Probability Density for Amazon Spot Instance Price

H M Dipu Kabir[1], Abbas Khosravi[1], M Anwar Hosen[1], Saeid Nahavandi[1], Rajkumar Buyya[2]
[1]Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Australia
[2]Cloud Computing and Distributed Systems (CLOUDS) Laboratory,
Department of Computing and Information Systems The University of Melbourne Melbourne Victoria Australia
{dkabir, abbas.khosravi, anwar.hosen, saeid.nahavandi}@deakin.edu.au, rbuyya@unimelb.edu.au

***Abstract*—In this paper, we develop a technique to calculate probability density of Amazon Spot Instance (SI) price. We consider both of the curve-fitting and historical similarities to compute the probability density of the SI price. Traditional SI bidding systems compute the bid price as a single value. That single value may not be suitable for all bidders. Different bidders have different urgency and preferences. The probability density of price may help users setting a bid-price considering both the urgency of the task and the condition of the market.**

***Keywords*—Amazon EC2, Spot Instance Management, Probability Density, Truthful Bidding.**

I. Introduction

Traditional point prediction algorithms provide an optimized value through the error optimization [1], [2]. Popularly applied error values are root-mean-square-error (RMSE), mean-absolute-percentage-error (MAPE), sum-squared-error (SSE) etc [3]. The resultant point prediction value is usually close to mean or the median. Therefore, bidding at the point prediction provide roughly 50% probability of winning the bid. Bidding high can potentially increase the price of the market and all bidders may suffer from that [4], [5]. Therefore, the bidder needs to know about all probable values of the price to complete their jobs with a good cost-efficiency tradeoff. When the probability density function (PDF) is available, the bidder can choose a value depending on the urgency of the task and how much he is willing to pay.

Heteroscedastic error values with the point prediction can express the uncertain condition when the probability distribution is Gaussian [6]–[8]. NN based prediction interval (PI) or construction of PIs from similar patterns can represent the uncertainty with closer precision. However, the PDF contains all possible values and represents the exact uncertainty [9].

Uncertainty upper bound or uncertainty lower bounds have applied to solve complex problems efficiently with a reasonable number of trials [10]–[13]. Bidders can easily deduct one value of certain assurance from the PDF before bidding. Therefore, we propose a method to calculate the PDF to help bidders.

II. Current Rules in Amazon EC2 Spot Market

It has been a decade since Amazon launched their very first publicly accessible EC2 with limited number servers on August 25, 2006. A lot of changes has been brought on the configuration and pricing of EC2 servers. Therefore we present current condition of the market to help future readers in understanding the situation of SI market during June 2018.

Key features of SIs, related to the bidding are mentioned here as bullet-points. Readers can visit the website of Amazon [14] for more information.

- Users bid for spare Amazon EC2 Instances known as spot instance (SI) and he gets access to SI when his bid is higher than the price of the SI.
- As bids of the same price are ranked in random order, there is a probability of getting the SI when the bid is equal to the current spot price.
- The price of SI varies over time; based on the number of several probable factors; such as- the number of users, the number and the value of bids, available spare servers and so on. The provider's probable pricing algorithm is mentioned in the second subsection.
- When the price of SI becomes higher than the bid, the user is notified with a two-minute warning.
- The user can increase the price within two minutes, after receiving the warning.
- Also, he can save the progress of the instance within two minutes and stay idle after that and the instance terminates after two minutes.
- The user can also stay idle without saving the progress and the instance terminates after two minutes.
- When the user loses the instance due to the price increment the partial hour is not charged.
- When the user releases the instance, the partial hour is charged.
- The user is charged with the price of closing time as a full hour when he is releasing the instance.
- The user is charged with the price of the hour-end time for running instances [14], [15].
- Amazon has recently included Hibernation feature to several SIs on Nov 28, 2017. During the price increase, the SI goes to the hibernation state instead of the termination.
- The user is not charged for the partial hour prior to the hibernation. The user needs to pay for the backup storage at standard Amazon Elastic Block Store (EBS) storage rates.
- The user may also terminate the hibernated job and cancel the bid during hibernation.

978-1-5386-6376-9/19/$31.00 ©2019 IEEE

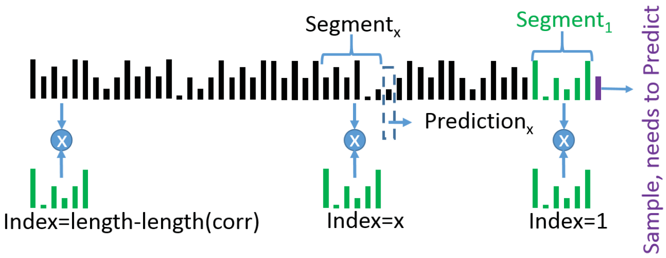

Fig. 1. Correlation of the string with recent samples for finding similar occurrences.

In order to manage instances reliability with the limited resources, Amazon has imposed some limits on the bidding of each user. The limitations are-

1) Some SI types are not available in all regions.
2) Each user account can bid for roughly 20 Spot instances per region. The limit is initially lower and the user can increase the limit by requesting.
3) The highest limit of the bid price is ten times the on-demand price.

In order to meet the requirements of bulk customers, Amazon introduces the Spot-fleet (SF), which is a collection of instances. Through the SF system, a user can bid for thousands of servers with a range of bids. That fleet also has several restrictions on the number of servers, bids, and regions. However, all of the limitations of SFs are not directly related to the bidding. Key features of SFs are as following-

1) The number of active SF in a region $\leq$ 1,000
2) The target capacity of an SF $\leq$ 3,000
3) The target capacity of all SF in a region $\leq$ 5,000
4) An SF request cannot span more than one region.
5) An SF request cannot span different subnets of a single region.
6) The SF supports the diversified strategy. Following the diversified strategy, the user can divide instances and bid at different prices.
7) Users can not bid into any pools with more than the on-demand price.

When the user can divide instances and bid at different prices, the price change varies the execution speed but the execution continues.

III. Probability Density Computation

A. Probability Density Through the Curve Fitting /Correlation

The method [16] of constructing probability density through historical similarities in the curve consists of searching similarities, normalizing predictions, providing empirical weight (relevance) for each similar events, drawing histograms and drawing cumulative probability distribution curves. Similar occurrences are searched through the direct correlation between recent samples and the training string. Fig. 1 presents the searching process with the indexing and Eqn. (1) represents the calculation of the normalized correlation value. Fig. 2 presents a graph which illustrates the locations of good matches using blue curves, and values of the next samples after each match through purple dots. These sample values denoted by purple

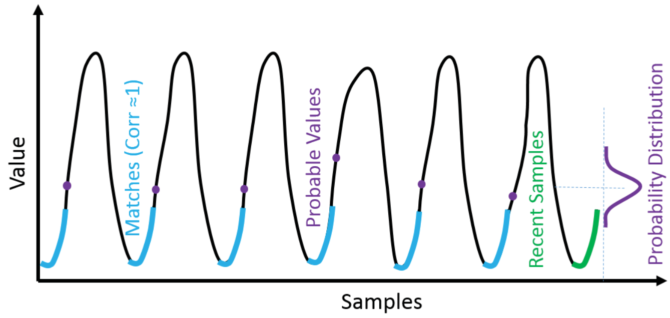

Fig. 2. A graph illustrating locations of good matches, values of the next samples after each match. These sample values are creating a probability distribution.

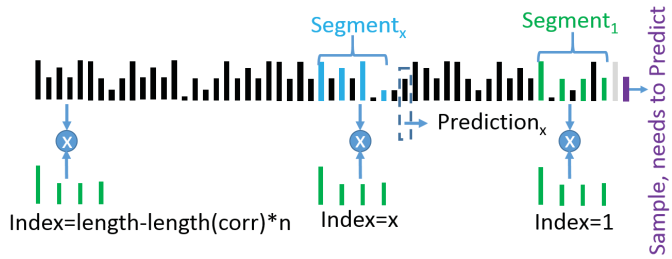

Fig. 3. Correlation based long term prediction distribution calculation.

dots creates a probability distribution. The probability distribution is shown to be a Gaussian one with no skewness in Fig. 2 and roughly presented by the purple curve on the right corner of the figure. Eqn. (1) presents the process of calculating the correlation.

$$Corr_{index} = \frac{\sum_{i=1}^{m} Segment_1(i) \times Segment_{index}(i)}{rms(Segment_1) \times rms(Segment_{index})} \quad (1)$$

Here, the sign $index$ conveys the same meaning as of Fig. 1, m is the length of each segment and rms means Root Mean Square. A higher value of m usually provides a slightly more accurate distribution with a higher execution time. Therefore, $m = 10$ is chosen for the balance between the accuracy and execution speed. The value of normalized correlation stays between -1 to +1 inclusive; +1 means the exact match, 0 means no match and -1 means the exact inverse match. After the search of similarities, the indexes which correspond to the best matches are selected for the formation of the cumulative probability distribution.

The correlation-based prediction system is designed in such a way that the prediction time needs to be equal to or an integer multiple of the sampling period. As the short-term prediction means the value after 5min and long-term prediction means the value after 1 hour, uniformly spaced samples at 5-minute intervals or lower intervals are required [17]. As samples are taken when the price of Amazon EC2 fluctuates, the samples are not uniformly spaced. Samples are uniformly spaced with 5 minutes interval by evaluation prices from the Amazon SI trace. Fig. 3 presents the correlation based long-term probability distribution calculation with downsampled segments. A downsampling of factor twelve is applied for the long-term (hourly) prediction, as the sampling period becomes one hour and the next sample becomes hourly prediction by doing so. However, the downsampling decreases the string size and results in fewer matches and a poor probability distribution function containing a few samples. Thus, the segments are downsampled by a factor (n) and the index is swept without skipping any value. Keeping the prediction time equals the sampling period, the searching can be performed from $index = 2$ to $index = (lengthofstring - lengthofthesegment)$. Similarly, when the prediction time is n times higher than the sampling period, searching can be performed from $index = n + 1$ to $index = (lengthofstring - lengthofthesegment * n)$. Correlating segments with $index < (n + 1)$ is possible but $prediction_x$ is not available for those points.

The normalized correlation equation compares only shapes of the segments. Two segments can have a similar waveshape with different amplitudes, resulting in the value of correlation close to +1. The prediction from that match needs to be normalized by dividing it by the RMS value of the corresponding segment and by multiplying it with the RMS value of the recent segment. Eqn. (2) presents the normalizing ratio (R_n). After the calculation of the ratio, Eqn. (3) is used for the ratio adjustment. Each prediction is given a weight or relevance score based on the value of the correlation and ratio. The empirical equation of the weight of prediction is presented as Eqn. (4).

$$R_n = \frac{rms(Segment_1)}{rms(Segment_{index})} \tag{2}$$

$$Prediction_{index} = Prediction'_{index} \times R_n \tag{3}$$

$$Relevance_{index} = Corr^{\eta}_{index} \times \frac{2}{R_n + 1/R_n} \tag{4}$$

Here, $Prediction'_{index}$ is the value of the next sample of the correlated segment ($Segment_{index}$) and $Prediction_{index}$ is the normalized version of that value. η is the weight parameter. The correlation parameter is given a higher weight with $\eta = 5$.

As the shape is more important compared to the ratio between the segments, the fifth order of the correlation is taken for the relevance calculation in (4). If the value of the correlation is slightly lower than one, the value is reduced while taking a high order. Fifth order is empirically taken with the adjustment of the ratio factor, combinedly defining the relevance. That relevance is used to plot the bar chart of the probability distribution. Fig. 4 presents the probability distribution of the price for Amazon EC2 $c4.4xlarge$ SI. The $Prediction_x$ points are distributed among 100 bars, uniformly spaced within the maximum and minimum values of predictions.

B. Limitations of the Curve Fitting based Prediction

Although the curve fitting based prediction can provide the user an idea of the recent trend [18], [19], it can not predict the exact value of the resource (SI) due to the limited search length. For example, the price of SI usually increases at the start of the peak hour and correlation-based prediction density function usually has a denser part at lower values due to the limitations of the search length. That increasing pattern varies from day to day based on various events; such as weekends, holidays, weekdays and the first day after holidays. Several aspects of that problem can be solved when each segment is longer than one week. However, correlating segments of one week is not computationally efficient. Also, the correlation of one week can not detect when the government holidays are. Many users may set automatic bidding for completing the job just before the holidays and their jobs are usually finished and SI's are returned during holidays/vacations. Also, curve fitting based techniques can not predict the exact price when it is becoming more than the on-demand price. When the predictions are normalized by the ratio, the value of one prediction on the probability density bar chart can be more than 10 times higher than the on-demand price, which is unrealistic.

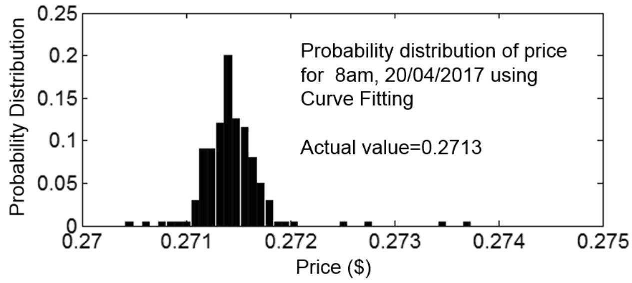

Fig. 4. Weighted distribution of the Amazon EC2 $c4.4xlarge$ SI price prediction using correlation based curve fitting.

When most of the users follow the curve fitting based technique, they become vulnerable to several market manipulation techniques; such as intentional market ramping [20]. There are small peaks at the EC2 price-traces during the off-peak hours while a number of users are bidding at a higher price due to the urgency of their task or carelessness. These price increases are usually for a very short span. That upward trend pattern can be similar to the upward trend of the start of peak hour. If most of the users are using the curve fitting based prediction, these upward trends can mislead them and potentially increase the market price. Similarly, a downward trend after the peak can shift the majority of the prediction distribution region below the minimum SI price. Therefore, a daily & weekly pattern based prediction with holiday consideration is required to know the exact value of the resource in order to solve these issues.

C. Probability Density Through the Daily and Weekly Patterns and Holiday Considerations

The daily and weekly patterns can predict the true value of the SI at a certain interval. The value of Amazon EC2 $c4.4xlarge$ is usually higher during the afternoon or the weekday evenings. However, peak hours may vary from instance to instance. Fig. 5 presents daily and weekly patterns of the Amazon EC2 $c4.4xlarge$ SI price. Data from the 10th April to the 19th April 2017 are downloaded to observe the situation. According to that figure, the price is higher during weekdays and the price usually has a daily peak during the evening. Moreover, when a user knows the daily and weekly pattern he can apply additional tricks. For example- he can choose a time when the price is expected to fall; as the user is charged with the end price, he would pay less. Also, when the task is not large enough he can bid at the rising edge. This means

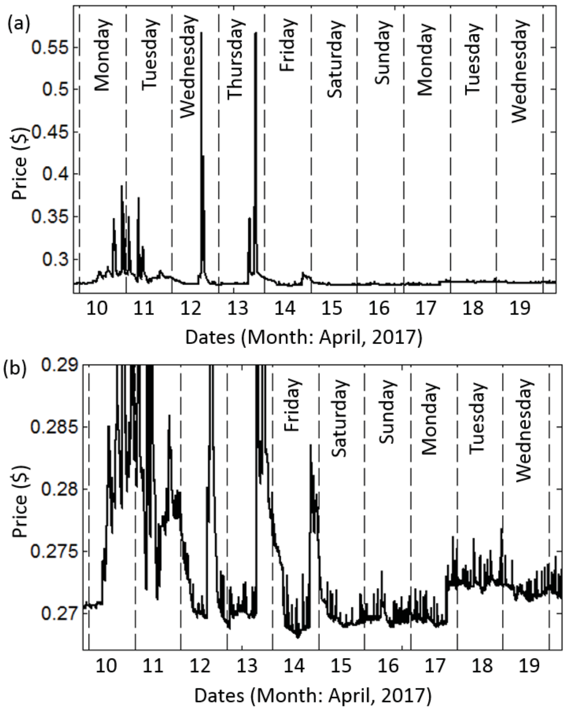

Fig. 5. Daily and weekly patterns of the Amazon EC2 $c4.4xlarge$ SI price (US-west location). Dotted vertical lines represent 12'O clock at night (Day-transition). Subplot (b) is a vertically magnified version of subplot (a).

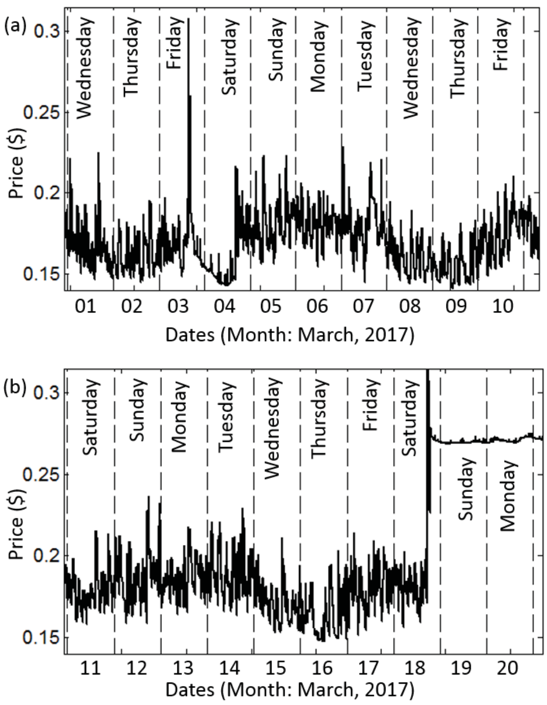

Fig. 6. Limitations of the Daily and Weekly Patterns- (a) Change in daily and weekly patterns and (b) unexpected price change for the Amazon EC2 $c4.4xlarge$ SI.

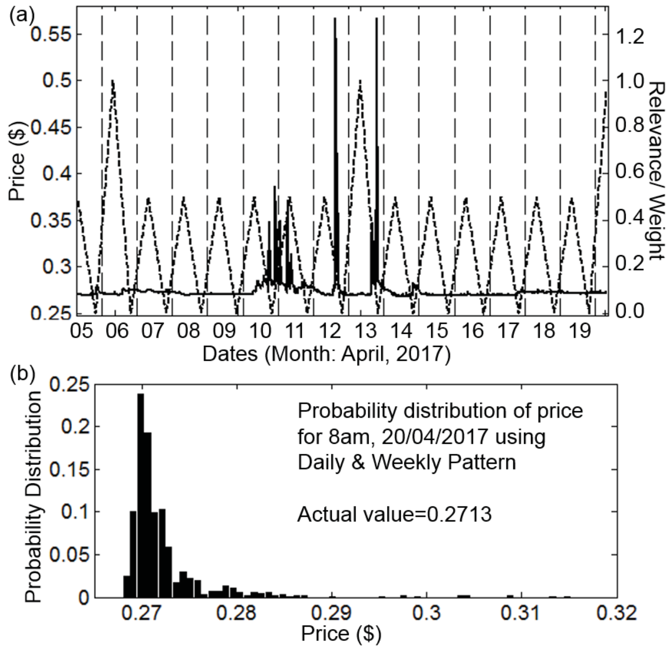

Fig. 7. Calculation of the probability density of Amazon EC2 $c4.4xlarge$ SI price. (a) The price of last 15 days presented by the solid line, thinner dotted lines are presenting day transition and thicker dotted lines are presenting corresponding weight calculated using Eqn. (5). (b) Bar chart of the probability density.

he might be able to finish his job within a small time and he might not be charged for that partial hour. However, these are risky decisions and the user may not be always successful by following such approach. Finally, the user can get an idea for the scheduling of his job. Such as, atomistic simulations and neural network training can be performed during off-peak hours. Some other applications like web-hosting and cloud brokers need to support a large number of users during office hours and they cannot be postponed due to the price. Therefore, they have to bid higher during the peak hours or they have to go for the on-demand instance.

D. Limitations of the Daily and Weekly Patterns and Holiday Considerations

Although the daily and weekly patterns can predict the true value of the SI, these patterns change from time to time. For example- according to the data of March 2017, shown as Fig. 6, minimum prices are found on Thursday and Saturday. We also checked that these Thursdays were not public holidays. Moreover, the daily pattern may not be repeated due to the worldwide expansion of the network. Asian users can submit their job in American EC2 servers and American users can submit their job in Asian EC2 servers. Moreover, many users work at night and most big IT companies have worldwide branches and employees staying in one location (say Europe) can work for other branches (Asia/USA/ South America etc.). Therefore, the daily pattern may not be continuously followed for the EC2 spot price prediction.

Besides the change in daily and weekly patterns, providers can change their strategy or the optimization equation at any time. Also, their optimization equation can be changed due to the number of users in other instances; such as reserved

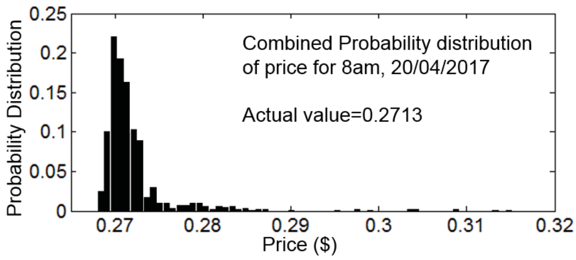

Fig. 8. Bar chart of the combined probability density.

instances, on-demand instances etc. Also, during maintenance, they have fewer servers available and that can shift the average price to a higher value. The opposite can happen when some new instances are installed. Before 18th March 2017, the price of Amazon EC2 $c4.4xlarge$ SI was roughly 17 cents per hour on average but the average price had increased to 26 cents per hour on that day. That increased average price continued until the end of our data (21st March 2017).

Datasets of several years are considered for the prediction of trends, such as the wind power or the electricity consumption. However, Amazon EC2 a user gets pricing information of the last 90 days. It is still enough for the price prediction with a slightly higher uncertainty. In fact, any major change in the available SI instance number shifts the price. Therefore, samples of the last 15 days are considered by providing those samples an empirical weight (relevance) and drawing the bar chart of the probability distribution. The formula for the empirical weight is provided as Eqn. (5).

$$Relevance_{Sample} = \frac{1 + Same_{Day}}{2} \times (1 - \frac{|\Delta T|}{12}) \quad (5)$$

Here, $Relevance_{Sample}$ is the relevance of the sample. $Relevance_{Sample}$ is used as the weight of prediction. $Same_{Day}$ is a function which returns 1 when the sample is taken on the same day of the week. ΔT is the time difference in hours with a range of -12 to 12.

While predicting for midnight and the day transition is considered as 12'O clock night, the day transition can provide irrelevant weights. Thus $Same_{Day}$ is 1 when the time is within the last 12 hours or time difference with the same time of previous weeks is less than 12 hours. Fig. 7(a) presents the SI price of the last 15 days, from 5 April 2017 to 19 April 2017, used for the prediction of the price at 8 am, 20 April 2017. The corresponding prediction weight is also depicted with the price curve. Fig. 8 presents the probability distribution of the price for 8 am 20 April 2017 according to the daily and weekly patterns.

E. Combined Probability Density for Bidding

Both of the curve fitting and the daily and weekly based predictions have advantages and limitations. Thus decisions are taken based on both distributions for the reliable completion of jobs. Two distributions are added in order to calculate the overall probability distribution. Fig. 7 (b) presents the bar chart of the combined probability density. When the task is not too urgent, the bid price is at the minima of the curve fitting or the daily and weekly patterns, whichever is lower. On the other hand, when the task is urgent enough, the bid price is at the maxima of the curve fitting or the daily and weekly patterns, whichever is higher.

IV. Popular Bidding Strategies

A lot of strategies have been developed by researchers over time. The consideration of probability density function (PDF) can potentially help to improve or to implement most of these strategies. The PDF represents the exact uncertainty of the price. Traditional point prediction cannot provide both of the social welfare and the reliable completion of the job. User's preference-based uncertainty upper/lower bounds can easily be deducted from the PDF.

1) Fixed Bidding Approaches: Many users do not have a sophisticated optimization algorithm for placing an exact bid. They usually bid with a constant value based on their philosophy. Some user bids at a very high price. They want to get possible discounts without any interruption. Some user bids at a very low price. Their tasks are not urgent and they want to get a highest possible discount including free partial hours. Some users bids at on-demand or near on-demand price to get possible discounts. Also, they may switch to on-demand when the price is higher than the on-demand [21].

2) Linear Optimization based Approaches: Tang et al proposed the AMAZING system for the cloud bidding that bids based on the Constrained Markov Decision Process (CMDP) and bargaining based on the condition of the user [22], [23]. However, the user's cannot rely on the current condition. They need to apply prediction algorithms before applying the linear optimization; solving linear equations with a large number of variables, changing over time is computationally expensive.

3) Game Theory based Approaches: Many researchers have treated spot bidding as a game between users. Some works have considered the Nash equilibrium of such a pricing game in cloud scenarios. However, the process becomes computationally complex with the considerations of different users entering & leaving at different times and associated uncertainties [24]–[27].

4) Probabiliy based Approaches: Time series analysis for the prediction of the number of incoming jobs and the spot price is vital for determining the condition of the user and the condition of the market. However, a few algorithms are performing that for the determination of the bid. Wang et al have determined the bid price using the Neural Network (NN) based prediction model [28]. Researchers are estimating the number of the incoming user through ARIMA model [29] and the Neural Network [30]. However, they are still relying on the point prediction.

5) Other Algorithms: As the bidding strategy of Amazon is close to the conventional uniform price auction [31] and the time series based forecast is helpful in finding truthful bidding. Many economic theories are being investigated to develop a truthful and reliable bidding system [32]–[34]. Many researchers are also trying to meet the urgent demand of completing tasks through overbooking [35], [36].

V. Conclusion

The probability density of price can potentially help bidders to bargain with the spot price. The user can easily understand uncertainty and take risks by picking one uncertainty bound based on the urgency of his task. The proposed correlation based probability density function is transparent as the user can check similar occurrences. In future, we will try to construct neural network-based probability density function.

References

[1] H. D. Kabir, A. Khosravi, M. A. Hosen, and S. Nahavandi, "Neural network-based uncertainty quantification: A survey of methodologies and applications," *IEEE Access*, 2018.

[2] C. Bennett, M. Moghimi, M. Hossain, J. Lu, and R. A. Stewart, "Applicability of load forecasting techniques for customer energy storage control systems," in *Power and Energy Engineering Conference (APPEEC), 2015 IEEE PES Asia-Pacific*. IEEE, 2015, pp. 1–5.

[3] H. M. D. Kabir, "A study on the grain dependent current variation of polycrystalline organic transistors," Master's thesis, Hong Kong University of Science and Technology, 2016.

[4] I. Amazon Web Services, "How AWS Pricing Works." [Online]. Available: https://aws.amazon.com/whitepapers/how-aws-pricing-works/

[5] M. D. Blog, "Amazon ec2 spot request volatility hits $1000/hour," Jun. 2017. [Online]. Available: https://moz.com/devblog/amazon-ec2-spot-request-volatility-hits-1000hour/

[6] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" in *Advances in Neural Information Processing Systems*, 2017, pp. 5580–5590.

[7] M. Rana, I. Koprinska, and V. G. Agelidis, "Univariate and multivariate methods for very short-term solar photovoltaic power forecasting," *Energy Conversion and Management*, vol. 121, pp. 380–390, 2016.

[8] M. Rana and I. Koprinska, "Neural network ensemble based approach for 2d-interval prediction of solar photovoltaic power," *Energies*, vol. 9, no. 10, p. 829, 2016.

[9] H. D. Kabir, A. Khosravi, M. A. Hosen, and S. Nahavandi, "Partial adversarial training for prediction interval," in *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2018, pp. 1–6.

[10] T. H. Nguyen, A. Yadav, B. An, M. Tambe, and C. Boutilier, "Regret-based optimization and preference elicitation for stackelberg security games with uncertainty." in *AAAI*, 2014, pp. 756–762.

[11] T. H. Nguyen, F. M. Delle Fave, D. Kar, A. S. Lakshminarayanan, A. Yadav, M. Tambe, N. Agmon, A. J. Plumptre, M. Driciru, F. Wanyama *et al.*, "Making the most of our regrets: Regret-based solutions to handle payoff uncertainty and elicitation in green security games," in *International Conference on Decision and Game Theory for Security*. Springer, 2015, pp. 170–191.

[12] A. Khosravi, S. Nahavandi, and D. Creighton, "Quantifying uncertainties of neural network-based electricity price forecasts," *Applied energy*, vol. 112, pp. 120–129, 2013.

[13] ——, "A neural network-garch-based method for construction of prediction intervals," *Electric Power Systems Research*, vol. 96, pp. 185–193, 2013.

[14] "Amazon ec2 spot instances," Jun. 2017. [Online]. Available: https://aws.amazon.com/ec2/spot/

[15] S. Yi, A. Andrzejak, and D. Kondo, "Monetary Cost-Aware Checkpointing and Migration on Amazon Cloud Spot Instances," *IEEE Transactions on Services Computing*, vol. 5, no. 4, pp. 512–524, 2012.

[16] H. M. D. Kabir, A. Hosen, S. Nahavandi, and A. Khosravi, "Prediction Interval with Examples of Similar Pattern and Prediction Strength," in *The 30th Annual IEEE Canadian Conference On Electrical and Computer Engineering, CCECE 2017*. IEEE, 2017, pp. 1–4.

[17] H. M. Dipu Kabir, S. B. Alam, M. I. Azam, M. A. Hussain, A. R. Sazzad, M. N. Sakib, and M. A. Matin, "Non-linear Down-sampling and Signal Reconstruction, without Folding," in *2010 Fourth UKSim Eur. Symp. Comput. Model. Simul.* IEEE, Nov 2010, pp. 142–146.

[18] M. Malvoni, M. Fiore, G. Maggiotto, L. Mancarella, R. Quarta, V. Radice, P. Congedo, and M. De Giorgi, "Improvements in the predictions for the photovoltaic system performance of the mediterranean regions," *Energy Conversion and Management*, vol. 128, pp. 191–202, 2016.

[19] M. Malvoni, M. G. De Giorgi, and P. M. Congedo, "Photovoltaic forecast based on hybrid pca–lssvm using dimensionality reducted data," *Neurocomputing*, vol. 211, pp. 72–83, 2016.

[20] J.-Y. Delort, B. Arunasalam, M. Milosavljevic, and H. Leung, "The impact of manipulation in internet stock message boards," *International Journal of Banking and Finance, Forthcoming*, pp. 1–19, 2009.

[21] A. W. Services, "Deciding on your spot bidding strategy," Jun. 2018. [Online]. Available: https://www.youtube.com/watch?v=WD9N73F3Fao

[22] S. Tang, J. Yuan, and X.-Y. Li, "Towards optimal bidding strategy for amazon ec2 cloud spot instance," in *Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on*. IEEE, 2012, pp. 91–98.

[23] Shaojie Tang, Jing Yuan, Cheng Wang, and Xiang-Yang Li, "A Framework for Amazon EC2 Bidding Strategy under SLA Constraints," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 1, pp. 2–11, Jan. 2014.

[24] D. Ardagna, B. Panicucci, and M. Passacantando, "A game theoretic formulation of the service provisioning problem in cloud systems," in *Proceedings of the 20th international conference on World wide web*. ACM, 2011, pp. 177–186.

[25] S. Sen, Y. Jin, R. Guérin, and K. Hosanagar, "Modeling the dynamics of network technology adoption and the role of converters," *IEEE/ACM Transactions on Networking (TON)*, vol. 18, no. 6, pp. 1793–1805, 2010.

[26] F. Teng and F. Magoulès, "A new game theoretical resource allocation algorithm for cloud computing," in *International Conference on Grid and Pervasive Computing*. Springer, 2010, pp. 321–330.

[27] H. Roh, C. Jung, W. Lee, and D.-Z. Du, "Resource pricing game in geo-distributed clouds," in *2013 Proceedings IEEE INFOCOM*. IEEE, Apr 2013, pp. 1519–1527.

[28] X. Wang, X. Wang, H. Che, K. Li, M. Huang, and C. Gao, "An Intelligent Economic Approach for Dynamic Resource Allocation in Cloud Services," *IEEE Transactions on Cloud Computing*, vol. 3, no. 3, pp. 275–289, Jul 2015.

[29] R. N. Calheiros, E. Masoumi, R. Ranjan, and R. Buyya, "Workload prediction using arima model and its impact on cloud applications qos," *IEEE Transactions on Cloud Computing*, vol. 3, no. 4, pp. 449–458, 2015.

[30] R. Mijumbi, S. Hasija, S. Davy, A. Davy, B. Jennings, and R. Boutaba, "Topology-aware prediction of virtual network function resource requirements," *IEEE Transactions on Network and Service Management*, vol. 14, no. 1, pp. 106–120, 2017.

[31] I. Menache, A. Ozdaglar, and N. Shimkin, "Socially optimal pricing of cloud computing resources," *Proceedings of the 5th International ICST Conference on Performance Evaluation Methodologies and Tools*, pp. 322–331, 2011.

[32] R. Pal and P. Hui, "Economic models for cloud service markets: Pricing and Capacity planning," *Theoretical Computer Science*, vol. 496, pp. 113–124, Jul 2013.

[33] B. Sharma, R. K. Thulasiram, P. Thulasiraman, and R. Buyya, "Clabacus: A Risk-Adjusted Cloud Resources Pricing Model Using Financial Option Theory," *IEEE Transactions on Cloud Computing*, vol. 3, no. 3, pp. 332–344, Jul 2015.

[34] S. Son and K. M. Sim, "A price- and-time-slot-negotiation mechanism for Cloud service reservations." *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society*, vol. 42, no. 3, pp. 713–28, Jun 2012.

[35] L. Tomás and J. Tordsson, "Improving cloud infrastructure utilization through overbooking," in *Proceedings of the 2013 ACM Cloud and Autonomic Computing Conference on - CAC '13*. New York, New York, USA: ACM Press, Aug. 2013, pp. 1–10.

[36] L. Tomas and J. Tordsson, "An Autonomic Approach to Risk-Aware Data Center Overbooking," *IEEE Transactions on Cloud Computing*, vol. 2, no. 3, pp. 292–305, Jul. 2014.