Auction–Based Pricing in Cloud Environment

Branka Mikavica
Faculty of Transport and Traffic Engineering, University of Belgrade, Serbia

Aleksandra Kostic-Ljubisavljevic
Faculty of Transport and Traffic Engineering, University of Belgrade, Serbia

INTRODUCTION

The rapid development of technologies triggered the emergence of cloud computing. Cloud computing enables network access to a scalable and elastic pool of shareable physical or virtual resources. Resources in the cloud environment are generally provisioned as cloud instances, i.e. Virtual Machines (VMs) with the required CPU, memory and I/O resources. The main advantages of cloud computing are scalability, low costs, dependability, service-driven, self-healing and multi-faceted. However, there are also numerous challenges including a vulnerability, bandwidth costs, transparency access, trust, reliability, security, availability, resource management, task scheduling, and performance (Gullhav & Nygreen, 2016; Sebastio, Gnecco & Bemporad, 2017). The dominant factor in supporting customers to use virtual resources and move towards cloud computing is pricing. Cloud customers pay for required cloud resources depending on instance type and pricing mechanism applied. In addition, different levels of guarantees in terms of availability and termination are possible. Pricing mechanisms in the cloud environment can be generally classified into static and dynamic. The static pricing mechanisms are the dominant scheme in the cloud environment and involve the charging the fixed price per hour and per instance. The dynamic pricing mechanisms include charging the usage of cloud resources as long as the cloud customers' bids are equal or exceed the price per instance and per time interval, determined by the cloud provider (Alzhouri & Agarwal, 2015). For example, Amazon Elastic Compute Cloud (EC2) uses three different pricing options for cloud resources: Reserved instances, On-demand instances, and Spot instances. Reserved instances are provided with long-term availability guarantees, On-demand instances are provided for specific time periods and are guaranteed to be available once assigned, and Spot instances are provided through an auction-like mechanism, but their availability depends on the time-varying Spot price. Reservation and On-demand pricing mechanisms are considered as static since prices for cloud resources are invariant for both under-utilization and over-utilization of the resources. Idle cloud resources can be offered as Spot instances. This purchasing option offers no guarantees on minimum sustained availability. Submitting a bid for the access to Spot instances, cloud customers specify the maximum price they are willing to pay. The actual price they pay is not the value of the bid, but the value of the Spot price. Cloud provider set Spot price depending on the availability of the idle resources and current demand. When the Spot price exceeds the value of the bid, the Spot instance is terminated. Cloud provider can also revoke Spot Instances when the supply of available Spot Instances decreases. Recently, a new variety of Spot instances, Spot Block instances, are introduced. These instances run continuously for a finite duration, up to 6 hours. When cloud customer submits a bid, two parameters need to be specified: the value of the bid (the maximum price willing to pay per hour) and the Block

DOI: 10.4018/978-1-7998-3473-1.ch008

duration parameter (the number of hours that Spot Block instance will run). The Spot Block instance is terminated only at the end of the Spot Block duration.

As a dynamic form of pricing, auction mechanisms can be an effective and promising solution for profit maximization of cloud providers. These mechanisms provide price variation depending on the changes in supply and demand by creating competition among cloud customers. Therefore, auctions can balance supply and demand for cloud resources. They support cloud customers for truthful bidding and allocate the resources to the customers that value them the most. Various auction-based allocation and pricing mechanisms can be used for cloud resources: Uniform price auctions, Second-price auctions, Combinatorial auctions, Double auctions etc. In this chapter, the analysis of different auction-based mechanisms for pricing and allocation of cloud resources is provided.

BACKGROUND

Auction mechanism is also referred as market-clearing mechanism that equates demand and supply. It is a widely adopted mechanism for resource allocation in telecommunications. For instance, auctions are used for efficient and effective spectrum allocation in mobile communication, and to trade-off electricity using spot marketing. There are three entities in an auction: seller, buyer and auctioneer (Baranwal, Kumar, Raza & Vidyarthi, 2018). Seller owns the commodities, offers them to the buyers and earns profit. In terms of cloud computing, cloud providers are sellers, while commodities are cloud resources (for instance, various types of VMs). Buyers are cloud customers who purchase the commodities, i.e. cloud resources from the cloud providers. Auctioneer is the entity that performs an auction. It is an intermediate agent that acts as the market maker. Since the auctioneer stores the repository of the historical transactions, it provides support for cloud providers to identify the new trends in cloud resource allocation. Auctions consist of two sub-processes: winner determination and payment mechanism. These processes are performed in six steps, as shown in Figure 1.

The auction starts with the collection of bids from various bidders by the auctioneer. Set of rules and constraints in an auction need to be specified clearly. Constraints might be bid expiry, resource availability, bidder's budget, time period for resource provisioning, etc. Bids that do not satisfy the given set of constraints are canceled in validation phase by the auctioneer. Step 2 refers to the auction closing. Auctioneer can close the auction and finish accepting the bids upon certain conditions (sufficient number of bids received, time limit specified for bidding, availability of resources, etc.). Once the auction is closed, any change in a bid or accepting new bids is not allowed. The value of each bid is computed after closing the auction (Step 3). Bids are than sorted, depending on their values. In Step 4, resources are ranked according to their intrinsic value. Depending on the predefined rules, some bidders are selected as winners in resource allocation (Step 5). Computational complexity of the winner determination increases with the increase in the number of the participants in the market. In Step 6, cloud customer pays according to the defined pricing mechanism. If the pricing mechanism is properly implemented, the auction is truthful, i.e. incentive compatible.

Various auction-based pricing and allocation mechanisms in the cloud environment are proposed (Kumar, Baranwal, Raza & Vidyarthi, 2017; Kumar, Baranwal, Raza & Vidyarthi, 2018; Lu, Yu, Zhu, Li, 2018). The current research on auction-based pricing mechanisms in the cloud environment is primarily focused on a static, single type resource with a single requirement. The issue closely related to pricing is cloud resources allocation. A comprehensive literature review of auction mechanisms for cloud resource allocation is provided by Sheikholeslami and Navimipour (2018). There are many possible classifications