2010 Sixth International Conference on Semantics, Knowledge and Grids

A Knowledge-based Continuous Double Auction Model for Cloud Market

Shifeng Shang, Jinlei Jiang, Yongwei Wu, Guangwen Yang, Weimin Zheng
Department of Computer Science and Technology, Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing,100084, China
Email:shangsf@gmail.com, {jjlei, wuyw, ygw, zwm-dcs}@tsinghua.edu.cn

***Abstract*---Recently, the Storage Networking Industry Association (SNIA) has released the first standard for cloud interoperability. With more and more standards for interoperability emerging, it can be expected that a global cloud resource exchange market will form. In such a market, it is challenging to present a dynamic pricing scheme to meet different requirements. To cope with the challenge, in this paper we first present a framework for constructing global cloud resource markets and then propose a knowledge-based continuous double auction (CDA) model that determines the price of cloud resources using a learning algorithm based on historical trading information. Experimental result shows that our model can attain high market efficiency as well as stable trading price.**

I. INTRODUCTION

Cloud computing, which refers to services (hardware such as CPUs and storage, platform and application) provisioning and consumption over the Internet in an on-demand approach, is becoming a hot topic both in academia and industry around the world. Academic efforts include Nimbus [1], Aneka[2][3], Open Nebula [4], Tsinghua Cloud[5]. Industrial services include Amazon EC2 (Elastic Compute Cloud) and S3 (Simple Storage Service) [6], Google App Engine (GAE) [7], Microsoft Azure [8], Rackspace [9], GoGrid [10], VPS.net [11], to name but just a few.

One attractive advantage of cloud computing is the pay-as-you-go billing model, which is believed of the potential to cut down the total cost of ownership. Typically, the price of a cloud computing resource consists of the following parts:

$P=P_{comp}+P_{storage}+P_{in}+P_{out}+P_{tran}$

P_{comp}: The price of the corresponding virtual machine instance. Amazon EC2 instances are grouped into three types: Standard, High-Memory and High-CPU [6].

$P_{storage:}$ The price of user data (the computing result through virtual instance) stored on the cloud.

P_{in},$P_{out:}$ The price of uploading data to or downloading data from the cloud, or transferring data between different regions of the same cloud vendor.

P_{tran}: the price of file operations within a virtual instance. For example, the P_{tran} of Amazon EC2 is $0.10 per 1 million file operations.

Table 1 shows the services prices of different vendors, where the basic configuration of compute is of 1 GB (=10^9 bytes) RAM, and 40 GB Disk. We can see that the prices of compute range from $0.06 to 0.12 with a maximal difference as much as $0.06 per hour.

Table I.
THE CLOUD SERVICES PRICES OF DIFFERENT VENDORS

Price Type	Amazon	Azure	Google	GoGrid	Rack-space
Compute CPU/hours	$0.085/ linux $0.12/ windows	$0.12	$0.10	$0.10	$0.06
Storage GB/month	$0.15	$0.15	$0.15 0.5GB free	$0.15 10GB free	$0.15
Data Upload GB	$0.10	$0.10	$0.12	$free	$0.08
Data Download GB	$0.17 $0.13 if>10TB	$0.15	$0.10	$0.29	$0.22

At present, most companies adopt a fixed rate pricing strategy, and user can get a great discount through pre-pay method. This can cut down cost further. Table 2 is the price comparison of different types of instance using different pricing methods; from the table we can see that the average difference can be as high as 2.7 times even for the same instance type.

Table II.
PRICES OF DIFFERENT AMAZON EC2 INSTANCES

Price Method Instance Type		On-Demand	Reservation	One Time Fee /Year
Standard	Small	$0.085/h	0.03/h	$227.50
	Large	$0.34/h	0.12/h	$910
	Extra Large	$0.68/h	0.24/h	$1,820
High-Memory	Extra Large	$0.50/h	0.17/h	$1,325
	Double Extra Large	$1.20/h	$0.42/h	$3,185
	Quadruple Extra Large	$2.40/h	$0.84/h	$6,370
High-CPU	Medium	$0.17/h	$0.06/h	$455
	Extra Large	$0.68/h	$0.24/h	$1,820

Though the pre-pay strategy can save money, it might be unfair to both resource provider and buyer. First, it may

978-0-7695-4189-1/10 $26.00 © 2010 IEEE
DOI 10.1109/SKG.2010.22

result in resource waste for applications that user only need to run once a month for hours. Secondly, the pre-pay method will potentially make users be locked a certain providers for long time with little chance to receive better and cheaper services from other vendors. Thirdly, in some situation, it is even expensive to use the fixed rate pricing model [12], which has a strong influence on the wide adoption of cloud computing.

Recently, the SNIA (Storage Networking Industry Association) has issued the first standard for interoperation among different cloud storage systems. Still, there are many other organizations (e.g., Cloud Security Alliance, Open Cloud Consortium, Open Grid Forum, the Distributed Management Task Force, etc) engaged in standards for cloud interoperation. These standards will provide the interoperability needed to enable vendors and users alike to take the next step towards widespread cloud computing [14]. At the same time, with the rapid advancement of QoS (quality of service) and the enhancement of cloud security, more and more users begin buying computing resources, storage resources and software resources from the cloud. It can be expected that in the near future, user can buy resource from the vendor of the highest performance-to-cost ratio, build applications using resources from different companies, or even exchange their free resources.

Since December, 2009, Amazon has tentatively established a one side auction market for users to consume resources at a lower and more flexible cost [13]. In this paper, we set out to establish a global double auction cloud resources trading market for both users and vendors. Such a market is further divided into futures market and spot market. Through the future market, providers can allocate resources more effectively in advance, and the idle cloud resource can be sold through spot market and get more revenue. The efficiency of cloud market trading mechanism must be higher to deal with a large number of transactions, and the pricing model must be flexible enough to meet different requirement.

The rest of the paper is organized as follows. Section II gives an overview of related work. In Section III the cloud resource market is introduced. Section IV explains the knowledge-based continuous double auction (CDA) trading mechanism and Section V is the experimental result. The last section is the conclusion and future work.

II. RELATED WORK

There exist many economy-based resource allocation models in grid research [15]. Wolski et al [16] have used the commodity market approach to allocate two types of resources (i.e., CPU and storage) in grid. Allocations are done when an equilibrium price is reached, or in other words, the demand equals the supply. The auction protocols are either one-to-many or many-to-many.

In one-to-many auctions, one agent initializes an auction and a number of other agents can make a bid. The English auction, Dutch auction, first-price auction, second-price auction (Vickrey auction) belong to this category. Popcorn [17] is an example of this approach.

In many-to-many auctions, several agents set up an auction and other agents can bid in the auction. The double auction is the most widely used auction protocol for many-to-many auctions. There are two types of double auctions, namely continuous double auction (CDA) and periodic double auction. [18] [19] are two examples of double auction model.

However, all the works above use virtual currency and pay more attention to the fairness to users, with no requirement on providers' price taken into consideration. In addition, most pricing mechanisms are static. In our opinion, the most important advantage of cloud computing is its capability to cut down the costs of users. Since it is the main aim of both vendors and users to maximize their revenue, the pricing strategy must be flexible enough to meet the different requirements of various users, and the market efficiency must be higher. To achieve this purpose, a global double auction market framework is established in this paper, and a knowledge-based CDA trading mechanism is proposed.

III. THE CLOUD RESOURCE MARKET

To facilitate cloud resources trading, we propose a uniform and fully competitive cloud market framework as shown in Fig. 1. The cloud resource trading market can be divided into the futures market and the spot market. There are three main types of participants in this framework, namely the cloud resource auctioneer (CRA), the cloud resource Buyer (CRB), and the cloud resource seller (CRS).

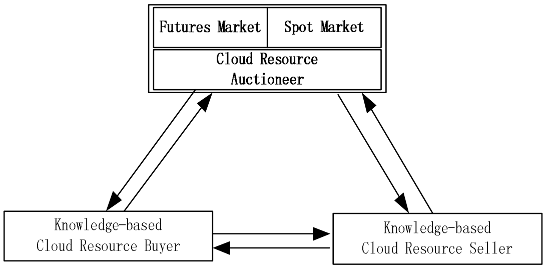
Futures Market
Spot Market
Cloud Resource Auctioneer
Knowledge-based Cloud Resource Buyer
Knowledge-based Cloud Resource Seller

Fig. 1 The Cloud Resource Market Framework

Futures Market

The futures market enables a cloud resource user to purchase a bundle of resources in advance so that they can get great discount. The trade of futures market auction should be held regularly with relatively stable intervals. The resources in futures market are cloud resource for a long term (e.g., three months, half a year, whole year, three years etc.) The participant of futures market mainly is cloud resource wholesaler, big consumer, or the user whose demand is stable or predictable.

Spot Market

The spot market sells short-duration resources for immediate use. For example, one might hold auctions every hour for resources that last two hours or weeks. The purpose of the spot market is to allow last-minute demand to be met by resources that have not been sold, yet would lead to waste if they were not sold immediately. All resources currently unused are auctioned. As mentioned above, resources bought from the futures market can also be entered into the spot market for immediate sale and consumption. Also, there are multiple possible mechanisms for the spot market, just like a person who has emergency to deal with or who goes to a hotel without reservation. Spot market provides more flexibility for users.

Spot market is a complement to futures market, providing another option for customers with flexible requirements on the time when their applications can run to obtain compute capacity on the fly. Additionally, Spot market can provide access to large amounts of additional capacity for applications with urgent needs, for example, image and video processing, scientific research data processing, financial data analysis, to name but just a few.

Cloud Resource Buyer (CRB)

CRB helps users to determine the most appropriate computing capacity according to the user's application requirement, budget, deadline and so on. CRB is also responsible for generating a bidding price within the user's SLA (Service-Level Agreement) requirement and submitting it to the CRA. For example, a typical purchase request might be: "I want to purchase the rights to 8 medium instances from 20:00 to 24:00 on Friday". After CRB receives the bid result from CRA, it forwards the resource request to CRS. Resources purchased can be re-entered into either the future market or the spot market.

Cloud Resource Seller (CRS)

Cloud resource seller (CRS) is a datacenter that sells its resources to users and profits from it. The resources include computing power, storage space, network bandwidth, etc. CRS is responsible for registering cloud resources and generating bidding prices to CRA. The CRS is also in charge of receiving the auction result from the CRA, and allocating resources according to the corresponding CRB's request. It is also CRS's duty to adjust the bidding price according to the situation of the supply and demand of cloud resources, and to charge CRBs for the cloud resources consumed.

Cloud Resource Auctioneer (CRA)

Double auction means the arrangement where cloud resources providers and a group of buyers interact to reach a mutually agreed price. The responsibility of cloud resource auctioneer is to collect the bids for resources made by the cloud resource buyers and cloud resource sellers. Based on the corresponding bidding information, the auctioneer determines the winning buyers and sellers according to different double auction mechanisms. Finally CRA returns the decision to participants. The CRA also provides insurance against a number of events that would harm both providers and user, for example, the availability of cloud resources is interrupted. Two forms of double auction are supported by CRA, namely clearing house auction and continuous double auction (CDA).

Clearing house auction is the simplest form of double auction. In this approach, the auctioneer collects all bids (demanding) and quotations (supplying). When the demand curve and a supply curve intersect, the intersection point specifies the market clearing (equilibrium) price and all possible trades occur simultaneously at that price.

CDA is a trading mechanism that a group of sellers and a group of buyers simultaneously announce sell orders and buy orders at any time, and may be retracted at any time. The cloud resource auctioneer maintains a public order book, and all sell orders is in an ascending list and buy orders is in a descending list. Trades are executed whenever a new order comes in and the highest bided price exceeds or equals to the highest asked price. CDA is practically important because its variants have been widely adopted in real-world stock markets like NASDAQ and NYSE or in trading markets like CME [22]. In this paper, a knowledge-based CDA is adopted, for its efficiency is higher.

IV. KNOWLEDGE-BASED TRADING MODEL

Previous research in grids focuses on using virtual currency to ensure the fairness of grid resource allocation, with little consideration of the price dynamic. Based on Gjerstad-Dickhaut algorithm [20][21], we proposed a knowledge-based continuous double auction trade model. We introduce a probability based on historical trading information, and use historical bids to determine the probability that future bids will succeed. With this probability agent can then adjust the bidding or quoting price or ask price automatically. Combine this probability with profit to estimate how to place bids to maximize expected profit. If there were many bids made at each price point than the probability could simply be the number of shouts accepted at a particular price point。This trading policy use the history of recent market activity (the bids and asks leading to the latest M trades) to estimate the probability for a bid or ask at price b or c to be accepted.

The knowledge-based CDA trading model can be defined as follows:

$$CDA=(P_s,P_B,S,U,R) \qquad (1)$$

- Ps={Ps1,…,Psi} is a queue of asking prices of Cloud resource seller, with PS1<Ps2<…<Psi.
- PB={PB1,…,PBj} is a queue of bidding prices of resource buyer, with PB1>PB2…>PBj
- S={S1,…,Sm} is a collection of cloud resource sellers.
- U= {U1,…,Un} is a set of cloud resource buyers.

- R is the type of cloud resource auctioned by the CDA. The cloud resource can be compute instance, storage resource, or software resource or a combination of them. For the simplicity, we assume that there is only one kind of resource.

The model works as follows.

- If $P_{Bi} \geqq P_{Si}$, a transaction happens between the seller that submitted P_{Si} and the buyer who submitted P_{Bi}. The trading price is defined by

  $$P=k\ P_{Bi}+(1-k)P_{Si} \qquad (2)$$

  Where k is a variable with typical value 0.5.

Once a transaction succeeds, the corresponding trading pairs are removed from the queue.

Let H_Q be the trade History of most Q rounds, that is,

$$H_Q=\{P_{S1},\ldots,P_{SQ}\} \qquad (3)$$

- When a Resource Seller submits a new quotation s, he uses the recent M trades in H_Q to calculate the probability that the seller will conduct a trade with a certain quotation.

For a seller, the probability that a quotation s would be accepted is defined as follows:

$$P(s)=\frac{T_{AH(s)}+T_{BH(s)}}{T_{AH(s)}+T_{BH(s)}+T_{RA(s)}} \qquad (4)$$

$T_{AH(s)}$ is the number of accepted quotations in the latest M trades with price no less than s; $T_{BH(s)}$ is the number of bids in the latest M trades with price no less than s; and $T_{RA(s)}$ is the number of unaccepted quotations with price less than s.

For a buyer, we use formula (5) to calculate the probability that a bidding b could be accepted.

$$P(b)=\frac{T_{BL(b)}+T_{AL(b)}}{T_{BL(b)}+T_{AL(b)}+T_{RB(b)}} \qquad (5)$$

$T_{BL(b)}$ is the number of accepted bids in the latest M trades with price no greater than b; $T_{AL(b)}$ is the number of quotations of seller in the latest M trades with price no greater than b; and $T_{RB(b)}$ is the number of unaccepted bids with price greater than b.

- After *P*(*s*) and *P*(*b*) are got, the algorithm uses interpolation method to solve

$$P(b)=A_3B^3+A_2B^2+A_1B+A_0 \qquad (6)$$

$$P(s)=A_3S^3+A_2S^2+A_1S+A_0 \qquad (7)$$

- The buyer agent gets a price that maximizes its expected surplus.

  P_{BQ+1}= （B| max (P (b) * (b– B))) (8)

  Using the same type of strategy for quotation, we get

  P_{SQ+1} = （B| max (P (b) * (s –S))) (9)

For the buyer, if $P_{BQ+1} \geqq P_{S1}$, then the trade happens; otherwise, it inserts P_{BQ+1} to an appropriate position in the current sorted bidding list. For the seller, if $P_{SQ+1} \leqslant P_{B1}$, then the trade happens; otherwise, it inserts P_{SQ+1} to an appropriate position in the current sorted quotation list.

V. EXPERIMENT AND RESULT

Overall profit is one of the measurements to evaluate the new distribution of resources, which is the sum of profits all the sellers and buyers obtain through the trading. For a seller, the profit is the difference between the cost of the commodity he or she wants to sell and the price at which the resource is sold. For a buyer, the profit is the difference between the price at which he or she buys a commodity and the actual price of that commodity. To better compare overall profits of different auctions, overall efficiency is used instead, which is defined as overall profit divided by theoretical overall profit. The latter refers to the overall surplus when the market is cleared at the equilibrium price. The equilibrium price is determined by the supply and demand curves of an auction, at which the total supply of sellers involved in transactions equals to the total demand of buyers making deals.

The overall efficiency of resources allocation is the most important feature of a market trading mechanism. It is defined as the percentage between actual overall profit E_a and E_t , that is,

$$A=\frac{E_a}{E_t}$$

The following metric measure the overall profit obtained through the auction:

$$E_b=\sum_1^n(P_b-P_i)$$

$$E_{bt}=\sum_1^n(P_b-P_t)$$

P_b is the valuation of buyer, P_i is the real trade price. E_b is the actual overall profit obtained by buyer. P_t is the trade price at the equilibrium point. E_{bt} is the theoretical overall profit of all buyer that all transactions happen at equilibrium price.

$$E_s = \sum_1^m (P_c - P_i)$$

$$E_{st} = \sum_1^m (P_c - P_t)$$

P_c is valuation of seller. P_i is the real trade price. The actual overall trade profit of seller is E_s . E_{st} is the theoretical overall profit of all resource provider.

E_{st} is the theoretical overall profit of all resource provider.

$$E_a = E_s + E_b$$

$$E_t = E_{st} + E_{bt}$$

So, the overall efficiency of resource allocation is:

$$A = \frac{\sum_1^n (P_b - P_i) + \sum_1^m (P_c - P_i)}{\sum_1^n (P_b - P_t) + \sum_1^m (P_c - P_t)}$$

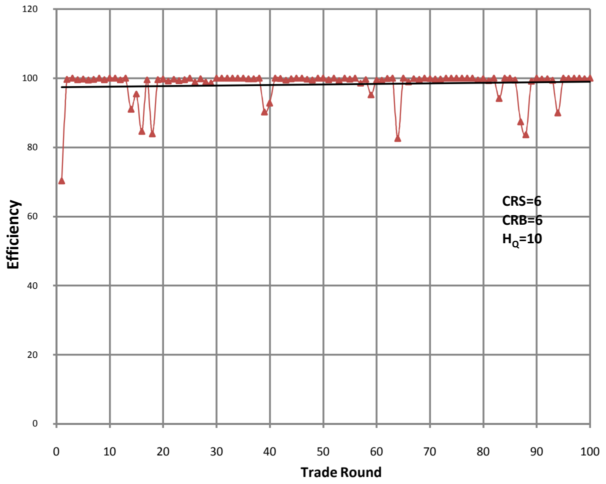

Fig. 2. The efficiency of our knowledge-based trading model in resource allocation

We have developed a simulator to test the related feature. We set the double auction round to 100, the number of cloud resource buyer and cloud resource seller both to 6, and the length of History Trade Activity Queue to 10. Fig. 2 is the experiment result of our simulation. The results show that our model is efficient in resource trading. The mean efficiency of resource trading is 97.770%. From the figure, we can also see that the trading price is more stable.

VI. Conclusion and Future Work

As the rapid advancement of cloud computing technology, more and more companies began to sell computing resources, storage resources and software resources online. With more and more standards for cloud interoperation emerging, it is significant to establish a global double auction market and a more flexible pricing policy to meet different requirements. In this paper, we present a global knowledge-based continuous double auction cloud resource market framework and discuss its pricing policies. In this market, the price of cloud resource is determined using a history trading information-based learning algorithm. It is shown that high market efficiency can be obtained, and the trading price is stable.

The future work includes applying our model to a real cloud resource environment and conducting experiments of larger scale to test the efficiency of our model.

Aknowledgement

This Work is supported by Natural Science Foundation of China (60803121, 60773145, 60911130371, 90812001, 60963005), National High-Tech R&D (863) Program of China (2009AA01A130, 2006AA01A101, 2006AA01A108, 2006AA01A111, 2006AA01A117).

References

[1]http://www.nimbusproject.org/,[10 Mar 2010]

[2]R. Buyya "Market-Oriented Cloud Computing: Vision, Hype, and Reality of Delivering Computing as the 5th Utility". in Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid. 2009: IEEE Computer Society.

[3] X. Chu, K. Nadiminti, C. Jin, S. Venugopal,and R. Buyya. Aneka:Next-Generation Enterprise Grid Platform for e-Science and e-Business Applications. In Proceedings of the3th IEEE International Conference on e-Science and Grid Computing (e-Science 2007),Bangalore, India, Dec. 2007.

[4] I. Lorente, Open Nebula Project. http://www.opennebula.org/ , [10 Mar 2010]

[5] http://grid.tsinghua.edu.cn/hpcgrid/GCD/cloud/cloud.htm

[6]Amazon Elastic Compute Cloud (EC2), http://aws.amazon.com/ec2/, [10 Mar 2010]

[7] http://www.microsoft.com/windowsazure/,[7 Mar 2010]

[8]Google App Engine, http://appengine.google.com, [10 Mar 2010]

[9]http://www.rackspace.com/index.php, [10 Mar 2010]

[10] http://www.gogrid.com/index.v2.php,, [10 Mar 2010]

[11]http://www.vps.net/ [10 Mar 2010]

[12]Jared Wilkening, Andreas Wilke, Narayan Desai, Folker Meyer1, Using Clouds for Metagenomics: A Case Study, Cluster 2009

[13]http://aws.amazon.com

[14]http://www.snia.org [2 April 2010]

[15]R. Buyya, D. Abramson, and S. Venugopal "The Grid Economy". Proceedings of the IEEE, 93(3): 698-714, IEEE Press, USA, March 2005.

[16]R. Wolski, J. Plank, J. Brevik, and T. Bryan. "G commerceMarket formulations controlling resource allocation on the computational grid", In In Proc. International parallel and Distributed Processing Symposium (IPDPS), April 2001.

[17] N. Nisan, S. London, O. Regev, and N. Camiel. “Globally distributed computation over the internet the popcorn project”, In ICDCS ’98: Proceedings of the The 18th International Conference on Distributed Computing Systems, page 592. IEEE Computer Society, 1998.

[18]S. Lalis and A. Karipidis. Jaws “An open market based framework for distributed computing over the internet”. In GRID, pages 36–46, 2000.

[19]B. Pourebrahimi, K. Bertels, G. Kandru, and S. Vassiliadis, “Market-based resource allocation in grids”, In proceedings of 2nd IEEE International Conference on e-Science and Grid Computing, Page(s): 80-88, 2006.

[20]Steven Gjerstad, John Dickhaut “Price Formation in Double Auctions”, GAMES AND ECONOMIC BEHAVIOR 22, 1]29_1998.

[21]Nie, Liu P. B., Rourke S. “Optimal integrate generation bidding and scheduling with risk management under a deregulated daily power market”, IEEE Trans on Power Systems, Vol. 19, No. 1, 2004, pp. 600-609.

[22]S. Parsons, M. Klein, and J. A. Rodriguez“A user's guide to auctions.Technical report, Department of Computer Science”, Brooklyn Collge, CityUniversity of New York, 2900 Bedford Avenue, Brooklyn, 11210 NY, May 2005.