Optimal Strategy Modelling in an Online Auction for the Rent of Computing Resources

Anna Ivashko[1,2] and George Safonov[2]

[1] Institute of Applied Mathematical Research of the Karelian Research Centre of the Russian Academy of Sciences, Pushkinskaya Str., 11, Petrozavodsk, 185910, Russia
[2] Petrozavodsk State University, Lenina Str., 33, Petrozavodsk, 185910, Russia
aivashko@krc.karelia.ru
jiri.safonov@gmail.com

Abstract. We consider the problem of optimal bid pricing in an auction for cloud computing resource allocation. The multistage model connected with the full-information best-choice problem is investigated. In this model, users set bids as a sequence of thresholds at each step to minimize the expected resource rent price. A form of the spot price distribution function is proposed based on real-life Amazon EC2 spot price dynamics. The optimal bid values, mean instance rent price and average step for buying the resource were found through simulations.

Keywords: Cloud Computing · Spot Instance · Mathematical Modeling · Best-choice Problem · Amazon EC2 · Threshold Strategies

1 Introduction

Cloud computing has a leading position in the IT-market, and is constantly increasing its market share. Public cloud computing resources are accessible, and allow users who require complex computing or other hardware and software tools to avoid software and hardware expenses and other costs. Also, these services respond flexibly and quickly to the changing computing needs of the users. Cloud services have also become increasingly popular in the modern world, where it became necessary to work remotely. Therefore, the demand for such resources has lately increased significantly. With cloud services available, a user only has to pay the rent of the computing resource he/she needs and can then start using it immediately.

One of the major computing resources providers on the web is the cloud computing platform Amazon EC2 [1]. Virtual machines on this platform are in the form of so-called "instances". Amazon EC2 offers its users various terms for renting instances. A new type of instances is Spot Instances (SI). SI are purchased through auction. Users can bid on unused Amazon EC2 capacity. The spot price changes all the time, and users can monitor purchase prices over previous periods. Therefore, the problem of optimal bid pricing in an auction for renting a certain computing resource is of high relevance.

In this paper, the following model of an auction for renting a spot resource (see Ivashko et al. [2]) is investigated. The user observes spot prices at discrete moments of time and wants to buy the instance within a predetermined time period. At each stage he/she can set their bid. If that bid exceeds the current spot price, the instance is sold to the user. The objective is to find the optimal strategy (minimal bid) that would minimize the cost of getting the instance within a specific time period. This model is reduced to the best-choice problem, which is well-known in the optimal stopping theory.

This paper is structured as follows. Section 2 offers a brief review of related works. In Section 3, we outline the spot auction mechanism. The mathematical model of spot instance auctions is considered, and the strategy for minimizing the expected spot price is described in Section 4. Next, in Section 5, we investigate the spot price dynamics and conduct parameter fitting for the distribution function of real data from real-like Amazon SI auctions. Finally, in Section 6, we present the findings and conclusions, and draft plans for the future.

2 Related Works

Different schemes based on competitive prices are used when buying/selling or renting resources. One can name various auctions and tenders, competition for computing resources or storage capacity. Cloud computing resources are gaining significance in the IT world. Therefore, a lot of research is going on in the area.

Research papers are dedicated to both exploring the existing pricing mechanisms and behavior of real data, and proposing new mechanisms. There are various approaches to analyzing cloud resource pricing. Different probabilistic and statistical pricing models are considered in papers by Kumar et al. [3], Abhishek et al. [4], and by Xu et al. [5]. Probabilistic models are grounded in the assumption that spot prices are independent and identically distributed random variables. Amazon EC2 spot price dynamics has been analyzed by Ben-Yehuda et al. [6], Cheng et al. [7], and by Javadi et al. [8]. It is also noted in these papers that spot prices have a random nature.

An important task is to model the competitive behavior of participants in various types of auctions. Online auction is the best example of modern markets. In an auction, it is necessary for a participant to determine his/her optimal strategy in order to increase the chance of winning.

Auction designs vary. Mechanisms for bidding in auctions for a computing resource are proposed in papers by Karunakaran et al. [9], Sowmya, Sundarraj [10], Tang et al. [11], and Wang et al. [12].

Optimal stopping models are often used to study the behavior of participants in socio-economic systems. The optimal stopping problems where the participants have to take the decision about choosing an item that fulfills certain criteria are also known as best-choice problems. One of the first to describe the classical statement of the best-choice problem was Moser [13]. Such problems often arise in economics, finance, sociology and politics.

The users' optimal strategies are often derived using the dynamic programming method, in which a complex problem can be solved by breaking it down into simpler sub-problems represented as a recursive sequence. This method has various applications in different spheres, such as the house-selling problem, the job-search problem, or the mate-choice problem, and is successfully used in economic contexts for best-choice problems and auctions.

In some schemes of online auctions for buying an item or resource participants do not know in advance the exact current price. Each customer can set the threshold price value (bid) at which he/she agrees to buy the resource. During the auction the deal is made if the price offered by a customer exceeds the current price. An important task is to analyze the behavior of participants in this type of auction. The optimal behavior of participants of various online auctions has been studied by the best-choice problem in papers by Harrall et al. [14], Mazalov, Ivashko [15], Babaioff et al. [16].

Ivashko et al. [2] applied the best-choice model to determine the optimal bid value in order to win the auction. The authors considered this probabilistic model for the cases where spot prices have a uniform distribution on [0,1] or a normal distribution. However, further studies prove that the distribution of spot prices has a more sophisticated structure.

In this paper, we use the model proposed by Ivashko et al. [2]. We assume that the price dynamics of each spot instance is modeled by a mixture of Gaussian distributions. This assumption is supported by studies in a paper by Javadi et al.[8]. We estimated the unknown parameters of the mixture of Gaussian distributions using the EM-algorithm. Optimal bid values in the auction were determined and numerically simulated.

3   Amazon Spot Auctions

Elastic Compute Cloud platform (Amazon EC2) [2] is one of the earliest and most popular web services offering the computing resources of the cloud computing system AWS (Amazon Web Services) for use. The AWS system is located in several data centers in different regions of the world so that the user can find the most technically available resource. Resources are available to users in the form of instances (virtual machines). A client pays a certain per hour fee. He/she can also choose the type of instances needed for the calculations and decide how to purchase instances. There are three alternative ways to buy instances. On-demand Instances are charged full price. A user can purchase an On-demand Instance when he/she needs it with no long-term commitments or upfront payments. Reserved Instances are rented for a fixed duration. They provide the user with a significant discount (up to 75%) compared to On-Demand Instance pricing. The Reserved Instance will always be available in the availability zone in which it was purchased. Spot Instances (SI) are idle EC2 instances available for less than the On-Demand price (up to 90% saving). A Spot Instance runs whenever capacity is available but, unlike On-demand and Reserved Instances, can be interrupted by the AWS system at any time.

SIs are in high demand among users who do not need to use the resources continuously. SIs are well-suited for data analysis, batch jobs, background processing, and optional tasks. SIs are rented through an auction for unused capacity. Customers bid the maximum cost of instance per hour of use they would be ready to pay. If their bids exceed the spot price, they win the auction. All winning customers pay the same price, which is equal to the value of the lowest winning bid. This closed auction is a Vickrey auction.

4   Optimal Strategy

This paper examines SIs and the auction for their allocation. The mathematical best-choice model proposed by Ivashko et al. [2] is used to determine the optimal bid value in order to win the auction.

The optimal stopping theory deals with the problems of choosing the time to make a decision based on observing sequentially the random values. The aim is to maximize the expected payoff. In the optimal stopping theory there is a class of best-choice problems. They arise from real choice processes. The best-choice problem with full-information is directly related to the problem of optimal bid pricing in a spot auction.

We describe the best-choice model as follows [2]. A customer willing to rent an SI within a certain time interval observes sequentially the spot prices X_1, X_2, $X_3, \ldots$ as a sequence of independent and identically distributed random variables from a known continuous cumulative distribution function $F(x)$ on the interval $[p_{min}, p_{max}]$. p_{min} corresponds to a minimal possible SI price at the auction. p_{max} is a maximal possible price not exceeding the on-demand price. There are n steps available to win a SI. The customer's strategy is a sequence of thresholds (bids). He/she makes the bid τ_i before i-th step, $i = 1, ..., n$. His/her aim is to minimize the expected spot price for the whole period of time.

Full-information best-choice problems with finite horizon are solved by the dynamic programming method. Let us find the optimal behavior of a customer. If the customer does not win a SI within the period n, he/she will have to buy it for the maximum price p_{max}. It makes sense for the customer to buy the SI at any price below or equal to p_{max}. So, the bid at the last step is $\tau_n = p_{max}$.

Because spot prices are independent and identically distributed random variables with a continuous cumulative distribution function $F(x)$, the expected SI price at the last step n is equal to $\int_{p_{min}}^{\tau_n} x dF(x)$. At the step $n-1$ the customer's bid τ_{n-1} is the minimum between the current SI price and the expected cost at the n-th step:

$$\tau_{n-1} = \mathbf{E}[\min\{X, \tau_n\}] = \int_{p_{min}}^{p_{max}} \min\{x, \tau_n\} dF(x) = \int_{p_{min}}^{\tau_n} x dF(x) + \int_{\tau_n}^{p_{max}} \tau_n dF(x).$$

Here, $\mathbf{E}[X]$ is the expectation of the random variable X.

Continuing the process, we get that the optimal bid value τ_i at each step i is rendered by the system of recurrent equations:

$$\tau_n = p_{max},$$
$$\tau_i = \int\limits_{p_{min}}^{p_{max}} \min\{x, \tau_{i+1}\} dF(x) = \int\limits_{p_{min}}^{\tau_{i+1}} x f(x) dx + \int\limits_{\tau_{i+1}}^{p_{max}} \tau_{i+1} f(x) dx,\ i = 1, ..., n-1,$$

(1)

where $f(x)$ is the probability density function of the distribution $F(x)$.

If the customer's bid τ_1 at the first step does not win, the next bid τ_2 is used at the second step. Continuing this process, the customer is guaranteed to get an instance during the time period n with the minimum expected spot price.

5 Optimal Strategy Modelling

We describe a method to derive the strategy for optimal bid pricing. We use real Amazon EC2 spot prices data [18]. Based on this history, we build the prices' probability density function and derive the optimal threshold bids using formula (1). Wolfram Mathematica platform is used for modelling.

Price histograms are plotted using spot price statistics. Figure 1 shows examples of price histograms for the instances *us-east-1b m4.xlarge linux/UNIX* and *us-east-1b m4.2xlarge linux/UNIX*.
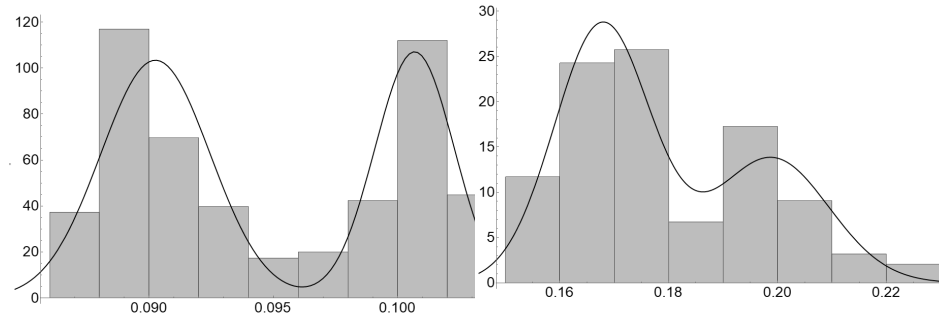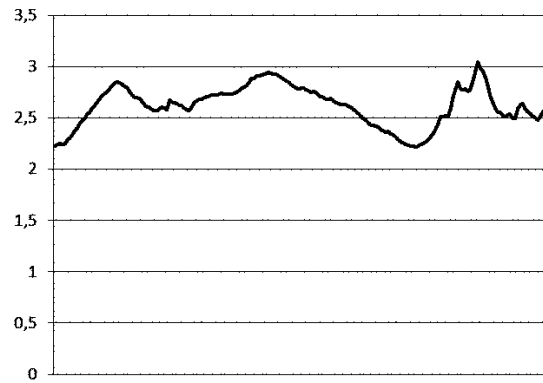

Fig. 1. Histograms and probability density functions of spot prices

Analyzing the data one can construct the prices probability density function, see Figure 1. We used the following truncated mixture of Gaussian distributions:

$$f_{b_1,b_2}(x) = \begin{cases} \omega_1 \dfrac{C}{\sqrt{2\pi}\sigma_1} e^{\frac{-(x-a_1)^2}{2\sigma_1^2}} + \omega_2 \dfrac{C}{\sqrt{2\pi}\sigma_2} e^{\frac{-(x-a_2)^2}{2\sigma_2^2}}, x \in [b_1, b_2], \\ 0, x \notin [b_1, b_2]; \end{cases}$$
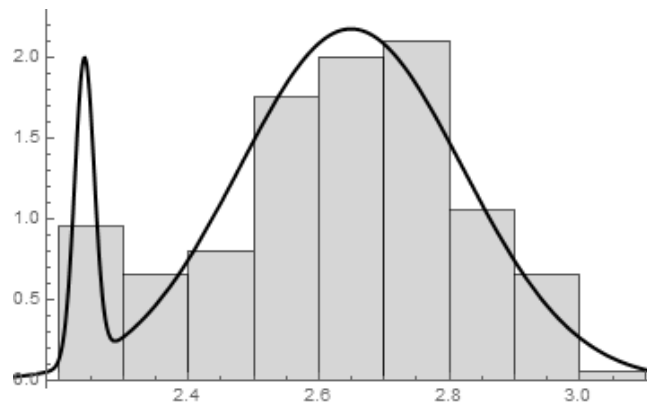
(2)

here $C = \dfrac{1}{\displaystyle\int_{b_1}^{b_2} \left(\dfrac{\omega_1}{\sqrt{2\pi}\sigma_1} e^{\frac{-(x-a_1)^2}{2\sigma_1^2}} + \dfrac{\omega_2}{\sqrt{2\pi}\sigma_2} e^{\frac{-(x-a_2)^2}{2\sigma_2^2}} \right) dx}$.

Take for example the optimal strategy modelling for the *us-east-1e p2.8xlarge Linux/UNIX* instance over the period of July12–August 30, 2019. We have $N = 200$ values of spot prices (see Figure 2).

Fig. 2. Spot pricing dynamics

Having the data we construct the histogram of prices (Figure 3). The step in the histogram is $1 + [\log_2 N]$, $N = 200$.

Fig. 3. Histogram and probability density function of spot prices for the *us-east-1e p2.8xlarge Linux/UNIX* instance

Using the data we get in formula (2) $b_1 = p_{min} = 2.216$, $b_2 = p_{max} = 3.048$, where p_{min} corresponds to the minimum price over the entire period of time, and p_{max} corresponds to the maximum price.

We tested the statistical hypothesis using Pearson's chi-squared test (χ^2). The null hypothesis H_0 stated that the frequency distribution of certain events observed in a sample (using histogram) is consistent with a mixture of two Gaussian distributions. The alternative hypothesis H_1 is that the frequency distribution does not have the form of a mixture of two Gaussian distributions. Having run the χ^2 test at a significance level $\alpha = 0.05$, we got $P\text{-}Value = 0.801189$. As $P\text{-}Value > \alpha$, so normally we would not reject the null hypothesis.

We estimate the unknown distribution parameters $a_1, \sigma_1, a_2, \sigma_2, \omega_1, \omega_2$ using an Expectation Maximization (EM) algorithm. The EM-algorithm is used to find maximum likelihood parameters of a statistical model in cases where the equations cannot be solved directly.

We consider the vector of the estimated parameters $\theta = (\omega_1, \omega_2; a_1, a_2; \sigma_1, \sigma_2)$.

The density function of the normal distribution has the form:

$$N(x, a_j, \sigma_j) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(x-a_j)^2}{2\sigma_j^2}}, \; j = 1, 2.$$

Let us consider two steps of the EM-algorithm:

The expectation (E) step:

$$g_{ij} = \frac{\omega_j N(x_i, a_j, \sigma_j)}{\sum\limits_{s=1}^{2} \omega_s N(x_i, a_s, \sigma_s)}, \; i = 1, ..., m;$$

The maximization (M) step:

$$\omega_j = \frac{1}{m} \sum_{i=1}^{m} g_{ij},$$

$$a_j = \frac{1}{m\omega_j} \sum_{i=1}^{m} g_{ij} x_i,$$

$$\sigma_j^2 = \frac{1}{m\omega_j} \sum_{i=1}^{m} g_{ij}(x_i - a_j)^2, \quad j = 1, 2,$$

here, m is the number of elements in the sample, g_{ij} is the probability that the observation x_i comes from the component j of the mixture of two Gaussian distributions.

We should iterate the steps E and M until convergence: until the value difference between a_1, a_2 at step h and a_1, a_2 at step $h-1$ is less than 10^{-4}.

The application of the EM-algorithm yielded the estimations shown in Table 1.

Next, we get the optimal threshold values (bids) for $n = 15$ steps of the model using formula (1) (Table 2). As one can see, the sequence of bids is increasing with i, $i = 1, ..., n$.

To illustrate the optimal behavior let us consider the time period September $5 - 28$, 2019 with the number of steps $n = 15$. In Figure 4 the optimal thresholds

ω_1	ω_2	a_1	a_2	σ_1	σ_2
0.0701	0.9299	2.2404	2.6492	0.0149	0.1707

Table 1. Parameter estimations

Table 2. Bid values generated by the optimal strategy for $n = 15$

i	1	2	3	4	5	6	7	8	9	10
τ_i	2.5772	2.5775	2.5779	2.5784	2.579	2.58	2.581	2.583	2.585	2.588

i	11	12	13	14	15
τ_i	2.592	2.598	2.607	2.622	3.048

(grey line) and spot prices (black line) are presented at each step. Here, the acceptance step is $i = 13$, and the bid value which exceeded the spot price is equal to 2.5828.
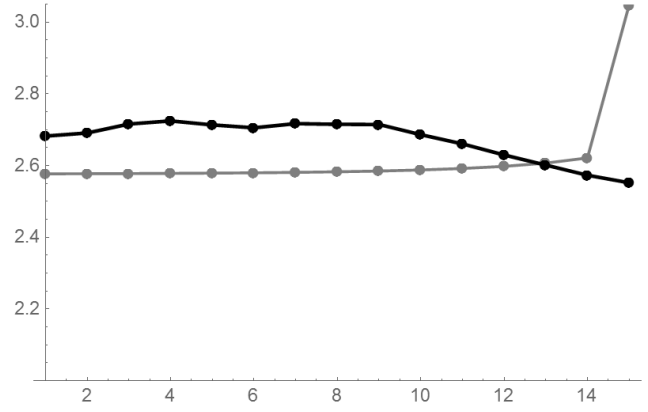
Fig. 4. Spot prices and threshold values for $n = 15$

In Table 3 the results on the optimized mean spot price and mean acceptance step are given for $n = 5$, $n = 10$, $n = 15$, and for 90 price values from time period September 7, 2019 – October 2, 2019.

As illustrated by Table 3, if the number of steps available in an auction increases, the mean acceptance step also grows, while the mean spot price decreases. As a result, the more steps are available for participating in the auction, the lower is the price at which a user can rent the resource.

Table 3. Simulations for different n

n	Mean spot price	Mean acceptance step
5	2.46043	2.3
10	2.45081	3.9
15	2.43388	4

6 Conclusion

This paper examines the problem of optimal bid pricing by cloud service users. The problem is connected with the full-information best-choice problem. Based on real data dynamics, the spot price distribution is modeled as a mixture of Gaussian distributions. Thus, we apply the theoretic model to real spot price dynamics. The optimal bid values, mean instance rent price and the acceptance step for buying the instance were found through simulations. These values allow winning the auction while minimizing the expected SI rent price.

In the future, new models for optimal bid pricing in spot auctions can be proposed. Various strategies can be compared to identify their respective benefits and limitations. We collected the price history during one year and we can use these data to further research. Also the model where the prices are dependent random variable can be considered. For this case a new model can be constructed.

References

1. Amazon Inc. Amazon Elastic Compute Cloud (Amazon EC2). http://aws.amazon.com/ec2
2. Ivashko, E.E., Tchernykh, A., Ivashko, A.A., Safonov, G.R.: Cost-efficient strategy in clouds with spot price uncertainity. Automation and Remote Control **81**(4), 731–745 (2020). https://doi.org/10.1134/S000511792004013X.
3. Kumar, D., Baranwal, G., Raza, Z., Vidyarthi, D. P.: A Survey on Spot Pricing in Cloud Computing. Journal of Network and Systems Management, 809–856 (2018). https://doi.org/10.1007/s10922-017-9444-x
4. Abhishek, V., Kash, I.A., Key, P.: Fixed and Market Pricing for Cloud Services. In:Proceedings of the 7th Workshop Economics of Networks System Computer (NetE- con 2012), IEEE Computer Society, Orlando, 20 March 2012, 157–162.
5. Xu, H., Li, B.: Maximizing revenue with dynamic cloud pricing: The infinite horizon case. 2012 IEEE International Conference on Communications (ICC), Ottawa, ON, 2929–2933 (2012). https://doi.org/10.1109/ICC.2012.6364013
6. Ben-Yehuda, A. O., Ben-Yehuda, M., Schuster, A., Tsafrir, D.: Deconstructing Amazon EC2 Spot Instance Pricing. Cloud Computing Technology and Science (Cloud-Com), 2011 IEEE Third International Conference on, Athens, 304–311 (2011). https://doi.org/10.1109/CloudCom.2011.48
7. Cheng, H.K., Li, Z., Naranjo, A.: Cloud computing spot pricing dynamics: Latency and limits to arbitrage. Information Systems Research **27**(1), 145–165 (2016).

8. Javadi, B., Ruppa K. Thulasiram, Rajkumar Buyya: Characterizing spot price dynamics in public cloud environments. Future Generation Computer Systems **29**(4), 988–999 (2013). ISSN 0167-739X, http://dx.doi.org/10.1016/j.future.2012.06.012.

9. Karunakaran, S., Sundarraj, R. P.: Bidding Strategies for Spot Instances in Cloud Computing Markets. in IEEE Internet Computing **19**(3), 32–40 (2015). https://doi.org/10.1109/MIC.2014.87

10. Sowmya, K., Sundarraj, R.P.: Strategic bidding for cloud resources under dynamic pricing schemes. Proceedings – 2012 International Symposium on Cloud and Services Computing, ISCOS 2012. art. no. 6481231, 25–30 (2013).

11. Tang, S., Yuan, J., Li, X.Y.: Towards Optimal Bidding Strategy for Amazon EC2 Cloud Spot Instance. Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on, Honolulu, HI, 91–98 (2012) https://doi.org/10.1109/CLOUD.2012.134

12. Wang, W., Liang, B., Li, B.: Revenue maximization with dynamic auctions in IaaS cloud markets. Quality of Service (IWQoS), IEEE/ACM 21st International Symposium on, Montreal, QC, 1–6 (2013) https://doi.org/10.1109/IWQoS.2013.6550265

13. Moser, L.: On a problem of Cayley. Scripta Math. **22**(5), 289–292 (1956).

14. Harrell, G., Harrison, J., Mao, G., Wang, J.: Online Auction and Secretary Problem. Int'l Conf. Scientific Computing. 241–244 (2015).

15. Mazalov, V.V., Ivashko, A.A.: Online Auction and Optimal Stopping Game with Imperfect Observation. Intelligent Information and Database Systems. ACIIDS 2020, LNCS, volume 12033. Springer, 145–156 (2020). https://doi.org/10.1007/978-3-030-41964-6_13

16. Babaioff, M., Immorlica, N., Kempe, D., Kleinberg, R.: Online auctions and generalized secretary problems. SIGecom Exch **7**(2), 1–11 (2008). http://doi.acm.org/10.1145/1399589.1399596

17. Amazon EC2 Spot Instances. http://aws.amazon.com/ec2/spot/

18. Spot Instance Pricing History Amazon Elastic Compute Cloud http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-spot-instances-history.html