How Much Should You Pay for Information?: Technical Report

Ioannis Vetsikas, Madalina Croitoru

How Much Should You Pay for Information? –Technical Report–

Ioannis A. Vetsikas[1], Madalina Croitoru[2]
[1] National Center for Scientific Research "Demokritos", Greece
[2] University of Montpellier 2, France

Abstract

The amount of data available greatly increases every year and information can be quite valuable in the right hands. The existing mechanisms for selling goods, such as VCG, cannot handle sharable goods, such as information. To alleviate this limitation, in this paper, we study mechanisms for selling goods that can be shared or copied. We present and analyze theoretically and experimentally efficient incentive compatible mechanisms for selling a single sharable good to bidders who are happy to share it.

1 Introduction

The evolution of the Web, and thus the facility of sharing data and putting data online has greatly improved, at least in the last decade. The data deluge can be noted in many day to day use cases: electronic journals access, music sharing, videos, social networks, open data initiatives etc. In the knowledge representation community (in a broad sense, and mainly in the database community) it is implicitly assumed that every answer to a query will be simply allocated to the user (unless constrained due to privacy restrictions); not to mention that the multiplicity of knowledge requesters was simply regarded as a simple extension of the individual case. However, in today's Web (Web of Data, Web of Science, Web of Knowledge, Semantic Web, Web 2.0 etc) information being given freely clearly does not always hold in practical applications where the requesters are in direct competition for information. The bottom line is that data, seen as an allocatable good, has the property of high cost production but negligible cost to copy. Studying implications of pricing information and allocating it thus becomes highly timely [1, 2]. The pricing information question on the web has also been investigated from a Linked Data perspective where information markets are being created [3, 4, 5]. The authors acknowledge the need to study the implications of mechanism design in this setting.

A very important issue is that data can easily be shared (for example, music or software). If we were to apply well known auction mechanisms (e.g. VCG) to selling pieces of information that can be shared, hence are infinitely copied, no profit would be made; this would happen because competition is what drives the prices up [6] and offering more items - here infinite - than buyers essentially removes any competition in this setting. And yet getting profit is usually the first goal of a seller. Now, [7] examines expanding auctions where more copies of the same good are offered as the competition increases. However, their mechanism does not handle infinite copies and it is likely not *incentive compatible (IC)*, meaning that bidder would have an incentive to lie which would break down the mechanism. For this reason, it is important to propose IC mechanisms for this problem. Some mechanisms have been proposed in [8, 9] but they analyze the mechanisms from a worse case view point. As shown by [3] this is not always relevant in practice where the distributions are not always a-priori known. Taking inspirations from these algorithm in this paper we present and analyze several incentive compatible mechanisms for selling a single sharable good to bidders who are happy to share it, aiming at creating competition by restricting the number of winners.

The only related work the examines this problem is [9]. However, this work is only concerned with examining the performance of the proposed mechanisms in the worst case, meaning how poor the performance becomes for any input even if this poor performance only occurs for extremely unlikely input. For a practical application, what would interest a company or an individual selling the information is the expected revenue that can be obtained from each mechanism. In view of this, in this paper, we thoroughly study mechanisms for selling goods that can be shared or copied as many times as necessary. We further the analysis of incentive compatible mechanisms, characterize a whole family of such mechanisms that can be used and evaluate the revenue obtained and the efficiency of these mechanisms, comparing also against the mechanisms of [9], showing that our mechanisms are better in average performance in most cases.

2 Incentive Compatible Mechanisms

In this section, we present formally the setting that we will address in this paper. We then present several incentive compatible mechanisms, starting from two baseline ones (mechanism $\mathcal{M}^{k+1}$ and $\mathcal{M}^{r}$) and then characterizing a family of such mechanisms, which generalizes the basic mechanism $\mathcal{M}^{A}$.[1] We subsequently evaluate these mechanisms in Section 3.

We consider a set of n bidders want shared access to a piece of information. The valuations of the bidders who want shared access are $\overrightarrow{v} = \{v_1, \ldots, v_n\}$.

Now, the good could be also allocated to all the buyers but in this case we would not be able to extract any profit. For example, if we apply the well known VCG mechanism to this setting, the goods will be sold to all the buyers at a price equal to 0, thus making no money whatsoever! This essentially happens because offering more items than buyers essentially removes any competition which would drive the prices up. There is a couple of straight forward ways to deal with this issue:

1. **Restricting the number of winners**. If the number of winners is fixed to a number k which is less than the actual number n of interested buyers, then this immediately will create competition between the bidders. By running an $(m +$

[1] This is a very basic and straight forward mechanism and a variation of this has been presented in [9].

$1)^{th}$ price auction, henceforth denoted as mechanism $\mathcal{M}^{k+1}$, we can guarantee that this is IC and provides some profit while ensuring that the bidders will all pay the same price.

2. **Setting a reserve price**. Instead of fixing the number of winners, a reserve price r is set; any bidder with valuation higher than r will buy the good. Now, the only IC mechanism in this instance is to make every winner pay r, henceforth denoted as mechanism $\mathcal{M}^r$; otherwise if the price paid depends on their bids, these bidders (who have valuations higher than r) will bid $r + \epsilon, \epsilon > 0$ instead, as they know that any bid above r will guarantee that they win.

A serious shortcoming of both these mechanisms ($\mathcal{M}^{k+1}$ and $\mathcal{M}^r$) is that the number of winners and the reserve price, respectively, should be selected optimally beforehand in order to maximize the revenue of the seller. This would need to rely on information such as prior knowledge of the distribution of the valuations. It cannot depend on the actual bidders valuations as then the mechanism would not be IC.

To alleviate this shortcoming, we need to design a mechanism that chooses the number of winners and subsequently the price that they pay so as to maximize the total revenue of the seller. Essentially, we should maximize:

$$\max_{j \in \{1,\ldots,n\}} j v^{(j)}$$

where $v^{(j)}$ are the valuations of $\overrightarrow{v}$ ordered from highest to lowest. A first attempt would be to select j as to maximize:

$$\max_{j \in \{1,\ldots,n\}} j v^{(j+1)}$$

instead and the price paid by the winners would be equal to $v^{(j+1)}$, the top bid that did not win. This is essentially the main idea from both the Vickrey (i.e. second price) auction [10] and the VCG mechanism. However, in this case *this mechanism is not IC*. The bidders can manipulate the price that they pay, and whether they win or not, by submitting a bid which is not their true value thus changing the number of winners j. It is relatively easy to check that neither a winner nor a loser can gain by increasing her bid. However, as shown next, both a winner and the $(j + 1) - th$ bidder whose bid sets the price can gain by lowering their bids:

Example 1 *Assume valuations* $\overrightarrow{v} = \{11, 9, 8, 5, 3\}$. *When all bidders declare their true values, then the mechanism selects* $j = 2$ *winners and they both pay* $v^{(3)} = 8$.

- *Any of the two top bidders would be able to profit by lowering her bid to* $v'_1 = 3.5$. *So if the first one lies, then the ordered set of valuations would be {9,8,5,3.5,3} and then the mechanism would select* $j = 4$ *winners all paying* $v^{(5)} = 3$.

- *The third highest bidder can also profit by lowering her bid to* $v' = 3.5$, *because then the ordered set of valuations would be {11,9,5,3.5,3} and then the mechanism would again select* $j = 4$ *winners all paying a price of* $v^{(5)} = 3$.

Why does this happen? While the price paid by a winner does not depend on her bid, the number of winners does, and therefore it is possible to indirectly manipulate the price paid. In fact, this is the reason why an IC mechanism must essentially ignore the bid of a bidder i when deciding whether bidder i is a winner and the price that she pays. The following mechanism $\mathcal{M}^A$ satisfies this requirement:

Definition 1 (IC Revenue Maximizing Mechanism $\mathcal{M}^A$)
For each bidder $i \in \{1, \dots, n\}$ do:
If $i > 1$ and $v^{(i)} = v^{(i-1)}$ then
decision is same as bidder with valuation $v^{(i-1)}$,
Else
Compute j^* such that $j^* = \arg\max_j jv_{-i}^{(j)}$,[2]
where $\overrightarrow{v_{-i}}$ is $\overrightarrow{v}$ without the valuation $v^{(i)}$
If $v^{(i)} < v_{-i}^{(j^*)}$, bidder with value $v^{(i)}$ does not win
otherwise, she is a winner and pays $v_{-i}^{(j^*)}$

Theorem 1 *Mechanism $\mathcal{M}^A$ is IC.*

As a variation of this mechanism has been presented in [9] and it is easy to prove that it is IC we will not do so here. *What we will focus on are the properties of this mechanism as they have not been analyzed in previous work* and they will be useful both in the experimental analysis we will conduct, as well as in generalizing it to the family of mechanisms we will later present.

We give two examples of how this mechanism works, the second of which contains tied valuations:

Example 2 *Assume valuations $\overrightarrow{v} = \{11, 9, 7, 5, 3\}$.*

- *For bidder 1: $\overrightarrow{v_{-1}} = \{9, 7, 5, 3\}$, therefore $j^* = 3$ and since $v_1 = 11 \geq 5 = v_{-1}^{(3)}$ she wins with payment 5.*
- *For bidder 2: $\overrightarrow{v_{-2}} = \{11, 7, 5, 3\}$, therefore $j^* = 3$ and since $v_2 = 9 \geq 5 = v_{-2}^{(3)}$ she wins with payment 5.*
- *For bidder 3: $\overrightarrow{v_{-3}} = \{11, 9, 5, 3\}$, thus $j^* = 2$ and since $v_3 = 7 < 9 = v_{-3}^{(2)}$ she does not win.*
- *For bidder 4: $\overrightarrow{v_{-4}} = \{11, 9, 7, 3\}$, thus $j^* = 3$ and since $v_1 = 5 < 7 = v_{-4}^{(3)}$ she does not win.*
- *Bidder 5 (the bidder with the lowest valuation) can never win in this mechanism.*

Therefore the two bidders with the highest valuations will win at a payment equal to the fourth highest bid of 5.

[2] In case that $\arg\max$ is a set, we define it to return the maximum element, meaning that if $\exists j_1 < j_2$ where $j_1 v_{-i}^{(j_1)} = j_2 v_{-i}^{(j_2)}$ are the maximizing terms, then $\arg\max$ will return j_2 and similarly if there are more maximizing terms.

Example 3 *Assume valuations* $\overrightarrow{v} = \{10, 10, 7, 7, 7, 5\}$.

- *For bidders* 1 *(and* 2*):* $\overrightarrow{v_{-1}} = \{10, 7, 7, 7, 5\}$*, therefore* $j^* = 4$ *and since* $v_1 = 10 \geq 7 = v_{-1}^{(4)}$ *they win with payment* 7*.*
- *For bidder* 3 *(**as well as** 4 **and** 5):* $\overrightarrow{v_{-3}} = \{10, 10, 7, 7, 5\}$*, therefore* $j^* = 4$ *and since* $v_3 = 7 \geq 7 = v_{-3}^{(4)}$ *they win (payment* 7*).*
- *Bidder* 6 *does not win.*

Therefore the five bidders with the highest valuations will win at a payment equal to the fifth highest bid of 7*.*

Now, notice that had we been able to use all the information available the optimal allocation in Example 2 would have been to sell the good to the top three bidders at a price equal to 5. However, the third bidder is not a winner meaning that some efficiency has been lost. On the other hand, in Example 3, bidders 3, 4 and 5 pay their valuation. It makes sense to examine the properties of this mechanism, regarding which of the bidders win and the price that they pay.

We will use the following lemma.

Lemma 1 *If the* i^{th} *term is the maximizing one for the optimization problem solved by mechanism* $\mathcal{M}^A$*, when it disregards bidder* i*'s bid, then it is also the maximizing term when the bid of bidder* $i' = i + 1$ *is disregarded. Furthermore, if the maximizing term is instead the* j*-th one (where* $j \neq i$*), when disregarding bidder* i*'s bid, then either it remains the maximizing term, when disregarding bidder* i'*'s bid (*$i' = i + 1$*), or the new* i^{th} *term is the maximizing one.*

Proof *Note that the vectors of valuations* $\overrightarrow{v_{-i}}$ *and* $\overrightarrow{v_{-i'}}$*, when* $i' = i+1$*, differ only by a single element. More specifically, the* i^{th} *highest value in each vector differs, and this is equal to* $v^{(i')}$ *in vector* $\overrightarrow{v_{-i}}$*, while it is equal to* $v^{(i)}$ *in vector* $\overrightarrow{v_{-i'}}$*. When the mechanism solves the optimization problem for determining whether bidder* i *is a winner, it looks for the highest term among terms* $jv_{-i}^{(j)}$*. Now as* $\overrightarrow{v_{-i}}$ *and* $\overrightarrow{v_{-i'}}$ *differ by a single element this means that all the terms* $jv_{-i}^{(j)}$*, with exactly one exception, are exactly the same as the terms* $jv_{-i'}^{(j)}$*. The terms that are different in the two optimization problems are:* $iv^{(i')}$ *(when examining bidder* i*) and* $iv^{(i)}$ *(when examining bidder* $i' = i + 1$*). Since* $v^{(i')} \leq v^{(i)}$*, this means that if* $iv^{(i')}$ *is the highest term in the optimization problem for bidder* i *then term* $iv^{(i)}$ *is certain to be the highest term in the optimization problem for bidder* i'*. The opposite is not necessarily true. We want to stress that this holds only for that specific term tied to the* $i' - th$ *bid. If another term was the maximum then it does not mean that it will remain maximum when disregarding the next bidder.*

Proving the second part of the lemma is easy, as the only term which changes is the i*-th one. We know that the* j*-th term is larger than any other term, when disregarding bidder* i*'s bid, therefore, when disregarding bidder* i'*'s bid, there is a single term that might be higher and this is the* i*-th one. Meaning that either the* j*-th or the* i*-th term is the maximizing one.* ■

We obtain the following theorem:

Theorem 2 *When using mechanism $\mathcal{M}^A$ some number j' of the top bidders will win, where $j' \leq j^* + 1$ and $j^* = \arg\max_j jv^{(j+1)}$. The price that they all pay is equal to $v^{(j^*+1)}$. Furthermore $j' \leq j^*$, when there are no ties in the bidders' valuations.*

Proof *For the optimization problem solved when determining whether bidder 1 wins (disregarding her bid), we can easily check that the solution is to select j^* bidders as winners, because $\forall j : v_{-1}^{(j)} = v^{(j+1)}$. Therefore bidder 1 is selected to buy the good and pays $v^{(j^*+1)}$. For the remaining bidders, there are two cases:*

- *The optimization problems select the same solution (meaning j^*) when disregarding each bidder $i = 2, \ldots, j^*$. Now, we know from Lemma 1 that when bidder $j^* + 1$ is disregarded, the solution will be j^*. Therefore, j^* winners will be selected and they will pay $v^{(j^*+1)}$. There is an exception to this, which is when there is a tie, i.e. $v^{(j^*+1)} = v^{(j^*)}$, because then bidder $j^* + 1$ will also win. But bidder $j^* + 2$ will not be a winner as from Lemma 1 we know that the $(j^* + 1)^{th}$ term is the maximizing one. Unless, of course, $v^{(j^*+2)} = v^{(j^*+1)}$, which is impossible though. Assume that it is so: $v^{(j^*+2)} = v^{(j^*+1)}$. Then $(j^*+1)v^{(j^*+2)} > j^* v^{(j^*+1)}$, which means that j^* could not have been chosen as the solution when solving the optimization problem of finding $\arg\max_j jv^{(j+1)}$, which is a contradiction. Therefore, in the case of a tie and only then, there will be exactly $j^* + 1$ winners.*

- *There exists $j' \leq j^*$, such that the optimization problems select the same solution (meaning j^*) when disregarding each bidder $i = 2, \ldots, j' - 1$, however, when bidder j' is disregarded, then a different solution is selected. In that case, the maximizing term can only be the j'-th term (from the second part of Lemma 1), which means the j' is not selected as a winner. Therefore less than j^* bidders will win. The winners will pay $v^{(j^*+1)}$.*

To complete the proof we need to check that there cannot be a case where a bidder does not win and a bidder with lower valuation does win. This again follows easily from the second part of Lemma 1. Assuming that it can happen we must have some bidder j'' who does not win while bidder $j'' + 1$ does win. However, we know that for bidder $(j'' + 1)$'s optimization problem the solution must either be the same as that for j'''s (she cannon win then), or be the j''-th term (and again she cannon win). So this leads to a contradiction. ■

We can observe that this mechanism has some very desirable properties beyond being simply IC: firstly, all the winners pay the same price, so there can be no envy really among them, and, secondly, they pay the price $v^{(j^*+1)}$ that is the one that maximizes the profit of the seller. However, in order for the profit of the seller and the efficiency of the system to be maximized it should be that j^* (or $j^* + 1$) bidders should win at this price. To alleviate this weakness of mechanism $\mathcal{M}^A$, we will examine some variations of it, *eventually generalizing it to a whole family of mechanisms*. Essentially, the maximization step $j^* = \arg\max_j jv_{-i}^{(j)}$ will be replaced via use of a voting protocol.

First, notice that the desired optimization $\max_{j\in\{1,\dots,n\}} jv^{(j)}$ and the maximization step $j^* = \arg\max_j jv_{-i}^{(j)}$ of mechanism $\mathcal{M}^A$ are the same optimization problem when bidder i is the one with the highest valuation (her valuation is ranked first). In other cases though, they can lead to different results and this is the reason for the inefficiency. We cannot use any knowledge of the bidder's value when deciding when she's a winner or not, not even the rank (i.e. how many other bidders have a higher valuation). Therefore, we propose to examine all possible cases for the rank of the valuation of bidder i and then aggregate the "optimal" number of winners in each case via a voting protocol.

Example 4 *Assume valuations* $\overrightarrow{v} = \{11, 9, 7, 5, 3\}$.

- *For bidder* 1*, it is* $\overrightarrow{v_{-1}} = \{9, 7, 5, 3\}$*. For her valuation we can assume the following cases:*
  1. $v_1 \geq 9$*. The set of valuations is* $\{v_1, 9, 7, 5, 3\}$.
  2. $9 > v_1 \geq 7$*. The set of valuations is* $\{9, v_1, 7, 5, 3\}$.
  3. $7 > v_1 \geq 5$*. The set of valuations is* $\{9, 7, v_1, 5, 3\}$*. In all three of these cases it is optimal to have* 3 *winners.*
  4. $5 > v_1 \geq 3$*. The set of valuations is* $\{9, 7, 5, v_1, 3\}$*, and it is optimal to either have* 3 *(when* $v_1 \geq 4$*) or* 4 *(when* $v_1 < 4$*) winners. Only when* $v_1 < 4$ *would bidder* 1 *win.*
  5. $v_1 < 3$*. The set of valuations is* $\{9, 7, 5, 3, v_1\}$ *and bidder* 1 *would never be a winner.*
- *For bidder* 2*, we obtain similar results to the above.*
- *For bidder* 3*:* $\overrightarrow{v_{-3}} = \{11, 9, 5, 3\}$*. We examine the cases:*
  1. $v_3 \geq 11$*. The set of valuations is* $\{v_3, 11, 9, 5, 3\}$.
  2. $11 > v_3 \geq 9$*. The set of valuations is* $\{11, v_3, 9, 5, 3\}$*. In both these cases it is optimal to have* 2 *winners.*
  3. $9 > v_3 \geq 5$*. The set of valuations is* $\{11, 9, v_3, 5, 3\}$*, and it is optimal to have* 2 *(when* $v_3 \geq 7.5$*) or* 3 *(when* $v_3 < 7.5$*) winners. Only in the second subcase would bidder* 3 *win.*
  4. $5 > v_3 \geq 3$*. The set of valuations is* $\{11, 9, 5, v_3, 3\}$*, and it is optimal to have* 3 *(when* $v_3 \geq 4$*) or* 4 *(when* $v_3 < 4$*) winners. Only in the second subcase would bidder* 3 *win.*
  5. $v_3 < 3$*. The set of valuations is* $\{11, 9, 5, 3, v_3\}$*, and bidder* 3 *is never selected as a winner.*
- *For bidder* 4 *the same analysis yields that in no subcase would she win and the same is true for bidder* 5.

To illustrate what we mean we re-examine the setting of Example 2.

Example 5 *Assume valuations* $\overrightarrow{v} = \{11, 9, 7, 5, 3\}$.

- *For bidder* 1*, it is* $\overrightarrow{v_{-1}} = \{9, 7, 5, 3\}$*. For her valuation we can assume the following cases:*

  1. $v_1 \geq 9$*. Then the whole set of valuations would be* $\{v_1, 9, 7, 5, 3\}$*, and it would be optimal to have* 3 *winners.*
  2. $9 > v_1 \geq 7$*. Then the whole set of valuations would be* $\{9, v_1, 7, 5, 3\}$*, and it would be optimal to have* 3 *winners.*
  3. $7 > v_1 \geq 5$*. Then the whole set of valuations would be* $\{9, 7, v_1, 5, 3\}$*, and it would be optimal to have* 3 *winners.*
  4. $5 > v_1 \geq 3$*. Then the whole set of valuations would be* $\{9, 7, 5, v_1, 3\}$*, and it would be optimal to either have* 3 *(when* $v_1 \geq 4$*) or* 4 *(when* $v_1 < 4$*) winners. Only in the second subcase (when* $v_1 < 4$*) would bidder* 1 *win.*
  5. $v_1 < 3$*. Then the whole set of valuations would be* $\{9, 7, 5, 3, v_1\}$*, and it would be optimal to either have* 4 *(when* $v_1 \geq 2.5$*) or* 2 *(when* $v_1 < 2.5$*) winners. In either case bidder* 1 *would not be selected as a winner.*

- *For bidder* 2*, we obtain similar results the analysis for bidder* 1*.*

- *For bidder* 3*:* $\overrightarrow{v_{-3}} = \{11, 9, 5, 3\}$*. For her valuation we can assume the following cases:*

  1. $v_3 \geq 11$*. Then the whole set of valuations would be* $\{v_3, 11, 9, 5, 3\}$*, and it would be optimal to have* 2 *winners.*
  2. $11 > v_3 \geq 9$*. Then the whole set of valuations would be* $\{11, v_3, 9, 5, 3\}$*, and it would be optimal to have* 2 *winners.*
  3. $9 > v_3 \geq 5$*. Then the whole set of valuations would be* $\{11, 9, v_3, 5, 3\}$*, and it would be optimal to have* 2 *(when* $v_3 \geq 7.5$*) or* 3 *(when* $v_3 < 7.5$*) winners. Only in the second subcase would bidder* 3 *win.*
  4. $5 > v_3 \geq 3$*. Then the whole set of valuations would be* $\{11, 9, 5, v_3, 3\}$*, and it would be optimal to either have* 3 *(when* $v_3 \geq 4$*) or* 4 *(when* $v_3 < 4$*) winners. Only in the second subcase would bidder* 3 *win.*
  5. $v_3 < 3$*. Then the whole set of valuations would be* $\{11, 9, 5, 3, v_3\}$*, and it would be optimal to either have* 4 *(when* $v_1 \geq 2.5$*) or* 2 *(when* $v_1 < 2.5$*) winners. In either case bidder* 3 *would not be selected as a winner.*

- *For bidder* 4*:* $\overrightarrow{v_{-4}} = \{11, 9, 7, 3\}$*. For her valuation we can assume the following cases:*

  1. $v_4 \geq 11$*. Then the whole set of valuations would be* $\{v_4, 11, 9, 7, 3\}$*, and it would be optimal to have* 3 *winners.*
  2. $11 > v_4 \geq 9$*. Then the whole set of valuations would be* $\{11, v_4, 9, 7, 3\}$*, and it would be optimal to have* 3 *winners.*
  3. $9 > v_4 \geq 7$*. Then the whole set of valuations would be* $\{11, 9, v_4, 7, 3\}$*, and it would be optimal to have* 3 *winners.*

4. $7 > v_4 \geq 3$. *Then the whole set of valuations would be* $\{11, 9, 7, v_4, 3\}$*, and it would be optimal to either have* 3 *(when* $v_4 \geq 14/3$*) or* 2 *(when* $v_3 < 14/3$*) winners.*
5. $v_4 < 3$. *Then the whole set of valuations would be* $\{11, 9, 7, 3, v_4\}$*, and it would be optimal to either have* 4 *(when* $v_1 \geq 2.5$*) or* 2 *(when* $v_1 < 2.5$*) winners.*

- *Bidder* 5 *(the bidder with the lowest valuation) can never win (unless there is a tie which we choose not to consider when designing our mechanisms, as it happens with very low probability), therefore we do not analyze her case.*

What can we observe from this example? Examining what happens each time we tried to solve the optimization problem for each bidder, the optimal number of winners changes as the valuation of that bidder is assumed to various ranges of values; of course the knowledge of this value is ignored in order to keep the mechanism IC, this is the reason why we need to examine all these possible cases. Now, note that considering only the case when this value is assumed to be higher than the highest among the remaining valuations and basing the decision on only that case, gives mechanism $\mathcal{M}^A$. However, this does not use the information from all the other cases where the bidder examined might still be a winner. Thus, we propose to use a voting protocol where, for each case where the bidder examined is selected to be a winner, votes would be cast for the number of winners that maximize the total profit.

We see that the decision regarding each bidders depends on the cases examined. Thus, we generalize the previous mechanism to consider all cases examined. To this end, we propose the following family of mechanisms $\mathcal{M}^*$ where each case (i.e. when the rank of the missing valuation is k) casts votes with weight w_k:

Definition 2 (Family of IC Mechanisms $\mathcal{M}^*(\overrightarrow{w}, \delta)$)
Select the function $\delta(< profit >, < max\_profit >)$
For each bidder $i \in \{1, \ldots, n\}$ *do:*
  If $i > 1$ *and* $v^{(i)} = v^{(i-1)}$ *then*
    decision is same as bidder with valuation $v^{(i-1)}$*,*
  Else
    Set $\overrightarrow{v_{-i}}$ *as* $\overrightarrow{v}$ *without the valuation* $v^{(i)}$
    Set $\psi_k = 0, \forall k = 1, \ldots, n$
    For $k = 1, \ldots, n-1$ *do*
      Assume that the missing valuation (denoted v*) is*
        $v_{-i}^{(k-1)} > v \geq v_{-i}^{(k)}$*, where* $v_{-i}^{(0)} = \infty$ *&* $v_{-i}^{(n)} = 0$
      Set the weight w_k
      Define the terms $t_l = \begin{cases} (l-1)v_{-i}^{(l)}; l < k, \\ lv_{-i}^{(l)}; l \geq k. \end{cases}$
      Among these terms, find the highest: l_1
        and the second highest: l_2
      The min and max values of term $(k-1)v$ *are resp.:*
        $t_{min} = (k-1)v_{-i}^{(k)}$ *and* $t_{max} = (k-1)v_{-i}^{(k-1)}$
      If $t_{l_1} \geq t_{max}$ *and* $l_1 \geq k$ *then*

$\psi_{l_1} = \psi_{l_1} + w_k$ *(full vote for best - weighted)*
If $t_{l_2} \geq t_{max}$ *and* $l_1 \geq k$ *then*
$\psi_{l_1} = \psi_{l_1} + w_k \delta(t_{l_2}, t_{l_1})$ *(partial vote for* 2^{nd} *best)*
If $t_{max} > t_{l_1} \geq t_{min}$ *and* $l_1 \geq k$ *then*
$\psi_{l_1} = \psi_{l_1} + w_k \big(\frac{t_{l_1} - t_{min}}{t_{max} - t_{min}} + \int_{t_{l_1}}^{t_{max}} \frac{\delta(t_{l_1}, x)}{t_{max} - t_{min}} dx \big)$
If $t_{max} > t_{l_2} \geq t_{min}$ *and* $l_2 \geq k$ *then*
$\psi_{l_2} = \psi_{l_2} + w_k \frac{t_{l_2} - t_{min}}{t_{max} - t_{min}} \delta(t_{l_2}, t_{l_1})$
Select $j^* = \arg\max \psi_j$
If $v^{(i)} < v_{-i}^{(j^*)}$*, bidder with value* $v^{(i)}$ *does not win*
otherwise, she is a winner and pays $v_{-i}^{(j^*)}$

The two lines that are presented in bold define the parameters that characterize the whole range of mechanisms that belong to this family of mechanisms $\mathcal{M}^*$. For example, mechanism $\mathcal{M}^A$, which we presented earlier, is derived from $\mathcal{M}^*$, by setting $\delta() = 0, w_1 = 1$ and $w_k = 0, \forall k > 1$. In this paper we will also use in our experiments, the following two mechanisms which are derived from $\mathcal{M}^*$:

- mechanism $\mathcal{M}^V$, in which $\delta(x, y) = \frac{x}{y}$ and $w_k = 1, \forall k$
- mechanism $\mathcal{M}^W$, in which $\delta(x, y) = \frac{x}{y}$ and $w_1 = 1$, while $w_k = (n - 2)\frac{v_{-i}^{(k)} - v_{-i}^{(k-1)}}{v_{-i}^{(1)} - v_{-i}^{(n-1)}}, \forall k > 1$

In both these mechanisms, the best option gets 1 vote while the second best option (regarding the number of winners) gets votes equal to the ratio of the second highest and the highest profits. However, in the first mechanism, the weights for all cases are 1, while in the second the votes are weighted depending on how likely each case $v_{-i}^{(k-1)} > v \geq v_{-i}^{(k)}$ is, which depends on the distance between the values $v_{-i}^{(k-1)}$ and $\geq v_{-i}^{(k)}$.[3] Let us give a couple of examples:

Example 6 *Assume valuations* $\overrightarrow{v} = \{11, 9, 7, 5, 3\}$.

- *For bidder* 1*, examining all cases (*$k = 1 \dots 4$*) we get that having three winners gets the most votes (*$\psi_3 = 3$*). So the bidder wins and pays* $v_{-i}^{(3)} = 5$.
- *For bidder* 2*, we obtain the same results and price.*
- *For bidder* 3*:* $\overrightarrow{v_{-3}} = \{11, 9, 5, 3\}$ *and we examine cases:*
  1. $k = 1 : v_3 \geq 11$*. The valuations are* $\{v_3, 11, 9, 5, 3\}$*. The best choices are* $l_1 = 2$ *and* $l_2 = 3$*, thus* $\psi_2 = \psi_2 + 1$ *and* $\psi_3 = \psi_3 + \frac{5}{6}$.
  2. $k = 2 : 11 > v_3 \geq 9$*. Similarly, the mechanism updates* $\psi_2 = \psi_2 + 1$ *and* $\psi_3 = \psi_3 + \frac{5}{6}$.

[3]Given that we do not use any prior information regarding the distribution of valuations, this is the most logical way to assign probabilities to each case. Essentially, as we do not assume any knowledge of the distribution of valuations, we will approximate it as a uniform distribution with bounds the lowest and highest values that are known at each step of the algorithm.

3. $k = 3 : 9 > v_3 \geq 5$. *The valuations are* $\{11, 9, v_3, 5, 3\}$*, thus* $l_1 = 3$*,* $l_2 = 4$*,* $t_{max} = 18$ *and* $t_{min} = 10$*. Therefore* $\psi_3 = \psi_3 + 0.967$ *and* $\psi_4 = \psi_4 + \frac{1}{5}$*.*
4. $k = 4 : 5 > v_3 \geq 3$. *The valuations are* $\{11, 9, 5, v_3, 3\}$*, thus* $l_1 = 4$*,* $l_2 = 3$ *(but* $l_2 < k = 4$ *so votes are cast), while* $t_{max} = 15$ *and* $t_{min} = 9$*. Therefore* $\psi_4 = \psi_4 + 0.9463$*.*

The tally of votes is: $\psi_2 = 2$*,* $\psi_3 = 2.63$ *and* $\psi_4 = 1.15$*, hence* $j^* = 3$*. Hence bidder* 3 *wins and pays* $v_{-i}^{(j^*)} = 5$*.*

- *For bidder* 4*:* $\overrightarrow{v_{-4}} = \{11, 9, 7, 3\}$*. Following the same reasoning,* $j^* = 3$ *(*$\psi_3 = 3$*) and therefore she does not win.*

To summarize, the three bidders with the highest valuations (11*,* 9 *and* 7*) would win and each pays a price equal to* 5*.*

Example 7 *Assume valuations* $\overrightarrow{v} = \{11, 9, 7, 5, 3\}$*.*

- *For bidder* 1*, it is* $\overrightarrow{v_{-1}} = \{9, 7, 5, 3\}$*. The following cases are examined by the mechanism:*

  1. $k = 1 : v_1 \geq 9$*. Then valuations are* $\{v_1, 9, 7, 5, 3\}$*. Best is to have* $l_1 = 3$ *winners and second best to have* $l_2 = 2$*, so* $\psi_3 = \psi_3 + 1$ *and* $\psi_2 = \psi_2 + \frac{14}{15}$*.*
  2. $k = 2 : 9 > v_1 \geq 7$*. The valuations are* $\{9, v_1, 7, 5, 3\}$*. Again,* $\psi_3 = \psi_3 + 1$ *and* $\psi_2 = \psi_2 + \frac{14}{15}$*.*
  3. $k = 3 : 7 > v_1 \geq 5$*. The valuations are* $\{9, 7, v_1, 5, 3\}$*. Now, the best choice is* $l_1 = 3$ *and second best is* $l_2 = 4$*, however* $t_{max} = 14$ *and* $t_{min} = 10$*. Therefore,* $\psi_3 = \psi_3 + 1$ *and* $\psi_4 = \psi_4 + \frac{12-10}{14-10}\frac{14}{15} = \psi_4 + \frac{7}{15}$*.*
  4. $k = 4 : 5 > v_1 \geq 3$*. The valuations are* $\{9, 7, 5, v_1, 3\}$*.* $l_1 = 4$ *and thus* $\psi_4 = \psi_4 + \frac{1}{2} + 2(\ln 15 - \ln 12) = \psi_4 + 0.9463$*. On the other hand,* $l_2 = 2 < k = 4$*, therefore no votes are cast.*

  The tally is that $j^* = 3 = \arg\max \psi_j$ *has received most votes, as* $\psi_3 = 3$*. Therefore the bidder with valuation* 11 *wins and pays* $v_{-i}^{(j^*)} = 5$*.*

- *For bidder* 2*, after using the same process, we obtain the same results and price.*

- *For bidder* 3*:* $\overrightarrow{v_{-3}} = \{11, 9, 5, 3\}$*. The following cases are examined by the mechanism:*

  1. $k = 1 : v_3 \geq 11$*. The valuations are* $\{v_3, 11, 9, 5, 3\}$*. The best choices are* $l_1 = 2$ *and* $l_2 = 3$*, thus* $\psi_2 = \psi_2 + 1$ *and* $\psi_3 = \psi_3 + \frac{5}{6}$
  2. $k = 2 : 11 > v_3 \geq 9$*. The valuations are* $\{11, v_3, 9, 5, 3\}$*. Again* $l_1 = 2$ *and* $l_2 = 3$*, thus* $\psi_2 = \psi_2 + 1$ *and* $\psi_3 = \psi_3 + \frac{5}{6}$*.*
  3. $k = 3 : 9 > v_3 \geq 5$*. The valuations are* $\{11, 9, v_3, 5, 3\}$*. In this subcase,* $l_1 = 3$ *and* $l_2 = 4$*, while* $t_{max} = 18$ *and* $t_{min} = 10$*. Therefore* $\psi_3 = \psi_3 + \frac{5}{8} + \frac{15}{8}(\ln 18 - \ln 15) = \psi_3 + 0.967$ *and* $\psi_4 = \psi_4 + \frac{1}{5}$*.*

4. *$k = 4 : 5 > v_3 \geq 3$. The valuations are $\{11, 9, 5, v_3, 3\}$. In this subcase, $l_1 = 4$ and $l_2 = 3$ (but $l_2 < k = 4$ so votes are cast), while $t_{max} = 15$ and $t_{min} = 9$. Therefore $\psi_4 = \psi_4 + \frac{1}{2} + 2(\ln 15 - \ln 12) = \psi_4 + 0.9463$.*

The total tally of votes is: $\psi_2 = 2$, $\psi_3 = 2.63$ and $\psi_4 = 1.15$, hence $j^* = 3$. Therefore, the bidder with valuation 7 wins and pays $v_{-i}^{(j^*)} = 5$.

- *For bidder 4: $\overrightarrow{v_{-4}} = \{11, 9, 7, 3\}$.*
- *Following the same reasoning, we find that $j^* = 3$, with $\psi_3 = 3$ votes and therefore the bidder with valuation 5 does not win.*

To summarize, we find that under mechanism $\mathcal{M}^V$, the three bidders with the highest valuations (11, 9 and 7) would win and each pays a price equal to 5. However, this does not mean that mechanism $\mathcal{M}^V$ work always best; as we will see in the experimental evaluation, when the valuations can take only a couple of possible values then mechanism $\mathcal{M}^A$ is better!

No Agents		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
M^{k+1}	PR	50,98	80,98	101,45	130,08	151,95	179,78	202,38	229,90	253,11	279,98	303,60	330,25	353,91	380,60	404,35	430,86	454,91	481,00
	EF	126,04	140,96	201,44	215,75	277,02	290,89	352,38	366,17	428,10	441,51	503,53	516,90	578,95	592,24	654,45	667,66	729,90	743,01
M^R	PR	76,53	101,97	127,42	152,97	178,47	204,00	229,41	255,06	280,65	306,08	331,59	357,09	382,43	408,12	433,35	459,04	484,30	509,89
	EF	113,30	152,93	188,62	226,44	264,26	302,00	339,58	382,62	421,02	459,14	497,35	528,66	573,64	604,17	641,56	679,58	726,45	754,81
M^A	PR	57,81	81,51	105,81	130,34	154,72	179,36	203,96	228,79	253,53	278,37	303,12	327,94	352,84	378,01	402,78	427,87	452,62	477,78
	EF	92,54	124,87	160,07	195,82	231,47	267,38	303,31	339,75	376,06	412,44	449,06	485,54	522,18	559,23	595,79	632,76	669,56	706,52
M^V	PR	57,21	83,54	109,05	134,50	159,75	184,87	209,87	235,13	260,11	285,11	310,09	335,06	360,09	385,27	410,02	435,20	460,05	485,08
	EF	114,05	155,94	194,53	232,74	270,46	307,85	345,12	382,78	420,11	457,36	494,84	532,21	569,50	607,15	644,37	681,95	719,41	756,91
M^W	PR	57,21	83,48	109,91	135,54	160,81	185,99	211,05	236,28	261,27	286,20	311,15	336,10	361,06	386,18	410,89	435,99	460,84	485,81
	EF	114,05	158,31	196,53	234,76	272,50	310,03	347,34	385,00	422,37	459,60	497,05	534,41	571,67	609,27	646,44	683,99	721,39	758,82
M^P	PR	53,599	74,13	95,6032	118,02	141,12	164,48	188,79	213,61	238,09	262,87	287,13	311,43	336,55	361,8	386,59	411,55	435,85	461,7
	EF	98,477	128,76	159,13	190,18	222,28	254,92	319,48	349,3	379,57	410,63	441,89	474,62	538,46	568,07	598,27	629,67	661,45	725,99

Figure 1: Table of experimental results for 3 up to 20 bidders. The valuation distribution used is Uniform$\{1, \ldots, 100\}$. (PR stands for profit; EF for efficiency) The error of the simulation is no larger than the first decimal point of each of the results presented in all the tables.

No Agents		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
M^{k+1}	PR	5,50	7,61	11,00	13,80	16,51	20,19	23,14	26,47	30,11	33,01	36,93	40,45	43,62	47,74	51,00	54,88	58,68	61,84
	EF	8,88	16,62	18,03	18,87	27,33	28,34	29,00	37,86	38,62	39,14	48,26	48,87	57,92	58,59	59,06	68,32	68,85	78,04
M^R	PR	15,01	19,99	25,00	29,98	35,00	39,98	44,99	50,04	54,98	60,01	65,02	69,99	75,01	80,04	85,01	90,00	95,00	99,97
	EF	15,01	19,99	25,00	29,98	35,00	39,98	44,99	50,04	54,98	60,01	65,02	69,99	75,01	80,04	85,01	90,00	95,00	99,97
M^A	PR	12,00	17,98	23,75	29,23	34,57	39,73	44,85	49,96	54,50	59,73	64,86	69,90	74,95	80,01	84,99	89,99	94,99	99,96
	EF	15,38	20,24	25,16	30,08	35,06	40,01	45,00	50,05	55,04	60,04	65,04	70,01	75,01	80,04	85,01	90,00	95,00	99,97
M^V	PR	12,00	11,98	18,75	25,48	31,95	32,74	39,77	46,46	47,42	54,56	61,22	67,41	73,29	78,91	84,28	89,54	94,71	99,79
	EF	15,38	20,99	25,78	30,55	35,39	40,89	45,64	50,49	55,93	60,69	65,50	70,32	75,22	80,18	85,10	90,05	95,03	99,99
M^W	PR	12,00	11,98	18,75	25,48	31,95	32,74	39,77	46,46	47,42	54,56	61,22	67,41	73,29	78,91	84,28	89,54	94,71	99,79
	EF	15,38	20,99	25,78	30,55	35,39	40,89	45,64	50,49	55,93	60,69	65,50	70,32	75,22	80,18	85,10	90,05	95,03	99,99
M^P	PR	10,00	14,56	19,58	24,86	30,38	35,92	41,49	47,11	52,54	57,99	63,37	68,64	73,89	79,11	84,22	89,30	94,38	99,39
	EF	15,63	20,67	25,68	30,62	35,58	40,49	45,42	50,41	55,29	60,26	65,23	70,16	75,14	80,15	85,11	90,08	95,07	100,04

Figure 2: Table of experimental results for 3 up to 20 bidders. The values have a 50% chance of being 1 or 10.

No Agents		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
M^{k+1}	*PR*	130,25	187,98	244,16	299,43	354,31	408,74	462,92	517,07	571,22	625,21	679,23	733,17	787,27	841,28	895,35	949,47	1003,55	1057,77
	EF	167,37	247,33	326,51	405,04	483,45	561,61	639,61	717,53	795,34	873,23	950,93	1028,57	1106,27	1183,89	1261,53	1339,12	1416,77	1494,28
M^R	*PR*	165,00	220,00	275,00	330,00	385,00	440,00	495,00	550,00	605,00	660,00	715,00	770,00	825,00	880,00	935,00	990,00	1045,00	1100,00
	EF	232,49	309,99	387,55	464,93	542,50	620,00	697,47	774,99	852,47	930,07	1007,53	1084,97	1162,50	1239,97	1317,49	1394,97	1472,52	1549,95
M^A	*PR*	140,25	192,74	246,78	301,91	355,85	409,94	463,81	519,89	573,39	629,98	684,54	738,76	793,01	845,24	899,57	954,10	1011,17	1061,95
	EF	177,37	247,13	318,75	392,68	466,62	537,40	609,97	689,65	761,12	842,80	918,95	985,07	1056,20	1123,85	1199,09	1276,60	1360,26	1425,02
M^V	*PR*	140,25	200,35	258,77	314,98	370,00	424,73	478,98	534,90	587,89	644,25	698,55	752,93	807,55	861,08	915,97	971,40	1028,09	1080,89
	EF	177,37	259,71	341,12	419,44	495,61	571,06	645,52	725,21	795,58	877,05	952,39	1025,53	1098,69	1171,27	1248,27	1328,42	1411,62	1481,12
M^W	*PR*	140,25	200,35	258,77	315,93	370,68	424,97	480,12	535,52	588,63	644,64	699,14	754,66	808,30	862,24	916,43	971,94	1028,38	1081,10
	EF	177,37	259,71	341,12	421,21	496,89	571,51	647,76	726,52	797,06	877,80	953,62	1031,55	1100,47	1173,99	1249,17	1329,42	1412,21	1481,52
M^P	*PR*	123,68	174,77	228,86	284,93	342,07	399,95	457,69	515,53	573,21	630,60	687,65	744,53	801,39	857,74	914,30	970,38	1026,55	1082,67
	EF	152,56	218,63	289,91	364,93	442,30	521,49	601,42	681,79	762,32	842,92	923,24	1003,60	1083,96	1163,91	1244,13	1323,87	1403,83	1483,58

Figure 3: Table of experimental results for 3 up to 20 bidders. The values have a 10% chance of being $55, 60, 65, \ldots, 95, 100$.

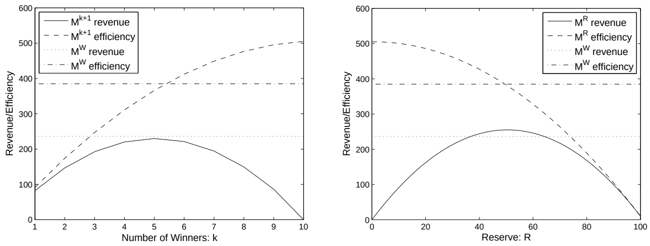

Figure 4: Revenue and efficiency of mechanisms $\mathcal{M}^{k+1}$ and $\mathcal{M}^R$ as the number of winners k and the reserve price R varies (versus those of $\mathcal{M}^W$). The number of bidder $n = 10$.

3 Experimental Evaluation

In this section, we conduct experiments to evaluate the performance of the mechanisms we presented in the previous section. Our goal is to compare the seller revenue and the efficiency (i.e. the sum of the valuations of all winners) of the different mechanisms $\mathcal{M}^A$, $\mathcal{M}^V$ and $\mathcal{M}^W$ as opposed to using the baseline mechanisms $\mathcal{M}^{k+1}$ and $\mathcal{M}^R$. We will also compare them with the Parameterized Random Sampling Optimal Price auction ($RSOP_r$) presented in Section 6.1 of [9]; we have implemented an improvement of this mechanism which does not allow one of the two randomized sets to be empty[4] and we assume knowledge of the distribution in order to select the optimal value of the parameter r for this mechanism. In the experiments we call this mechanism $\mathcal{M}^P$ (i.e. probabilistic). Note that in [9], another algorithm is proposed for this problem: the Random Sampling Profit Extraction auction (RSPE); however the performance of this mechanism is very poor as it sacrifices half the profit (in most cases), therefore we chose not to include it in our experiments.

Now, there is very little research and knowledge on what real distributions of the valuations for data are like. Some information on current data markets is given in [3], however very little is known about the real values for such data. Obviously, different distributions would affect to some degree the performance of the different mechanisms. In view of this, we present here three sets of experiments each performed with a different valuation distribution. We explain why we select each, in turn, before presenting the results of the simulation.

Experiment Set 1: We simulate n bidders whose valuations are i.i.d. random

[4] The original algorithm which did not impose this restriction had much worse performance, and it would always perform worse than the other algorithms, therefore it does not provide a good enough benchmark without this (minor) modification, as this significantly improves the mechanism's performance.

variables drawn from the uniform distribution on $\{1, \ldots, 100\}$. The number of bidders n varies from 3 to 20. For the baseline mechanisms $\mathcal{M}^{k+1}$ and $\mathcal{M}^R$, we calculate beforehand the best values for k and R respectively that maximize the expected revenue using the knowledge of the distribution from which the bids are drawn; *for the other mechanisms no such knowledge is necessary*.

The results of these experiments are presented in the table of Figure 1. The best revenue among the mechanisms $\mathcal{M}^A$, $\mathcal{M}^V$ and $\mathcal{M}^W$ (of family $\mathcal{M}^*$) is consistently obtained by mechanism $\mathcal{M}^W$. Its revenue is actually better than that of mechanism $\mathcal{M}^{k+1}$ (with k set optimally to maximize revenue), because mechanism $\mathcal{M}^W$ adjusts the number of winners based on the actual bids submitted rather than choosing the same number regardless of the input. On the other hand, mechanism $\mathcal{M}^R$ (with R set optimally) clearly outperforms the other mechanisms, because it uses its knowledge of the expected valuations to set a threshold (a reserve price) R which must be paid by all winning bidders. In this way it balances the revenue from each winner against the number of bidders. However, this mechanism is very dependent on knowing the distribution of valuations, as setting the reserve R to the wrong value will reduce very significantly the revenue obtained. In fact, to examine this effect we present in Figure 4, the revenue and efficiency of mechanisms $\mathcal{M}^{k+1}$ and $\mathcal{M}^R$ for different values of k and R respectively. We observe that the revenue obtained from $\mathcal{M}^R$ degrades significantly if R deviates by more than 15 from its optimal value. What is worse, if mistakenly the valuations were assumed to be between 1 and 50 (or, even worse, between 1 and 200), which would have approximately halved (or doubled resp.) the value for R, then very little revenue would be obtained!

Regarding the efficiency of the mechanisms, we observe that all of them perform similarly. The only exception is mechanism $\mathcal{M}^A$, because, as we've seen in Example 2, sometimes the mechanism reduces the number of winners which significantly impacts its efficiency.

Finally, we notice that mechanism $\mathcal{M}^P$ performs worst than any other mechanism (even $\mathcal{M}^{k+1}$ in many cases). This happens because the mechanism splits the problem into two separate problems and uses the solution (i.e. the best price) of one to impose the cutoff price for the other; however, the solutions for these two problems are not always very close (or identical which would be the optimal case) and this leads to a loss of revenue and some efficiency.

Experiment Set 2: In the previous experiment set, we assumed that the valuations could take a continuum of values. The extreme opposite of this is that only two values are possible, thus we assume here that the values have a 50% chance of being either 1 or 10. While we do not believe that this could be realistic (to have so small a number of possible values), this case is suggested in [9] as the case when the deterministic algorithms would fail to produce good results.

The results of these experiments are presented in the table of Figure 2. We observe that the performance of the baseline mechanisms is similar in broad terms to the previous experiment set. Regarding the other mechanisms, now mechanism $\mathcal{M}^A$ performs best, better than mechanisms $\mathcal{M}^V$ and $\mathcal{M}^W$. This is not entirely surprising as the possible values are only two, which means that in almost all cases the optimal decision would be to select the bidder with value 10 and make them pay 10, which matches the maximum profit that can ever be extracted from any (not necessarily IC) mechanism; the other more complex mechanisms $\mathcal{M}^V$ and $\mathcal{M}^W$ try to be cleverer, but that is unnecessary and they suffer a bit because of this.

Furthermore, mechanism $\mathcal{M}^P$ actually shines in this case, even if it is outperformed by mechanism $\mathcal{M}^A$: because of having only two possible values, and the solutions for the two problems solved by the mechanism are almost always identical, therefore this leads to almost maximal revenue and efficiency.

Experiment Set 3: We mentioned that the second distribution is probably not realistic for real data markets. However, this does pose the question what happens in an intermediate case, where there are relatively few possible values (but still not as few as only two). To this end, we assumed for the third experiment set we conducted, that the values have a 10% chance of being $55, 60, 65, \ldots, 95, 100$ (10 possible values in total).

The results of these experiments are presented in the table of Figure 3. In this case, we notice that the observations of the different mechanisms performance are close to those made for the first experiment set. In particular, disregarding the baseline mechanisms, mechanism $\mathcal{M}^W$ performs best in this set closely followed by mechanism $\mathcal{M}^V$. Mechanism $\mathcal{M}^P$ lacks in performance to a substantial degree (the exception being when the number of bidders n approaches 20) and so does mechanism $\mathcal{M}^A$.

To summarize our observations from all the experiment sets, we notice that the baseline mechanism $\mathcal{M}^R$ is overall consistently the best, but it relies significantly on selecting the best reserve R. Our proposed mechanisms (and in particular $\mathcal{M}^W$ and $\mathcal{M}^V$) are typically the best among the other mechanisms. The exception to this is when there are very few (two or close) possible valuations when it is more advantageous to use mechanisms $\mathcal{M}^A$ primarily and $\mathcal{M}^P$ secondary. However, we remind the reader that mechanism $\mathcal{M}^P$ also relies on using knowledge of the valuation distribution in order to select the optimal parameter r, albeit to a lesser extend than mechanism $\mathcal{M}^R$.

4 Conclusions

In this paper, we studied mechanisms for selling sharable information goods. We presented and analyzed several IC mechanisms, including a family of such mechanisms, for selling a single sharable good to bidders who are happy to share it; furthermore, we analyzed the properties of these mechanisms via simulations (for the most part).

There are still a number of avenues for future work. The most important extension is to examine whether we can generalize our mechanisms to the case where several goods are sold to bidder who want to buy bundles of these and are willing to share or would want each good exclusively. Furthermore, for the single unit case examined in this paper, the mechanisms of family $\mathcal{M}^*$ restrict the number of winners, in a similar manner to mechanism $\mathcal{M}^{k+1}$, the difference being not using prior knowledge; in this spirit, our second extension will examine new mechanisms that estimate a reserve price (like $\mathcal{M}^R$, the highest revenue mechanism we considered) without using prior knowledge about the valuation distribution.

References

[1] H. R. Varian, "Buying, sharing and renting information goods," *The Journal of Industrial Economics*, vol. 48, no. 4, pp. 473–488, 2000.

[2] F. Linde and W. Stock, *Information Markets: A Strategic Guideline for the I-Commerce*. Walter de Gruyter, 2011.

[3] S. F. Stahl, F. and G. Vossen, "Marketplaces for data: An initial survey," no. 14, 2012.

[4] P. Koutris, P. Upadhyaya, M. Balazinska, B. Howe, and D. Suciu, "Query-based data pricing," in *PODS*, M. Benedikt, M. Krötzsch, and M. Lenzerini, Eds. ACM, 2012, pp. 167–178.

[5] C. Li, D. Y. Li, G. Miklau, and D. Suciu, "A theory of pricing private data," in *ICDT*, W.-C. Tan, G. Guerrini, B. Catania, and A. Gounaris, Eds. ACM, 2013, pp. 33–44.

[6] V. Krishna, *Auction theory*. Academic Press, 2002.

[7] O. Shehory and E. Dror, "Computationally efficient and revenue optimized auctioneer's strategy for expanding auctions," in *Proceedings of AAMAS'06*, ser. AAMAS '06, 2006, pp. 1175–1182.

[8] M. Babaioff, R. Kleinberg, and R. Paes Leme, "Optimal mechanisms for selling information," in *Proceedings of the 13th ACM Conference on Electronic Commerce*, ser. EC '12. New York, NY, USA: ACM, 2012, pp. 92–109. [Online]. Available: http://doi.acm.org/10.1145/2229012.2229024

[9] A. V. Goldberg, J. D. Hartline, and A. Wright, "Competitive auctions and digital goods," in *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2001, pp. 735–744.

[10] W. Vickrey, "Counterspeculation, auctions, and competitive sealed tenders," *The Journal of Finance*, vol. 16, no. 1, pp. 8–37, 1961.