Spot Price Characterization: What Are the Right Features to Model?

[Extended Abstract]

Qianlin Liang, Cheng Wang, Bhuvan Urgaonkar
Department of Computer Science and Engineering
The Pennsylvania State University

1. INTRODUCTION

Amazon EC2 has been offering spot instances [2] since 2009 and a large segment of its “tenant” workloads has come to embrace these [6]. The appeal of spot instances lies in their low prices - up to one-tenth of the prices of on-demand instances of equivalent (advertized) capacities. Unlike an on-demand instance, whose price changes very slowly (over months or years), a spot instance has a highly dynamic price that may change as frequently as once every few minutes. To procure a spot instance, a tenant needs to place a bid in the concerned marketplace. Following this, whenever the dynamic price for the requested type of spot instance (simply the spot price henceforth) falls below the bid, an instance is allocated to the tenant. However, when the bid falls below the spot price, a warning is issued to the tenant following which the instance is reclaimed/revoked by EC2. The tenant may choose to use this warning period (2 minutes as of this writing) to save all or some of that instance’s state. From a tenant’s point of view, a spot instance is a virtual machine (VM) that is cheaper than its on-demand counterpart but appears to have poorer availability.

To benefit from the low prices of spot instances, a tenant must effectively deal with two sources of complexity: (i) it must predict relevant aspects of spot price, and (ii) it must combine these predictions with its application-specific trade-offs 1 to devise online instance procurement algorithms. Both these issues have received a lot of attention recently. In this paper, our focus is on (i). Our findings also have implications for (ii) and will inform our future work as we discuss in Section 4.

A key challenge in using spot instances, as already identified by many recent papers, is the poor predictability of spot prices. Researchers have proposed prediction techniques for spot prices with varying degrees of modeling complexity. Among the simpler ones are techniques based on (dynamically updated) empirical probability distributions [3, 4, 7]. More sophisticated techniques based on auto-regressive time-series [1, 8, 10], Markovian models [9, 5], etc., have also been proposed. A key requirement for cost-effective procurement of a spot instance is the ability to predict its service contiguity, i.e., the contiguous duration for which it is likely to be available. Regardless of the modeling technique used, all existing approaches (to our knowledge) are based on creating a statistical model for the raw spot price itself (with underlying assumptions about its stationarity).

We believe this to be an untenable exercise because our extensive analysis of spot price traces indicates that these are best considered non-stationary. For example, Figure 1 shows that the first two moments of spot price for many different marketplaces vary significantly over a 90-day period. Since existing techniques are based on modeling raw spot prices directly (despite their high dynamism), they are either ineffective or computationally non-scalable when used in online control wherein the prediction of some other properties of spot prices (rather than mere raw values) are desired.

What Should be Modeled? Rather than modeling the exact spot price values, we argue that the focus should be on the features we identify below.

- **Feature I**: Since spot prices tend to be significantly smaller than on-demand prices (of equally-sized VMs) during periods when a bid is successful, and since EC2 charges a tenant based on the spot price during such periods (not based on the bid), attempting to predict spot prices very accurately is of little value. A visual inspection of the 90-day long spot price time-series in Figure 1 and how these prices compare with a bid that equals to the on-demand price clarifies this. In particular, it suffices that we predict the average spot price during such periods with reasonable accuracy (since that is what will determine our costs).

- **Feature II**: Tenants want the spot instance to be available for long enough time to maintain the service contiguity, i.e., they are interested in how long a successful bid is likely to last. An effective predictor should not overestimate this quantity - doing so may render a control scheme overly optimistic in its estimation of the cost vs. performance trade-off.

- **Feature III**: The raw cross-correlations among markets, used by prior works [7, 4] to capture simultaneous bid failures, may not be informative for decision-making. As we will show in Section 2.2 and 3.2, simultaneous bid failures crucially depend on the chosen markets and bids.

The rest of this paper is organized as follows. In Section 2, we present our prediction approach. In Section 3, we evaluate proposed approach with extensive real-world traces. We discuss future directions and conclude in Section 4.

2. OUR APPROACH

We model as a random variable $L(b)$ the length of a contiguous period during which the spot price is less than or equal to a bid $b$. In other words, $L(b)$ captures the lifetime of a spot instance using bid $b$. We denote as $\bar{p}(b) = E[p_t | L(b)]$ a random variable for the average spot price $p_t$ during a period when the bid $b$ is successful, which serves to estimate the cost of a spot instance procured by placing a bid $b$. 

---

1These trade-offs would be between costs, on the one hand, and over-heads of possible revocations (either in the form of fault-tolerance mechanisms or loss of performance/correctness), on the other.
2.1 Our prediction technique

Our technique employs empirical probability distributions computed over recent sliding time windows (H most recent time slots, e.g., days) for making predictions of \( L(b) \) and \( \tilde{p}(b) \). \( H \) must be chosen such that temporal locality\(^2\) holds for these quantities. Figure 2 clarifies these quantities. Large \( L(b) \) and small \( \tilde{p}(b) \) imply long service continuity and low costs, thereby encouraging the use of spot instances using bid \( b \). We use a small percentile (e.g., 5th) of the recently constructed distribution of \( L(b) \) - denoted as \( \hat{L}(b) \) - as our prediction in the ongoing horizon. The reasoning behind this choice is that if the statistical properties of \( L(b) \) do not change much over \( H \), we expect that with a very high probability, bid \( b \) would be successful for at least \( L(b) \) time units. We use average of \( \hat{p}(b) \) during the relevant \( H \) as its predictor (denoted as \( \tilde{p}(b) \)).

**Assessment Metrics:** We say that an over-estimation of \( L(b) \) has occurred when \( L(b) > \hat{L}(b) \). This represents a scenario wherein the tenant was likely overly ambitious in using spot instances. We further define \( L(b) \) over-estimation rate as the fraction of \( L(b) \) predictions that result in over-estimation, denoted as \( f(b) \). The assessment metric for \( \tilde{p}(b) \) should capture the extent of its deviation from actual values. Therefore, we compute \( \xi(b) = (\hat{p}(b) - \tilde{p}(b))/\tilde{p}(b) \) and define as relative deviation of \( \tilde{p}(b) \) the mean value of \( \xi(b) \) for all occurrences of \( \tilde{p}(b) \) in the relevant \( H \). Lower values are better for both.

In Section 3, we focus on the evaluation of the right choice of history window size (for model training), which turns out to be dependent on both the market and the bid. None of the prior works (to our knowledge) have analyzed these idiosyncrasies.

2.2 Simultaneous revocations

When placing a bid, tenants may want to avoid picking spot markets with the likelihood of simultaneous revocations (the spot instances may be terminated simultaneously due to coincident bid failures). Prior works, e.g., [7, 4], suggest bidding across markets where there are no significant statistical correlations among the “raw” history spot price traces.

However, such raw correlations might not be informative for tenants’ decision making. Let us consider an illustrative example in Figure 3. We generate synthetic spot prices for two markets wherein the cross-correlation between the two markets’ spot prices is low and the tenant might be tempted to use both markets if its decision is only based on the raw correlation. However, it is obvious that the bid failures from the two markets are highly correlated under bid 1 but not correlated under bid 2. Therefore, it may be imprudent for the tenants to make decisions solely based on the raw correlations without considering the actual bids. More specifically, what the tenant really needs is measurements of simultaneous revocations, *conditioned on bids*. Furthermore, the statistical correlation of bid failures across markets may not be very informative for decision-making regarding bid placement. Instead, a tenant might find it more beneficial to learn the absolute time durations of simultaneous revocations, i.e., the total amount of time that a bid fails in two candidate markets within the history window.

A more informative metric that we propose is based on characterizing simultaneous revocations conditioned on pre-specified bids. Under a given bid, we denote as \( A \) and \( B \) the sets of time periods when the bid fails in two spot markets under comparison, respectively. Denote as \( T(A) \) and \( T(B) \) the corresponding lengths/sizes of \( A \) and \( B \). \( T(A \cap B) \) and \( T(A \cup B) \) represent the time durations of coincident bid failures and total bid failures, respectively. Denote as \( \xi(b) = (\hat{p}(b) - \tilde{p}(b))/\tilde{p}(b) \) the mean value of \( \xi(b) \) for all occurrences of \( \tilde{p}(b) \) in the relevant \( H \). Lower values are better for both.

### Figure 1: Sample spot price timeseries collected during the 90-day period (2015-07-08 to 2015-10-06) and are chosen due to their very different properties. The green “*” and error bars represent the moving average and standard deviations over each 2-day interval.

2.### Evaluation and Lessons

**Experiment Setup:** We evaluate our technique with dozens of spot price traces of which we show four 90-day spot price traces for VM types of m3.large and c3.large in availability zones us-east-1c and us-east-1d in Figures 1. We denote as m3.large-c the spot market for m3.large in us-east-1c (same notation rule for other markets). For each trace, we pick bid \( b \) from \{0.5d, d, 2d, 5d, 10d\}, where \( d \) is the corre-
3.1 Evaluation of Our Prediction Technique

We vary VM types, markets, bids, history window size $H$ and show the assessment metrics $f(L(b)$ over-estimation rate) and $\xi(p(b)$ relative deviation) in Table 1. Validation of prediction. Under most of (market, bid) pairs, the optimal (lowest) $f$ and $\xi$ are below 10%, which demonstrates the good predictive power of our technique. For $c3.large-c$, $f$ and $\xi$ are much higher than those from other markets; therefore, it might be better not to use this market temporarily until its predictability gets improved.

Table 1: The assessment metrics $f(b)$ and $\xi(b)$ under different bid values and history window sizes ($H$ in days). The shaded cells represent the optimal window size that minimizes $f(b)$ and $\xi(b)$. “-c” and “-d” represent the markets.

<table>
<thead>
<tr>
<th>Bid</th>
<th>$f(b)$</th>
<th>$\xi(b)$</th>
<th>$H$ (days)</th>
<th>$f(b)$</th>
<th>$\xi(b)$</th>
<th>$H$ (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c3.large-c</td>
<td>0.0241</td>
<td>0.0989</td>
<td>0.0588</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>c3.large-d</td>
<td>0.0241</td>
<td>0.0989</td>
<td>0.0588</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>m3.xlarge</td>
<td>0.0241</td>
<td>0.0989</td>
<td>0.0588</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

What is the right choice for history window size? We observe that (i) $H = 7$ days seem to be the best choice (minimized $f$ and $\xi$) cases we examine, (ii) the optimal $H$ varies across markets and bids, implying the necessity for considering different markets separately when determining history window size, instead of blindly choosing a window size that has to work for all markets, and (iii) changing bid values may not affect $f$ and $\xi$ much (e.g., $m3.large-c$), possibly due to the fact that the $L(b)$ and $p(b)$ do not vary much when spot price exceeds bid price.

4. DISCUSSIONS AND FUTURE DIRECTION

What is the right percentile for prediction? Recall that we use a percentile of $L(b)$ in the history window as our prediction. As this percentile decreases, we may be more

5. REFERENCES