



Contents lists available at ScienceDirect

## Simulation Modelling Practice and Theory

journal homepage: [www.elsevier.com/locate/simpat](http://www.elsevier.com/locate/simpat)

# On optimization of simulation execution on Amazon EC2 spot market

Bogumił Kamiński\*, Przemysław Szufel

Warsaw School of Economics, Al. Niepodległości 162, 02-554 Warszawa, Poland

## ARTICLE INFO

### Article history:

Available online xxxx

### Keywords:

Cloud based simulation  
Simulation of the cloud  
Cost optimization  
Amazon EC2  
Spot prices

## ABSTRACT

Large scale simulations require considerable amounts of computing power and often cloud services are utilized to perform them. In such settings the execution costs can be significantly decreased through the use of the Amazon spot price market. Its downside is that Amazon can interrupt the user's computations when her bid price is too low. This poses a problem in finding an on-line bidding algorithm that balances the computation cost and the simulation experiment completion time.

We identify key drivers governing the spot prices on Amazon EC2 and using these insights propose an adaptive bidding strategy that simultaneously minimizes the computation cost and the delays due to computation termination. We show that bidding close to a spot price and dynamically switching between instances is a strategy that is efficient and simple to implement in practice.

In the paper we present a simulator of the EC2 spot pricing mechanism. The simulator can be easily used to develop and test other bidding strategies on Amazon spot price market.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The goal of this paper is to propose an algorithm for a cost and time optimization for running simulations on public computational clusters with a spot pricing mechanism. The algorithm is implemented in Python and is ready for application to real-life computationally intensive simulations executed on the Amazon Elastic Compute Cloud (Amazon EC2).

Amazon is the largest cloud computing provider and offers many server types in eight regions around the globe. Amazon offers its customer three pricing mechanisms: on-demand, reserved instances and spot pricing. In this paper we focus on the spot pricing mechanism. It offers the lowest prices on average at the expense of the risk of an abrupt termination of computations when the user's bid price becomes lower than the current spot price. For applications where the up-time is crucial this is a severe limitation. However, when running simulations one can decide to allow for breaks in computation if the cost reduction is significant enough. Amazon is fully aware of this situation and EC2 documentation [1] discusses four main state-of-the-art architectures that are designed for scientific computing on spot instances.

However, customers wanting to take full advantage of spot pricing face a complex ecosystem. Firstly, the spot prices are volatile and there is a possibility of a simulation termination. In such a situation, if the computation state is not saved (i.e. the simulation has not been check-pointed) the results are lost. Secondly, there are over 100 different server types to choose

\* Corresponding author. Tel.: +48 (22) 564 60 00.

E-mail address: [bkamins@sgh.waw.pl](mailto:bkamins@sgh.waw.pl) (B. Kamiński).

<http://dx.doi.org/10.1016/j.simpat.2015.05.008>

1569–190X/© 2015 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

from, but, on the other hand, the number of spot instances that a user can initiate in parallel is limited and controlled by Amazon. Thirdly, when considering available computation power one has to take into account the booting time of the virtual machine – see [2] for analysis of dependencies between instance booting time and instance type. Finally, the Amazon EC2 spot billing mechanism has several nuances that can be exploited to improve the cost-performance of computations. Due to this complexity the spot pricing analysis attracts a lot of attention in the literature.

Researchers consider the spot market for computing power from a provider's perspective as well as from a user perspective (a detailed literature review is provided in Section 2; here we give its summary). Analysis of the cloud provider's side focuses on designing a pricing mechanism that maximizes profits, see for example [3]. Other papers also include market design postulates. For instance Vanmachlen et al. [4] propose an extension of the existing spot market with the capability of bidding for future prices for computing power (futures market).

The problem of the computational grid pricing from the end-user's perspective focuses on cost-reduction. Several papers discuss how cloud utilization is becoming an important approach in cost-optimizing for large scale scientific computing – an overview of cloud scientific computing applications is given in [5]. Mattess et al. [6] point out that using cloud computing and spot instances is an efficient cost management technique for peak loads in local computational clusters. Efficient utilization of spot instances requires a user to find the optimal bidding strategy. Javadi et al. [7] analyze hourly and weekly patterns in spot price and time between price changes and they propose a mixture of Gaussian distributions to model the price patterns. In [8] a model of bidding strategies with service level agreement constraints is presented and the results show that low bids lead to long deadlines for job execution. The analysis carried in [9] shows that check-pointing times below one hour can significantly decrease costs incurred due to job termination, although in some instance types check-pointing time can be extended above one hour. Tang et al. [10] propose a bidding strategy for Amazon spot prices that minimizes costs for given time constraints. Kushwaha et al. [11] perform an extensive simulation analysis of the spot market analysis on US East and South Asia regions. However they only consider bidding levels close to the on-demand price, do not consider adaptive bidding nor check-pointing of the computation state.

The Amazon EC2 spot instances are particularly suited for scientific computing, e.g. in [12] a genome sequencing is considered – the authors analyze dependencies between bid levels, number of virtual processors and sequencing time and costs. Another approach discussed in [13] is to utilize reserved instances for long-term running process and to use spot instances for short-term processes.

The existing approaches have several limitations. Firstly, some authors (for instance Javadi et al. [7]) perform statistical analysis only. However, due to the complex spot pricing mechanism this information is insufficient for decision making regarding bidding decisions.

Secondly, Amazon is constantly changing its pricing policy. For example, major recent changes have included around a 30% price reduction on April 1, 2014, introduction of 2 min spot instance termination notice on January 6, 2015 [14] and introduction of new C4 instances on January 11, 2015. The changes in the pricing scheme may lead to previous analyses no longer being valid. For instance, earlier results which claimed that bidding too low leads to a small increase in savings but huge increases in computational time (e.g. Andrzejak et al. [8] point out that “(...) *bidding low prices reduces the monetary cost typically only by about 10% but can lead to extremely high execution times (or, equivalently, realistic deadlines) – up to 400x the task length (...)*”), are no longer true under the current price patterns. Thirdly, some papers (cf. [10]) assume that the spot price changes hourly and a decision to perform calculations at a given price is made hourly as well. Moreover, the above papers do not consider the possibility of obtaining free time by bidding close to the current low price and having a good chance of an out-of-bid situation. Finally, the papers do not analyze how the need for check-pointing influences spot bidding strategies.

In this paper we focus only on an end-user perspective. Our aim is to prepare a realistic model of the spot pricing market with flexible characteristics for simulation experiment features. In this way, we remove all the limitations of models proposed in the literature. We propose a cloud simulation framework that emulates the Amazon EC2 spot billing mechanism exactly and takes into an account the following factors that significantly determine the costs of running large scale simulations:

- (1) ability to bid for many different instances and switch between them;
- (2) instance booting delay;
- (3) real time reaction to instance termination by Amazon;
- (4) exact modeling of the Amazon billing mechanism (in particular, obtaining computing power at no cost when calculations are terminated by Amazon);
- (5) varying time between subsequent simulation state check-pointing events.

The proposed optimizer fulfills the above requirements and is designed in such a way that it can be applied to guide real-life bidding on Amazon EC2.

The remainder of the paper is organized as follows. After the Introduction in Section 2 we discuss the literature related to our research. Next, in Section 3, we shortly describe technical characteristics of how simulations in Amazon EC2 are performed. In Section 4, we provide the details of Amazon EC2 billing rules and outline the design of our cloud simulator which is implemented in Python (source code available at <http://bogumilkaminski.pl/pub/cloudspotsim.zip>). Next, in Section 5, we perform an initial analysis of cost-time trade-off in simulation execution. In Section 6, we discuss a mathematical formalism

Download English Version:

<https://daneshyari.com/en/article/10348709>

Download Persian Version:

<https://daneshyari.com/article/10348709>

[Daneshyari.com](https://daneshyari.com)