Journal of Computers Vol. 29 No. 6, 2018, pp. 66-78
doi:10.3966/199115992018122906006

Multi-instance Combination Decision in Infrastructure-as-a-Service Cloud under Cloud Users' Demands Fluctuation

Mengdi Yao[1*], Donglin Chen[1], Zhong Wu[2], Qipin She[1,3]

[1] Institute of Economic, Wuhan University of Technology, Wuhan 430070, China
vicmengdiyao@163.com

[2] Physical Education and Equestrian School, Wuhan Business University, Wuhan430056, China
7849800@qq.com

[3] City College, Wuhan University of Science and Technology, Wuhan 430083, China
524037171@qq.com

Received 14 July 2017; Accepted 17 July 2017

Abstract. Infrastructure-as-a-Service (IaaS) cloud providers offer diverse pricing plans, namely on-demand, reservation, and spot pricing plan to meet distinct preferences of customers. This allows cloud users to purchase cloud instances based on their own demand and to generate less cost accordingly. In this paper, we propose a model of multi-instance combination decision for minimizing customer's cost under cloud demands fluctuation with taking risk cost into consideration. We discuss the condition and reason of multi-instance optimal decision of cloud customers. And through case, we analysis the relationship between the optimal decision and fluctuation forms, fluctuation range, effective working time and risk sensitive factor of cloud users' future demand respectively. Our extensive simulations driven by large-scale Google cluster-usage traces have shown that this model can minimize customer's cloud customer's cost which bring a series of economy effect on cloud computing investment of enterprises.

Keywords: digital signal processing, e-commerce, operating systems, RFID

1 Introduction

The Cloud computing has emerged as an important ICT (information and communication technology) innovation that could potentially revolutionize the way of computing resource which are consumed and provided. In emerging economies, such innovation is regarded as the new way to provide information infrastructure that has the potential for further economic upgrading [1]. According to a forecast from International Data Corporation (IDC), the worldwide spending on public cloud services is expected to surpass $107 billion in 2017 [2]. Thanks to the economies of scale, cloud service providers can maintain large-scale data centers and offer different types of virtual machines at a relatively low cost, also can bring about significant business value for enterprises [3], such as reducing IT investment cost, increasing the competitiveness of core products. Therefore, cloud computing cost management has become a primary concern for most enterprises, particularly SMEs (Small Medium Enterprise) [4]. According to cloud users' fluctuating demand, how to combine multi-instances to meet cloud users' diverse needs of cloud services to decrease investment cost of cloud service will become a key point for these enterprises.

In order to meet customers' demands, many cloud providers, such as Amazon EC2 (the largest cloud provider), offer various pricing plans, i.e. reserved instance, spot instance and on-demand instance. On-demand instance charges customers for compute capacity based on actual usage, without requiring any contractual long-term commitments [2]. The customers who use on-demand instance are served on a

* Corresponding Author

based-effort basis. Reserved instance allows cloud users to confirm the reservation capacity in advance according to the amount used during the period. The service quality is same as the on-demand instance, but the cost is lower than on-demand instance. In 2009, AmazonEC2 introduced the spot instance. It offers cloud users idle resource in a low discount through competitive bidding. According to the supply and demand situation, Amazon publishes the instance price periodically. When the highest bid of users is higher than the instant price, cloud users could acquire cloud service. Otherwise the service is automatically terminated. The cloud users don't have independent control over the instance's lifetime. In other words, spot instance has advantage of the lowest price, there exists risk of losing bidding or service interruption at the same time [5].

For most cloud users, the primary reason to choose cloud service is pursuing its economy, so cloud users will purchase reserved instance and spot instance as many as possible to avoid paying for expensive on-demand instance. However, the future demand of customers fluctuates with the change of business demand which makes cloud instance combination decision difficult. According to statistics, using reserved instance totally could save customers'49% cost compared to on-demand instance. While the usage rate is lower than 50%, on the contrary, the average unit running cost is close to and then exceed on-demand instance gradually. The unit price of spot instance is 52.3% lower than on-demand instance, but it exists risk of losing auction and service interruption which brings about temporary shortage of resources. In this case, cloud users should purchase on-demand instance to make up the service, which could bring about potential risk cost. Therefore, according to future demand fluctuation and risk sensitive level, how to help enterprise balance the cost optimization and risk avoidance, make the optimization combination decision of three instance has practical significance.

In summary, our main contributions are:

We formulate the optimal multi-instance combination decision that minimizes acquisition cost of cloud customers.

According to the feature of losing bidding or service interruption of spot instance at any time, we take the risk cost into consideration, and establish a new cost formula of cloud customers.

We evaluate our proposed decision process and decision model through simulations, driven by cluster traces provided by Amazon EC2.Using simulation results. We analysis the optimal combination decision under cloud users' demand fluctuation, service time, risk sensitive factor which can give a reasonable multi-instance combination decision suggestion for cloud customers.

The paper is organized as follows. Section 2 is a literature review; Section 3, we discuss cloud users' multi-instance combination decision process. And based on the process, we propose the multi-instance combination decision model in section 4.The section5 is the numerical examples under nonlinear and linear stochastic demand and time. we make an evaluation under the environment of Matlab in section 6. In section 7, we discuss the managerial implication from the economy and technology aspect. The conclusion and future outlook of multi-instance are discussed in section 8.

2 Related Works

With the research of cloud computing from technology to application, the resource allocation based on system can't meet business needs. The market economy theory was brought into the field of cloud computing resource allocation [6-7]. Especially the spot instance has been launched into the cloud market, it has attracted attention of a, amount of scholars. Andrzejak et al. [6] researched spot instance decision which satisfy the restrain of Service Level Agreement (SLA), and put forward a possibility model which could optimize cost, performance and usability according to users' demand and dynamic environment change. Mattess et al. [8] researched bidding decision of spot market, and concluded that spot instance can increase the reliability and decrease the whole cost when the bidding price is relatively high. Yi et al. [9] proposed a dynamic inspection mechanism to make the cost of spot instance reach the lowest when the reliability is the highest.

But all the researches mentioned above are focused on the acquisition of spot instance, there is little attention of multi-instance combination decision. In order to solve the problem of reserved instance's excessive reservation, Yuan et al. [10] proposed a lease instance pricing model. The lease instance refers to the reserved instance that has been reserved unused. Cloud users could submit the idle reserved instance to cloud providers again to obtain redress. Chaisiri et al. [11] came up with the robust cloud resources supply algorithm (RCRP) and optimal algorithm of cloud resources supplies (OCRP), and

proposed optimal acquisition decision of reserved instance and demand instance. Experimental results show that the algorithm could reduce cloud users' cost. Based on this, he proposed two virtual server supply algorithms to make the short term and long term purchase cost lowest respectively with considering the spot instances. Aiming at lacking of users' future demand, Wang et al. [12] proposed two practical online algorithms, i.e. one deterministic and another randomized, which dynamically combine the two instance options online without any knowledge of future demand. These algorithms achieve minimizing IaaS cloud providers' cost. Then Ran et al. [13] put forward another new cloud brokerage service that reserves a large pool of instances from cloud providers and serves users with price discounts to achieve minimizing its service cost. However, all the researches mentioned above didn't consider the risk cost which was brought about by losing bidding or service interruption of spot-instance. In this paper, we introduce risk cost, propose multi-instance combination decision model to make cloud users' total cost lowest under cloud users' fluctuation demand.

3 Cloud Users' Multi-instance Combination Decision Process

Our model divides the total cost of cloud users' multi-instance combination decision into two parts: one part is purchase cost and usage cost of these three types instance; the other part is risk cost which was brought about by losing auction or service interruption of spot instance. According to the features of different instance types and customers' demand fluctuation, multi-instance acquisition is a good way to optimal cost management for cloud computing investment of most enterprises. The reserved instance need be reserved in advance, therefore, users have to determine purchase capacity and purchase form (this paper assumes that the purchase form is one-year contract).Second, the price of on-demand instance is much higher than spot-instance. Aiming at saving cost, users usually give priority to choosing spot-instance, and then purchase on-demand instance for resource shortage after the spot-instance ratio was determined. Lastly, there is risk of losing auction or service interruption for spot-instance which could bring about resource unavailability. So cloud providers need invoke on-demand instance to meet cloud users' demand. Fig. 1 shows the multi-instance combination decision process of cloud users.

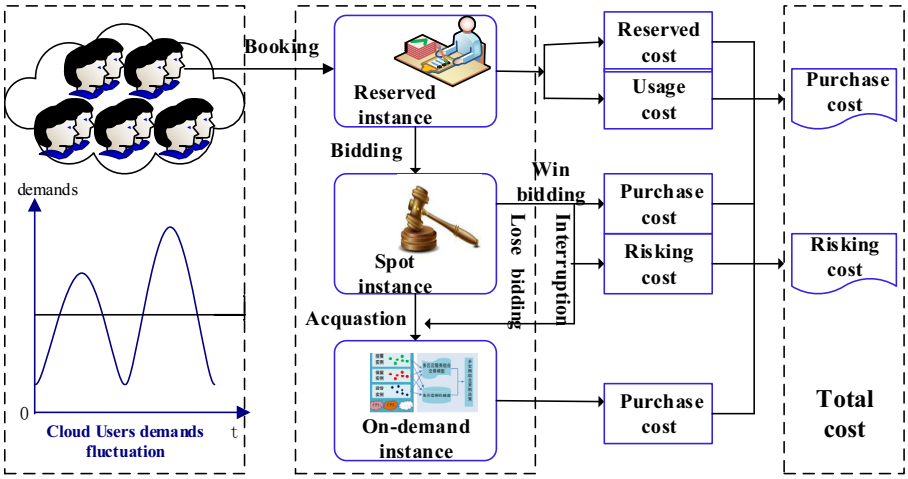


Fig. 1. Cloud users' multi-instance combination decision process

The following assumptions has been used to explain the combination decision process.

(1) Based on working time, every day's demand of cloud users is divided into two parts: one part is basic demand of non-working time (T0); the other part is business demand of working time. Similar to Armbrust et al. [14], we assume that the basic demand is stable, and the daily business demand is random fluctuation.

(2) For the stable basic demand, cloud providers could provide the reserved instance to satisfy cloud users' demand. Owing to the fee is stable, we don't take this part into consideration. What's more, for the fluctuation business demand, we adapt combination decision of three different instances to meet cloud users' demand.

4 Multi-instance Combination Decision Models

The work most relevant [13] to ours focused on combination decision of on-demand and reserved instance, which considers the provisioning cost contains the cost of on-demand and reserved instance. However, the spot instance with the lowest price can reduce the total purchase cost [15]. Purchasing spot instance in a lowest price can minimize cost but also generates risk of service interruption in any time. When the cloud users purchase spot instance, the risk cost should be calculated. Therefore, the total cost of multi-instance combination decision consist of two parts: (1) acquisition cost, i.e. explicit cost (EC); (2) risk cost, i.e. implicit cost (IC). Table 1 shows all the variables and parameters in this paper.

Table 1. Font sizes of headings. Table captions should always be positioned *above* and center the tables. The final sentence of a table caption should end without a period

P_0	The unit reserved cost of reserved instance
P_1	The unit usage cost of reserved instance
P_2	The unit bidding price of spot instance, that is the unit usage cost of spot instance
P_3	The unit price of on-demand instance, that is the unit usage cost of on-demand instance
T	The business operation cycle, the time slot is set as t, so t∈ [0,T]
f(t)	Every hour's random business demand, so the total business demand $F(T)=\int_0^T f(t)\cdot Td(T)$
Q	The reserved capacity of reserved instance
q	The usage capacity of reserved instance
β	The purchase rate of spot instance owing to the reserved instance couldn't meet the resource demand
θ	Under the situation of losing bidding or service interruption, the acquiring probability of spot instance
λ	Risk sensitivity factor, which refers to damage degree caused by service interruption or delay of cloud users, $\lambda \in [0,1]$

Inspired by the capacity control and revenue management method put forward by Doyle et al. [16], the capacity of the customer's demand is denoted by $F(t)$. It contains three types instance's capacity. So the acquisition cost of these three instances are as follows:

(1) The pricing mechanism of reserved instance obeys two-part tariff. The cloud provider charges cloud users for an upfront reservation fee at first, the rest usage fee enjoys a heavy discount. The purchase capacity of reserved instance is Q, usage capacity is q. So the purchase cost of reserved instance is calculated as reserved cost plus period usage cost. The purchase cost of reserved instance RC is denoted as follows:

$$RC = p_0 Q + p_1 q$$

(2) The customers' demand that reserved instance's capacity couldn't satisfy is $F(T)-q$, the purchase rate of spot-instance is β, and its purchase cost is defined as SC:

$$SC = p_2 \beta F(T) - q$$

(3) Due to the risk of losing auction or service interruption of spot-instance, the customer's demand that reserved instance and spot-instance couldn't satisfy is $F(T)-q-(F(T)-q)*\beta*\theta$. Therefore, cloud customers adopt buying on-demand instance, and its purchase cost is defined as OC:

$$OC = (1-\beta\theta)F(T) - qp_3$$

So, the purchase cost, i.e. explicit cost is calculated as:

$$EC = RC + SC + OC = p_0 Q + p_1 F(T) + F(T) - q[\beta p_2 + (1-\beta\theta)p_3 - p_1]$$

According to economy and opportunity cost theory, risk cost which was bring about by spot instance's service interruption can be regarded as a kind of implicit cost of opportunity cost. For risk cost, there are two computing methods: fixed value method and proportional method. Fixed value method refers to the loss severity is independent of jobs and users just loss fixed amount cost. While the proportional method refers to loss severity is dependent on application itself, that is to say, once the job failed, the loss cost is

depending on the value of the work itself and the failure sensitivity [17]. In this paper, we adopt proportional method to calculate loss cost. Therefore, the risk cost IC is $IC = F(T) - q\beta(1-\theta) \times p_2 \times \lambda.$

So the total cost of multi-instance combination decision TCO is including purchase cost (i.e. explicit cost) and risk cost, that is

$$TCO = EC + IC = (RC + OC + SC) + IC$$

So,

$$\begin{aligned} TCO &= p_0 Q + p_1 F(T) + F(T) - q[\beta p_2 + (1-\beta\theta)p_3 - p_1] + F(T) - q\beta(1-\theta)p_2\lambda \\ &= p_0 Q + p_1 F(T) - q[\beta p_2 + (1-\beta\theta)p_3 + \beta(1-\theta)p_2\lambda - p_1] \end{aligned}$$

5 Multi-instance Combination Decision under the Nonlinear and Linear Relationship Between Stochastic Demand and Time

Due to workload traces of IaaS cloud provider is strictly confidential, it difficult for us to gain public data of workload information. Recently, Google (a large cloud provider) has published a data set about their customers' demand. Here, the Google typically serves two different customer groups, which can be distinguished from the relationship between their instance demand and time. One kind of cloud customers such as universities which need cloud service to build teaching management and experimental platform and large IT enterprises. Their cloud demand varies with time. The other kind of customers such as the SME which runs their jobs for a short time, so their cloud demand and time is linear relationship. Here, we discuss these multi-instance combination decision under nonlinear and linear stochastic business demand and time.

5.1 Multi-instance Combination Decision under Nonlinear Stochastic Business Demand and Time

We assume that the daily random business demand $f(t)$ satisfies that $f(t) = \alpha \times \cos(2\pi/365) \times t + \alpha,$, and $\alpha > 0$, as shown in Fig. 2.

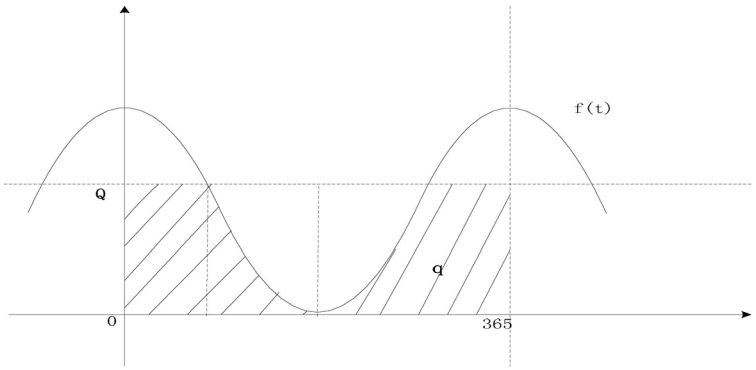


Fig. 2. Random business demand of future instance

Take one year as a cycle, day as a unit, T=365, the total business demand of future instance $F(T)$ is

$$F(T) = \int_0^T f(t) \times T_1 dt = \int_0^{365} \left[\alpha \times \cos\left(\frac{2\pi}{365} \times t \right) + \alpha \right] \times T_1 = 365\alpha T_1.$$

The purchase capacity of reserved instance is Q, so the usage capacity of reserved instance q is equal to the shadow part in Fig. 2.

Then, the purchase cost (EC), risk cost (IC) and total cost were perceptively:

$$EC = p_0 Q + p_1 \times 365 \times \alpha \times T_1 + (365/\pi) \times \left[\sqrt{\alpha^2 - (Q-\alpha)^2} - (Q-\alpha) \times \arccos \frac{Q-\alpha}{\alpha} \right] \times [\alpha p_2 + (1-\beta\theta) p_3 - p_1] \times T_1$$

$$IC = \beta(1-\theta) \times p_2 \times \omega\lambda \times (365/\pi) \times T_1 \times \left[\sqrt{\alpha^2 - (Q-\alpha)^2} - (Q-\alpha) \times \arccos \frac{Q-\alpha}{\alpha} \right]$$

$$TC0 = \left[\sqrt{\alpha^2 - (Q-\alpha)^2} - (Q-\alpha) \times \arccos \frac{Q-\alpha}{\alpha} \right] \times \frac{365}{\pi} [\beta p_2 + (1-\beta\theta) p_3 + \beta(1-\theta) p_2 \lambda - p_1] p_0 Q + p_1 \times 365 \times \alpha \times T_1$$

Taking Amazon EC2 (the largest cloud provider) as an example, according to the official pricing of Amazon, the reserved cost of reserved instance per year is $p_0 = 227.5$, the usage cost of reserved instance per hour is $p_1 = 0.04$, the purchase cost of unit on-demand instance per hour is $p_3 = 0.095$, the price of spot instance is $p_2 \in [0.035, 0.095]$. We choose 15220 time points data (data comes from cloudxchange.org).The occurrence frequency of different prices is shown in Fig. 3 and the relationship between spot instance price and acquiring probability θ is shown in Fig. 4.

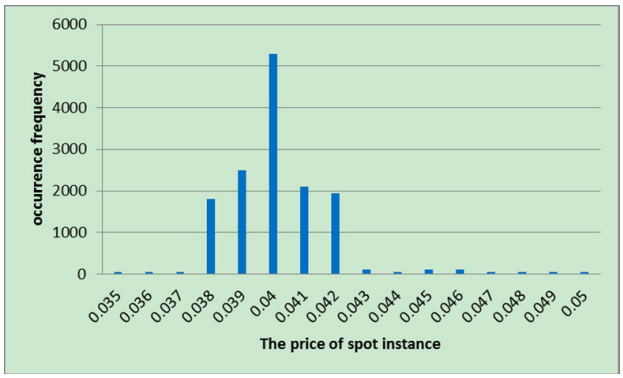


Fig. 3. Occurrence frequency of different spot-instance prices

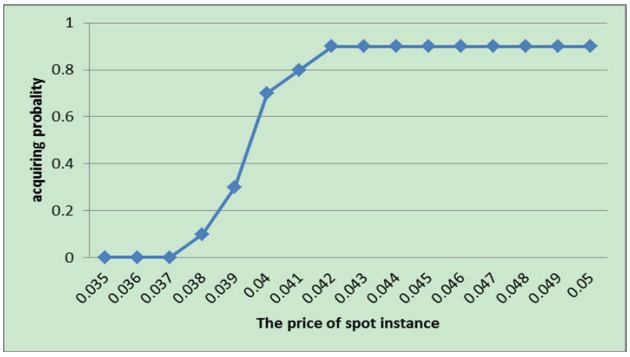


Fig. 4. The relationship between spot instance price and acquiring probability

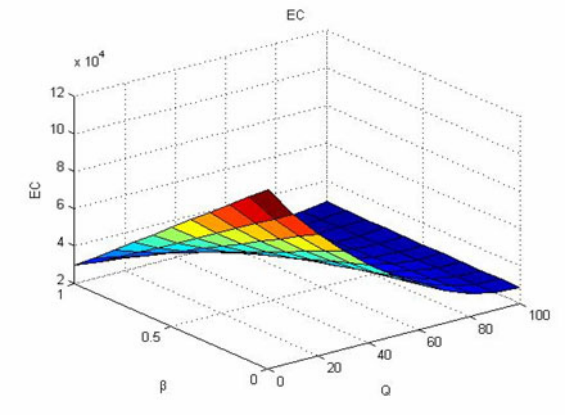
It's obvious that the main range of spot instance price's fluctuation is [0.038, 0. 042], when the spot instance price $p_2 = 0.042$, the acquiring probability $\theta \approx 0.9$. When it continues to increase, the range of θ is small. In order to simply calculation, we choose $p_2 = 0.042$, $\theta = 0.9$. Through the investigation and analysis of the university's cloud resource demand, we set $\alpha = 50$, $\lambda = 0.6$, $T_0 = 14$, $T_1 = 10$, so:

$$EC = 227.5 \times Q + 5480 + \frac{365}{\pi}(0.055 - 0.043/\beta) \times 10 \times \left[\sqrt{Q(100-Q)-(Q-50)} \times \arccos\frac{Q-50}{50} \right]$$

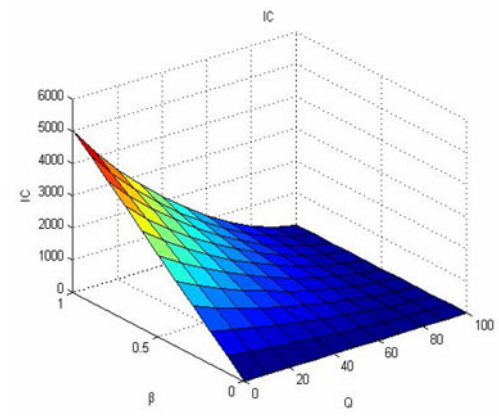
$$IC = \left[\sqrt{Q(100-Q)-(Q-50)} \times \arccos\frac{Q-50}{50} \right] \times 0.0025 \times \beta \times \frac{365}{\pi} \times 10$$

$$TCO = 227.5 + 5480 + 10(0.055 - 0.0405\beta) \times \frac{365}{\pi} \left[\sqrt{Q(100-Q)-(Q-50)} \times \arccos\frac{Q-50}{50} \right]$$

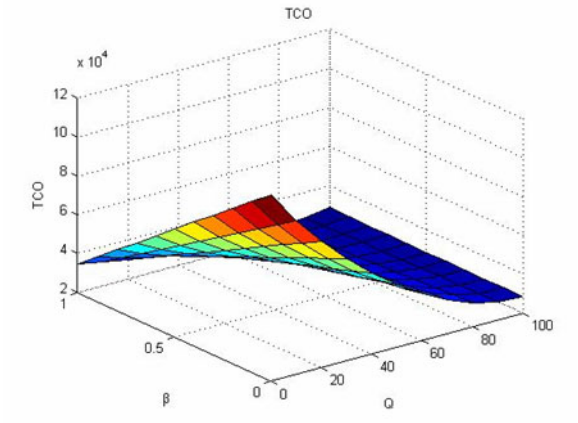
Here we use Matlab to calculate the change trend of purchase cost, risk cost and total cost (Z axis) as the reserved capacity of reserved instance Q (X axis) and the acquisition rate of spot instance β (Y axis) change.



(a) Purchase cost



(b) Risk cost



(c) Total cost

Fig. 5. The purchase cost, risk cost and total cost under the nonlinear stochastic business demand and time

From the Fig. 5, we find that:

Purchase cost.

(1) When the reserved capacity of reserved instance Q keeps constant, purchase cost gradually decreases as the purchase rate of spot instance β increases. The key reason is that the bidding price of spot instance is much lower than the price of on-demand instance. Cloud users can adopt the spot-instance as much as possible to reduce the purchase cost when the reserved instance could satisfies cloud customers' demand.

(2) When purchase rate of spot-instance β keeps constant, the purchase cost increases and then decreases as the reserved capacity of reserved instance Q increases. For a limited cloud demand, low reserved capacity refers to relative high utilization rate, so the average unit operation cost of reserved instance is relatively low, and the total cost is low. While the reserved capacity of reserved instance continue increasing which leads to the decrease of utilization rate, so the instance's average unit operation cost and the total cost increase.

According to the results, when $Q=50$, $\beta=1$, the purchase cost EC achieves the minimum value EC=$\$2.488\times 104$.

Risk cost. When reserved capacity of reserved instance Q keeps constant, risk cost decreases along with the induce of purchase rate of spot instance β, while the purchase rate of spot-instance β is constant, the risk cost decreases as reserved capacity Q increases. In addition, when the purchase rate of spot instance is equal to one or without reserved instance, cloud customers need pay for risk cost. The main reason is that under the situation of losing bidding and service interruption, the acquisition probability of spot instance θ are determined, the change of risk cost is just related to the purchase capacity of spot-instance. When β decreases or Q increases, the purchase capacity of spot-instance decreases, and when $\beta=1$ or $Q=0$, it becomes 0 and risk cost is 0.

Total cost. The total cost shows the same change trend of purchase cost, that is, when reserved capacity of spot instance Q keeps constant, the total cost decreases with the increasing of purchase rate of spot instance β. When the spot instance's purchase rate β keeps constant, total cost decrease as the reserved capacity of spot instance Q increases and increases afterwards.

The reason is that the risk cost is much lower than purchase cost, total cost is monotone decreasing function. When the reserved capacity of spot instance is determined, total cost decrease as purchase rate of spot instance β increases. When the purchase rate of spot instance keeps constant, the purchase cost decreases and then increases, but the risk cost is continue decreasing. Moreover, risk cost is too low; the total cost continues maintaining the trend of decreases at first and then increases.

The computing results show that, when reserved capacity of reserved instance $Q=70$, $\beta=1$, the total cost reaches minimum value, that is TCO=$\$2.596$*104.

From above analysis we can know, risk cost is relatively low, so the total cost is a monotone decreasing function, and when purchase rate of spot instance $\beta=1$, total cost is minimum. Therefore, in this paper we set $\beta=1$.

In previous works, if we just combined the on-demand instance and reserved instance (Wang et al., 2015a), we can calculate purchase cost and total cost under the nonlinear relationship which is compared with our model.

When $\beta=0$, $Q=100$, the purchased cost is minimum, min $(EC)=TCO=\$3.452*10^4$.

Similar to Sadashiv et al. (2014), if cloud customers purchase these three instance as a combination without considering the risk cost, we can calculate purchase cost and total cost.

When $Q=50$, $\beta=1$ the purchase cost is minimum, min $(EC)=\$2.488*10^4$. And the total cost is $\$2.6481*10^4$.

5.2 Multi-instance Combination Decision Under Linear Stochastic Business Demand and Time

In this section, we make an assumption that the random business demand of every month $f(t)$ follows $f(t)=at(a>0)$, as shown in Fig. 6.

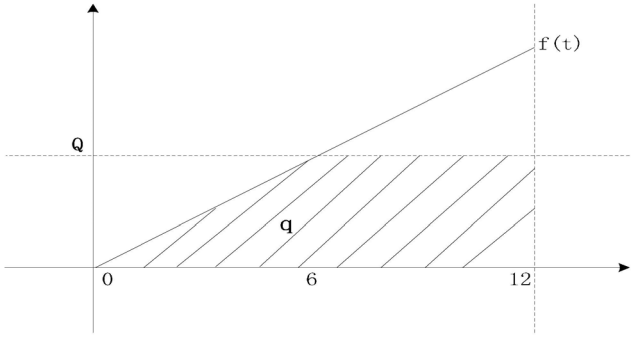


Fig. 6. The random business demand of multi- instance

The random business demand cycle is one year and the unit time slot is one month, so the time slot T=12, min $t \in [0,12]$. The demand function of cloud customers' multi-instance min $F(T)$ can be written as follows:

$$F(T) = \int_0^T f(t) \times T_1 dt = \int_0^{12} f(\alpha \times t) \times T_1 dt = 72\alpha T_1$$

The reserved capacity of reserved instance Q obeys that: $0 < Q < 12$, the capacity of reserved instance q is:

$$q = (12 - \frac{Q}{a} + 12) \times Q \times \frac{1}{2} \times T_1 = (12Q - \frac{Q^2}{2a}) \times T_1$$

Take Amazon EC2 as an example, the type of instance is ec2-eu-west-1.linux.m1.small, and the pricing lists are as follows: $P_0 = \$227.5$, $P_1 = \$0.004$, $P_2 = \$0.095$, $\theta = 0.9$. We assume that $\alpha = 10$, $\lambda = 0.6$ (Take one month as the unit time slot unit), so $T_0 = 14 \times 30$, $T_1 = 30 \times 10$. The purchase cost EC, risk cost IC, total cost TCO are respectively:

$$EC = 227.5Q + 8640 + (720 - 12Q + Q^2/20) \times (0.055 - 0.043\beta) \times 300$$
$$TC = F(T) - q\beta(1-\theta) \times p_2 \omega\lambda = 0.0025\beta(720 - 12Q + Q^2/20) \times 300$$
$$TCO = 227.5Q + (720 - 12Q + Q^2/20) \times 300(0.055 - 0.0405\beta) + 8640$$

The trends of purchase cost, risk cost and total cost as the reserved capacity of reserved instance Q and the purchase rate of spot instance β changes are shown in Fig. 7. The 3D results' z axis respects the EC (purchase cost), IC (risk cost) and TCO (total cost) respectively. And the x axis indicates the changes range of purchase rate of spot instance β, and the y axis respect the changes range of the reserved capacity of reserved instance Q.

As Fig. 7 indicates, the random business demand and time is liner relationship, the change trends of purchase cost, risk cost and total cost are same as multi-instance combination decision under nonlinear stochastic business demand and time and according to the computing results:

When $Q = 60, \beta = 1$, the purchase cost is minimum, $\min(EC) = \$2.942 \times 10^4$

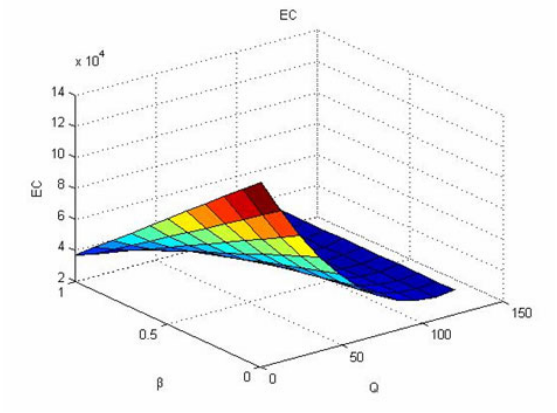
When $Q = 72, \beta = 1$, the total cost is minimum, $\min(TCO) = \$3.053 \times 10^4$

In previous works, if we just combined the on-demand instance and reserved instance (Wang et al. [4]), we can calculate purchase cost and total cost which is compared with our model.

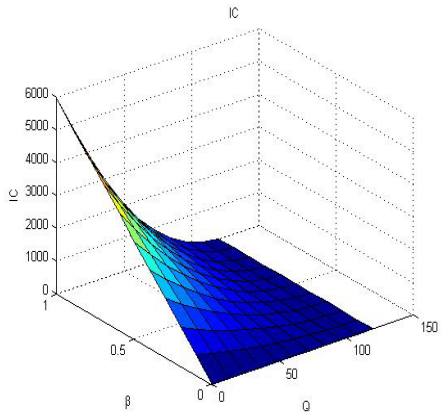
When $\beta = 1, Q = 108$, the purchased cost is minimum, $\min(EC) = TCO = \$3.52 \times 10^4$

Similar to Sadashiv et al. [15] (2014), if cloud customers purchase multi-instance in a combination method without considering risk cost, we can calculate purchase cost and total cost.

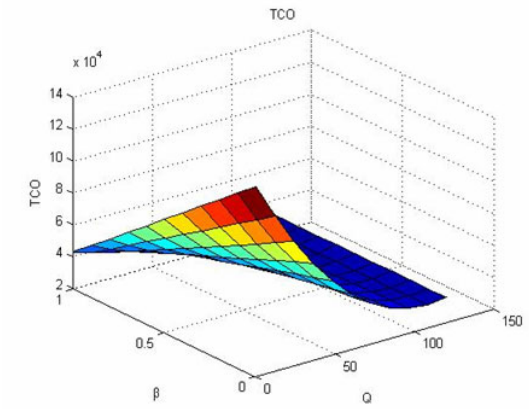
When $Q = 60, \beta = 1$, the purchase cost is minimum, $\min(EC) = \$2.942 \times 10^4$, the total cost is $\$3.0903 \times 10^4$.



(a) Purchase cost



(b) Risk cost



(c) Total cost

Fig. 7. The purchase cost, risk cost and total cost under the linear stochastic business demand and time

It is concluded that our model can minimize the purchase cost of cloud customers' multi-instance and the risk cost under linear and nonlinear cloud demand and time.

6 Evaluation

This section makes a further analysis for the optimal portfolio decision under the situation of the cloud users' demand fluctuation α, service time T1and risking sensitive factor λ changes. When the value of β is determined, the main task is to analyze the optimal reserved capacity of reserved instance under different situation.

Finding 1. The relationship between optimal portfolio decision and cloud demand fluctuation

According to above model, cloud demand fluctuation range of business time slot is 2a, and the variable value: $\alpha = 0.6, T_0 = 14, T_1 = 10$. Fig. 8. shows the relationship between optimal portfolio decision and cloud demand fluctuation.

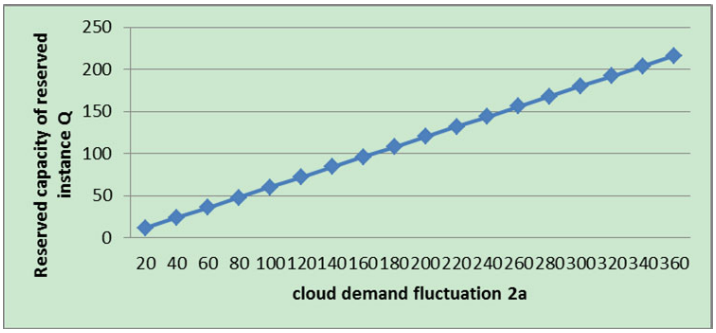


Fig. 8. The changing trend between reserved capacity of reserved instance and cloud demand fluctuation

It shows that the reserved capacity of reserved instance Q increases as the business time increases. This is because that total business demand F(T) is the function of a, the change trend of F(T) is same as the reserved capacity of reserved instance Q.

Finding 2. The relationship between optimal portfolio decision and business time

From above model, the business time T1, and we set $\alpha = 50, \lambda = 0.6$. Fig. 9 shows the relationship between optimal portfolio decision and business time.

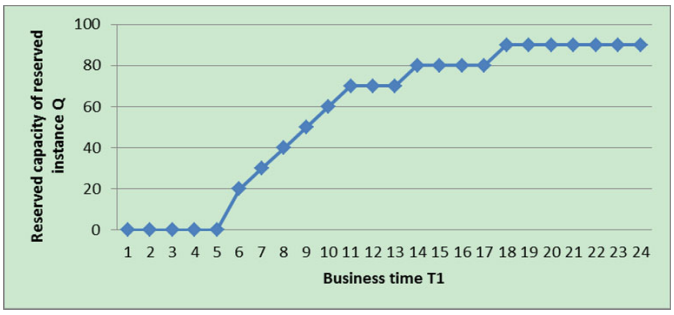


Fig. 9. The trend between reserved capacity of reserved instance and business time

We find that when $T_1 \le 5, Q = 0$, as $T_1 Q$ increases gradually; after $T_1 \ge 11, Q$ grows slowly; $T_1 \ge 18$, the reserved capacity of reserved instance keep 90 at all time. So it is conclude that when the business time is small, utilization rate of reserved instance is low, it's not good choice to purchase the spot-instance and on-demand instance. However, the total demand increases as business time increases, and the utilization rate of reserved instance and its reserved capacity show the same trend as business time. But, the reserved capacity is too large which leads to the utilization rate decreases, i.e., although the business time reaches whole day (24 hours), reserved capacity is just near but not equal to the most daily demand.

Finding 3. The relationship between optimal portfolio decision and risk sensitive factor

Here, we set $\alpha = 50, T_1 = 10$. Fig. 10. shows the relationship between optimal portfolio decision and risk sensitive factor λ. The reserved capacity of reserved instance increases as risk sensitive factor λ increases. Although the increasing extent is small, the overall trend is escalation. Due to the risk cost is relatively higher, cloud users are more likely to purchase reserved instance to avoid risk with non-risk.

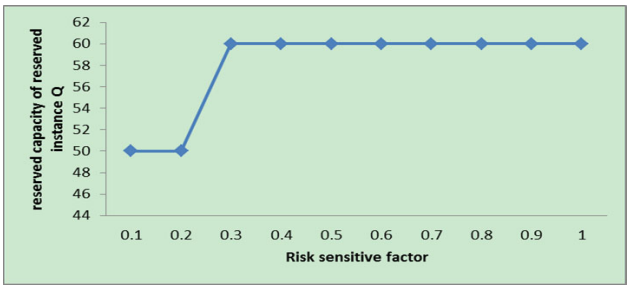


Fig. 10. The trend between reserved capacity of reserved instance and risking sensitive factor

In addition, we also conclude when the random business demand and time is linear relationship, the change trend between optimal portfolio decision and cloud users demand fluctuation, business time and risking sensitive factor are still same as Finding1-Finding 3. It is clear that these conclusions is of high adaptability.

7 Managerial Implications

The application of multi-instance optimal combination decision has economic implication for cloud customers On the one hand, we provide multi-instance combination decision as a part of a toolbox for cloud customers, who need to purchase their cloud instance. Especially, the cloud customers can freely choose the optimal acquisition decision to decrease cost. As we all known, different enterprises have their own cloud demand, and many of them are just beginning to adapt cloud service, their demand and history data cannot be traced by themselves. The optimal combination model gives them reasonable purchase suggestion and reduces adaptation cost of cloud service which also decreases the risk of spot instance.

8 Conclusion

From cloud users' perspective, we introduce risk cost and establish a new combination decision model based on three types instance model under cloud users' demand fluctuation. According to change trend of cost, we analyze the condition and cause of optimal decision, then make a further discussion about the relationship between optimal decision and demand fluctuation, business time, risk sensitive factor. Through analyzing optimal solution under linear and nonlinear customer demand and time, we prove that the conclusion is high adaptability which can give a good suggestion for the cloud users to reduce cost when purchasing cloud service.

One possible direction for future research is to compare the proposed methods with other approach, and research the service combination decision under multi-providers and multi-service demand and different levels instance. The proposed approach could be valuable under the situation of multi-providers and multi-service demand.

Acknowledgements

This work is financially supported by the National Natural Science Foundation of China (No. 71172043), the Fundamental Research Funds for the Central Universities (No. 2014-yb-017), Humanity and Social Science Youth foundation of Ministry of Education of (No. 14YJCZH165).

References

[1] J. Yu, X. Xiao, Y. Zhang, From concept to implementation: the development of the emerging cloud computing industry in China, Telecommunications Policy 40(2)(2016) 130-146.

[2] V. Abhishek, I.-A. Kash, P. Key, Fixed and market pricing for cloud services, in: Proc. 2012 IEEE Conference on Computer Communications Workshops, 2012.

[3] A.N. Toosi, On the economics of infrastructure as a service cloud providers: pricing markets and profit maximization, [dissertation] Parkville, Australia: University of Melbourne, 2014.

[4] W. Wang, B. Liang, B. Li, Optimal online multi-instance acquisition in IaaS clouds, IEEE Transactions on Parallel and Distributed Systems 26(12)(2015) 3407-3419.

[5] J. He, Y. Wen, J. Huang, D. Wu, On the cost–QoE tradeoff for cloud-based video streaming under Amazon EC2's pricing models, IEEE Transactions on Circuits and Systems for Video Technology 24(4)(2014) 669-680.

[6] A. Andrzejak, D. Kondo, S. Yi, Decision model for cloud computing under sla constraints, in: Proc. 2010 IEEE International

Symposium on Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), 2010.

[7] L. Wu, S.-K. Garg, R. Buyya, Sla-based resource allocation for software as a service provider (saas) in cloud computing environments, in: Proc. 2011 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), 2011.

[8] M. Mattess, C. Vecchiola, R. Buyya, Managing peak loads by leasing cloud infrastructure services from a spot market., in: Proc. 2010 12th IEEE International Conference on High Performance Computing and Communications (HPCC), 2010.

[9] S. Yi, D. Kondo, A. Andrzejak, Reducing costs of spot instances via checkpointing in the amazon elastic compute cloud, in: Proc. 2010 IEEE 3rd International Conference on Cloud Computing (CLOUD), 2010.

[10] Q. Yuan, Z. Liu, J. Peng, X. Wu, J. Li, F Han, A leasing instances based billing model for cloud computing, in: Proc. 2011 International Conference on Grid and Pervasive Computing, 2011.

[11] S. Chaisiri, B.-S. Lee, D. Niyato, Robust cloud resource provisioning for cloud computing environments, in: Proc. 2010 IEEE International Conference on Service-Oriented Computing and Applications (SOCA), 2010.

[12] W. Wang, D. Niu, B. Liang, B. Li, Dynamic cloud instance acquisition via IaaS cloud brokerage, IEEE Transactions on Parallel and Distributed Systems 26(6)(2014b) 1580-1593.

[13] Y. Ran, J. Yang, S. Zhang, H. Xi, Dynamic iaas computing resource provisioning strategy with qos constraint, IEEE Transactions on Services Computing 10(2)(2017) 190-292.

[14] M. Armbrust, A. Fox, R. Griffith, A view of cloud computing, Communications of the ACM 53(4)(2010) 50-58.

[15] N. Sadashiv, D.-K. SM, R. Goudar, Hybrid spot instance based resource provisioning strategy in dynamic cloud environment, in: Proc. 2014 International Conference on, High Performance Computing and Applications (ICHPCA), 2014.

[16] J. Doyle, V. Giotsas, M.-A. Anam, Y. Andreopoulos, Cloud instance management and resource prediction for computation-as-a-service platforms, in: Proc. 2016 IEEE International Conference on Cloud Engineering (IC2E), 2016.

[17] J.-P. Hughes, L.-J. Mester, Who said large banks don't experience scale economies? evidence from a risk-return-driven cost function, Journal of Financial Intermediation 22(4)(2013) 559-585.

[18] M. Wu, J. Yang, Y. Ran, Dynamic instance provisioning strategy in an iaas cloud, in: Proc. 2013 32nd Chinese on Control Conference (CCC), 2013.