Double Auction for Resource Allocation in Cloud Computing

Zhichao Zhao[1], Fei Chen[2], T-H. Hubert Chan[1], Chuan Wu[1]
[1]*The University of Hong Kong, Hong Kong*
[2]*Noah's Ark Lab, Huawei Technologies, Hong Kong*
[1]{*zczhao,hubert,cwu*}*@cs.hku.hk*, [2]*chenfei100@huawei.com*

Keywords: cloud resource allocation, double auction, truthful mechanism

Abstract: Cloud computing has become more and more popular as more companies choose to deploy their services and applications to the cloud. Particularly, trading unused cloud resources provides extra profits for companies with rapidly changing needs. Cloud market enables trading additional resource between buyers and sellers, where a buyer may have different valuations for different instances of the same resource due to factors such as geographical location, configuration, etc. In this paper, we study double auctions with non-identical items for cloud resource allocation, and develop a framework to decompose the design of truthful double auctions. We propose two auctions based on the framework that achieve: (i) truthfulness; (ii) individual rationality; and (iii) budget balance. We prove that the social welfare is constant-competitive to the (not necessarily truthful) optimal auction under certain distributions. We run simulations to investigate the social welfare achieved by our auctions. We use different probability distributions to capture various scenarios in the real world. Results show that our mechanisms generally achieve at least half of the optimal social welfare, while one auction achieves over a 0.9 fraction of the optimal in some circumstances.

1 INTRODUCTION

Motivation. Cloud computing has gained more attention in recent years (Zaman and Grosu, 2011; Zhang et al., 2014; Zhao et al., 2014). It enables on-demand delivery of computing resources such as computation and data storage via the Internet. Compared to the traditional approach where enterprises run their own IT infrastructure, it has the advantage of lower maintenance cost and better scalability for both hardware and software. As an example, Amazon EC2 (Amazon, 2015d) and Google Compute Engine (Google, 2015) both provide resizable compute capacity in the cloud. Both services provide a wide variety of configurations consisting of different CPU capacities, RAM sizes and storage volumes. Furthermore, they provide virtual servers in multiple locations with different charges, satisfying customers' need for low network latency worldwide. Cloud customers normally buy long term contracts to reduce the average cost. For example, Amazon EC2 reserved instances (Amazon, 2015b) have a much lower price for long period contracts than short period ones.

However, enterprises' demand might change rapidly, which motivates the need for dynamic reallocation of cloud resources. A user can (1) offer unused cloud resource at a lower market price; and (2) purchase additional cloud resource when the demand increases. For instance, the aforementioned Amazon EC2 has a marketplace (Amazon, 2015a) that enables second-hand cloud resource trading for reserved EC2 instances. Another company UCX, the Universal Compute Xchange, also enables trading of computer resources (UCX, 2015) that are measured and transacted on UCX.

One issue that remains is how to allocate cloud resources between multiple buyers and sellers to maximize resource utilization. As sellers and buyers aim to maximize their individual utility, it is important to design truthful mechanisms to deal with both sides.

Objectives. We aim at designing double auctions with non-identical items for cloud resource allocation, where both sellers and buyers are agents (bidders) and a buyer's valuations on the items may depend on factors such as geographical distance. We consider the following properties (Dong et al., 2014):

- Truthfulness: Every agent has no incentive to lie about its valuation to maximize its own utility.
- Individual rationality: Honest agents always have non-negative utilities. That is, each buyer pays at

[1]This research was partially supported by the Hong Kong RGC under the grant 17202715.

most its bidding price and each seller gets paid at least its bidding price.
- Budget balance: The party running the auction has non-negative profit.
- Social Welfare: The auction is targeted to maximize the social welfare, which is defined as the difference between sum of winning buyers' valuations and sum of winning sellers' valuations. Intuitively, this means that allocation should be made to buyers with higher valuations and sellers with lower valuations. The performance ratio is defined to be the ratio of the social welfare of the auction to that of an (not necessarily truthful) optimal auction.

In this paper we assume each agent is either a buyer or a seller, but not both.

Limitation of Previous Work. To our knowledge, most previous work that proposes auctions for cloud computing considers one-sided auctions, where only the buyers are agents (Zhang et al., 2014; Zhao et al., 2014). Auctions have also been studied in spectrum allocation settings. In some of the work spectrum owners act as the auctioneer (Zheng et al., 2014; Khaledi and Abouzeid, 2013), hence truthfulness of the sellers is not considered. In some of the work (Zhou and Zheng, 2009; Dong et al., 2014; Yang et al., 2014), double auctions are considered and truthfulness of both the sellers and the buyers are guaranteed. However spectrums are considered to be identical, which ignores the differences between items. Researchers do consider heterogeneous double auction in spectrum settings recently (Feng et al., 2012). However it relies heavily on the spectrum conflict graph to filter out some of the bids. In (Feng et al., 2012), buyers are grouped into buyer groups using the conflict graph and only a single bid is considered by each group for winner determination. When applied to auctions in cloud settings its performance would decrease, due to the fact that cloud resource cannot be shared simultaneously.

Another drawback of previous work lies in theoretical proofs. Some of the work does not have theoretical proofs (Feng et al., 2012), hence there is no guarantee on the social welfare achieved by the auction. In some other work (Zheng et al., 2014) the competitive ratio depends on the conflict graph, which would not be satisfactory in cloud resource relocation setting. To the best of our knowledge, there is no known truthful auction that can achieve optimal social welfare. Hence, we use a different approach and prove that our auction is constant competitive to the (possibly untruthful) optimal auction under certain distributions. Though this is not a guarantee that our auction works in all cases, it is a good indicator that it will work in real applications.

Challenges. Most of previous work on double auction focuses on identical items. When each buyer has the same valuation for all sellers, one can simply try to assign a subset of sellers with smallest valuations to a subset of buyers with largest valuations (Deshmukh et al., 2002). The presence of non-identical items significantly complicates the design of auctions with required properties. The one-sided VCG auction returns a maximum matching and allocates items accordingly. On the other hand, a (possibly untruthful) optimal double auction would indeed find a maximum weight matching, where valuations of sellers are subtracted from the edge weights. Hence, one would expect to find some matching in a truthful double auction. However, it could be a challenging problem to analyze truthfulness (and budget balance) of auctions that make use of bids from both sides at the same time. To overcome this difficulty, we decompose the double auction into multiple steps and strictly control the information used in each step.

Our Approach. One typical approach in designing truthful auctions is to use VCG type of mechanisms that are originally devised in the context of one-sided auction. One useful gadget used in our approach is a modified version of VCG auction (Fu, 2013). In this setting there is a reserve price for each buyer, which is the minimum amount the buyer should pay if chosen. VCG auction with reserve price has also been used in (Khaledi and Abouzeid, 2013) to determine the payment. However, in their setting the spectrum owner acts as the auctioneer, whose truthfulness is not considered. On the contrary, in our setting the sellers participate as bidders whose truthfulness needs to be considered.

To design double auctions for non-identical items, we introduce a framework consisting of two phases. In the first phase, we identify a subset of "good" sellers, which we call candidates. We set prices for the candidates ensuring truthfulness for all the sellers. In the second phase, we use a truthful one-sided auction on the buyers whose needs are served by the candidates. To achieve budget balance, we adopt the aforementioned variant of VCG auction where the reserve price is guaranteed to be at least the prices for candidates.

With this framework we give our double auction. We use the idea of learning (Balcan et al., 2005). Particularly, we partition both sellers and buyers into two subsets A_1, A_2 and B_1, B_2. Then we use A_2, B_2 to "learn" a good reserve price r_2 and apply it to the auction between A_1 and B_1. Similarly we use A_1, B_1 to learn another reserve price r_1 and apply it to the auction between A_2 and B_2. Our evaluation results show that the auction have good performance in general.

Our Contribution. We summarize our contributions in this paper as follows.

- We consider double auctions in cloud resource allocation setting. We propose a framework for designing double auctions which decomposes the task into two phases.
- We design a double auction under the framework we propose. The auction is proved to satisfy truthfulness, individual rationality and budget balance.
- We prove the performance of Auction is constant competitive to the (possibly untruthful) optimal auction under certain probability distributions.
- We evaluate auction with probability distributions that correspond to scenarios in real applications. Our numerical results show that our auctions achieve good social welfare.

2 RELATED WORK

Truthful auction has been an interesting topic in game theory. One important result in the literature is VCG auction (Vickrey, 1961; Clarke, 1971; Groves, 1973). The auctioneer tries to sell a set of items to agents (buyers) who have valuations for different outcomes (all possible assignments). It is proved that VCG is truthful and maximizes the social welfare, which is the sum of valuations of buyers.

Maximizing the auctioneer's profit has also been considered. For example, in (Fiat et al., 2002) the authors designed an auction that achieves at least $\frac{1}{4}$ of the optimal profit. They also considered a different type of auction, cancellable auction, where the auctioneer cancels the auction if the profit is smaller than some predetermined value.

One-sided auctions have been extensively used for cloud provider to determine the allocation of resources. For example, (Zhang et al., 2014; Zaman and Grosu, 2011; Zhao et al., 2014) studied auctions for resource allocation problems in cloud computing settings.

Double auctions with identical items have also been considered. A third party (instead of the seller) acts as the auctioneer who, in a high level, trades with sellers and buyers based on bids from both sides. Under this setting, McAfee's mechanism (McAfee, 1992) is proved to be truthful, individual rational and budget balanced. In (Deshmukh et al., 2002) a truthful double auction was given that aims to maximize the auctioneer's profit.

Auctions have been considered for spectrum allocation. In (Zheng et al., 2014; Khaledi and Abouzeid, 2013), the spectrum owners act as the auctioneer and trade with the buyers. In (Zheng et al., 2014), the competitive ratio is shown to be $\Omega(\frac{1}{m \cdot \Delta_{max}})$, where Δ_{max} is the max degree of the conflict graph and m is the number of items. Double auctions have also been studied in spectrum allocation. In (Zhou and Zheng, 2009; Yang et al., 2014; Dong et al., 2014; Feng et al., 2012), double auctions are applied to spectrum allocation, where spectrum interference is taken into consideration to maximize spectrum utilization.

3 PRELIMINARIES

3.1 The Problem Model

We consider a sealed-bid double auction with non-identical items for cloud resource allocation. A *seller* is a cloud service provider with one unit of resource. A *buyer* is a user who is willing to buy one unit of resource. For example, in the above mentioned Amazon EC2 Reserved Instance Marketplace (Amazon, 2015a), customers with redundant cloud EC2 instances are willing to sell them for profit. They act as sellers in our settings. Due to increasing needs, other companies or individuals may want to expand their cloud resource capabilities. They act as buyer and trade with the sellers under the help of an auctioneer.

Let A be the set of sellers and let B be the set of buyers. Let $n := |A|$ and $m := |B|$ be the number of sellers and buyers, respectively. Each seller i has a valuation $v_i \geq 0$ for its resource, the lowest price at which it would still sell his resource. Each buyer j has a valuation $w_{i,j} \geq 0$ for the resource provided by seller i, the highest price at which it would still buy the resource from seller i. We note that the resources from the sellers are essentially non-identical items in the auction as $w_{i,j}$ and $w_{i',j}$ can be unequal when $i \neq i'$. This setting captures practical scenarios where each user has different preferences on the service providers. For instance, when a user is seeking a server to host a website, the valuation could depend on network latency, server capacity and etc.

We introduce some useful notations for double auction. We shall use *agent* to refer to either a seller or a buyer. Agents submit bids to an *auction*, where the bids may not be the same as the true valuations. In particular, a seller i submits $\widehat{v}_i$, and a buyer j submits $\widehat{w}_{i',j}$ for all sellers i'. Let $\widehat{v} := (\widehat{v}_i)_{i \in A}$ and $\widehat{w} := (\widehat{w}_{i,j})_{(i,j) \in A \times B}$. The auction decides (i) an allocation $x : A \times B \to \{0,1\}$, where $x_{i,j}$ indicates whether seller i is allocated to buyer j; and (ii) payment $p^s : A \to \mathbb{R}_+$ to sellers and payment $p^b : B \to \mathbb{R}_+$ from buyers. Observe that the allocation x, payment p^s and p^b are functions of $\hat{v}$ and $\hat{w}$. A valid allocation satisfies the following properties:

- $\sum_{j\in B} x_{i,j} \leq 1$ for all i;
- $\sum_{i\in A} x_{i,j} \leq 1$ for all j.

Essentially, the inequalities guarantee that an item is assigned to at most one buyer and a buyer can get at most one item. Define $x_i^s := \sum_{j\in B} x_{i,j}$ and $x_j^b := \sum_{i\in A} x_{i,j}$. A seller i is a *winner* if $x_i^s = 1$, and a buyer j is a winner if $x_j^b = 1$. We require that $p_i^s = 0$ if $x_i^s = 0$ and that $p_j^b = 0$ if $x_j^b = 0$; in other words, the payment is zero whenever the involved agent is not a winner. We call the amount $\sum_{j\in B} p_j^b - \sum_{i\in A} p_i^s$ the auctioneer's *profit*. The *social welfare* of the auction is defined as $S := \sum_{i\in A, j\in B} x_{i,j} \cdot w_{i,j} - \sum_{i\in A} x_i^s \cdot v_i$.

Let G be the complete bipartite graph between sellers A and buyers B, where edge $\{i, j\}$ between seller i and buyer j has a weight $w_{i,j} - v_i$. Then, an allocation can be represented by a matching in G. For a matching M, let $w(M)$ be the sum of weights of edges in M. We say an agent x is *covered* in M if there exists agent y such that $\{x, y\} \in M$. When the allocation is represented by M, the auction achieves social welfare $S = w(M)$.

The *utility* for seller i is defined as $u_i^s := p_i^s - x_i^s \cdot v_i$. The utility for buyer j is defined as $u_j^b := \sum_i x_{i,j} \cdot w_{i,j} - p_j^b$. Note that $x_i^s = 0$ implies $u_i^s = 0$ and that $x_i^b = 0$ implies $u_i^b = 0$. Let $u^s := (u_i^s)_{i\in A}$ and $u^b := (u_j^b)_{j\in B}$. Then u^s and u^b are functions of $\hat{v}$ and $\hat{w}$. Also note that in the double auction agents submit the bids to maximize their individual utilities.

We now define some economic properties that are investigated in this paper. For a vector $V := (V_1, \ldots, V_n)$, let $(V_{-i}, \hat{V}_i)$ be the vector obtained from V by replacing V_i with $\hat{V}_i$.

Definition 3.1 (Truthfulness)**.** *An auction is* truthful *if for any v and w, the utilities for agents are such that $u_i^s(v,w) \geq u_i^s((v_{-i}, \hat{v}_i), w)$ and $u_j^b(v,w) \geq u_j^b(v, (w_{-(i,j)}, \hat{w}_{i,j}))$ for all $i \in A$, $j \in B$, $\hat{v}_i$ and $\hat{w}_{i,j}$.*

Definition 3.2 (Individual Rationality)**.** *An auction achieves* individual rationality *if the utilities for agents are such that $u_i^s \geq 0$ for all $i \in A$ and $u_j^b \geq 0$ for all $j \in B$.*

Definition 3.3 (Budget Balance)**.** *An auction achieves* budget balance *if the auctioneer's profit is non-negative, i.e., $\sum_{j\in B} p_j^b - \sum_{i\in A} p_i^s \geq 0$.*

Our target is to design double auctions satisfying truthfulness, individual rationality and budget balance, with the goal to maximize the social welfare S. Since the auctions designed in this paper are truthful, we shall use v and w to denote the bids submitted by agents.

3.2 One-sided Auction

We describe the one-sided auction setting that is general enough for our use in this paper. There is a set Y of m agents and a set Z of non-identical items. There is also a collection $\mathcal{T}$ of possible *outcomes*, each of which specifies an allocation of the items among the agents. An agent $j \in Y$ has a valuation $w_j(T)$ for outcome $T \in \mathcal{T}$. Given bids from the agents, an auction determines an outcome T and payment p on the agents, where the charge for agent j is $p_j(T)$. We say an agent is a *winner* in an outcome T if it is allocated a positive number of items, and the set $W(T)$ of winners is the *winning set* of T. Each agent j aims to maximize its *utility*, which is $u_j(T) := w_j(T) - p_j(T)$ if $j \in W(T)$ and 0 otherwise. If an auction returns an outcome T, then it achieves *social welfare* $\sum_{j\in W(T)} w_j(T)$.

4 OUR SOLUTIONS

A natural idea to design a double auction for non-identical items would be to find some maximum matching to optimize the social welfare. Taking sellers' valuations into consideration, we would construct some maximum weight matching with edge weights w' where $w'_{i,j} = w_{i,j} - v_i$ for seller i and buyer j. However this procedure is not budget balanced. Indeed it is a challenging task to prove both budget balance and truthfulness when we make use of bids from both sellers and buyers at the same time (such as setting $w'_{i,j} = w_{i,j} - v_i$). To overcome this difficulty, we divide the interaction between sellers and buyers into several steps, while restricting the information used in each step. We introduce a framework consisting of two phases extending the approach for identical items (Dong et al., 2014).

(i) In the first phase, we construct a subset A_0 of sellers, called *candidates*, from which winners are finally selected. We set reserve prices p for agents in A_0 such that each seller in A_0 has non-negative utility and no seller has the motivation to be dishonest, even if some candidates are not finally selected. We specify a subset B_0 of buyers to serve each candidate (which can be different), guaranteeing that the subset serving a candidate does not influence its selection and reserve price.

(ii) In the second phase, we run a one-sided auction, which is truthful and individual rational, on B_0 to be served by the sellers A_0.

Note that only a subset of sellers are considered to be candidates. The intuition is that some seller can be too expensive to be allocated to any buyers. More importantly, the payment for the candidates should be no smaller than the bid of any candidate, and hence an expensive candidate would make it difficult to achieve budget balance. Also only a subset of buyers is used to serve a candidate since the information from other buyers may be used in its selection and reserve price. Throughout the procedure no seller has the motivation to be dishonest as the input to the one-sided auction is independent of the bids from A_0, and no buyer has the motivation to be dishonest as a buyer is either discarded (if not in B_0) or served by a truthful auction. Individual rationality and budget balance can also be easily checked.

4.1 VCG Auction with Reserve Price

Recall that the *VCG auction* (Vickrey, 1961; Clarke, 1971; Groves, 1973) is a one-sided auction that selects the outcome T maximizing the social welfare. We use $p_j(T)$ to denote its charge to the selected agents. The VCG auction with *reserve price* (VCG-RP) (Fu, 2013) sets a reserve price r_j for each agent, hence charging j at price $\max\{p_j(T), r_j\}$. Assuming that the sellers are always honest (i.e., sellers are not agents), then the double auction reduces to a one-sided auction, where all possible comes $\mathcal{T}$ is the matchings between A and B. We adopt VCG-RP to calculate the allocation as described below. We use M to denote the returned matching and use $p^b_{\{j\}}$ to denote the charges.

When we invoke VCG-RP we always guarantee the payments to the sellers $p^s_{\{i\}}$'s are the same and that we will use a higher reserve price r so that we can achieve budget balance.

Proposition 4.1 ((Fu, 2013)). *The VCG-RP auction is truthful.*

4.2 Our Double Auction

When the sellers are heterogeneous, a seller i may be more favorable than i' even if $v_i > v_{i'}$, when a large portion of buyers have higher valuations for i than for i'. In the auction, we exploit the valuations from both sellers and buyers to determine the candidates (from sellers).

We apply the idea of learning from (Balcan et al., 2005). We partition both sellers and buyers into two subsets, denoted by A_1, A_2 and B_1, B_2. Then we use A_1, B_1 to calculate an optimal matching M_1, and use A_2, B_2 to calculate an optimal matching M_2. Intuitively, if the partitioning is done randomly the two matchings should be close. Hence whatever learned from one side can be applied to the other in winner determination. We learn two thresholds p_1 and p_2 from

---

VCG-RP (A, B, w, r)

Input: Set A of sellers, set B of buyers with bids w, reserve price r

Remove all edges $\{i, j\}$ with $w_{i,j} < r$ from the complete bipartite graph between A and B. Let G be the resulting graph.

Compute a maximum weight matching M^* of G.

foreach $j \in B$ *covered in* M^* **do**

  Suppose $\{i, j\} \in M^*$. Let G' be the graph obtained from G with j and edges incident to j removed.

  Compute a maximum weight matching M' of G'.

  Set $p_j := \max\{w(M') - (w(M^*) - w_{i,j}), r\}$.

end

foreach $j \in B$ *not covered in* M^* **do**

  Set $p_j := 0$

end

Output: M^* and $(p_j)_{j \in B}$

M_1 and M_2. Concretely, this can be the average of all seller bids or the maximum of all seller bids. Then we use VCG-RP with reserve price p_1 between A_2 and B_2, ignoring sellers above p_1. Similarly, we use VCG-RP with reserve price p_2 between A_1 and B_1, ignoring sellers above p_2. Note that in determining p_1, p_2 we can use various policies, e.g., maximum and average. They both use similar ideas in guaranteeing truthfulness, individual rationality and budget-balance. We focus on using average in our proof for simplicity. In our simulations, we evaluate both variants.

We prove that this auction satisfies truthfulness, individual rationality and budget-balance. We also show that our auction is efficient on some distribution. Precisely, let $\mathcal{D}$ be the uniform distribution over $[a, b]$ and $\mathcal{D} + \Delta$ be the uniform distribution over $[a + \Delta, b + \Delta]$. We consider sampling seller side valuations from $\mathcal{D}$ and buyer side valuations from $\mathcal{D} + \Delta$. We prove our auction is constant competitive under this distribution for any $a, b, \Delta \geq 0$.

Theorem 4.1. *Auction is truthful, individual rational and budget balanced.*

The details of the proof are omitted due to space limit. The proof for individual rationality and budget balance is straight forward. For truthfulness, the idea is to show no participant can benefit from reporting a dishonest value.

Theorem 4.2 (Constant Approximation for Social Welfare). *Suppose the parameters $a, b, \Delta \geq 0$ are fixed. Moreover, there are n sellers whose valuations are sampled independently from $\mathcal{D} := [a, b]$ uniformly*

```
Auction (A, B, v, w)
Input: Sellers A = {1,...,n} with bids v,
       buyers B = {1,...,m} with bids w
Randomly partition A (B) into equally sized
subsets, denoted by A_1 and A_2 (B_1 and B_2).
foreach k ∈ {1,2} do
    Calculate a maximum weight matching M_k
    between A_{3−k} and B_{3−k}, using w_{i,j} − v_i as
    edge weights.
    r_k ← Avg(v_i : i ∈ M_k) OR max(v_i : i ∈ M_k)
    A'_k ← {i ∈ A_k : v_i ≤ r_k}.
    Run (M, p') ←VCG-RP (A'_k, B_k, w, r_k)
    Allocation:
    foreach i ∈ A_k and j ∈ B_k do
        Set x_{i,j} := 1 if {i, j} ∈ M and x_{i,j} := 0
        otherwise.
    end
    Payment:
    foreach i ∈ A_k do
        Set p^s_i := r_k if i is covered in M, and
        p^s_i := 0 otherwise.
    end
    foreach j ∈ B_k do
        Set p^b_j := p'_j.
    end
end
```

at random. Furthermore, there are $m \geq 2n$ buyers whose valuations for each seller are sampled independently from $\mathcal{D}+\Delta := [a+\Delta, b+\Delta]$ uniformly at random. Then, for all $0 < \varepsilon \leq \frac{1}{8}$, with probability at least $1-2n\exp(-\frac{1}{48}\varepsilon^2 n)$, Auction produces an assignment that achieves $\frac{1-4\varepsilon}{8}$ fraction of the optimal social welfare.

The proof is omitted due to space limit.

5 PERFORMANCE EVALUATION

In this section we simulate the proposed auctions. We compare our algorithms to TAHES from (Feng et al., 2012). Note that TAHES was originally designed for spectrum auction, and we adapt it to our setting by setting the conflict graph to be a complete graph. We generate instances with valuations of sellers and buyers taken from various probability distributions, and study the impact on the social welfare achieved by different auctions.

5.1 Simulation setup

Note that the design goal of our algorithm is to maximize the social welfare $S = \sum_{i,j} x_{i,j} \cdot w_{i,j} - \sum_i x_i^s \cdot v_i$. Since it is unclear what is the optimal truthful auction, we compare the social welfare to the optimal (possibly untruthful) auction, which will output the maximal weighted matching between sellers A and buyers B, weighted by $w'_{i,j} = w_{i,j} - v_i$. Then, the *performance ratio* of an auction $\mathcal{A}$ is defined to be $\mathrm{E}[S(\mathcal{A})]/S^*$, where S^* is the maximal matching and $S(\mathcal{A})$ is the social welfare by the auction.

We employ several distributions (for valuations of sellers and buyers) to simulate different situations in the real world. We use $[\alpha, \beta]$ to denote the (real number) interval from α to β. We use $[\alpha, \beta; \delta]$ to denote the values from α to β with step size δ.

(a) Uniform distribution. We test our auctions in a distribution similar to that in Theorem 4.2. Let $n := |A|$ and $m := |B|$ be the number of sellers and buyers respectively. There are four parameters α^s, β^s for sellers and α^b, β^b for buyers. For each seller $i \in A$, we sample v_i uniformly from $[\alpha^s, \beta^s]$. For each seller-buyer pair $\{i, j\} \in A \times B$, we sample $w_{i,j}$ uniformly from $[\alpha^b, \beta^b]$.

We choose different combinations of parameters to study their influences. The parameters n and m control the number of participants. We fix n, then adjust the number of buyers m to see the influence of the ratio of sellers and buyers. Note the two intervals $[\alpha^s, \beta^s]$ and $[\alpha^b, \beta^b]$ control how sellers' and buyers' valuations diverge. We fix the sellers' valuations $[\alpha^s, \beta^s]$ and change the buyers' valuation distributions to study how different distributions affects the performance.

(b) Geo-related distribution. In a real world scenario where companies rent cloud infrastructure to host websites as buyers, they typically choose a server location near their target clients. This is because the network latency is the most influential factor in such applications and a nearby server normally provides a lower latency network. We simulate our auction using Amazon EC2 as an example cloud provider considering the case when network latency is a critical issue.

We obtain the statistics on network latencies between cities from (Reinheimer and Roberts, 2015). Then we combine it with the locations of Amazon EC2 servers from (Amazon, 2015c) to estimate the network latency from Amazon servers. In order to simulate a real world sellers' and buyers' geographical distribution, we sample seller locations from Amazon availability zones and buyer locations from the checkin locations from Brightkite in (Amazon, 2015c). We assume that a buyer prefers service with lower latency and its valuation decreases linearly with the network latency. We use real Amazon selling price (using t2.large) as the sellers' bidding price. While for buyers' bidding price we use the formula $w_{i,j} = \max\{\Delta - \frac{1}{\alpha} \cdot d_{i,j}, 0\}$, where $d_{i,j}$ is the network latency from seller (server) i to buyer j.

5.2 Evaluation

In this section, we evaluate the performance of the auctions under uniform and geo-related distributions. We run both auctions on the distributions to see how the parameters affect the performance ratio. We consider two variants of Auction, namely Auction (max) and Auction (avg). The difference is Auction (max) uses the maximum of seller bids as the reserve price, whereas the other one uses the average of seller bids.

(1) Impact of valuations under uniform distribution. We fix the sellers' sampling interval to be $[0,100]$ and vary the buyers' sampling interval. In particular, for each $\beta^b \in [60,140;10]$, the valuation for each buyer is sampled uniformly from $[0,\beta^b]$. The performance ratio is shown in Figure 1(a), where the x-axis shows the choice of β^b.

Our auctions outperforms TAHES. Recall that $\beta^s = 100$, and thus when β^b reaches 100 the sellers and buyers have about the same valuations on average. When β^b goes above 100, the performance ratio of Auction (max) increases while that of Auction (avg) decreases slightly. The reason of this phenomenon is that Auction (avg) throws away many of the sellers, and it damages the performance when buyers' values are high enough. On the other hand, Auction (max) can dynamically adjust the number sellers to be thrown, resulting better performance.

(2) Impact of the number of agents under uniform distribution. We fix the number of sellers to be 30 and vary the number of buyers in $[20,90;10]$. Then, we sample both the sellers' and buyers' valuations uniformly from $[0,100]$. This setting corresponds to a market where there are more buyers than sellers each with a similar valuation for the resources to be traded. The performance ratio is shown in Figure 1(b).

The performances of both variant of Auction increases slightly with the number of buyers. Auction (max) has the highest performance ratio which is slightly over 0.8. All our auctions have a relatively high performance compared to TAHES. This is likely because we can take advantage of multiple bids from a single buyer compared to the TAHES auction.

(3) Impact of valuations under geo-related distribution. Next, we study the impact of valuations sampled under geo-related distribution on the performance ratio. Recall that the valuation of buyer j for seller i is calculated as $w_{i,j} = \max\{\Delta - \frac{1}{\alpha} \cdot d_{i,j}, 0\}$. We simulate a market with more buyers than sellers. Specifically, we fix the number of sellers and buyers to be 30 and 60 respectively and focus on the impact of the choice of Δ and α. Note that a typical Amazon EC2 price (using t2.large) would be 0.1 USD to 0.2 USD per hour, and network latency varies from 5 ms to 500 ms. Hence a reasonable choice of Δ would be from 0.2 to 0.4, and α from 100 to 1000.

First, we fix α to be 800 and investigate the effect of the choice of Δ. It can be seen that with a larger Δ, buyers tend to have relatively higher valuations. We vary the Δ in $[0.14,0.4;0.02]$. The performance ratio is shown in Figure 1(c). It can be seen that with small Δ, all auctions benefits from higher Δ values. After Δ reaches 0.2, Auction (avg) decreases slightly. However the ratio of Auction (max) keeps increasing and reaches 0.95. This is because Auction (max) can adjust its threshold automatically. Next we fix $\Delta = 0.2$. Note that the weigh of network latency is especially important in such circumstances. We vary the α value from $[100,1000;100]$. Note that α adjusts the weight of network latency. With a lower α the network latency becomes more important. The performance ratio is shown in Figure 1(d).

From the numerical result, we can see that with the increase of α the performance of all auctions increases. This is likely because with larger α the buyers have higher valuations, hence more edges can be matched in the auctions. Auction (avg) reaches its limit when the ratio reaches 0.7. However Auction (max) keeps increasing as α increases and remains the best auction.

6 CONCLUSION

In this paper we have developed techniques for the design of truthful double auctions with non-identical items for cloud resource allocation. We propose a framework in which an auction can "learn" a good subset of agents and appropriate payment policy. Under certain distribution, we prove theoretically that our auction achieves at least $\frac{1}{8}$ of the optimal social welfare. Then our auctions are evaluated on different probabilistic distributions. Particularly, we use data on network latencies between cities together with Amazon EC2 server locations to simulate real world situations where a buyer's valuations depend on its latency to the sellers (items). Our performance evaluation shows that the auctions achieve performance ratios at least 0.5 in most cases, and Auction (max) reaches a ratio of over 0.9 in certain cases.

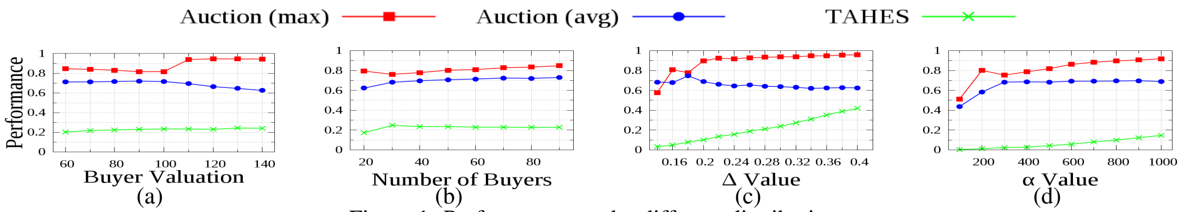

Figure 1: Performances under different distributions

REFERENCES

Amazon (2015a). Amazon EC2 reserved instance marketplace. http://aws.amazon.com/ec2/purchasing-options/reserved-instances/marketplace.

Amazon (2015b). Amazon EC2 reserved instances. http://aws.amazon.com/ec2/purchasing-options/reserved-instances.

Amazon (2015c). AWS global infrastructure. https://aws.amazon.com/about-aws/global-infrastructure.

Amazon (2015d). Elastic compute cloud (EC2) cloud server & hosting AWS. https://aws.amazon.com/ec2.

Balcan, M., Blum, A., Hartline, J. D., and Mansour, Y. (2005). Mechanism design via machine learning. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), 23-25 October 2005, Pittsburgh, PA, USA, Proceedings*, pages 605–614. IEEE Computer Society.

Clarke, E. H. (1971). Multipart pricing of public goods. *Public choice*, 11(1):17–33.

Deshmukh, K., Goldberg, A. V., Hartline, J. D., and Karlin, A. R. (2002). Truthful and competitive double auctions. In *Algorithms - ESA 2002, 10th Annual European Symposium, Rome, Italy, September 17-21, 2002, Proceedings*, pages 361–373.

Dong, W., Rallapalli, S., Qiu, L., Ramakrishnan, K. K., and Zhang, Y. (2014). Double auctions for dynamic spectrum allocation. In *2014 IEEE Conference on Computer Communications, INFOCOM 2014, Toronto, Canada, April 27 - May 2, 2014*, pages 709–717.

Feng, X., Chen, Y., Zhang, J., Zhang, Q., and Li, B. (2012). TAHES: truthful double auction for heterogeneous spectrums. In Greenberg, A. G. and Sohraby, K., editors, *Proceedings of the IEEE INFOCOM 2012, Orlando, FL, USA, March 25-30, 2012*, pages 3076–3080. IEEE.

Fiat, A., Goldberg, A. V., Hartline, J. D., and Karlin, A. R. (2002). Competitive generalized auctions. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 72–81.

Fu, H. (2013). VCG auctions with reserve prices: Lazy or eager. *EC, 2013 Proceedings ACM*.

Google (2015). Compute engine - google cloud platform. https://cloud.google.com/compute.

Groves, T. (1973). Incentives in teams. *Econometrica: Journal of the Econometric Society*, pages 617–631.

Khaledi, M. and Abouzeid, A. A. (2013). A reserve price auction for spectrum sharing with heterogeneous channels. In *22nd International Conference on Computer Communication and Networks, ICCCN 2013, Nassau, Bahamas, July 30 - Aug. 2, 2013*, pages 1–7. IEEE.

McAfee, R. (1992). A dominant strategy double auction. *Journal of Economic Theory*, 56(2):434 – 450.

Reinheimer, P. and Roberts, W. (2015). Global ping statistics. https://wondernetwork.com/pings.

UCX (2015). UCX expands trade of cloud (IaaS) on global exchange. http://ucxchange.com/ucx-expands-trade-of-cloud-iaas-on-global-exchange.

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37.

Yang, D., Zhang, X., and Xue, G. (2014). PROMISE: A framework for truthful and profit maximizing spectrum double auctions. In *2014 IEEE Conference on Computer Communications, INFOCOM 2014, Toronto, Canada, April 27 - May 2, 2014*, pages 109–117.

Zaman, S. and Grosu, D. (2011). Combinatorial auction-based dynamic VM provisioning and allocation in clouds. In *IEEE 3rd International Conference on Cloud Computing Technology and Science, CloudCom 2011, Athens, Greece, November 29 - December 1, 2011*, pages 107–114.

Zhang, L., Li, Z., and Wu, C. (2014). Dynamic resource provisioning in cloud computing: A randomized auction approach. In *2014 IEEE Conference on Computer Communications, INFOCOM 2014, Toronto, Canada, April 27 - May 2, 2014*, pages 433–441.

Zhao, J., Li, H., Wu, C., Li, Z., Zhang, Z., and Lau, F. C. M. (2014). Dynamic pricing and profit maximization for the cloud with geo-distributed data centers. In *2014 IEEE Conference on Computer Communications, INFOCOM 2014, Toronto, Canada, April 27 - May 2, 2014*, pages 118–126.

Zheng, Z., Wu, F., Tang, S., and Chen, G. (2014). Unknown combinatorial auction mechanisms for heterogeneous spectrum redistribution. In Wu, J., Cheng, X., Li, X., and Sarkar, S., editors, *The Fifteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc'14, Philadelphia, PA, USA, August 11-14, 2014*, pages 3–12. ACM.

Zhou, X. and Zheng, H. (2009). TRUST: A general framework for truthful double spectrum auctions. In *INFOCOM 2009. 28th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, 19-25 April 2009, Rio de Janeiro, Brazil*, pages 999–1007.