2014 IEEE Annual Symposium on Foundations of Computer Science

Mechanism Design for Crowdsourcing: An Optimal 1-1/e Competitive Budget-Feasible Mechanism for Large Markets

Nima Anari,
University of California, Berkeley.
anari@eecs.berkeley.edu

Gagan Goel,
Google NY.
gagangoel@google.com

Afshin Nikzad,
Stanford University.
nikzad@stanford.edu

***Abstract*—In this paper we consider a mechanism design problem in the context of large-scale crowdsourcing markets such as Amazon's Mechanical Turk (MTRK), ClickWorker (CLKWRKR), CrowdFlower (CRDFLWR). In these markets, there is a requester who wants to hire workers to accomplish some tasks. Each worker is assumed to give some utility to the requester on getting hired. Moreover each worker has a minimum cost that he wants to get paid for getting hired. This minimum cost is assumed to be private information of the workers. The question then is - if the requester has a limited budget, how to design a direct revelation mechanism that picks the right set of workers to hire in order to maximize the requester's utility?**

We note that although the previous work (Singer (2010); Chen et al. (2011)) has studied this problem, a crucial difference in which we deviate from earlier work is the notion of *large-scale* markets that we introduce in our model. Without the large market assumption, it is known that no mechanism can achieve a competitive ratio better than 0.414 and 0.5 for deterministic and randomized mechanisms respectively (while the best known deterministic and randomized mechanisms achieve an approximation ratio of 0.292 and 0.33 respectively). In this paper, we design a budget-feasible mechanism for large markets that achieves a competitive ratio of $1 - 1/e \simeq 0.63$. Our mechanism can be seen as a generalization of an alternate way to look at the *proportional share* mechanism, which is used in all the previous works so far on this problem. Interestingly, we can also show that our mechanism is optimal by showing that no truthful mechanism can achieve a factor better than $1 - 1/e$; thus, fully resolving this setting. Finally we consider the more general case of submodular utility functions and give new and improved mechanisms for the case when the market is large.

***Keywords*-Crowdsourcing, Truthful Mechanisms, Large Markets, Budget-feasibility.**

I. INTRODUCTION

Crowdsourcing is a recent phenomenon that is used to describe the procurement of a large number of workers to do certain tasks. These tasks can be of a variety of natures, and - to give a few examples - include image annotation, data labeling for machine learning systems, consumer surveys, rating search engine results, spam detection, product reviews, etc. There are several platforms (such as Amazon's Mechanical Turk (MTRK)) that facilitate and automate various steps involved in setting up and executing crowdsourcing tasks.

A key challenge in these online labor markets is to be able to properly price the tasks. Since the requester (the one who wants to procure workers) is usually budget constrained, pricing the tasks too high can result in lower output for the requester. On the other hand, pricing the tasks too low can disincentivize workers to work on the tasks. This makes pricing a non-trivial step for the requester when setting up a crowdsourcing task. One idea - to make pricing more automated and to prevent economic loss from poor pricing - is to design a direct revelation mechanism that solicits bids from workers to report their cost of participation, and based on this decide which workers to hire and how much to pay them.

A simple model that captures the above problem is as follows: There is a set S of workers. Worker i has a private cost c_i and provides utility u_i to the requester on getting hired. We want to design a truthful mechanism that decides which workers to recruit and how much to pay them. The goal is to maximize the requester's utility without violating her budget constraint.

For the above model, Singer (2010) gave an incentive-compatible mechanism that achieves an approximation ratio [1] of $1/6$ compared to the *offline optimum* that knows the costs of the workers. Later on Chen et al. (2011) improved the approximation ratio to $\frac{1}{2+\sqrt{(2)}} \simeq 0.292$ (and to $1/3$ for randomized mechanisms). Chen et. al. also showed that no deterministic mechanism can achieve an approximation ratio better than $\frac{1}{1+\sqrt{(2)}} \simeq 0.414$; Singer (2010) showed that no randomized mechanism can achieve an approximation ratio better than 0.5.

Our work is motivated by the following observation:

[1]we use the terms approximation factor and competitive ratio interchangeably

DOI 10.1109/FOCS.2014.36

Most of the crowdsourcing tasks are *large-scale* in nature in terms of the number of workers involved. On the other hand if one looks at the impossibility result of Chen et al. (2011), they involve only a small number of workers (specifically, only 3 workers). Thus, this leads to a natural open question - *Do these lower bounds extend to the case of large markets? or can one design better mechanisms for this important case of large markets?*

In this paper, we seek to understand the above question. We show that one can significantly improve the approximation ratio for the case of large markets. We give a mechanism that achieves an approximation ratio of $1 - 1/e \simeq 0.63$ for large markets. In addition, we show that our mechanism is the best possible mechanism by showing that no truthful budget-feasible mechanism can achieve a factor better than $1 - 1/e$. Finally, we look at the more general case of submodular utility functions.

A. The Model

We define the model abstractly: Consider a reverse auction scenario with one buyer and n sellers, where the set of sellers is denoted by S. Each seller $i \in S$ owns a single item (denoted by item i) and has a *private* cost c_i for it. The buyer derives a utility of u_i from item i. The buyer has a limited budget B, and its goal is to buy a subset of items that maximizes her utility without exceeding her budget.

Note that if the sellers are not strategic and the costs are known to the buyer, then this is the well-known *knapsack* optimization problem. However, the cost c_i is assumed to be a private information of seller i. Thus we are interested in designing direct-revelation mechanisms where the buyer solicits bids from the sellers, and then computes which sellers to buy from and how much to pay them. More formally, a mechanism $\mathcal{M}$ consists of two functions $A : (\mathbb{R}_+)^n \to \{0,1\}^n$ and $P : (\mathbb{R}_+)^n \to (\mathbb{R}_+)^n$. The allocation function $A(\cdot)$ takes as input the costs of n sellers and reports the set of winners. The payment function $P(\cdot)$ takes as input the costs of n sellers and reports how much is paid to each seller. We will use functions $A_i : (\mathbb{R}_+)^n \to \{0,1\}$ and $P_i : (\mathbb{R}_+)^n \to \mathbb{R}_+$, for each $i \in S$, to refer to the restriction of functions $A(\cdot)$ and $P(\cdot)$ to seller i. In other words, A_i, P_i represent the i-th element of the output of functions A, P, respectively.

The mechanism $\mathcal{M} = (A, P)$ should satisfy the following properties:

1) Budget Feasibility: The sum of the payments made to the sellers should not exceed B, i.e., $\sum_i P_i(\mathbf{c}) \leq B$ for all $\mathbf{c}$.
2) Individual rationality: A winner $i \in S$ is paid at least c_i.
3) Truthfulness/Incentive-Compatibility: Reporting the true cost should be a dominant strategy of the sellers, i.e. for all non-truthful reports $\bar{c}_i$ from seller i, it holds that

$$P_i(\bar{c}_i, \mathbf{c_{-i}}) - c_i \cdot A_i(\bar{c}_i, \mathbf{c_{-i}}) \leq$$
$$P_i(c_i, \mathbf{c_{-i}}) - c_i \cdot A_i(c_i, \mathbf{c_{-i}})$$

Among all mechanisms that satisfy the above properties, we are interested in the ones that give high utility to the buyer. Note that no mechanism can achieve utility larger than $U^*(\mathbf{c}, \mathbf{u})$, where $U^*(\mathbf{c}, \mathbf{u})$ is the utility of the knapsack optimization problem assuming costs of the sellers are known to the buyer. When there is no risk of confusion, we also denote $U^*(\mathbf{c}, \mathbf{u})$ by U^* for brevity.

We say a mechanism $\mathcal{M}$ is an α-approximation (for $\alpha \leq 1$) if it gives utility at least $\alpha \cdot U^*(\mathbf{c}, \mathbf{u})$ for any $\mathbf{c}$ and $\mathbf{u}$.

Indivisible vs Divisible Items. Note that the above description is given for indivisible items, however, we can define the above problem for divisible items as well. For instance, if the item being sold by a seller is his own time, then it can be modeled as a divisible item. For fraction $x \leq 1$ of a divisible item, the cost of seller i is $x.c_i$ and the utility obtained by the buyer is $x \cdot u_i$. The allocation function for divisible items is defined as $A : (\mathbb{R}_+)^n \to [0,1]^n$.

More general utility functions. An interesting generalization of the above model is when the utility function over the set of items is a submodular function rather than an additive function. We denote this function by $U : 2^S \to \mathbb{R}_+$ (for additive functions, $U(T) = \sum_{i \in T} u_i$, for $\forall T \subseteq S$). We assume that the utility function is known to the buyer.

1) The Large Market Assumption: Crowd-sourcing systems are excellent examples of *large markets*. Informally speaking, a market is said to be large if the number of participants are large enough that no single person can affect the market outcome significantly. Our results take advantage of this nature of crowdsourcing markets to give better mechanisms.

We define the **large market assumption** as follows: We assume that in our model, the cost of a single item is very small compared to the buyer's budget B. More formally, let $c_{\mathsf{max}} = \max_{i \in S}\{c_i\}$. Then, the large market assumption is defined as below.

The Large Market Assumption: $c_{\mathsf{max}} \ll B$.

In other words, we define the *largeness ratio* of the market to be $\theta = \frac{c_{\max}}{B}$ and analyze our mechanisms for when $\theta \to 0$.

This assumption - also known as the *small bid to budget ratio assumption* - is used in other large-market problems as well (for instance, see Mehta et al. (2007) for a similar definition with application in online advertising). All the mechanisms that we present in the main body of the paper (mechanisms for additive utility functions) will be analyzed under this assumption. The mechanisms that we design for submodular utility functions work under a different large market assumption which is explained below.

An Alternative Assumption: We also suggest another definition for large markets, the discussion of which will be deferred to the full version of this paper. Our mechanisms for submodular utility functions work under this assumption; moreover, we can slightly modify our mechanisms for additive utility functions so that they work under this assumption as well, while preserving their approximation ratio. We define this assumption below.

Let $u_{\max} = \max_{i \in S} u_i$ and U^* be the utility of the optimum solution (i.e. the maximum utility that the buyer can achieve when the costs are known to her). This large market assumption states that:

An Alternative Large Market Assumption: $u_{\max} \ll U^*$.

In other words, we define the largeness ratio of the market to be $\theta = \max_{i \in S} \frac{u_i}{U^*}$ and analyze our mechanisms for when $\theta \to 0$.

We note that our impossibility result for additive utilities (Section VI) holds for either of the two definitions.

B. Our Results

In this paper, we design optimal budget-feasible mechanisms for *large markets*. To the best of our knowledge, we are the first ones to study the case of large markets. We list our results below:

1) If the items are divisible, we design a deterministic mechanism which satisfies all the required properties (i.e. budget feasibility, individual rationality and truthfulness) and has an approximation ratio of $1 - 1/e$ (Section V). Note that previously, no mechanism was known for the case of divisible items. In fact, one can show that no bounded approximation ratio is possible for divisible items if the large market assumption is dismissed.
2) If the items are indivisible, we can modify our mechanism and give a randomized truthful mechanism for this case which achieves approximation ratio $1 - 1/e$. The proof is deferred to the full version of this paper.
3) In Section VI, we show that the above results are optimal by proving that no truthful (and possibly) randomized mechanism can achieve approximation ratio better than $1 - 1/e$. Our impossibility result holds for both cases of divisible and indivisible items.
4) For the case of submodular utility functions, we design deterministic mechanisms that achieve approximation ratios of $\frac{1}{2}$ and $\frac{1}{3}$ with exponential and polynomial running times respectively. Note that we only consider the case of indivisible items for submodular utility functions. The discussion of our results on submodular functions is deferred to the full version of this paper.

As we saw in Section I-A1, one could define a notion of θ-large market, i.e. a market with largeness ratio θ. To gain a better understating of the problem, we focus on large markets (i.e. when $\theta \to 0$) and state our main theorems for this setting. However, our mechanisms do not need "very large" markets to perform well; for instance, in the knapsack problem with additive utilities, the approximation ratio [2] is $(1 - 1/e) \cdot (1 - 6\theta/5)$ when all the items have equal utilities (Section V-B). Thus, say for $\theta = 1/20$ and $\theta = 1/40$ (which are reasonable assumptions in many settings) we get approximation factors 0.592 and 0.613 respectively.

Also we point out that the above results have applications beyond crowdsourcing - for instance, see Singer (2011) for application in marketing over social networks, and Horel et al. (2013) for application in experiment design. Singer (2011) provides a truthful mechanism with approximation ratio ≈ 0.032 and Horel et al. (2013) provides an approximately truthful mechanism with approximation ratio ≈ 0.077. For both these settings, large market assumption is a very reasonable assumption to make; thus, our results apply to these applications as well. In particular, our results give fully truthful mechanisms for these applications with approximation ratios $\frac{1}{2}, \frac{1}{3}$ (for exponential and polynomial running time respectively) in large markets.

C. Related Work

The most relevant related work is that of Singer (2010) and Chen et al. (2011). Singer (2010) first introduced this model (without the large market assumption). For the case of additive utilities and indivisible items, he

[2] we didn't try to optimize the dependence on θ in our analysis as we focus on the main ideas for the sake of better understanding.

gave a deterministic mechanism with an approximation ratio of $1/6$. Chen et al. (2011) later improved it to $1/(2+\sqrt{2})$, and also gave a randomized mechanism with an approximation ratio of $1/3$. They gave a lower bound of $1/(1+\sqrt{2})$ and $1/2$ for deterministic and randomized mechanisms respectively. For the case of submodular utilities, Singer (2010) gave a randomized mechanism with an approximation ratio of $1/112$ which was improved to $1/7.91$ by Chen et al. (2011). Chen et al. (2011) also gave an exponential time deterministic mechanism for submodular utility functions with an approximation ratio of $1/8.34$.

Dobzinski et al. (2011) looked at the more general sub-additive utility functions and gave a $1/log^2(n)$ and $1/log^3(n)$ approximation ratio for randomized and deterministic mechanisms respectively. Singla and Krause (2013a) design budget-feasible mechanisms for adaptive submodular functions with applications in community sensing.

In another work, Bei et al. (2012) study this problem in Bayesian setting. Singer (2011) looks at the application of this model in marketing over social networks. Horel et al. (2013) study the application of this model in experiment design.

Another related model that has been inspired from crowdsourcing applications is when the workers arrive online. A sequence of papers model this as an online learning problem. See Singla and Krause (2013b); Badanidiyuru et al. (2012); Singer and Mittal (2013) for more details.

Finally, we note that our assumption for large markets is similar to the assumption made in other application areas; notably in the Adwords problem as studied by Mehta et al. (2007). See Goel and Mehta (2008); Devanur and Hayes (2009); Feldman et al. (2010, 2009) for other models motivated by online advertising where they make similar assumptions.

D. Roadmap

The readers are encouraged to read this section before proceeding further. We begin by presententing a $\frac{1}{2}$-approximate mechanism (in section II) using a simple *proportional share mechanism* which was proposed in Singer (2010); Chen et al. (2011). Then, in section III, we build our new ideas that help us find the optimal mechanism for the case of the additive utilities. This section also defines a simple version of our mechanism that is not truthful but is envy-free. We define our truthful mechansim in section IV. Both our envy-free and truthful mechanisms are parameterized by a single input variable (which we call *standard allocation rule*). Most of the work in the later sections goes in finding (and prooving) the optimal standard allocation rule. In particular, in section V, we find the standard allocation rule for which our mechanism provides an approximation ratio of $1 - 1/e$ in large markets. In section VI, we complement this result by showing that no truthful mechanism can achieve approximation ratio better than $1 - 1/e$. In the full version of this paper, we further adapt our mechanism to the case of indivisible items, and present two mechanisms for submodular utility functions which have exponential and polynomial running times and approximation ratios $\frac{1}{2}$ and $\frac{1}{3}$, respectively.

II. A Simple $\frac{1}{2}$-Approximate Truthful Mechanism

In this section, we briefly explain the previous mechanism designed for this problem for the additive utility functions that gives an approximation factor of $\frac{1}{2}$ in large markets.

Definition 1: Cost-per-utility rate of a seller i is equal to c_i/u_i.

A natural approach to this problem tries to find a single *payment-per-utility rate* (denoted by rate r) at which all the winning sellers get paid. In other words, this approach picks a single number r and makes a payment of $r.u_i$ to seller i if she wins and pays her 0 otherwise. For brevity, we sometimes call the payment-per-utility rate r simply the *rate* r when there is no risk of confusion.

Individual rationality implies that a seller i is willing to sell her item at rate r iff $r \geq c_i/u_i$. Initially the buyer declares a very large rate r, and then sees which sellers are willing to sell at this rate. If the total cost to buy from all these sellers at rate r is higher than the budget B, then the buyer decreases the rate r. More formally, a natural descending price auction for this problem works as follows:

1) Let A denote the set of active sellers, and initially set $A = S$.
2) Start with a very high rate r.
3) Verify if all the active sellers can be paid with rate r, i.e. whether $\sum_{i \in A} r.u_i \leq B$ or not.
4) If the payment is feasible, then allocate the subset A, make the payment and stop.
5) If the payment is not feasible then decrease r slightly; update A accordingly by removing the sellers i for whom $c_i/u_i > r$; go to Step 3.

The above auction captures the main idea behind the *proportional share* mechanisms designed in Singer

(2010); Chen et al. (2011)[3], although they describe it in a forward auction format. It is not hard to see that the above mechanism is truthful, budget-feasible, and in large markets achieves an approximation ratio of $\frac{1}{2}$ (with small modifications, this can be converted to a randomized $\frac{1}{2}$-approximation for arbitrary markets as well Chen et al. (2011)).

III. Our Approach

In this section, we give a high level overview of our approach. Subsections III-A and III-B are preliminary sections and must be read before proceeding further. Also for rest of the paper we will assume that the sellers' items are divisible, unless we explicitly talk about indivisible items.

A. A Notion of An Allocation Rule

To build our new ideas, we first introduce and formalize the notion of an allocation rule.

An allocation rule $f : \mathbb{R}^+ \to [0, 1]$ is a function which determines how much to buy from a given seller. The domain of allocation rules is the *cost per utility rate*; meaning, given a (u_i, c_i) pair of seller i, the allocation rule f says that we should buy $f(\frac{c_i}{u_i})$ fraction of seller i's item. We do not enforce using the same allocation rule for all sellers.

We say an allocation rule $f : \mathbb{R}^+ \to [0, 1]$ is a ***Standard Allocation Rule*** if f is a decreasing function such that $f(0) = 1$ and $f(e - 1) = 0$.[4]

For any standard allocation rule $f : \mathbb{R}_+ \to [0, 1]$, we can define an associated family of allocation rules

$$\mathcal{F}(f) = \{f_r : \mathbb{R}_+ \to [0, 1]\}_{r>0}$$

where f_r denotes an allocation rule which is same as f except that it is stretched along the horizontal axis with ratio r, i.e. $f_r(x) = f(x/r)$ for all $x \geq 0$.

As we will see later, any single standard allocation rule f and its corresponding family of allocation rules $\mathcal{F}(f)$ will uniquely specify our mechanism. At a high level, our mechanism will work as follows: we will pick the largest positive r such that $f_r \in \mathcal{F}(f)$ is budget-feasible; meaning the sum of the payments with allocation rule f_r does not exceed B. However, note that we have not yet defined a payment rule given an allocation rule f_r - we define it next.

[3] It is worth pointing out that for submodular utilities, they need to use an additional trick: constructing a (sorted) list of sellers in a greedy manner before running the auction.

[4] The choice of $e - 1$ is just for simplifying the future calculations; it can be replaced with any other constant.

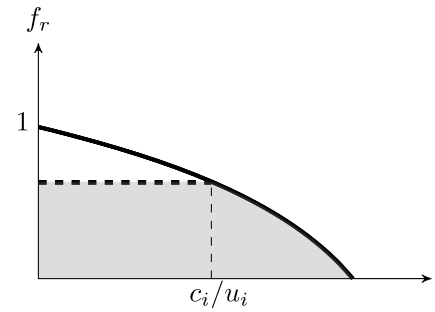

Figure 1: With allocation rule f_r, the payment to seller i is defined to be u_i times the shaded area under the curve.

B. Payment Rule

Recall that given a function $f_r \in \mathcal{F}(f)$ and (u_i, c_i) pair for a seller i, the value of $f_r(\frac{c_i}{u_i})$ only tells us what fraction of seller i's item to buy. But how much should we pay seller i in order to give incentives to seller i to report its cost truthfully to the mechanism? We compute these payments based on the well known Myerson's characterization of the truthful mechanisms [Myerson (1981)].

Let the payment rule for seller i is denoted by $P_{i,r} : \mathbb{R}_+ \to \mathbb{R}_+$ for an allocation rule f_r. Here $P_{i,r}$ maps the reported cost of seller i into its payment.

To define $P_{i,r}$, we do the following thought process: Let's divide seller i's item into u_i different pieces. Note that now the seller's cost for each piece is $\frac{c_i}{u_i}$. Thus function f_r can now be seen as mapping the cost of a single piece into the fraction of that piece that we will buy. Let $Q_r(x) : \mathbb{R}_+ \to \mathbb{R}_+$ denote the function that maps the cost for a single piece into a payment for that piece. Now, Myerson's characterization [Myerson (1981)] says that the payment for each piece is given by the following formula:

$$Q_r(x) = x \cdot f_r(x) + \int_x^\infty f_r(y)\, \mathrm{d}\, y.$$

Intuitively, $Q_r(x)$ represents the area under the curve as seen in Figure 1. Going forward, we will call the function Q_r a *unit-payment rule*. Note that $P_{i,r}(x)$ and Q_r are related by the following formula:

$$P_{i,r}(x) = u_i \cdot Q_r(x/u_i),$$

Thus, to summarize, for an allocation rule f_r, we buy $f_r(c_i/u_i)$ units of her item, and pay her $P_{i,r}(c_i)$ amount of money.

Remark: We make a remark that the above payment rule is truthful only if the allocation rule f_r that is offered to seller i does not depend on the private information (cost c_i in this case) of the seller i. If the allocation rule f_r does depend on the private information of the seller, then the mechanism may or may not be truthful. In the next section we give a mechanism in which the allocation rule f_r for a seller i depends on its reported costs. Later in section IV, we give a mechanism where the allocation rule f_r for a seller i doesn't depend on its reported cost, thus our payment rule will ensure that the resulting mechanism is truthful.

C. First Attempt: A Parameterized Class of Envy-Free Mechanisms

In this section we describe a mechanism (denoted by Mechanism Envy-Free(f)) that is not always truthful, but it will form the basis of our truthful mechanism. Moreover, some structural results about this mechanism will be useful while analyzing our truthful mecahanism, thus we will be talking about this mechanism throughout the paper. This mechanism is parameterized by the choice of a standard allocation rule f. The mechanism described in this section *offers* a single allocation rule $f_r \in \mathcal{F}(f)$ to all the sellers, thus it is envy-free (although it may not be truthful).

Definition 2: We say that an allocation rule f_r is a *budget-feasible* allocation rule if $\sum_{i \in S} P_{i,r}(c_i) = B$, i.e. the payments defined with respect to f_r sum up to B.

Now given any standard allocation rule f, the mechanism starts with a very large scaling ratio $r = \infty$ so that we are guaranteed to have $\sum_{i \in S} P_{i,r}(c_i) > B$.

Then, the mechanism decreases r until the rule f_r becomes a *budget-feasible* rule (say at $r = r^\star$). The mechanism stops at this point and uses $f_{r^\star}$ and $\{P_{i,r^\star}\}_{i \in S}$ to determine the allocations and payments. The ratio $r^\star$ is also called the *stopping rate* of the mechanism. We define this process formally in Mechanism Envy-Free(f).

Mechanism Envy-Free: Parameterized by a standard allocation rule f

input : Budget B, (u_i, c_i) pair for each seller i
output: A scaling ratio $r^\star$

$r \leftarrow \infty$;
while f_r *is not a budget-feasible rule* **do**
  Decrease r slightly;
end
$r^\star \leftarrow r$;
Output the scaling ratio $r^\star$;

One can easily see that the above mechanism is budget-feasible, individually rational, and envy-free; however, it may not be truthful. Also, the efficiency of the above mechanism depends on the choice of function f. Thus, an important question is: What is the optimal choice of function f? Let's first understand the performance of the above mechanism for a simple choice of function f.

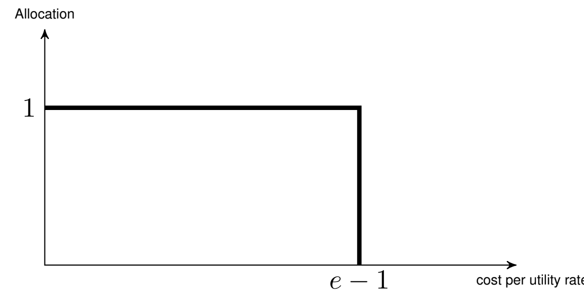

Figure 2: The Uniform Standard Allocation Rule

Definition 3: A standard allocation rule $f : \mathbb{R}^+ \rightarrow [0,1]$ is called a uniform standard allocation rule if $f(x) = 1$ for $x < e - 1$, and $f(x) = 0$ otherwise. Figure 2 depicts this curve.

One can show that the above envy-free mechanism when run using a uniform standard allocation rule, mimics the simple factor $\frac{1}{2}$ mechanism presented earlier. Thus, it turns out to be truthful as well for this choice of standard allocation rule. However for more general allocation rules, the above envy-free mechanism might not be truthful. Thus before we answer the harder question about the optimal choice of function f, we next describe the truthful version of the above envy-free mechanism.

IV. A Parameterized Class of Truthful Mechanisms

We use a simple trick to convert Mechanism Envy-Free(f) to a truthful mechanism. The idea is to define, for each seller i, an allocation rule which does not depend on c_i. In particular, we define the allocation rule for seller i to be f_{r_i}, where r_i will be chosen independently of c_i. For finding r_i, we run Mechanism Envy-Free(f) on the instance which is obtained by setting c_i to be 0 while keeping cost of the other sellers intact; r_i would be the stopping rate of the mechanism Envy-Free(f). The formal definition of the truthful mechanism appears in Mechanism Truthful(f).

In Lemma 2, we prove that Mechanism Truthful(f) is individually rational, truthful, and budget-feasible for any given standard allocation rule f. First, we state the following useful lemma.

Lemma 1: For any seller $i \in S$ we have $r^\star \geq r_i$.

Mechanism Truthful(f): Parameterized by a standard allocation rule f

input : Budget B, (u_i, c_i) pair for each seller i

foreach $i \in S$ **do**
  $temp \leftarrow c_i$;
  $c_i \leftarrow 0$;
  $r_i \leftarrow$ Mechanism Envy-Free(f);
  $c_i \leftarrow temp$;
end
foreach $i \in S$ **do**
  Allocate $f_{r_i}(c_i)$ from seller i;
  Pay $P_{i,r_i}(c_i)$ to seller i;
end

Proof: The proof is based on the fact that $P_{i,r}(x)$ is an increasing function of r (for a fixed x) and is a decreasing function of x (for a fixed r). The proof is by contradiction, suppose $r^\star < r_i$. Let $c'_j = c_j$ for all $j \in S \backslash \{i\}$ and let $c'_i = 0$. Observe that

$$B = \sum_{j \in S} P_{j,r^\star}(c_j) \leq \sum_{j \in S} P_{j,r^\star}(c'_j) < \sum_{j \in S} P_{j,r_i}(c'_j),$$

where the first inequality is due to the fact that $P_{j,r^\star}(x)$ is a decreasing function of x and the second inequality is due to the fact that $r^\star < r_i$. However, note that the above inequalities imply that $B < \sum_{j \in S} P_{j,r_i}(c'_j)$, which contradicts with the budget feasibility of Mechanism Envy-Free(f): see that $\sum_{j \in S} P_{j,r_i}(c'_j)$ represents the payment of Envy-Free(f) when the costs are $c'_1, \ldots, c'_n$, and so it can not be larger than B. ■

Lemma 2: Mechanism Truthful(f) is individually rational, truthful, and budget-feasible.

Proof: Note that the allocation and payment rules for seller i, i.e. f_{r_i}, P_{i,r_i}, do not depend on the cost reported by her. This fact, along with the fact that f_{r_i} is a monotone rule (decreasing function) implies individual rationality and truthfulness. The proof is almost identical to the proof of Myerson's Lemma and we do not repeat it here.

The proof for budget feasibility needs a bit more work. Let p_i, p'_i denote the payments to seller i respectively in Mechanism Truthful(f) and Mechanism Envy-Free(f), i.e. $p_i = P_{i,r_i}(c_i)$ and $p'_i = P_{i,r^\star}(c_i)$. The lemma is proved if we show that $p_i \leq p'_i$, since we have $\sum_{i \in S} p'_i = B$.

To see $p_i \leq p'_i$, note that $P_{i,r}(x)$ is an increasing function of r (for a fixed x). So, since we have $r^\star \geq r_i$ due to Lemma 1, it must be the case that $P_{i,r_i}(c_i) \leq P_{i,r^\star}(c_i)$. ■

V. A $(1-1/e)$-Approximate Optimal Truthful Mechanism

So far, we have introduced a parameterized class of individually rational, truthful, and budget-feasible mechanisms for the problem: Passing any standard allocation rule f to Mechanism Truthful(f) fixes the mechanism which we denote by $\mathcal{M}_f$. Our goal in this section is to find the *most efficient* mechanism in this class. Formally, given a standard allocation rule f, we denote the approximation ratio of $\mathcal{M}_f$ by $\mathcal{R}_f$ and define it as:

$$\mathcal{R}_f = \inf_I \frac{\mathcal{U}_f(I)}{\mathcal{U}^\star(I)},$$

where the infimum is taken over all instances I of the problem [5] . Here $\mathcal{U}_f(I)$ denotes the utility obtained by $\mathcal{M}_f$ in instance I, and $\mathcal{U}^\star(I)$ denotes the optimum utility in instance I.

The most efficient allocation rule f, is the one which maximizes $\mathcal{R}_f$. Our goal, in this section and Section VI, is to find the most efficient allocation rule and its corresponding approximation ratio. Formally, we prove the following theorem.

Theorem 1: Under the large market assumption, and with additive utilities, the most efficient standard allocation rule for Mechanism Truthful(f) is $f(x) = \ln(e-x)$, for which we get $\mathcal{R}_f = 1 - 1/e$, i.e. it has an approximation ratio $1 - 1/e$.

We prove this theorem in two parts: In the first part we show that $\mathcal{R}_f \geq 1 - 1/e$ for $f(x) = \ln(e-x)$; this is proved in the current section. In the second part, we show that $\mathcal{R}_g \leq 1 - 1/e$ for any standard allocation rule g. This fact can be seen as a consequence of our hardness result in Section VI, which states that no truthful mechanism can achieve approximation ratio better than $1 - 1/e$. We also provide a more direct (alternative) proof in the full version of this paper that shows our choice of $f(x) = \ln(e-x)$ is optimal among all possible choices of the standard allocation rules.

A. *Finding an optimal f for the (non-truthful) Mechanism Envy-Free(f)*

In this section, we prove that Mechanism Envy-Free(f) has approximation ratio $1 - 1/e$ for $f(x) = \ln(e-x)$. Note that the Mechanism Envy-Free(f) is not truthful, however its analysis will be helpful when

[5]If we are focused on large markets, we take only instances I for which the largeness ratio is smaller than some threshold, and take the limiting approximation factor as the threshold goes to 0.

analyzing our truthful mechanism in Section V-B. Here, we analyze Mechanism Envy-Free(f) assuming that the true costs are known; later, in Section V-B, we use this result to prove that Mechanism Truthful(f) has approximation ratio $1-1/e$ for the same choice of f.

1) Preliminaries: We use g_r to denote the inverse of an allocation rule f_r, i.e. $g_r(x) = f_r^{-1}(x)$. Given an allocation rule f_r, we also write an alternative definition of its corresponding unit-payment rule Q_r. This definition, rather than being in terms of $\frac{c_i}{u_i}$, would be in terms of $f_r(\frac{c_i}{u_i})$. This alternative definition is denoted by G_r, and is defined such that $Q_r(\frac{c_i}{u_i}) = G_r(f(\frac{c_i}{u_i}))$. For instance, if a seller owns an item with utility 1, then we pay her $G_r(x)$ when a fraction x of her item is allocated. To be more precise, for $y = f_r(\frac{c_i}{u_i})$ we define

$$G_r(y) = \int_0^y g_r(x)\,\mathrm{d}\,x = Q_r(\frac{c_i}{u_i}).$$

We also denote g_1 and G_1 respectively by g and G.

Proposition 1: Given the standard allocation rule $f(x) = \ln(e-x)$, it is straight-forward to verify that $g(x) = e - e^x$ and $G(x) = ex - e^x + 1$. Also, $f_r(x) = \ln\left(\frac{er-x}{r}\right)$.

From now on in this section, we assume that $f(x) = \ln(e-x)$. Next, we prove a useful inequality in the following lemma which will be used in the analysis of Envy-Free(f).

Lemma 3: For any x, α such that $0 \le x, \alpha \le 1$ we have

$$G(x) - \alpha \cdot g(x) \le e \cdot (x - \alpha \cdot (1 - 1/e)).$$

Proof:

$$\begin{aligned}
&\alpha(e^x - 1) \le e^x - 1 \\
\Rightarrow\ &\alpha e^x - e^x + 1 \le \alpha \\
\Rightarrow\ &\alpha e^x - e^x + 1 + e(x-\alpha) \le \alpha + e(x-\alpha) \\
\Rightarrow\ &G(x) - \alpha \cdot g(x) \le e \cdot (x - \alpha \cdot (1-1/e)).
\end{aligned}$$

The last line follows from the definition of g and G. ■

2) Approximation Ratio of Mechanism Envy-Free(f): In the following lemma, we prove the efficiency of Mechanism Envy-Free(f) when all sellers report true costs.

Lemma 4: If sellers report true costs, then Mechanism Envy-Free(f) has approximation ratio $1 - 1/e$.

Proof: Observe that w.l.o.g. we can assume $r^\star = 1$: If $r^\star \neq 1$, then we can construct a new instance which is *similar* to the original instance and has stopping rate 1. More precisely, there exists some $\beta > 0$ such that if we multiply the budget and the reported costs by β, the stopping rate becomes equal to 1. Note that this operation will not change the optimal solution or the solution of Envy-Free(f) and can be performed w.l.o.g.

Now, suppose that a fraction x_i of item i is allocated by Envy-Free(f). Since $r^\star = 1$, we can use Lemma 3 to write the following inequality for each $i \in S$:

$$G(x_i) - \alpha_i \cdot g(x_i) \le e \cdot (x_i - \alpha_i \cdot (1 - 1/e)),$$

where α_i is the fraction that is allocated from seller i in the optimal solution (recall that we are are comparing Envy-Free(f) with the optimum fractional solution). The above inequality can be multiplied by u_i on both sides and be written as:

$$u_i \cdot (G(x_i) - \alpha_i \cdot g(x_i)) \le$$
$$u_i \cdot e \cdot (x_i - \alpha_i \cdot (1 - 1/e)).$$

By adding up these inequalities, we get:

$$\sum_{i \in S} u_i \cdot (G(x_i) - \alpha_i \cdot g(x_i)) \le$$
$$e \cdot \sum_{i \in S} u_i \cdot (x_i - \alpha_i \cdot (1 - 1/e))\,. \qquad (1)$$

Now, we show that if

$$0 \le \sum_{i \in S} u_i \cdot (G(x_i) - \alpha_i \cdot g(x_i))\,, \qquad (2)$$

then the lemma is proved using (1) and (2). First we show why (1) and (2) prove the lemma, and then in the end, we prove (2) itself.

Observe that (1) and (2) imply that

$$0 \le \sum_{i \in S} u_i \cdot (x_i - \alpha_i \cdot (1 - 1/e))\,. \qquad (3)$$

Now, let U denote the utility gained by Envy-Free(f) and $U^\star = \sum_{i \in S} u_i \alpha_i$ denote the utility of the optimum (fractional) solution; see that (3) implies

$$(1 - 1/e) \cdot U^\star = \sum_{i \in S} \alpha_i u_i \cdot (1 - 1/e) \le \sum_{i \in S} x_i u_i = U,$$

This would prove the lemma.

So, it only remains to show that (2) holds: First observe that $\sum_{i \in S} u_i \cdot G(x_i) = B$, since the sum represents the payment of Envy-Free(f). Also, see that $\sum_{i \in S} \alpha_i u_i \cdot g(x_i) \le B$, since this sum is a lower bound on the cost of the optimal solution, which is at most B. ■

B. Special case: Analyzing our truthful mechanism for unit utilities

In this section, we prove that Mechanism Truthful(f) has approximation ratio $1 - 1/e$ in large markets when it uses the standard allocation rule $f(x) = ln(e - x)$ for the special case when all the utilities are equal to 1.

In other words, we will show that approximation ratio approaches $1-1/e$ as θ, the market's largeness ratio, approaches 0 for the case of unit utilities. The proof for the case of general utilities is intricate and appears in the full version of the paper.

Note that the assumption of unit utilities imply $c_i/u_i = c_i$ for any seller i. Next, we state two lemmas before proving the approximation ratio. For simplicity in the analysis, w.l.o.g., assume that $c_1 \leq c_2 \leq \ldots \leq c_n$.

Lemma 5: $r_1 \geq r_2 \geq \ldots \geq r_n$.

Lemma 6: Let $u^\star(b) : \mathbb{R}_+ \to \mathbb{R}_+$ denote the maximum utility that the buyer can achieve with budget b (when the items are divisible). Then, $u^\star(b)$ is a concave function.

Proofs for both of these lemmas are straight-forward and are deferred to the full version of the paper.

Lemma 7: Mechanism Truthful(f) has approximation ratio $1-1/e$ when all the items have utility equal to 1.

Proof: Recall that $U^\star = u^\star(B)$ and let U denote the utility achieved by Mechanism Truthful(f). We need to show that $(1-1/e)\cdot U^\star \leq U$. Instead of showing that $U = \sum_{i\in S} f_{r_i}(c_i)$ is large enough compared to $U^\star$, we show that $\sum_{i\in S} f_{r_n}(c_i)$ is large enough compared to $U^\star$; the lemma then would be proved since we have $f_{r_n}(c_i) \leq f_{r_i}(c_i)$ for all $i \in S$. To see why $f_{r_n}(c_i) \leq f_{r_i}(c_i)$, it is enough to note that $r_n \leq r_i$ due to Lemma 5 which implies $f_{r_n}(c_i) \leq f_{r_i}(c_i)$.

We consider two cases for the proof: In Case 1 we assume $c_{\mathsf{max}} \leq \bar{c}$, and in Case 2 we assume otherwise, where the number $\bar{c}$ is the cost at which $f_{r_n}(\bar{c}) = 1 - 1/e$, more precisely, this happens at $\bar{c} = r_n(e - e^{1-1/e})$.

Case 1: In this case, observe that we have $f_{r_n}(c_i) \geq 1-1/e$ for all $i \in S$, which implies $f_{r_i}(c_i) \geq 1-1/e$. This just means $U \geq (1-1/e)n \geq (1-1/e)U^\star$.

Case 2: Let $U_n = \sum_{i\in S} f_{r_n}(c_i)$, we will show that

$$U_n \geq (1-1/e)\cdot(1-o(1))\cdot U^\star. \quad (4)$$

To prove this, consider an auxiliary instance in which, instead of budget B, we have a reduced budget $B' = \sum_{i\in S} Q_{r_n}(c_i)$. Note that if we run Mechanism Envy-Free(f) on the auxiliary instance, then its stopping rate is r_n, and so, the utility gained by the mechanism is exactly U_n. Let $U^\star_{\mathsf{aux}}$ denote the optimal utility in the auxiliary instance. Then, by applying Lemma 4 on the auxiliary instance, we have $U_n \geq (1-1/e)\cdot U^\star_{\mathsf{aux}}$. So, if we show that

$$U^\star_{\mathsf{aux}} \geq (1-o(1))\cdot U^\star \quad (5)$$

then (4) holds and the proof is complete.

We use Lemma 6 to prove (5): First, we show that $B' \geq (1-o(1))\cdot B$; then, applying Lemma 6 would imply that $u^\star(B') \geq (1-o(1))\cdot u^\star(B)$, which is identical to (5) by definition. So all we need to complete the proof is showing that $B' \geq (1-o(1))\cdot B$.

To this end, we prove that $B' \geq (1-\alpha\cdot\frac{c_{\mathsf{max}}}{B})\cdot B$, where α is a constant with value $(e-e^{1-1/e})^{-1} \approx 6/5$. This would prove the Lemma due to the large market assumption. First, observe that

$$B = Q_{r_n}(0) + \sum_{i\in S\setminus\{n\}} Q_{r_n}(c_i) \leq Q_{r_n}(0) + B'$$
$$\Rightarrow B' \geq B - Q_{r_n}(0) \geq B - r_n. \quad (6)$$

Now, recall that in Case 2, we have $c_{\mathsf{max}} \geq \bar{c}$, which implies

$$B \geq \bar{c}\cdot\frac{B}{c_{\mathsf{max}}} = r_n(e-e^{1-1/e})\cdot\frac{B}{c_{\mathsf{max}}}$$
$$\Rightarrow c_{\mathsf{max}}\cdot(e-e^{1-1/e})^{-1} \geq r_n. \quad (7)$$

Combining (6) and (7) implies $B' \geq (1-\alpha\cdot\frac{c_{\mathsf{max}}}{B})\cdot B$ with the promised value for α. ■

VI. Impossibility Result: On why $1-1/e$ is the best approximation possible

In this section we show that no truthful (and possibly) randomized mechanism achieves approximation ratio better than $1-1/e$. The proofs are deferred to the full version of the paper and we include only an outline of the proof together with the instance used in the proof. We show a stronger claim by allowing the mechanism to satisfy the budget feasibility in expectation, i.e. we prove that no truthful mechanism that is budget-feasible in expectation can achieve a ratio better than $1-1/e$.

Proof Outline.: We construct a bayesian instance of the problem and prove that no budget-feasible truthful mechanism for this instance can achieve approximation ratio better than $1-1/e$; this also implies that no mechanism for the prior-free setting can achieve ratio better than $1-1/e$ [6]. The proof is done in two steps. First, we show that for any truthful mechanism for this instance, there exists a simple posted price mechanism that achieves at least the same revenue. The posted price mechanism simply offers the same price p to every seller and pays p to any seller who accepts the offer and 0 to others. In the second step of the proof, we show that for any choice of p, such mechanisms cannot achieve a ratio better than $1-1/e$. The proof that we present, w.l.o.g. analyzes the market in expectation: budget feasibility is satisfied

[6]This is so because an α-approximate mechanism in the prior-free setting is also α-approximate in Bayesian setting

in expectation; also, the utility of the mechanisms are computed in expectation.

We present the instance used in the proof below and leave the rest of the proof to the full version of the paper.

The Impossibility Instance.: We construct a bayesian instance of the problem in which all the sellers have unit utility and their costs are drawn i.i.d. from a distribution with cumulative distribution function F, defined as follows:

$$F(x) = \begin{cases} 1/e & \text{if } x = 0, \\ \frac{1}{e(1-x)} & \text{if } 0 < x <= 1 - 1/e. \end{cases}$$

In other words, $F(x)$ denotes the probability that the cost of a seller is at most x. Let $\mathcal{D}$ be the distribution defined by F and let $\overline{c}$ denote the expected cost of a seller sampled from $\mathcal{D}$, i.e. $\overline{c} = \mathbb{E}_{x\sim\mathcal{D}}[x]$. We define the budget to be $B = \overline{c} \cdot N$ where $N \geq 1$ is an arbitrary integer denoting the number of sellers.

Acknowledgments

We acknowledge the comments and ideas of an anonymous reviewer that helped us generalize our impossibility result to the case of bayesian setting.

References

Amazon's mechanical turk platform. https://www.mturk.com/.

Clickworker, virtual workforce. http://www.clickworker.com.

Crowdflower. http://www.crowdflower.com/.

Badanidiyuru, A., Kleinberg, R., and Singer, Y. 2012. Learning on a budget: Posted price mechanisms for online procurement. In *EC*.

Bei, X., Chen, N., Gravin, N., and Lu, P. 2012. Budget feasible mechanism design: from prior-free to bayesian. In *STOC*.

Chen, N., Gravin, N., and Lu, P. 2011. On the approximability of budget feasible mechanisms. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '11. SIAM, 685–699.

Devanur, N. R. and Hayes, T. P. 2009. The adwords problem: online keyword matching with budgeted bidders under random permutations. In *ACM Conference on Electronic Commerce*.

Dobzinski, S., Papadimitriou, C. H., and Singer, Y. 2011. Mechanisms for complement-free procurement. In *ACM Conference on Electronic Commerce*. 273–282.

Feldman, J., Henzinger, M., Korula, N., Mirrokni, V. S., and Stein, C. 2010. Online stochastic packing applied to display ad allocation. In *ESA (1)*. 182–194.

Feldman, J., Korula, N., Mirrokni, V. S., Muthukrishnan, S., and Pál, M. 2009. Online ad assignment with free disposal. In *WINE*. 374–385.

Goel, G. and Mehta, A. 2008. Online budgeted matching in random input models with applications to adwords. In *SODA*.

Horel, T., Ioannidis, S., and Muthukrishnan, S. 2013. Budget feasible mechanisms for experimental design. *CoRR abs/1302.5724*.

Mehta, A., Saberi, A., Vazirani, U. V., and Vazirani, V. V. 2007. Adwords and generalized online matching. *J. ACM 54,* 5.

Myerson, R. B. 1981. Optimal Auction Design. *Mathematics of Operations Research 6*, 58–73.

Singer, Y. 2010. Budget feasible mechanisms. In *FOCS*. 765–774.

Singer, Y. 2011. How to win friends and influence people, truthfully: Influence maximization mechanisms for social networks. In *WSDM*.

Singer, Y. and Mittal, M. 2013. Pricing mechanisms for crowdsourcing markets. In *WWW*. 1157–1166.

Singla, A. and Krause, A. 2013a. Incentives for privacy tradeoff in community sensing. In *AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*.

Singla, A. and Krause, A. 2013b. Truthful incentives in crowdsourcing tasks using regret minimization mechanisms. In *WWW*. WWW '13. 1167–1178.