Future Generation Computer Systems 29 (2013) 988–999

Contents lists available at SciVerse ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Characterizing spot price dynamics in public cloud environments

Bahman Javadi [a,*], Ruppa K. Thulasiram [b], Rajkumar Buyya [c]

[a] School of Computing, Engineering and Mathematics, University of Western Sydney, Australia

[b] Computational Financial Derivatives (CFD) Laboratory, Department of Computer Science, University of Manitoba, Winnipeg, Canada

[c] Cloud Computing and Distributed Systems (CLOUDS) Laboratory, Department of Computing and Information Systems, The University of Melbourne, Australia

ARTICLE INFO

Article history:
Received 3 February 2012
Received in revised form
6 May 2012
Accepted 11 June 2012
Available online 14 July 2012

Keywords:
Cloud computing
Spot instances
Spot price
Statistical model
Amazon's EC2

ABSTRACT

The surge in demand for utilizing public Cloud resources has introduced many trade-offs between price, performance and recently reliability. Amazon's Spot Instances (SIs) create a competitive bidding option for public Cloud users at lower prices without providing reliability on services. It is generally believed that SIs reduce monetary cost to the Cloud users, however it appears from the literature that their characteristics have not been explored and reported. We believe that characterization of SIs is fundamental in the design of stochastic scheduling algorithms and fault tolerant mechanisms in public Cloud environments for the spot market. In this paper, we have done a comprehensive analysis of SIs based on one year price history in four data centers of Amazon's EC2. For this purpose, we have analyzed all different types of SIs in terms of spot price and the inter-price time (time between price changes) and determined the time dynamics for spot price in hour-in-day and day-of-week. Moreover, we have proposed a statistical model that fits well these two data series. The results reveal that we are able to model spot price dynamics as well as the inter-price time of each SI by a mixture of Gaussians distribution with three or four components. The proposed model is validated through extensive simulations, which demonstrate that our model exhibits a good degree of accuracy under realistic working conditions.

1. Introduction

Due to the surge in demand for using utility computing systems like public Cloud resources, many trade-offs between price and performance have emerged. One particular type of Cloud service, which is known as Infrastructure-as-a-Service (IaaS) provides raw computing with different capacity and storage in the form of Virtual Machines (VMs) with various prices on a pay-as-you-go basis. For instance, Amazon provides on-demand and reserved VM instances, which are associated with a fixed set price [1]. However, Amazon can increase or decrease these prices based on their own local policy. There are 64 different types of instances with various capacities and prices under two operating systems (i.e. 32 for Linux and 32 for Windows) which are made available by Amazon in four data centers as illustrated in Table 1 (sorted by their prices).[1] In this Table, the prices are given for the Linux operating system and the instances labeled with '*m*1', '*m*2', and '*c*1' are standard, high-memory, and high-CPU instances, respectively.

In December 2009, Amazon released a new type of instances called Spot Instances (SIs) to sell the idle time of Amazon's EC2 data centers [2]. The price of an SI, *spot price*, depends on the type of instance as well as VM demand within each data center. In fact, spot instances are an alternative to the other two classes of instances which offer a low price but less reliable and competitive bidding option for the public Cloud users. Therefore, another aspect, *reliability*, has been added to the existing trade-offs to make utility computing systems more challenging than ever.

In order to utilize SIs, the Cloud users provide a *bid* which is the maximum price to be paid for an hour of usage. Whenever the current price of an SI is equal or less than the user bid, the instance is made available to the user. If the price of an SI becomes higher than the user's bid, *out-of-bid* event (failure), the VM(s) will be terminated by Amazon automatically and the user does not pay for any partial hour. However, if the user terminates the running VM(s), she has to pay for the full hour. Amazon charges users per hour by the market price of the SI at the time of VM creation.

There are a number of works on how to utilize SIs to decrease the monetary cost of utility computing for Cloud users [3–5]. However, a thorough statistical analysis and modeling of SIs have not appeared in the literature, the focus of our research in this study. In this paper, we provide a comprehensive analysis of all SIs in terms of spot price and the inter-price time (time between price changes) in four Amazon data centers (i.e. us-west, us-east, eu-west, and ap-southeast). Moreover, we propose a statistical model to capture the volatile spot prices in Amazon's data centers. The main contributions of this paper are as follows:

* Corresponding author. Tel.: +61 2 9685 9181; fax: +61 2 9685 9245.
E-mail address: b.javadi@uws.edu.au (B. Javadi).

[1] Amazon now has seven data centers around the world, but the four major data centers are considered in this research.

0167-739X/$ – see front matter
doi:10.1016/j.future.2012.06.012

Table 1
Prices of on-demand instances in different data centers of Amazon (prices given in cents).

Instances	us-west	us-east	eu-west	ap-southeast	EC2 compute unit	Memory (GB)	Storage (GB)
*m*1.small	9.5	8.5	9.5	9.5	1	1.7	160
*c*1.medium	19	17	19	19	5	1.7	350
*m*1.large	38	34	38	38	4	7.5	850
*m*2.xlarge	57	50	57	57	6.5	17.1	420
*m*1.xlarge	76	68	76	76	8	15	1690
*c*1.xlarge	76	68	76	76	20	7	1690
*m*2.2xlarge	114	100	14	114	13	34.2	850
*m*2.4xlarge	228	200	228	228	26	68.4	1690

- We provide statistical analysis for all SIs in Amazon's EC2 data centers. We also determine the time correlation in spot price in terms of hour-in-day and day-of-week.
- We model spot price and the inter-price time of each SI with a mixture of Gaussians distribution. A model calibration algorithm is also proposed to deal with an observed price trend in the real price history.
- We validate and verify the accuracy of our proposed model through simulation under realistic working conditions.

We believe that results of this research will be significantly helpful in the design of stochastic scheduling algorithms and fault tolerant mechanisms (e.g. checkpointing and replication algorithms) for the spot market in public Cloud environments. In addition, although Amazon is the only provider of SIs at the moment, some research has been conducted to analyze the free computing resource markets [6,7]. So, this model can be used by other resource providers that look to offer such a service in the near future.

The paper is structured as follows. In Section 2, we describe the processes that we model in this paper. We discuss related work in Section 3. We examine the pattern of spot price in Section 4. In Section 5, we present the global statistics for all SIs. We then illustrate distribution fitting for spot price and the inter-price time in Section 6. In Section 7, we propose an algorithm for model calibration. We discuss the validation of the proposed models through simulation in Section 8. In Section 9, we summarize our contributions and describe future directions.

2. Modeling approach

In this section, we describe two variables that we are going to analyze and model. In Amazon's data centers, SIs have two variables (i.e. spot price and inter-price time) specified by the Cloud provider and one variable (user's bid) determined by users. In this study, we focus on the analysis and modeling of spot price and the inter-price time as two highly volatile system variables. These variables are illustrated in Fig. 1 where P_i is the price of an SI at time t_i. So, the inter-price time is defined as $T_i = t_{i+1} - t_i$. Therefore, the time series of spot price (P_i) and the inter-price time (T_i) are analyzed and modeled in the following sections.

The traces that we use in this study are one year price history of all Amazon SIs from the first of February 2010–mid-February 2011. We use the first 10 months (Feb-2010–Nov-2010) in the modeling process. These 10-month traces along with the last 2 months are used for the model validation purpose. The spot price history is freely provided by Amazon per SI for each data center and also available through other third parties such as [8]. We do not use data prior to February 2010 due to an algorithm issue reported in [9] for prices. Moreover, we only use the SIs with Linux operating systems from all data centers. Due to the similarity of the results, we present our findings for only two data centers (i.e. eu-west and us-east). Interested readers can refer to the extended version of this paper [10] for more discussion about other data centers.

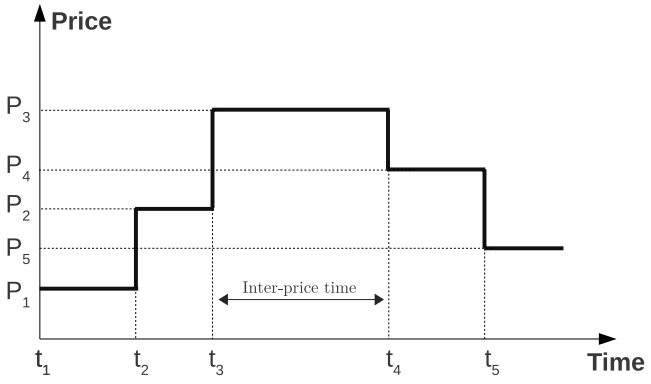

Fig. 1. Spot price and the inter-price time of Spot Instances.

3. Related work

Although the current literature shows that SIs are a good alternative for on-demand or reserve instances in terms of monetary cost, the characteristics of SIs are still not clear to users and researchers in the community. Wee [11] considered SIs as computing resources with real-time pricing. Focusing on the real price history of SIs, this paper concluded that still users need more monetary incentive to shift their workload into SIs. Another work that investigated the behavior of spot prices is presented in [12], where the authors used reverse engineering to construct a price model based on the Auto-Regressive (AR) model for SIs.

Our work is different in several aspects. We provide statistical analysis of all SIs and study their behavior in terms of hour-in-day and day-of-week. Moreover, we propose to devise a statistical model for spot price as well as inter-price time. In addition, the simulation results reveal that we are able to model behavior of SIs by a mixture of Gaussians with three or four components.

In the following, we briefly review the other related work mainly investigating the usage of SIs to decrease the monetary cost of utility computing. Yi et al. [3,4] introduced some checkpointing and migration mechanisms for reducing the cost of SIs. They used the real price history of EC2 spot instances and showed how the adaptive checkpointing and migration schemes could decrease the monetary cost and improve the job completion times. Chaisiri et al. [13] proposed two provisioning algorithms based on stochastic programming, robust optimization, and sample-average approximation to optimized the provisioning cost for long-term and short-term planning. Moreover, in [14], a resource allocation policy to run deadline constrained jobs on SIs in a cost-effective manner is proposed.

In [15], a decision model for the optimization of performance, cost and reliability under SLA constraints while using SIs is proposed. They used the real price history and workload models to demonstrate how their proposed model can be used to bid optimally on SIs to reach different objective with desired levels of confidences. Mazzucco and Dumas [16] considered a case where a web service is deployed on SIs and proposed a bidding schema and