite of the argument about the perspective of cloud computing, the undebatable phenomenon is that, the market of cloud computing has been so large that most of the established IT incumbents (Microsoft, Google, Amazon and etc.) has been involved and all the participants and players in IT-related industry are forced to take a serious consideration of the emergence of this phenomenon and carefully evaluate both the advantages and disadvantages of cloud computing.

Yoo (2011) and Kim et al. (2010) summarize the debates about cloud computing and give some policy implications as well. They enumerate some important benefits of adopting cloud computing, such as avoiding up-front payment, increasing reliability and etc., among all those benefits, the main advantage of cloud computing, in their paper, is aggregate demand, which means that because the cloud essentially gathers a huge amount of cloud users, whose demands are imperfectly correlated, the variability of the aggregate demand will be reduced. Therefore, the underlying hardware (the processor ability) can be utilized more efficiently and the demand peaks can be handled with less cost.

In the work of Chandran et al. (2010), they point out several drawbacks and potential risk of cloud computing: loss of control, security and etc. The major issue is that the server of cloud users

*Thanks very much for the comments of Jacques Cremer, Zonglai Kou, Patrick Rey for useful comments and thanks for the financial support from Florence School of Regulation, European University Institute, most of the works are done there.

are supervised and maintained by a third party – the cloud service provider instead of themselves, this agency problem leads a possibility of adverse selection and moral hazard behavior of cloud service providers. To figure out the trade-offs of adopting cloud computing, it is meaningful to use economic model to analyze the relationship between cloud users and cloud service providers further.

Current literature mainly concerns the legal or secure perspective of cloud computing (Sluijs, Larouche, and Sauter 2011, Molnar and Schechter 2010, Maier-Rigaud 2011), especially on the privacy policy, security and proper right of the data. For the agency problem, Kim and Moskowitz (2011)first introduce principal-agent framework to analyze the relationship between cloud users and cloud service providers. In their model, the cloud users are the principal and the cloud service providers are the agents with the limited liability constraint, there is no additional asymmetric information between them so it is a pure moral hazard problem. The result is aligned with Holmostrom (1979), that the principal should implement an incentive scheme and leave some rents to the agents.

In this paper, we further explore the interaction between a public could service provider (CSP) and a website owner who may or may not delegate his website to the public cloud. If the website owner (WO) decides to run and maintain the severs by an in-house server (a private cloud), he can only purchase limited computing capacity, and all the market demand that exceeds this capacity will not be served; instead, if the WO decides to delegate his server onto the public cloud and sign contract with the CSP, potentially all the market demand can be satisfied due to the auto-scalability property (Jie, Jie, and Ying 2009) of public cloud computing and almost unlimited resources of CSP, but the moral hazard problem of the CSP will prevent the WO from achieving his first-best profit. Our results shows that the WO intends to run his server by himself when the per-customer cost are neither too large nor too small. The reason is that, when the per-customer cost is small, the WO can easily purchase sufficient computing capacity to fulfill the market demand; otherwise when the per-customer cost is large, auto-scalability property of the cloud computing is not attractive enough for the WO to migrate his website onto the cloud.

After considering the two polar cases: pure in-house sever and pure public cloud, we also study the hybrid cloud, in which case the WO builds an in-house server to satisfy basic market demand and delegate the extra demand to the CSP, the most interesting finding is that, the optimal scale of the in-house server without the option of cloud computing is not necessarily larger than that with such an option. The underlying tradeoff is that, with the option of public cloud, the substitution

effect between in-house server and cloud computing leads the WO to decrease the scale of in-house server; however, to squeeze the message space of the CSP and to decrease possible distortion yield by asymmetric information, the WO also has incentives to increase the scale.

We then extend our model into a dynamic environment, by adopting a recursive envelop theorem method (Pavan, Segal, and Toikka 2010), we find that, under some plausible assumptions, the distortion led by asymmetric information vanishes along with the time, therefore from the long run, it is optimal for the WO to decrease the scale of in-house server and migrate more onto the cloud. This results explain part of the trend that cloud computing becomes more and more popular not only within IT firms but also among those who have a comparative large IT department.

This paper is organized as follows: in section 2, we present the model setting and discuss the polar cases and hybrid cloud as well. In section 3, we extend the model into a dynamic environment and discuss several extensions, and section 4 concludes.

2 The Basic Model

We have three kinds of players: one cloud service provider (CSP), one WO and a continuum of consumers. Consumers want to visit the website to view some content. The website can either be built in-housely by the WO or be delegated to the CSP. In the following subsection, we characterize the interaction between consumers and the website.

2.1 Interaction Between Consumers and the Website

The number of consumers θ is a stochastic variable and is randomly drawn from a distribution $F(\theta)$ with a support $[\underline{\theta}, \overline{\theta}]$.

On the side of website, making the website work properly requires sufficient computing capacity s, which represents the synthetic data process ability of a certain amount of CPU time, hard disk volume and etc.

Every customer visiting the website yields a revenue r to the WO, the visiting of customers contains two sub-periods, on-peak and off-peak. the computing capacity cannot be altered between sub-periods. All the customers will initialize their first visiting in the on-peak period, denote θ as the numbers of customers and s as the capacity of in-house servers, If the realized customer number $\theta \leq s$, all the consumers are served; otherwise If $\theta > s$, only s consumers are randomly picked

from the consumer pool and the rest is rejected. Among these rejected customers, a proportion β of them will attempt to access the website in the off-peak sub-period, we assume that β is small enough so that there are always excessive capacity in the off-peak sub-period[1]. the total consumer number will thus be $s + \beta(\theta - s)$.

We also assume that, if the customer demand is known to the WO, fully accommodating all the consumers is profitable, that is, β must satisfy:

Assumption 1 $\beta < 1 - \frac{\kappa}{r}$.

This assumption ensures that whatever the consumer demand is, it is always profitable for the WO to accommodate all the consumers in their first visiting.

Moreover, we assume that severs need appropriate monitor. With the minimum monitor effort, the website incurs some loss L and the extend of loss can be decreased by exerting the monitor effort e at a cost $C(e) = \frac{e^2}{2}$. We normalize the initial loss equal to L and minimum effort cost 0. the loss after exerting effort is $(1 - e)L + \varepsilon$, where ε is a random noise with zero mean. In another word, the monitor effort cannot be perfect monitored and there is thus a potential moral hazard problem.

In a word, given θ, L and s, the WO's expected gross revenue $\pi = r(\min\{\theta, s\} + \beta \max\{\theta - s, 0\}) - (1 - e)L$ and the net profit is:

$$V = \pi - \frac{e^2}{2} - \kappa s + \varepsilon$$

To avoid trivial cases, we assume that the participation constraint of the WO is always satisfied. In the following subsection, we discuss two polar cases: the WO relies on solely on in-house servers and cloud servers respectively.

2.2 Polar Cases

If the WO builds the website in-house, she cannot observe the market demand ex-ante, therefore, she must decide the capacity s before she learns θ.

The WO's expected profit in-house is

$$V^i(\theta) = \int_{\underline{\theta}}^{s} r\theta f(\theta) d\theta + \int_{s}^{\bar{\theta}} r[s + \beta(\theta - s)] f(\theta) d\theta - (1 - e)L - \frac{e^2}{2} - \kappa s$$

[1] we also assume out the case that the WO purchase little computing capacity and extremely relies on re-visiting consumers.

The WO chooses s and e to maximize the expected profit. The FOCs are:

$$\begin{aligned} r(1-\beta)[1-F(s^i)] &= \kappa \\ e^i &= L. \end{aligned}$$

The second-order condition of s^i is $-r(1-\beta)f(s^i) < 0$, which indicates that s^i is indeed the maximum. Clearly, $s^i = F^{-1}[1 - \frac{\kappa}{r(1-\beta)}]$, we have the following proposition:

Proposition 1 *The optimal in-house website capacity decreases with the computing capacity cost κ and the re-visiting rate β, while it increases with the per-consumer profitability r.*

The reason for κ and r is simple: investing more ex-ante decreases the risk of capacity shortage but increases the risk of capacity waste, and a larger marginal cost of server κ amplifies the risk of waste and thus pushes the optimal level downgrades, while a larger per-consumer profitability r increases the attraction of accommodating more possible consumers. As for the re-visiting rate β, a larger β implies that the WO losses less when the true consumer numbers exceeds s, thus makes a large pre-investment on the computing capacity unnecessary.

Now consider full cloud computing scenario, the timing is as follows:

- The WO offers a contract menu $\{T(\theta), \pi(\theta)\}$to the CSP, if rejected, the game is over, and both parties obtain 0 payoff; if accepted:
- The CSP observes the consumers' request θ and secretly decides the allocation of computing capacity $\hat{\theta}$ and the effort level e.
- The WO realizes a gross profit $w(\hat{\theta}, \theta)$,
- The CSP receives a payment $T(\hat{\theta})$.

One of the prominent advantages of cloud service is auto-scalability(Jie, Jie, and Ying 2009), that is, it is comparatively easy and quick for the WO to adjust the computing capacity on the cloud. Therefore, by delegating servers on the cloud, the WO not only avoids the up-front investment on the in-house website and no longer have problems about computing capacity, Instead, the WO now faces an agency problem, since with the monitor tools provided by the CSP, she can not learn the true consumer number but only observe $\hat{\theta}$ the signal sent by the CSP. Thanks to the revelation principle, we can restrict our attention to direct mechanism, that is, in equilibrium, the

CSP truthfully reveals the information to the WO. Moreover, to use the local incentive constraint instead of the global one, we add the following assumption:

Assumption 2 *The hazard rate of θ is everywhere non-increasing in the field of definition.*

Now we start to solve the equilibrium. the maximization problem of the WO is

$$\max \int_{\underline{\theta}}^{\overline{\theta}} T(\hat{\theta}) f(\theta) d\theta$$
$$s.t. \theta \in \arg\max_{\hat{\theta}} w(\hat{\theta}, \theta)$$

in which $w(\hat{\theta}, \theta)$ refers to the payoff of the CSP:

$$w(\hat{\theta}, \theta) = \pi(\hat{\theta}) - T(\hat{\theta}) - \frac{e^2(\theta, \hat{\theta})}{2} - \kappa\hat{\theta}.$$

Here $e(\theta, \hat{\theta})$ is defined by the equation

$$\pi(\hat{\theta}) = r[\hat{\theta} + \beta(\theta - \hat{\theta})] - (1 - e(\theta, \hat{\theta}))L.$$

According to the envelop theorem, the information rent R satisfies:

$$\frac{\partial R}{\partial \theta} = \frac{\partial w(\theta, \hat{\theta})}{\partial \theta}|_{\hat{\theta}=\theta} = \frac{r\beta}{L} e(\theta),$$

It follows that

$$R(\theta) = \int_{\underline{\theta}}^{\theta} \frac{\partial R}{\partial i} di = \int_{\underline{\theta}}^{\theta} \frac{r\beta}{L} e(i) di.$$

The WO's expected profit is the expected total surplus minus the expected information rent transferred to the CSP:

$$V^c = \int_{\underline{\theta}}^{\overline{\theta}} [w(\theta) + T(\theta)] f(\theta) d\theta - \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\theta}}^{\theta} \frac{r\beta}{L} e(i) di dF(\theta)$$

Integrated by part[2], the WO's profit becomes:

$$V^c = \int_{\underline{\theta}}^{\overline{\theta}} [r\theta - (1 - e)L - \frac{e^2}{2} - \kappa\theta - \frac{r\beta}{L} \frac{1 - F(\theta)}{f(\theta)} e] f(\theta) d\theta$$

Through the first-order condition, it is easy to get the second best effort:

$$e^{SB} = L - \frac{r\beta}{L} \frac{1 - F(\theta)}{f(\theta)}.$$

[2] $\int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\theta}}^{\theta} e(i) di dF(\theta) = F(\theta) \int_{\underline{\theta}}^{\theta} e(i) di |_{\underline{\theta}}^{\overline{\theta}} - \int_{\underline{\theta}}^{\overline{\theta}} F(\theta) e(\theta) d\theta = \int_{\underline{\theta}}^{\overline{\theta}} [1 - F(\theta)] e(\theta) d\theta.$

The analytical solution of the second-best effort shows standard properties "no distortion at the top, downward distortion at the bottom" in incentive theory (**?**). An interesting observation is that the distortion diminishes along with β, the re-visiting rate. The underlying logic is that when β is close to 0, the CSP can hardly earn extra benefit by mis-reporting the consumer number to the WO, therefore, from the perspective of the WO, the necessity of downward distorting the effort level decreases.

Substituting into the expression of V, the WO's profit thus is:

$$V^{c*} = \int_{\underline{\theta}}^{\overline{\theta}} [(r-\kappa)\theta - L + \frac{1}{2}(e^{SB})^2] f(\theta) d\theta.$$

It is clear that in such a case, the firm can fully satisfy the customers' request at the price of paying the CSP enough information rent.

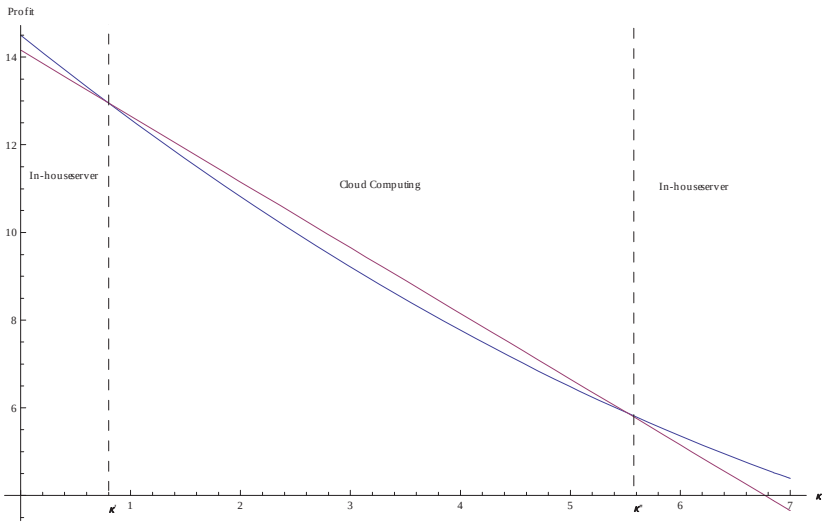
According to the analysis of these two polar cases, Depending on κ and $F(\theta)$, either regime can yield a competitively higher revenue for the firm.

Proposition 2 *The WO is more likely to adopt cloud computing when the capacity cost κ is neither too large nor too small.*

Proof. See Appendix 1. ■

Compared with in-house servers, cloud computing offers higher capacity flexibility to the WO while it also brings asymmetric information about the real number of consumers and related moral hazard on server monitor as well. When κ becomes extremely large, to accommodate more consumers is less profitable to the WO, which implies that the flexibility is essentially not so attractive, the WO, consequently chooses in-house server; Alternatively, when κ becomes extremely small, the efficiency loss that rises from the downward distortion of the monitor effort becomes more significant, thus the WO still choose in-house servers. Here we present a numerical example to illustrate this trade-off:

Example 1 $\theta \in U[1,2], F(\theta) = \theta - 1. r = 70, \beta = 0.1, L = 40.$



The x-axis is the unitary cost of capacity κ, and the y-axis is the profit of WO. The red line represent the profit under cloud computing while the blue line represents the profit under in-house servers.Transparently, when $\kappa < \kappa'$ or $\kappa > \kappa''$, cloud computing yields higher profit for the WO, while in-house servers generate higher payoff when κ located in between.

2.3 Hybrid Cloud

In many circumstance, the website owner can adopt a hybrid form of cloud, which includes both in-house servers and cloud servers, in Mazhelis et al. (2012), hybrid cloud is more likely to be cost-efficient. Now we consider such a hybrid case: the firm can build an in-house server with capacity s to satisfy a basic demand, and use the cloud service as a complementary to accommodate the demand that exceeds the capacity of in-house servers (if any). The corresponding conditional distribution function $F(\theta|\theta > s) = \frac{F(\theta)-F(s)}{1-F(s)}, f((\theta|\theta > s)) = \frac{f(\theta)}{1-F(s)}$[3]. The new timing is as follows:

- The WO builds her own capacity s
- The WO offer a contract $\{T(\theta), \pi(\theta)\}$with the CSP, if rejected, the game is over, and both parties obtain 0 payoff; if accepted:

[3]Although the total consumer number are no longer directly observable by the CSP, Signing a contract on a reported θ is equivalent to signing a contract on a reported excess demand. The reason is that in equilibrium, the WO's capacity are rationally expected by the CSP, therefore, once the CSP observes an excess demand, the CSP can deduct the total consumer demand and reported to the WO.

- The CSP observes the consumers' request θ and decides the allocation of computing capacity $\hat{\theta}$ (reported).
- The WO choose his monitor effort level e^i and the CSP chooses e^c.
- The CSP realizes a profit $w(\hat{\theta}, \theta)$,
- The WO receives a payment $T(\hat{\theta})$.

Since the WO and the CSP exert their monitor effort independently, therefore Accordingly, the profit of the WO can be divided by two parts: the profit from in-house servers is

$$V^i = \int_{\underline{\theta}}^{s} [r\theta - (1-e^i)L - \frac{(e^i)^2}{2}] f(\theta) d\theta - \kappa s,$$

while the profit for the CSP is the solution of the following problem:

$$\begin{aligned} V^c &= [1-F(s)] \int_s^{\overline{\theta}} T(\theta) f(\theta|\theta > s) d\theta \\ s.t. \theta &\in \arg\max_{\hat{\theta}} w(\theta, \hat{\theta}) \end{aligned}$$

in which $w(\hat{\theta}, \theta)$ refers to the payoff of the CSP:

$$w^h(\hat{\theta}, \theta) = \pi(\hat{\theta}) - T(\hat{\theta}) - \frac{(e^c)^2(\theta, \hat{\theta})}{2} - \kappa(\hat{\theta} - s).$$

Here the effort exerted by the CSP, $e(\theta, \hat{\theta})$ is defined by the equation

$$\pi^h(\hat{\theta}) = r[\hat{\theta} + \beta(\theta - \hat{\theta})] - (1 - e^c(\theta, \hat{\theta}))L.$$

Applying the envelop theorem, the information rent must satisfy:

$$\frac{\partial R}{\partial \theta} = \frac{\partial w(\theta, \hat{\theta})}{\partial \theta}|_{\hat{\theta}=\theta} = \frac{r\beta}{L} e^c(\theta),$$

which is equivalent to that of the polar case, correspondingly,

$$R(\theta) = \int_s^{\theta} \frac{\partial R}{\partial i} di = \int_s^{\theta} \frac{r\beta}{L} e^c(i) di.$$

The expected profit of the WO can now be reformulated as

$$\begin{aligned} V^c &= [1-F(s)] \int_s^{\overline{\theta}} [(r\theta - \kappa(\theta - s) - (1-e^c)L - \frac{(e^c)^2}{2} - \int_s^{\theta} \frac{r\beta}{L} e^c(i) di] f(\theta|\theta > s) d\theta \\ &= \int_s^{\overline{\theta}} [(r-\kappa)\theta + \kappa s - (1-e^c)L - \frac{(e^c)^2}{2} - \frac{r\beta}{L} \frac{1-F(\theta)}{f(\theta)} e^c] f(\theta) d\theta \end{aligned}$$

Therefore, the total expected profit of the WO under a hybrid cloud scheme is $V^h = V^i + V^c$. More formally, the WO solves the following maximization problem:

$$\max_{s,e^i,e^c} \int_{\underline{\theta}}^{s} [r\theta - \kappa s - (1-e^i)L - \frac{(e^i)^2}{2}]f(\theta)d\theta \qquad (*)$$
$$+ \int_{s}^{\overline{\theta}} [(r-\kappa)\theta - (1-e^c)L - \frac{(e^c)^2}{2} - \frac{r\beta}{L}\frac{1-F(\theta)}{f(\theta)}e^c]f(\theta)d\theta$$

According to the F.O.Cs, the equilibrium effort levels of WO and CSP are:

$$\begin{aligned} e^{i*} &= L \\ e^{c*} &= L - \frac{r\beta}{L}\frac{1-F(\theta)}{f(\theta)} \end{aligned}$$

Therefore equation (*) can be simplified

$$\max_{s,e^i,e^c} \quad \int_{\underline{\theta}}^{s} (r\theta - \kappa s - L + \frac{L^2}{2})f(\theta)d\theta$$
$$+ \int_{s}^{\overline{\theta}} [(r-\kappa)\theta - L + \frac{(e^{c*}(\theta))^2}{2}]f(\theta)d\theta$$

Consequently, the optimal in-house building s^* is characterized by the following equation

$$[(r-\kappa)s - L + \frac{L^2}{2}]f(s) - \kappa F(s) - [(r-\kappa)s - L + \frac{(e^{c*}(s))^2}{2}]f(s) = 0,$$

simplifying the equation above leads to

$$\underbrace{[\frac{L^2}{2} - \frac{(e^{c*}(s))^2}{2}]f(s^*)}_{\text{Squeeze Effect(Efficiency Gain)}} + \underbrace{[1-F(s^*)]\kappa}_{\text{Substitute Effect}} = \kappa.$$

By such a re-formulating, the RHS is κ, which is the same as the F.O.C. in the pure in-house scheme, while the LHS can be separated into two parts: the first part is the difference between First-best profit and Second-best profit of the WO, which implies that by marginally increasing the capacity of in-house servers, the WO can decrease the information rent and take in charge of these marginal consumers with a higher effort level, thus there exists an efficiency gain which stipulate the WO to invest more on the in-house capacity; for the information perspective, this also means that the WO can squeeze the message space of the CSP, so we call it squeeze effect; on the other hand, by assumption 1, unlike in the pure in-house servers case, a proportion of the excessive consumers, if any, is always rejected by the servers due to limited capacity, adopting the cloud can

accommodate all the excessive consumers. Therefore, the WO also has an intention to build less in-house capacity. The optimal in-house capacity s^* is a balance of these two effects, which can be either larger or smaller than s^c, the capacity under pure in-house server case. We conclude these results in the following proposition:

Proposition 3 *Hybrid cloud yields higher expected profit than both polar cases. Depending on parameters, the optimal in-house server capacity under hybrid cloud scheme s^* can be either larger or smaller than s^i, the optimal capacity under pure in-house scheme. More specifically,*

- *if $\kappa \to r(1-\beta)$, $s^* > s^i$;*
- *if $L \to \infty$, $s^i > s^*$;*

Proof. See appendix. ■

3 Extensions

3.1 Endogenizing the Cloud Price

In our basic model, we assume that the cloud price is exogenous and equal to the cost κ, now we try to figure out the price strategy of the CSP under hybrid cloud scheme[4]. As for the timing, we add an additional step before s is determined,

- The CSP determines unitary cloud price p.
- The WO builds her own capacity s
- The WO offers a contract $\{T(\theta), \pi(\theta)\}$with the CSP, if rejected, the game is over, and both parties obtain 0 payoff; if accepted:
- The CSP observes the consumers' request θ and decides the allocation of computing capacity $\hat{\theta}$.
- The WO choose his monitor effort level e^i and the CSP chooses e^c.

[4]Since in the pure cloud scheme, the CSP always sets the unitary price up to the point that the WO is indifferent in choosing in-house servers or cloud servers. See appendix for some illustrative simulations.

- The CSP realizes a profit $w(\hat{\theta}, \theta)$,
- The WO receives a payment $T(\hat{\theta})$.

Respectively, the profit of the WO from in-house server and cloud server are:

$$\begin{aligned} V^i &= \int_{\underline{\theta}}^{s} [r\theta - (1-e^i)L - \frac{(e^i)^2}{2}] f(\theta) d\theta - \kappa s, \\ V^c &= \int_{s}^{\overline{\theta}} [(r-p)\theta + ps - (1-e^c)L - \frac{(e^c)^2}{2} - \frac{r\beta}{L}\frac{1-F(\theta)}{f(\theta)} e^c] f(\theta) d\theta \end{aligned}$$

while $\max_{s,e^i,e^c} V^i + V^c$ yields:

$$[\frac{L^2}{2} - \frac{(e^{c*}(s^*))^2}{2}] f(s^*) + [1 - F(s^*)]p = \kappa$$

Clearly, a higher p weakens the substitute effect, thus increases s^*, that is, $\frac{\partial s^*}{\partial p} > 0$. Subsequently, the maximization problem with respect to p is

$$\max_p \int_{s^*}^{\overline{\theta}} (p-\kappa)(\theta - s^*) f(\theta) d\theta + R(\theta),$$

where R is the information rent which has nothing to do with p. The F.O.C. leads to

$$(1 - \frac{\partial s^*}{\partial p})(p-\kappa) s^* f(s^*) + \int_{s^*}^{\overline{\theta}} \theta f(\theta) d\theta - (p-\kappa)\frac{\partial s^*}{\partial p} F(s^*) - s^*[1 - F(s^*)] = 0$$

The equilibrium price can be represented as a cost plus a margin,

$$p^* = \kappa + \frac{\int_{s^*}^{\overline{\theta}} (\theta - s^*) \frac{f(\theta)}{f(s^*)} d\theta}{\frac{\partial s^*}{\partial p}[s^* + \frac{F(s^*)}{f(s^*)}] - s^*}$$

We conclude our findings in the following proposition:

Proposition 4 *The cloud price margin is more likely to be low when:*

- *The expected excessive demand* $\int_{s^*}^{\overline{\theta}} (\theta - s^*) f(\theta) d\theta$ *is low.*
- $\frac{\partial s^*}{\partial p}$ *is large, that is, the optimal capacity is sensitive the price.*
- *the slope of hazard rate* $\frac{F(\theta)}{f(\theta)}$ *with respect to* θ *is steep.*

The first and second conditions are straightforward, since more demand induces higher price, while increasing price leads to a high level of in-house capacity and decreases the demand of cloud delegation; therefore, if a small change of price leads a relative large shift of optimal in-house capacity, the CSP would rather to keep the price margin in a low level. For the third condition, although the consumer demand is stochastic, a steeper hazard rate implies that the number of customers is more likely to be realized in a low level, so that the optimal price of the cloud service is also kept in a low level.

3.2 Separating Elements

Computing capacity might be too abstract to contract with, in reality, computing capacity is regarded as a composition of several elements, including data storage, network bandwidth and etc. Therefore, in this subsection, we discuss the situation that these elements can be separately deal with. Particularly, this also give us the opportunity to establish a micro-foundation of "randomly dropping", when the customer demand exceeds the accommodation ability of one of the elements.

We assume that, to accommodate customers, the WO needs enough storage space l and sufficient fast network access speed, which is determined by the network bandwidth μ that she has purchased, and lack of either elements leads to a failure of accommodation. The cost functions of l and μ are linear, namely, the unitary cost of storage space and network bandwidth are respectively κ^l and κ^μ. On the customer side, we assume that each customers has an impatience variable i; a customer with an impatience value i will abandon his visiting to the website if the waiting time of the website $\omega > i$. i is a stochastic variable and satisfies a continuous and differentiable distribution $g(i), i \in [\underline{i}, \bar{i}]$. As before, we assume that in the next sub-period, a small proportion β of the abandoning visitors will return and initialize a second try to access the website, The waiting time ω is a function of network bandwidth, $\omega = \frac{1}{\mu-\theta}$. Therefore, given μ, l and θ, the number of customers that finally access the website is $\min\{l, [1 - G(\frac{1}{\mu-\theta})]\theta\}$. To make our model tractable, we assume that in equilibrium, the reasonable ranges of μ and l always satisfy the following assumption,

Assumption 3 $1 - G(\omega) > \theta g(\omega)\omega^2 > \frac{\kappa^\mu}{(1-\beta)r-\kappa^l}$ *for any* θ.

The first inequation implies $\frac{\partial\{[1-G(\frac{1}{\mu-\theta})]\theta\}}{\partial\theta} > 0$, so that additional customers will not crowd out the existing ones; while the second inequation means that $\frac{\partial\{[1-G(\frac{1}{\mu-\theta})]\theta\}}{\partial\mu}[(1-\beta)r - \kappa^l] > \kappa^\mu$, which

implies that if $G(\frac{1}{\mu-\theta}) > 0$, marginally increasing μ and l so as to accommodate more impatient customers is always profitable. Therefore, in equilibrium, $G(\frac{1}{\mu-\theta}) = 0$, if θ is certain.

We now follow the same route as in the basic model, first we calculate the pure in-house and cloud scheme and then the hybrid one.

Lemma 1 *Given the storage capacity l, the network bandwidth μ is a corner solution $\mu(l) = l + \frac{1}{i}$.*

Proof. Obviously, build excessive network bandwidth is not profitable for the WO, since the storage capacity defines the upper-limit of customers, therefore the maximal value of $\mu = l + \frac{1}{i}$. However, by assumption 3, building less network bandwidth $\mu' = l' + \frac{1}{i}$, where $l' < l$, the profit of the WO is

$$V^i = \int_{\underline{\theta}}^{l'} r\theta f(\theta)d\theta + \int_{l'}^{l''(\mu')} r\{[1-(1-\beta)G(\frac{1}{\mu-\theta})]\theta f(\theta)d\theta + r[1-F(l)]l - (1-e)L - \frac{e^2}{2} - \kappa^l l - \kappa^\mu \mu$$

and

$$\frac{\partial V^i}{\partial l'} = \int_{l'}^{l''(\mu')} r(1-\beta)g(\omega)\omega^2\theta f(\theta)d\theta > 0,$$

therefore, in equilibrium, $l' = l$. ■

According to lemma 1, in the pure in-house scheme,

$$V^i = \int_{\underline{\theta}}^{l} r\theta f(\theta)d\theta + \int_{l}^{\bar{\theta}} r[s + \beta(\theta - l)]f(\theta)d\theta - (1-e)L - \frac{e^2}{2} - \kappa^l l - \kappa^\mu \mu(l)$$

The solutions includes

$$\begin{aligned} r(1-\beta)[1-F(l^{i*})] &= \kappa^l + \kappa^\mu, \\ e^{i*} &= L. \end{aligned}$$

This solution is quite similar to the baseline model, except that the cost of computing capacity in RHS now is replaced by the sum of the cost of network bandwidth and storage.

For the full cloud scheme, μ and l can also be functions of θ, CSP now has incentives to under-provide network capacity to manipulate the consumers that are observed by the WO, denote $\alpha(\theta) = 1 - G(\frac{1}{\mu(\theta)-\theta})$,and $l(\theta) = \alpha(\theta)\theta$, we now prove that in equilibrium, $\alpha(\theta) = 1$.

Lemma 2 *In equilibrium, $\alpha(\theta) = 1$.*

Proof. Suppose that in a candidate equilibrium $\alpha(\theta) < 1$, after observing $\alpha(\theta) + \beta(1 - \alpha(\theta))$, the WO learns θ and receive a payment $T(\theta)$. However, due to assumption 3, the CSP can deviate by

increasing $\alpha(\theta) = 1$, which creates a strictly higher output, and pay WO $T(\theta) + \eta$, where η is an arbitrary small but positive number. By doing so, both CSP and WO earns strictly higher payoff, therefore any candidate equilibrium involves $\alpha(\theta) < 1$ cannot survive. ■

If we denote $\tilde{\theta} = \alpha\theta$, Lemma 2 can be regarded as an alternative form of revelation principle. Now we start to solve the equilibrium. the maximization problem of the WO is

$$\max \int_{\underline{\theta}}^{\overline{\theta}} T(\tilde{\theta}) f(\theta) d\theta$$
$$s.t. 1 \in \arg\max_{\alpha} w(\alpha, \theta)$$

in which $w(\tilde{\theta}, \theta)$ refers to the payoff of the CSP:

$$w(\tilde{\theta}, \theta) = \pi(\tilde{\theta}) - T(\tilde{\theta}) - \frac{e^2(\tilde{\theta}, \theta)}{2} - \kappa^l l(\tilde{\theta}) - \kappa^\mu \mu(\tilde{\theta}).$$

Here $e(\tilde{\theta}, \theta)$ is defined by the equation

$$\pi(\tilde{\theta}) = r(1-\beta)\tilde{\theta} + r\beta\theta - (1 - e(\tilde{\theta}, \theta))L.$$

According to the envelop theorem, the information rent R satisfies:

$$\frac{\partial R}{\partial \theta} = \frac{\partial w(\tilde{\theta}, \theta)}{\partial \theta}|_{\tilde{\theta}=\theta} = \frac{r\beta}{L} e(\theta),$$

It follows that

$$R(\theta) = \int_{\underline{\theta}}^{\theta} \frac{\partial R}{\partial i} di = \int_{\underline{\theta}}^{\theta} \frac{r\beta}{L} e(i) di.$$

The same calculation yields

$$V^c = \int_{\underline{\theta}}^{\overline{\theta}} [r\theta - (1-e)L - \frac{e^2}{2} - (\kappa^l + \kappa^\mu)\theta - \frac{1}{\underline{i}} - \frac{r\beta}{L} \frac{1 - F(\theta)}{f(\theta)} e] f(\theta) d\theta$$

and the same second-best effort level:

$$e^{SB} = L - \frac{r\beta}{L} \frac{1 - F(\theta)}{f(\theta)}.$$

For the hybrid case, it is also similar as the baseline model[5], the F.O.C. condition

simplifying the equation above leads to

$$[\frac{L^2}{2} - \frac{(e^{c*}(s^*))^2}{2}] f(s^*) = F(s^*)(\kappa^l + \kappa^\mu).$$

[5] See appendix.

Proposition 5 *Define $\kappa = \sum_{j=l,\mu} \kappa^j$,separating contracts elements of computing capacity does not alter the monitor effort level and the optimal in-house capacity under all three schemes.*

This proposition is true as long as all the elements of computing capacity are perfectly complementary since by the property of perfect complementary, fixing one of the elements is enough to capture the upper-bond of all the other elements. Imperfect substitution among elements is of future interest.

3.3 Dynamic Migration

The relationship between the CSP and WO can varies more than one period, in this subsection, we try to characterize dynamic equilibria when the CSP and WO are both engaged into a long-term, contingent contract.. Assuming that the contract last for 2 periods: In the first period, the WO decides to build s_1, and the CSP learns the in-house information θ_1. In the second period, the customer demand is $\theta_2 = \gamma\theta_1 + \varepsilon$, where ε satisfies a continuous and differentiable distribution $H(\varepsilon)$. To facilitate our analysis, we define $F_2(\theta_2|\theta_1)$ as the cumulative distribution function of θ_2 when θ_1 is revealed. Also, we assume that there is a discount rate of the profit δ from period 1to period 2, and a discount rate of the computing capacity λ, that is, the computing capacity $s^2 = s_2 + (1-\lambda)s^1$, where s^1 is the computing capacity inherited for the previous period and s_2 is the incremental capacity built in the beginning of period 2. To be consistent, we still assume that the in-house servers can only be built BEFORE the realization of the customer demand of each period and the price of cloud p is exogenuously given. The Timing is as follows:

1. The WO builds her capacity s^1
2. The WO offer a contract $\{T_1(\theta_1), \pi_1(\theta_1), T_2(\theta_1,\theta_2), \pi_2(\theta_1,\theta_2)\}$to the CSP, if rejected, the game is over, and both parties obtain 0 payoff; if accepted:
3. The CSP observes the consumers' request θ_1 and decides the allocation of computing capacity $\hat{\theta}_1$.
4. The WO choose his monitor effort level e_1^i and the CSP chooses e_1^c.
5. The CSP realizes a profit $w_1(\hat{\theta}_1, \theta_2)$,
6. The WO receives a payment $T_1(\hat{\theta}_1)$.

7. The same procedure from step 3 to 6 in the second period.

We are interested in the hybrid case, so that we can figure out how the in-house capacity varies along with the time. The expected profit of the WO in the beginning of the first period is:

$$\begin{aligned}
&\max_{s^1,e^i,e^c} \int_{\underline{\theta}}^{s^1} [r\theta_1 - (1-e_1^i)L - \frac{(e_1^i)^2}{2}] f(\theta_1) d\theta_1 - \kappa s^1 \\
&+ \int_{s^1}^{\bar{\theta}} [r\theta_1 - p(\theta_1 - s^1) - (1-e_1^c)L - \frac{(e_1^c)^2}{2} - R(\theta_1)] f(\theta_1) d\theta_1 \\
&+\delta \int_{\underline{\theta}}^{\bar{\theta}} \{ \int_{\theta_2}^{s^2} [r\theta_2 - (1-e_2^i)L - \frac{(e_2^i)^2}{2}] f_2(\theta_2|\theta_1) d\theta_2 - \kappa[s^2 - (1-\lambda)s^1] \} f(\theta_1) d\theta_1 \\
&+\delta \int_{\underline{\theta}}^{\bar{\theta}} \{ \int_{s^2}^{\bar{\theta}_2} [r\theta_2 - p\max\{(\theta_2 - s^2), 0\} - (1-e_2^c)L - \frac{(e_2^c)^2}{2} - R(\theta_1, \theta_2)] f_2(\theta_2|\theta_1) d\theta_2 \} f(\theta_1) d\theta_1
\end{aligned}$$

s^1 and s^2 are determined by

$$\begin{aligned}
[\frac{L^2}{2} - \frac{(e_2^{c*}(s^{2*}))^2}{2}] f_2(s^{2*}|\theta_1) + [1 - F_2(s^{2*}|\theta_1)]p &= \kappa, \\
[\frac{L^2}{2} - \frac{(e_1^{c*}(s^{1*}))^2}{2}] f(s^{1*}) + [1 - F(s^{1*})]p &= [1 - (1-\lambda)\delta]\kappa
\end{aligned}$$

the in-house capacity of the last period is little special compared with the static model, however, in the first period, since part of the in-house servers can be utilized in the subsequent period, in another word, there is externalities between periods, therefore, the WO intends to build more in-house capacity than that of the static model. No surprisingly, the incentives of building in-house servers in the first period increases when the profit discount rate δ becomes larger or the capacity discount rate λ becomes smaller.

Moreover, using the recursive envelop theorem developed by Pavan et al. (2010) and a series of subsequent works ((Garrett and Pavan 2010), (Pavan 2007) and (Pavan, Segal, and Toikka 2010)), the monitor effort shows a vanish trend along with the time, the analytical solutions of the efforts are[6]:

$$\begin{aligned}
e_1^{c*} &= L - \frac{r\beta}{L} \frac{1 - F(\theta_1)}{f(\theta_1)}, \\
e_2^{c*} &= L - \gamma \frac{r\beta}{L} \frac{1 - F(\theta_1)}{f(\theta_1)}.
\end{aligned}$$

[6] See appendix for the details.

The diminishing distortion leads to less efficiency gains by building in-house servers, therefore for the dynamic point of view, the WO has less incentives to invest on in-house capacity than she does in the first period, this offers a justification of the recent trend of migration, that is, more and more enterprises, although they have fairly large in-house capacity, begin to migrate their services on the cloud. Moreover, when the distortion of the monitor effort in the second period is negligible, the in-house capacity satisfies $F(s^{2*}|\theta_1) = 1 - \frac{\kappa}{p}$, where s^2 defines the minimum in-house capacity that the reserved by the WO.

Proposition 6 *Along with the increase of time, the WO has less incentives to invest on the in-house capacity and, if $\delta\lambda$ is small enough, the WO invests more on capacity than that in static model, and keep the capacity no less than $s^{\min}$ in the subsequent period, where $s^{\min}$ is defined by $F_2(s^{\min}|\theta_1) = \max\{0, 1 - \frac{\kappa}{p}\}$.*

4 Concluding Remarks and Furture Research

Cloud computing is an increasingly important component of the information technology, and capture more and more attention from both academia and industry. However, most of related works are from the legal side, while the formal economic analysis is still comparatively rare. In this paper, we focus on the perspective of asymmetric information and moral hazard, and develop an model to analyze the underlying the interaction of website and cloud service provider. Our most surprising result is that the website might build more in-house capacity when cloud is an available option, while this can explain why some enterprises, especially big ones, still invest on their in-house capacity and simultaneously adopting cloud technology. However, our dynamic model predict that, along with the time increases, the efficiency gain from the in-house capacity will become less, so that migration to the cloud is a natural trend in the long run. Particularly, if as claimed, the CSP owns scale economics, which means probability the unitary cost of computing capacity, our results implies a clear trend that full delegation is the future.

From the policy perspective, our result implies that a long-term, contingent contract might leads to more equilibrium output than the current "pay-as-you-go" business pattern. One evidence of our implication is that, at least for large firms, the CSP has now offered more tailored and specialized contract instead of unified pricing.

However, we still omit two important perspectives that may affect our analysis. First of all,

evidence indicates that there is tough competition among CSPs, but we do not explicitly model the competition. the competition can be added with the lock-in effect in the future work, where an incumbent CSP already has a deal with the WO, while a new CSP with lower unitary cost may want to come in, then the switch decision of the WO is of great interest; secondly, in realities, the elements of computing capacity are contracted with great details. Although in one of our extension we also consider separating elements contract, our implicit assumption of perfect complementarity, which save us from multi-screening variables, prevent us from extending our results to more general framework. Therefore, another possible next step is to release the assumption of perfect complementarity and check the robustness of our results.

5 Appendix

5.1 Proof of Proposition 2

First of all, all functions here are continuous and differentiable, and the WO's profit under either regime is a decreasing function with respect to κ.

$$\begin{aligned}\frac{\partial V^{c*}}{\partial \kappa} &= -\int_{\underline{\theta}}^{\overline{\theta}} \theta f(\theta) d\theta < 0,\\ \frac{\partial V^{i*}}{\partial \kappa} &= -s^* < 0.\end{aligned}$$

Therefore the only thing left is to prove that in-housing building is superior to cloud computing when κ has extreme values.

If $\kappa \to 0$, $s \to \bar{\theta}$,

$$\begin{aligned}\lim_{\kappa\to 0} V^{i*} &= \int_{\underline{\theta}}^{s} r\theta f(\theta) d\theta - L + \frac{1}{2}e^{*2}\\ \lim_{\kappa\to 0} V^{c*} &= \int_{\underline{\theta}}^{\overline{\theta}} [r\theta - L + \frac{1}{2}(e^{SB})^2] f(\theta) d\theta\end{aligned}$$

Since $e^* > e^{SB}$,

$$\begin{aligned}\lim_{\kappa\to 0} V^{c*} &= \int_{\underline{\theta}}^{\overline{\theta}} [r\theta - L + \frac{1}{2}(e^{SB})^2] f(\theta) d\theta\\ &< \int_{\underline{\theta}}^{\overline{\theta}} (r\theta - L + \frac{1}{2}e^{*2}) f(\theta) d\theta\\ &= \lim_{\kappa\to 0} V^{i*}.\end{aligned}$$

If $\kappa \to r$, the profit of WO will only rise from the effort,

$$\begin{aligned}\lim_{\kappa\to r} V^{c*} &= \int_{\underline{\theta}}^{\overline{\theta}} [-L + \frac{1}{2}(e^{SB})^2] f(\theta) d\theta\\ &< \int_{\underline{\theta}}^{\overline{\theta}} (-L + \frac{1}{2}e^{*2}) f(\theta) d\theta\\ &= \lim_{\kappa\to r} V^{i*}.\end{aligned}$$

5.2 Proof of Proposition 3

First we prove that hybrid cloud yield the highest expected profit for the WO. In equilibrium, since pure-cloud scheme can be regards as a specific hybrid cloud with in-house server mandatorily being 0, by definition of the optimality, the profit under hybrid cloud is higher than that under pure cloud. Conversely, the pure in-house scheme can be regarded as a specific hybrid cloud where delegating excessive consumers yields zero profit, since delegating to the cloud will only take place when it generates non- negative profit, we can draw a conclusion that hybrid cloud scheme does yield no less profit than pure in-house scheme.

Under pure in-house server scheme, the F.O.C. is:

$$r(1-\beta)[1-F(s^i)] = \kappa.$$

While under hybrid scheme, the F.O.C. is

$$[\frac{L^2}{2} - \frac{(e^{c*}(s^*))^2}{2}]f(s^*) + [1-F(s^*)]\kappa = \kappa.$$

If $\kappa \to r(1-\beta)$, assuming $s^* = s^i$, the second F.O.C. cannot establish since $[\frac{L^2}{2} - \frac{(e^{c*}(s^*))^2}{2}]f(s^*) > 0$; therefore, considering that $1 - F(s^*)$ is a decreasing function of s^*, in equilibrium, $s^* > s^i$.

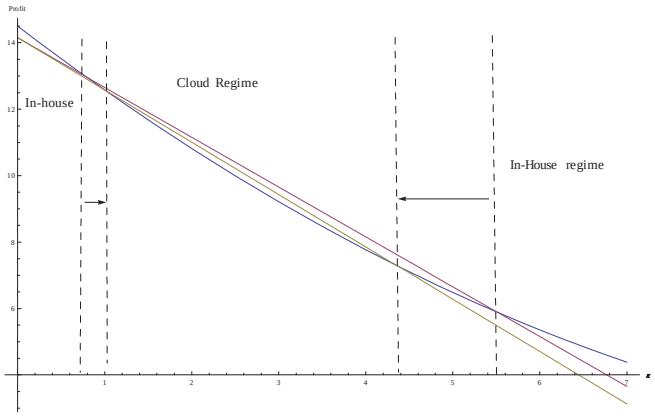
Otherwise, if L is large enough, $\frac{e^{c*}(s^*)}{e^*}$ is an increasing function of L, since:

$$\frac{\partial \frac{e^{c*}(s^*)}{e^*}}{\partial L} = 2\frac{r\beta}{L^3}\frac{1-F(s^*)}{f(s^*)} > 0.$$

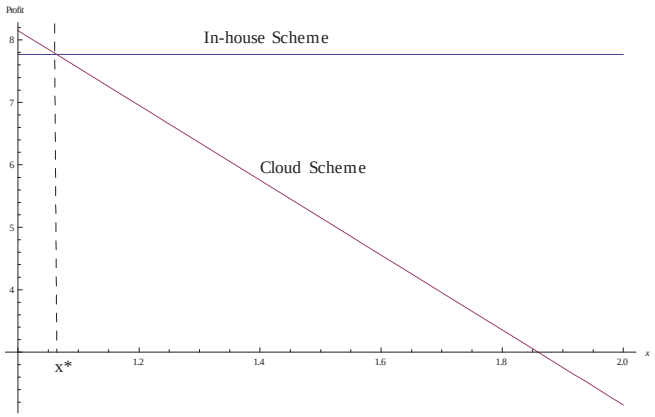
Therefore, when $L \to \infty$, $\frac{e^{c*}(s^*)}{e^*} \to 1$, by assumption 1, $s^* < s^i$.

5.3 Pricing under Pure Cloud Scheme

Although it is very hard to have an analytical solution, from simulation, we can capture the trade-off faced by the CSP. Assuming $\theta \in U[1,2], F(\theta) = \theta - 1.r = 70, \beta = 0.1, L = 40$, the figure below illustrate the situation when the cloud price $p = \kappa$ (bertrand) and $p = 1.05\kappa$ (monopolistic competition):



As it is shown, the cloud regime shrinks when the price of cloud increases. If the cost κ is common knowledge ex-ante, the CSP will charge less price margin when κ is close to 0 or large enough, since otherwise the WO will choose in-house servers instead. If κ is in between, the CSP will increase the price up to the point that the profit line of in-house scheme intersects with the new profit line of cloud scheme. The graph below shows the equilibrium multiplier x of the cloud price ($p = x * \kappa$) when $\kappa = 4$(other paramters unchanged).



5.4 Separating Elements with Hybrid Scheme

The timing is as follows:

- The WO builds her own capacity s
- The WO signs a contract $\{T(\theta), \pi(\theta)\}$with the CSP.

- The CSP observes the consumers' request θ and decides the allocation of μ and l.
- The WO choose his maintenace effort level e^i and the CSP chooses e^c.
- The CSP realizes a profit $w(\tilde{\theta}, \theta)$,
- The WO receives a payment $T(\tilde{\theta})$.

Since the WO and the CSP exert their monitor effort independently, therefore Accordingly, the profit of the WO can be divided by two parts: the profit from in-house servers is

$$V^i = \int_{\underline{\theta}}^{l} [r\theta - (1-e^i)L - \frac{(e^i)^2}{2}]f(\theta)d\theta - (\kappa^l + \kappa^\mu)l - \kappa^\mu \frac{1}{\underline{i}}$$

while the profit for the CSP is the solution of the following problem:

$$\begin{aligned} V^c &= [1-F(l)]\int_l^{\overline{\theta}} T(\theta)f(\theta|\theta > l)d\theta \\ s.t.\tilde{\theta} &\in \arg\max_{\hat{\theta}} w(\theta, \tilde{\theta}) \end{aligned}$$

in which $w(\tilde{\theta}, \theta)$ refers to the payoff of the CSP:

$$w^h(\tilde{\theta}, \theta) = \pi(\hat{\theta}) - T(\tilde{\theta}) - \frac{(e^c)^2(\theta, \tilde{\theta})}{2} - (\kappa^l + \kappa^\mu)(\tilde{\theta} - l) - \kappa^\mu \frac{1}{\underline{i}}.$$

Here the effort exerted by the CSP, $e(\theta, \tilde{\theta})$ is defined by the equation

$$\pi^h(\tilde{\theta}) = r[\tilde{\theta} + \beta(\theta - \tilde{\theta})] - (1 - e^c(\theta, \tilde{\theta}))L.$$

Applying the envelop theorem, the information rent must satisfy:

$$\frac{\partial R}{\partial \theta} = \frac{\partial w(\theta, \tilde{\theta})}{\partial \theta}|_{\tilde{\theta}=\theta} = \frac{r\beta}{L}e^c(\theta),$$

which is equivalent to that of the polar case, correspondingly,

$$R(\theta) = \int_l^{\theta} \frac{\partial R}{\partial i}di = \int_l^{\theta} \frac{r\beta}{L}e^c(i)di.$$

The expected profit of the WO can now be reformulated as

$$\begin{aligned} V^c &= [1-F(l)]\int_l^{\overline{\theta}} [(r\theta - (\kappa^l + \kappa^\mu)(\theta - l) - (1-e^c)L - \frac{(e^c)^2}{2} - \int_l^{\theta} \frac{r\beta}{L}e^c(i)di]f(\theta|\theta > l)d\theta \\ &= \int_l^{\overline{\theta}} [(r-\kappa)\theta + (\kappa^l + \kappa^\mu)l - (1-e^c)L - \frac{(e^c)^2}{2} - \frac{r\beta}{L}\frac{1-F(\theta)}{f(\theta)}e^c]f(\theta)d\theta \end{aligned}$$

Therefore, the total expected profit of the WO under a hybrid cloud scheme is $V^h = V^i + V^c$. More formally, the WO solves the following maximization problem:

$$\max_{s,e^i,e^c} \int_{\underline{\theta}}^{l} [r\theta - (\kappa^l + \kappa^\mu)l - (1-e^i)L - \frac{(e^i)^2}{2}]f(\theta)d\theta$$
$$+ \int_{l}^{\overline{\theta}} [(r - (\kappa^l + \kappa^\mu))\theta - (1-e^c)L - \frac{(e^c)^2}{2} - \frac{r\beta}{L}\frac{1-F(\theta)}{f(\theta)}e^c]f(\theta)d\theta$$

According to the F.O.Cs, the equilibrium effort levels of WO and CSP are:

$$\begin{aligned} e^{i*} &= L \\ e^{c*} &= L - \frac{r\beta}{L}\frac{1-F(\theta)}{f(\theta)} \end{aligned}$$

Therefore equation (*) can be simplified

$$\max_{s,e^i,e^c} \quad \int_{\underline{\theta}}^{l} (r\theta - (\kappa^l + \kappa^\mu)l - L + \frac{L^2}{2})f(\theta)d\theta$$
$$+ \int_{l}^{\overline{\theta}} [r - (\kappa^l + \kappa^\mu)]\theta - L + \frac{(e^{c*}(\theta))^2}{2}]f(\theta)d\theta$$

Consequently, the optimal in-house building s^* is characterized by

$$[\frac{L^2}{2} - \frac{(e^{c*}(l))^2}{2}]f(l^*) = F(l^*)(\kappa^l + \kappa^\mu).$$

5.5 Solution of the Dynamic Problem

The maximization problem of this dynamic problem is

$$\max_{s^1,e^i,e^c} \int_{\underline{\theta}}^{s^1} [r\theta_1 - (1-e_1^i)L - \frac{(e_1^i)^2}{2}]f(\theta_1)d\theta_1 - \kappa s^1$$
$$+ \int_{s^1}^{\overline{\theta}} [r\theta_1 - p(\theta_1 - s^1) - (1-e_1^c)L - \frac{(e_1^c)^2}{2} - R(\theta_1)]f(\theta_1)d\theta_1$$
$$+\delta \int_{\underline{\theta}}^{\bar{\theta}} \{\int_{\theta_2}^{s^2} [r\theta_2 - (1-e_2^i)L - \frac{(e_2^i)^2}{2}]f_2(\theta_2|\theta_1)d\theta_2 - \kappa[s^2 - (1-\lambda)s^1]\}f(\theta_1)d\theta_1$$
$$+\delta \int_{\underline{\theta}}^{\bar{\theta}} \{\int_{s^2}^{\bar{\theta}_2} [r\theta_2 - p\max\{(\theta_2 - s^2), 0\} - (1-e_2^c)L - \frac{(e_2^c)^2}{2} - R(\theta_1,\theta_2)]f_2(\theta_2|\theta_1)d\theta_2\}f(\theta_1)d\theta_1$$

As before, the information rent can be represented

$$\begin{aligned} R(\theta_1) &= \int_l^{\theta_1} \frac{\partial R}{\partial i} di = \int_l^{\theta_1} \frac{r\beta}{L} e^c(i) di \\ R(\theta_1,\theta_2) &= \int_{\underline{\varepsilon}}^{\bar{\varepsilon}} \int_{s^2}^{\theta_2} \frac{\partial R}{\partial i} h(\varepsilon) d\varepsilon di = \int_{\underline{\varepsilon}}^{\bar{\varepsilon}} \int_{s^2}^{\theta_2} \frac{r\beta}{L} e^c(i) h(\varepsilon) d\varepsilon di \\ &= \int_{\underline{\varepsilon}}^{\bar{\varepsilon}} \int_{s^2}^{\gamma\theta_1+\varepsilon} \frac{r\beta}{L} e^c(i) h(\varepsilon) d\varepsilon di \end{aligned}$$

Integrate by parts yields

$$\begin{aligned} &\max_{s^1,e^i,e^c} \int_{\underline{\theta}}^{s^1} [r\theta_1 - (1-e_1^i)L - \frac{(e_1^i)^2}{2}] f(\theta_1) d\theta_1 - \kappa s^1 \\ &+ \int_{s^1}^{\bar{\theta}} [r\theta_1 - p(\theta_1 - s^1) - (1-e_1^c)L - \frac{(e_1^c)^2}{2} - \frac{r\beta}{L} \frac{1-F(\theta)}{f(\theta)} e_1^c(\theta_1)] f(\theta_1) d\theta_1 \\ &+ \delta \int_{\underline{\theta}}^{\bar{\theta}} \{ \int_{\theta_2}^{s^2} [r\theta_2 - (1-e_2^i)L - \frac{(e_2^i)^2}{2}] f_2(\theta_2|\theta_1) d\theta_2 - \kappa[s^2 - (1-\lambda)s^1] \} f(\theta_1) d\theta_1 \\ &+ \delta \int_{\underline{\theta}}^{\bar{\theta}} \{ \int_{s^2}^{\bar{\theta}_2} [r\theta_2 - p \max\{(\theta_2 - s^2), 0\} - (1-e_2^c)L - \frac{(e_2^c)^2}{2} - \gamma \frac{r\beta}{L} \frac{1-F(\theta)}{f(\theta)} e_2^c] f_2(\theta_2|\theta_1) d\theta_2 \} f(\theta_1) d\theta_1 \end{aligned}$$

Taking the first derivative yields

$$\begin{aligned} e_1^{c*} &= L - \frac{r\beta}{L} \frac{1-F(\theta_1)}{f(\theta_1)}, \\ e_2^{c*} &= L - \gamma \frac{r\beta}{L} \frac{1-F(\theta_1)}{f(\theta_1)}. \end{aligned}$$

References

Fox, A., and R. Griffith (2009): "Above the clouds: A Berkeley view of cloud computing," *Dept. Electrical Eng. and Comput. Sciences, University of California, Berkeley, Tech. Rep. UCB/EECS*, 28.

Garrett, D., and A. Pavan (2010): "Dynamic managerial compensation: On the optimality of seniority-based schemes," .

Hölmstrom, B. (1979): "Moral hazard and observability," *The Bell Journal of Economics*, pp. 74–91.

Jie, Y., Q. Jie, and L. Ying (2009): "A profile-based approach to just-in-time scalability for cloud applications," pp. 9–16.

Kim, A., and I. S. Moskowitz (2010): "Incentivized Cloud Computing: A Principal Agent Solution to the Cloud Computing Dilemma," .

Maier-Rigaud, F. (2011): "Network Neutrality: A Competition Angle," *Competition Policy International*, 2.

Mazhelis, O., and P. Tyrväinen (2012): "Economic aspects of hybrid cloud infrastructure: User organization perspective," *Information Systems Frontiers*, 14(4), 845–869.

Molnar, D., and S. Schechter (2010): "Self hosting vs. cloud hosting: Accounting for the security impact of hosting in the cloud," .

Pavan, A. (2007): "Long-term contracting in a changing world," *Available at SSRN 1620663*.

Pavan, A., I. Segal, and J. Toikka (2010): "Dynamic mechanism design: Incentive compatibility, profit maximization and information disclosure," *Profit Maximization and Information Disclosure (May 1, 2009)*.

Sluijs, J., P. Larouche, and W. Sauter (2011): "Cloud Computing in the EU Policy Sphere," .

Yoo, C. S. (2011): "Cloud computing: Architectural and policy implications," *Review of Industrial Organization*, 38(4), 405–421.

Author contacts:

Tong Wang

Centre for Competition Policy

University of East Anglia

NR4 7TJ Norwich

UK

Email: Tong Wang (NBS) <T.Wang3@uea.ac.uk>