2018 18th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing

Location, Location, Location: Exploring Amazon EC2 Spot Instance Pricing Across Geographical Regions

Nnamdi Ekwe-Ekwe and Adam Barker
School of Computer Science, University of St Andrews, UK
Email: {nnee, adam.barker}@st-andrews.ac.uk

***Abstract*— Cloud computing is a ubiquitous part of the computing landscape. For many companies today, moving their computing infrastructure to the cloud reduces time to deployment and saves money. Spot Instances, a subset of Amazon's cloud computing infrastructure (EC2), expands upon this. They allow a user to bid on spare compute capacity in EC2 at heavily discounted prices. If other bids for the spare capacity exceeds the user's own, the user's instance is terminated.**

In this paper, we conduct one of the first detailed analyses of how location affects the overall cost of deployment of a Spot Instance. We analyse pricing data across all available AWS regions for 60 days for a variety of Spot Instances. We relate the pricing data we find to the overall AWS region and examine any patterns we see across the week. We find that location plays a critical role in Spot Instance pricing and that pricing differs, sometimes markedly, from region to region.

We conclude by showing that it is indeed possible to run workloads on Spot Instances with low risk of termination and a low overall cost.

I. INTRODUCTION

Amazon EC2 allows developers to provision on-demand compute resources, configure them to their needs, and scale resources up or down depending on application requirements.

Developers have three primary ways to pay for what they provision. They can pay "On-Demand" which allows the developer to pay for what they use by the hour with no "long-term commitments" [2]. They can also pay via "Reserved Instance[s]" allowing them to commit to using a set of instances for either a "1 or 3-year term" [2]. Finally, developers can pay for resources using a "Spot Instance" model [2] allowing them to bid on spare compute capacity with savings of up to "90% off the On-Demand price" [2].

Spot Instances are very useful for specific use cases - for example, users that suddenly have "urgent computing needs for large amounts of additional capacity" [2]. The user has access to both a sizeable quantity of compute resources *and* at low prices. The only caveat with this approach, however, is that the application that the user is running must be fault-tolerant due to the potential of an instance being terminated at any time.

The majority of research already done into Spot Instances [4, 5, 7] focuses on modelling Amazon's pricing strategy and how to determine the best bids to make. There is a distinct lack of research that considers *where* (in terms of region and availability zone) a user can *reliably* (low risk of termination) deploy a spot instance in order to *minimise cost*.

We address this by:

- Conducting one of the first detailed analyses of how location affects the overall cost of deployment of a spot instance.
- Analysing pricing data across all available Amazon Web Services (AWS) regions for a variety of spot instance types. We relate the data we find to the overall AWS region as well as to the Availability Zone (AZ) within that region.
- Checking whether (for any pricing differences we find) those differences are *reliable* (low standard deviation) and as a result, whether we can be confident in the ensuing bids we make.

Please note that in this paper, the words "instance" and "instance type" are used interchangeably.

II. EC2 SPOT PRICING DATA

Amazon provides spot price data via its API endpoint [3] for a period of up to 90 days from when a request is made.

```
{
 "Timestamp": "2017-06-25T00:20:56.000Z",
 "ProductDescription": "Linux/UNIX",
 "InstanceType": "d2.2xlarge",
 "SpotPrice": "0.177800",
 "AvailabilityZone": "eu-west-2a"
}
```

In the above example price point (JSON format), the instance type (*d2.2xlarge*) was $0.177800 (per hour) at 00:20:56 on the 25th of June 2017. The instance was running Linux/UNIX and was based in the *eu-west-2a* Availability Zone. For our analysis, we used a data timeline of 60 days. At the time of carrying out this research, there were issues with retrieving the full 90 days worth of data from some zones. Additionally, not all Amazon regions provided a Spot Instance capability, so our dataset comprised of data from only 4 AWS Regions - the EU, US, Asia Pacific and Canada.

68 different instance types can be launched on EC2. We restricted the dataset based on the most popular instance types available: small, medium and large instance types [6]. Spot instances are not available for small instance types, and so our final dataset consisted only of medium and large instances. We then randomly picked a variety of instances (of varying compute power) from each of the AWS EC2 instance type categories available [1].

0-7695-6410-0/18/$31.00 ©2018 IEEE
DOI 10.1109/CCGRID.2018.00059

Our final selection of data consisted of *m3.medium*, *m4.large*, *c4.large*, *c3.large*, *r3.large*, *r4.large* and *i3.large* instance types. The total pricing data over the 60 days for all the above instances was 4,112,000 data points.

III. Analyses

We focused on exploring the pricing data in relation to i) the *instance type* and ii) the *region/availability zone*.

We ran three main analyses.

- First, we ran an average price analysis of all instance types within an AWS region over the full 60 days.
- Second, we ran an average price analysis for each instance in our list of instance types for every AWS region.
- Finally, we ran a histogram analysis plotting the frequency of all price points in a particular AZ for each of the instances deployed in that zone.

In this paper, we have included the first two analyses broken down by day of the week and in AWS regions only. We show the mean price across both these metrics. We also calculated the standard deviation and related the average price analysis to these to examine the price volatility. In this context, volatility means the propensity for the price to change across a time metric. The lower the standard deviation, the lower the volatility and the more confident we could be in the mean price metrics reported. Conversely, the higher the standard deviation, the less confident we could be in the price's reliability.

For the weekly analysis, we divided the week into 7 portions with 0 standing for Monday and 6 for Sunday.

IV. Results

A. Average Price Analysis of All Instance Types within an AWS region

1) EU: In general, across the EU, prices were quite volatile (Table I). *m3.medium* instances' pricing were the lowest and most consistent throughout the week (Figure 1(a)). We saw slightly more volatility in *r4.large* and *m4.large* instance types. The end of the week seems to be the best time to use the two aforementioned instance types - offering the user the cheapest and most reliable pricing. *i3.large* instances were significantly more expensive than all the other instances throughout the week. This, coupled with a very high standard deviation (as also seen in *c4.large* and *m3.large* instance types) shows that the EU is not the best place to deploy such instance types.

2) US: In the United States, we see significant and pronounced price volatility across nearly all of the instance types. We see both very high levels of standard deviations (Table II) as well as high average price metrics throughout the week. Only two instances (Figure 1(b)) display significantly low levels of standard deviations - *m3.medium* and *r3.large* ones. *m4.large* instances display the most volatile pricing in the region, closely followed by *i3.large* and *m3.large* instances. The type of instance does not seem to have as large an impact on price as the location does. *m3* instance types are the least powerful in our dataset, yet we see both very high and volatile pricing across the board. Conversely, we see reliable and low price metrics reported for *r3.large* instances, some of the most powerful.

Instance Type	Mean±Standard Deviation($p/h)
c3.large	0.077±0.077
c4.large	0.068±0.104
i3.large	0.181±0.496
m3.large	0.078±0.088
m3.medium	0.063±0.051
m4.large	0.052±0.054
r3.large	0.081±0.082
r4.large	0.080±0.055

TABLE I
MEAN AND STANDARD DEVIATION RESULTS OF ALL INSTANCE TYPES IN THE EU OVER 60 DAYS (TO 3DP)

Instance Type	Mean±Standard Deviation($p/h)
c3.large	0.074±0.104
c4.large	0.079±0.107
i3.large	0.081±0.115
m3.large	0.088±0.102
m3.medium	0.063±0.055
m4.large	0.084±0.118
r3.large	0.081±0.066
r4.large	0.090±0.127

TABLE II
MEAN AND STANDARD DEVIATION RESULTS OF ALL INSTANCE TYPES IN THE US OVER 60 DAYS (TO 3DP)

Instance Type	Mean±Standard Deviation($p/h)
c3.large	0.076±0.050
c4.large	0.096±0.139
i3.large	0.087±0.167
m3.large	0.085±0.056
m3.medium	0.063±0.047
m4.large	0.089±0.067
r3.large	0.088±0.058
r4.large	0.076±0.049

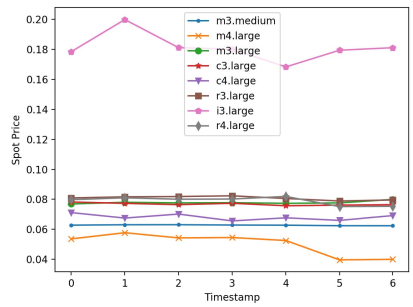
TABLE III
MEAN AND STANDARD DEVIATION RESULTS OF ALL INSTANCE TYPES IN AP OVER 60 DAYS (TO 3DP)

Instance Type	Mean±Standard Deviation($p/h)
c4.large	0.043±0.040
i3.large	0.072±0.045
m4.large	0.027±0.032
r4.large	0.029±0.033

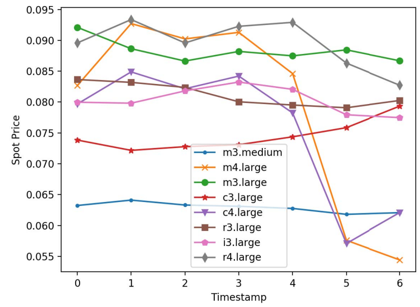
TABLE IV
MEAN AND STANDARD DEVIATION RESULTS OF ALL INSTANCE TYPES IN CA OVER 60 DAYS (TO 3DP)

3) Asia-Pacific: The Asia-Pacific (AP) region, in comparison to the EU and US regions, shows both low price points and standard deviations in many of the instances (Table III). We see very low and reliable pricing in the *m3.medium*, *r4.large* and *c3.large* instances (Figure 1(c)) - the latter two instances being some of the most powerful in our dataset. In AP in general, the more powerful the instance type, the more favourable the pricing. There are exceptions to this however, for example *i3.large* and *c4.large* instances.

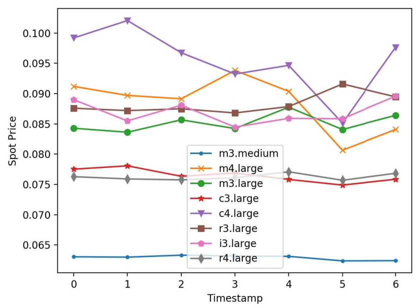
4) Canada: The Canada region hosts the least amount of instance types from our original selection - only the more



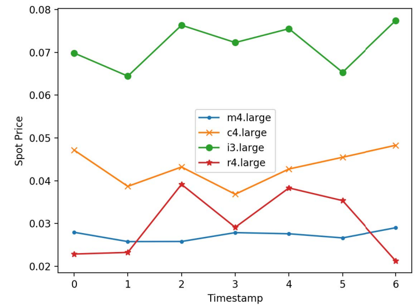
(a) Average Pricing By Day Of Week: EU



(b) Average Pricing By Day Of Week: US



(c) Average Pricing By Day Of Week: Asia-Pacific



(d) Average Pricing By Day Of Week: Canada

Fig. 1. Average Price Analysis of All Instances across all AWS regions: By Day of Week

powerful *c4.large*, *m4.large*, *i3.large* and *r4.large* instance types are present.

The Canada region exhibits some of the cheapest pricing across the week for the different instance types (Figure 1(d)). The *m4.large*, *r4.large* and *c4.large* instances had standard deviations of \$0.032, \$0.033 and \$0.040 respectively, the lowest across all four regions (Table IV). The *i3.large* instance type had mean prices of around \$0.07 per hour with the lowest standard deviation recorded across all four regions at \$0.045 (Table IV). By doing a cross-comparison with the same results in the AP, US and EU regions, we can see that it would be much cheaper to deploy an *i3.large* instance in the Canada region than in any of the others. Based on the overall results from Canada, the more powerful the instance type, the cheaper and more reliable it is to deploy there.

B. Average Price Analysis of Each Instance for every AWS region

This analysis plotted every single instance type's pricing for every single AWS region. This allowed us to perform cross-regional comparisons and offered us another view of the price data we had gathered. For some of the regions, some instance types were not deployed and as a result do not appear on some graphs.

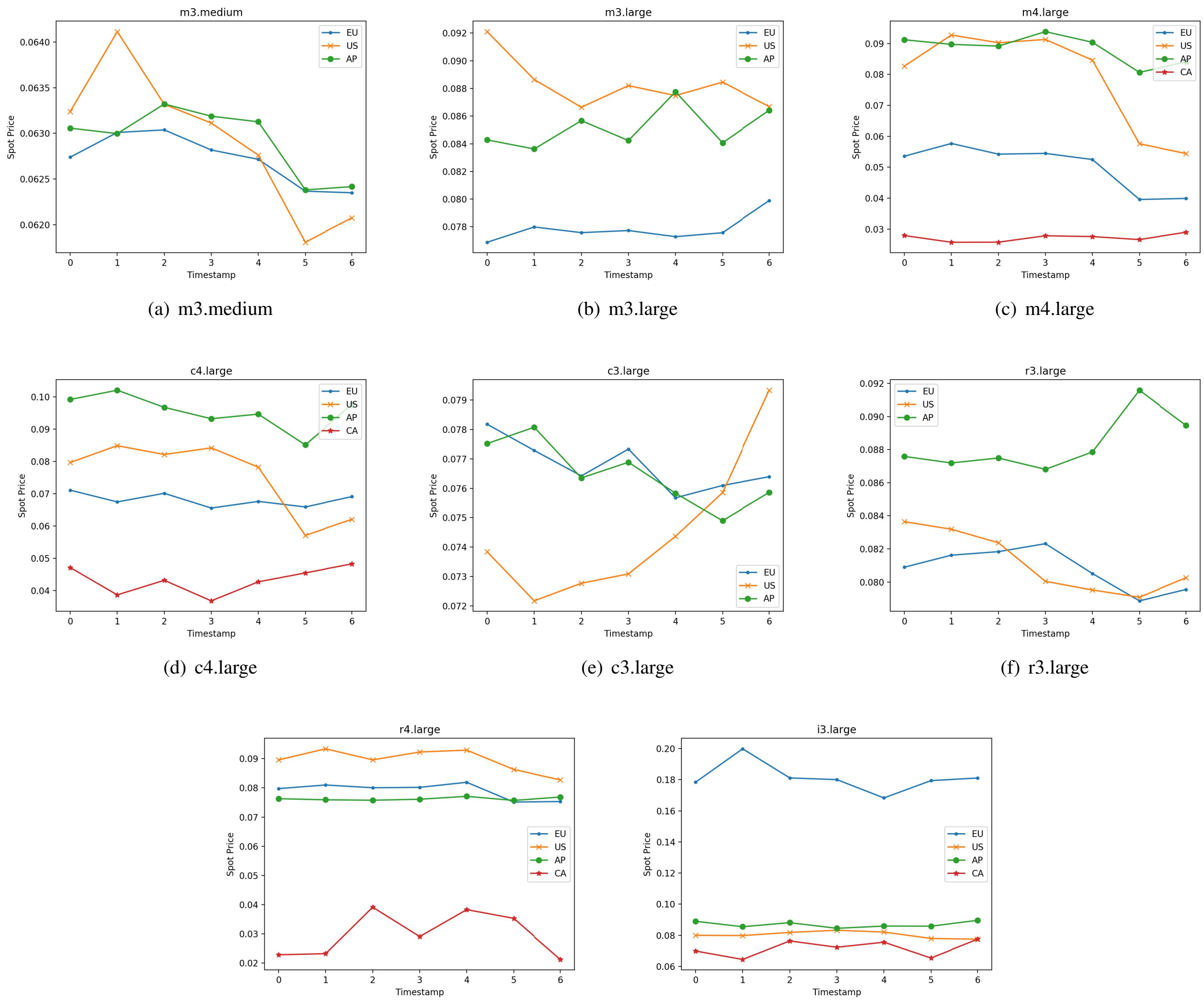
In any instance type that is deployed in Canada, we see both its lowest and most reliable price metrics. If we take a look at Figure 2, we see this very clearly. The less powerful the instance, the more reliable it is to deploy in the EU or US - in the "m" instance types we generally see cheaper and more reliable pricing in those two regions.

For example, if one was to deploy an *m4.large* instance and only had a choice between the AP and EU regions, one would obtain much more reliable pricing in the EU with a lower mean price and standard deviation than in AP. Some anomalies do occur, for example *m4.large* instance types being amongst the cheapest to run in the EU but the most expensive in the US. With all these different price metrics available, users can begin to make informed decisions about where to deploy an instance based on their use case.

V. CONCLUSIONS

In this paper, we have explored and gleaned very interesting insights from Spot Price data. We have seen the impact that a coarse-grained AWS *region* plays in affecting spot prices. We have analysed pricing volatility and whether we can be confident in the reliability of the average price point shown (from the standard deviation).

Based on this subset of analyses, we see a clear pattern. Overall, the more powerful the instance, the cheaper and more reliable the pricing one would get in Canada than in any of the other three regions. This is closely followed by the Asia Pacific region. On the other hand, the less powerful the instance, the cheaper and more reliable (in terms of pricing) it is to deploy in the EU and US regions.



(a) m3.medium (b) m3.large (c) m4.large

(d) c4.large (e) c3.large (f) r3.large

(g) r4.large (h) i3.large

Fig. 2. Average Price Analysis of Each Instance for every AWS region: By Day of Week

References

[1] Amazon. *Amazon EC2 Instance Types - Amazon Web Services (AWS)*. URL: https://aws.amazon.com/ec2/instance-types/.

[2] Amazon. *Amazon EC2 Pricing - AWS*. URL: https://aws.amazon.com/ec2/pricing/.

[3] Amazon. *describe-spot-price-history - AWS CLI 1.11.188 Command Reference*. URL: https://docs.aws.amazon.com/cli/latest/reference/ec2/describe-spot-pricehistory.html.

[4] Artur Andrzejak, Derrick Kondo, and Sangho Yi. "Decision Model for Cloud Computing under SLA Constraints". In: *2010 IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems* (2010). DOI: https://doi.org/10.1109/MASCOTS.2010.34.

[5] Orna Agmon Ben-Yehuda et al. "Deconstructing Amazon EC2 Spot Instance Pricing." In: *ACM Transactions on Economics and Computation* 1.3 (2013). DOI: http://dx.doi.org/10.1145/2509413.2509416.

[6] Channel Futures. *2nd Watch Identifies Most Popular AWS Services, Instances — Channel Futures*. URL: http://www.channelfutures.com/cloud-services/2nd-watch-identifies-most-popular-aws-services-instances.

[7] Liang Zheng et al. "How to Bid the Cloud". In: *SIGCOMM '15 Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication* 45.4 (2015), pp. 71–84. DOI: http://dx.doi.org/10.1145/2829988.2787473.