

# An Efficient and Truthful Pricing Mechanism for Team Formation in Crowdsourcing Markets

Qing Liu\*, Tie Luo†, Ruiming Tang\*, Stéphane Bressan§

\*School of Computing, National University of Singapore

†Institute for Infocomm Research, A\*STAR, Singapore

§National University of Singapore and CNRS, Image & Pervasive Access Lab (IPAL UMI), Singapore

{liuqing,tangruiming,steph}@nus.edu.sg, luot@i2r.a-star.edu.sg

**Abstract**—In a crowdsourcing market, a requester is looking to form a team of workers to perform a complex task that requires a variety of skills. Candidate workers advertise their certified skills and bid prices for their participation. We design four incentive mechanisms for selecting workers to form a valid team (that can complete the task) and determining each individual worker’s payment. We examine profitability, individual rationality, computational efficiency, and truthfulness for each of the four mechanisms. Our analysis shows that TruTeam, one of the four mechanisms, is superior to the others, particularly due to its computational efficiency and truthfulness. Our extensive simulations confirm the analysis and demonstrate that TruTeam is an efficient and truthful pricing mechanism for team formation in crowdsourcing markets.

## I. INTRODUCTION

Future crowdsourcing platforms need to support collaboration [1]. In a collaborative crowdsourcing market, a requester is looking to form a team of workers who can perform a task that requires a set of skills. Interested candidate workers advertise their certified skills and bid prices for their participation. This differs from usual team formation problems (e.g. [2], [3], [4]) in that it not only considers the skills required by the task and possessed by the workers, but also involves economic incentives and several criteria that guarantee profitability for requesters and workers, social welfare, and truthfulness [5]. A large body of research on crowdsourcing markets have confirmed the intuition that financial incentives increase workers’ interest [6] and effort level [7] as well as the attractiveness to experienced workers [8]. However, to the best of our knowledge, existing pricing models (e.g. [9], [10], [11]) for crowdsourcing platforms only consider individual workers without taking into account teamwork which is crucial in collaborative environments [1].

In this paper, we design task allocation and pricing mechanisms for selecting a team of workers in crowdsourcing markets. We start by presenting two baseline mechanisms which pay the selected workers the same as their bids. The first baseline mechanism looks for the lowest-cost team by enumerating all possible teams in a brute force manner. This mechanism is *profitable* for both the requester and selected workers but is not computationally efficient. It also does not ensure *truthfulness* or *incentive-compatibility*, which means that workers may not bid their true costs but can cheat to gain higher payments. We refer to this baseline mechanism as OPT as it minimizes total payment if all workers bid truthfully. We show that OPT is profitable, individually rational but not efficient nor truthful.

The second baseline mechanism follows a greedy approach to form a team of workers with low bids and high total expertise.

We refer to this mechanism as GREEDY and show that it is efficient, profitable, individually rational but not truthful.

Both baseline mechanisms do not prevent workers from over-bidding, which necessitates the need of designing mechanisms in which workers will bid their true costs.

We first adapt the celebrated Vickrey-Clarke-Groves auction to our problem, and refer to this mechanism as VCG. We show that it is profitable, individually rational, truthful but not efficient.

Finally, we design a mechanism that combines the greedy selection rule and a special payment scheme. A selected worker receives a payment that equals to the highest bid she could have placed and been selected. We refer to this mechanism as TruTeam. It possesses all the above desirable properties, i.e., efficiency, profitability, individual rationality, and truthfulness.

Using synthetic scenarios, we evaluate the properties and performance of these four mechanisms. The results show that TruTeam is an efficient, truthful task allocation and pricing mechanism for team formation in crowdsourcing markets. In summary, this paper makes the following contributions:

- To the best of our knowledge, this is the first study on team formation in collaborative crowdsourcing markets.
- We formulate the problem of team formation in crowdsourcing as an task allocation and pricing mechanism design problem.
- We design two baseline and two refined mechanisms and prove profitability, individual rationality, prove or disprove truthfulness for each of the four mechanisms. In addition, we also evaluate their computational complexity.
- We show via both analysis and extensive simulations that TruTeam is an efficient, profitable, individually rational and truthful mechanism for team formation in crowdsourcing markets.

## II. RELATED WORK

### A. Pricing Mechanisms

Various models and techniques have been proposed for pricing on crowdsourcing platforms. Budget-feasible mechanisms [12] maximize a requester’s profit under a budget constraint while satisfying other properties such as truthful bidding. In [9], the authors designed a framework for task allocation and pricing in an online environment. The framework aims to maximize the number of tasks performed under a given budget or to minimize payments for a given number of tasks. The authors of [10] designed a no-regret posted price mechanism which bridges the gap between procurement auctions and multi-armed

bandits. That mechanism satisfies budget feasibility, achieves near-optimal utility for the requester, and also guarantees that workers bid their true costs. In [11], workers and tasks are modeled as a bipartite graph where an edge  $(m, n)$  in the graph represents worker  $n$  is willing to perform task  $m$ . The authors designed a payment mechanism that ensures budget feasibility and one-way-truthfulness while achieving near-optimal utility. Recent work [13] considers variable rather than fixed payment in crowdsourcing, as a function of the best worker's effort, and achieves a higher utility than the optimum achieved by fixed payment. This idea was then extended to heterogeneous crowdsourcing environments [14] to tackle non-uniform knowledge possessed by different workers.

### B. Task Allocation and Team Formation

The task allocation problem is related to the job scheduling problem which aims at minimizing the load of the machines that have maximal work load. An extended version of the job scheduling problem, in which each job needs to be performed on a set of machines, was proved to be NP-hard [15].

The authors of [16] studied a new variant of the task allocation problem in which the workers are connected in a social network. The workers are assumed to only have local knowledge about resources and hence each task can only be assigned to its neighboring workers. The authors proved this problem to be NP-hard and proposed a max-flow network model to solve it.

Team formation in a non-crowdsourcing environment (e.g. social networks) was studied by [17], [4], assuming workers need to communicate when performing a task and communication is costly. The selected team has the skills to complete a task and has minimum communication cost. A different metric, the workload of workers, was studied by [2] when selecting a team to complete a given task. The authors of [3] considered both workload and communication costs when selecting a team to perform a task.

In this paper, we consider pricing mechanisms for the team formation problem, which bridges the gap between budget-feasible mechanisms and the traditional team formation problem. We do not consider workload or communication costs but rather focus on truthfulness which is more important in crowdsourcing environments. In addition, our problem is more general than prior work on budget-feasible mechanisms where a task is always assigned to a single worker.

## III. MODEL

### A. Requester and Workers

In our model, a single *requester* posts her task to a crowdsourcing platform. The task has a value  $v$  which is the requester's revenue if the task is completed. There is a set of  $n$  available workers  $W = \{w_1, \dots, w_n\}$ , and the task needs a subset of workers,  $S \subseteq W$ , to collaborate. When a worker signs up to participate in the task, she should report to the requester what skills she has and how much she expects to be paid. Then the requester selects a set of workers and decides on the payment for each of them.

**Cost and bid of a worker.** We assume that each worker's cost of doing the task is private information. Each worker  $w_i$  has a non-negative cost  $c_i \in \mathcal{R}_{\geq 0}$  to perform the task, and bids  $b_i \in \mathcal{R}_{\geq 0}$  when she signs up to do the task.

We assume that a worker cannot lie about the skills that she has. Some existing crowdsourcing platforms such as MTurk [18] provided qualification test to ensure the validity of workers' skills. A platform can also use work history information to verify one's skill.

**Utility of the requester.** The requester's utility  $U_R$  is the revenue obtained from the completed task, subtracting the payment to the selected workers.

$$U_R = \begin{cases} v - \sum_{w_i \in S} p_i & \text{if task is completed} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where  $S$  is the set of selected workers and  $p_i$  is the payment to worker  $w_i$ .

**Utility of a worker.** Worker  $w_i$ 's utility  $u_i$  is the payment she receives subtracting her cost of performing the task.

$$u_i = \begin{cases} p_i - c_i & \text{worker } w_i \text{ is selected} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Both the requester and the workers aim to maximize their respective utilities.

### B. Skill Profiles

The skill profile of the given task is an  $l$ -dimension vector  $s_* = (s_*[1], \dots, s_*[l])$ , where  $s_*[i] = \{0, 1\}$  represents that the  $i$ th skill is required (1) or not (0). We assume that a maximum of  $l$  skills are required for any task. The skill profile of a worker  $w_i$  is also an  $l$ -dimension vector  $s_{w_i} = (s_{w_i}[1], \dots, s_{w_i}[l])$ , similarly defined but representing what skills  $w_i$  possesses.

The skill profile of a team  $s_T = (s_T[1], \dots, s_T[l])$  is defined by a logical OR of the skill profiles of all the individual workers in the team  $T$ :

$$s_T[j] = \bigvee_{w_i \in T} s_{w_i}[j], \quad j = 1, \dots, l \quad (3)$$

The team that can complete the task is the team that has all the required skills of the task, i.e.,

$$s_T[j] \geq s_*[j], \quad j = 1, \dots, l \quad (4)$$

### C. Desirable Properties

**Computational Efficiency.** The task allocation and pricing mechanism can be executed in polynomial time.

**Individual Rationality.** No worker is worse off if she is selected to do the task. In other words, each selected worker receives a payment no less than her true cost.

**Profitability.** The utility of the requester is non-negative.

**Truthfulness (Incentive Compatibility).** Bidding her true cost is each worker's dominant strategy. Formally, if  $u_i$  and  $u'_i$  are the utilities of worker  $w_i$  when bidding truthfully and untruthfully, respectively, then a truthful mechanism guarantees that  $u_i \geq u'_i$  regardless of what other workers bid.

**Theorem 1.** *An auction mechanism is truthful if and only if [12]:*

- *The allocation rule is monotone: If worker  $w_i$  wins the auction by bidding  $b_i$ , she also wins by bidding  $b'_i \leq b_i$ .*
- *Each winner is paid the threshold price: Worker  $w_i$  will not win the auction if she bids higher than this price.*

### D. Design Objectives

We want to design a task allocation and pricing mechanism for the requester to select a team to complete a task she posts, with

the objective of minimizing the total payment to the selected workers, i.e.,

$$S = \arg \min_T \sum_{w_i \in T} p_i \quad (5)$$

$$s.t. s_T[j] \geq s_*[j], \quad j = 1, \dots, l \quad (6)$$

In addition, the mechanism should satisfy computational efficiency, individual rationality, profitability and truthfulness.

#### IV. MECHANISMS

##### A. Optimal Mechanism

This mechanism selects the team with the lowest total bid that is able to complete the task, by taking a brute-force approach to attempt all the possible  $2^n - 1$  teams (excluding the empty set) of the  $n$  workers. We refer to this mechanism as OPT.

**Lemma 1.** *OPT is individually rational, profitable, but not computationally efficient or truthful. Time complexity of OPT is  $O(2^n)$ .*

The proof is deferred to [19] due to space constraint, as is also the case for Lemma 2 and Theorem 2.

##### B. Greedy Mechanism

This heuristic mechanism selects the worker with the minimum cost per *marginal skill contribution* until a team that can complete the task is formed or all the workers have been considered. It pays each selected worker her bid.

In the above,  $w_i$ 's *marginal skill contribution*,  $\Delta_i(S)$ , is defined with respect to an existing worker set  $S$  as the number of uncovered skills that  $w_i$  can cover if selected into  $S$ :

$$\Delta_i(S) = s_{S \cup \{w_i\}} \cdot s_* - s_S \cdot s_* \quad (7)$$

where  $s_S$  is the skill profile of team  $S$ , and  $s_S \cdot s_*$  is the inner product of vectors  $s_S$  and  $s_*$ . In each iteration, the GREEDY mechanism always selects the worker who has the lowest cost per marginal skill contribution,  $\frac{b_i}{\Delta_i(S)}$ .

**Lemma 2.** *GREEDY is computationally efficient, individually rational, profitable, but not truthful. Time complexity of GREEDY is  $O(n^2)$ .*

##### C. VCG-based Mechanism

We adapt the traditional Vickrey-Clarke-Groves (VCG) mechanism [5] to our problem, and design a mechanism called VCG.

**Allocation rule.** VCG selects the team with the lowest total cost, i.e.,

$$S = \arg \min_T \sum_{w_i \in T} c_i \quad (8)$$

$$s.t. s_S[j] \geq s_*[j], \quad j = 1, \dots, l$$

**Payment rule.** VCG pays each selected worker  $w_i$  the difference between the optimal welfare (for the other workers) if  $w_i$  was not participating and welfare of the other workers with respect to the selected team:

$$p_i = \left( \min_T \sum_{w_j \in T \wedge w_i \notin T} c_j \right) - \sum_{w_j \in S \wedge j \neq i} c_j \quad (9)$$

$$s.t. s_T[j] \geq s_*[j], \quad j = 1, \dots, l$$

where  $S$  is defined in (8).

**Lemma 3.** *VCG is individually rational, profitable, truthful, but not computationally efficient. Time complexity of VCG is  $O(n2^n)$ .*

*Proof.* Suppose  $u_i, u'_i, p_i, p'_i$  are the corresponding utilities and payments of  $w_i$ , when she bids truthfully and untruthfully, respectively. Let  $S, S'$  denote the selected team when  $w_i$  bids truthfully and untruthfully, respectively.

$$\begin{aligned} u'_i &= p'_i - c_i \\ &= \left( \min_T \sum_{w_j \in T \wedge w_i \notin T} c_j \right) - \sum_{w_j \in S' \wedge j \neq i} c_j - c_i \\ &= \left( \min_T \sum_{w_j \in T \wedge w_i \notin T} c_j \right) - \sum_{w_j \in S'} c_j \\ &\leq \left( \min_T \sum_{w_j \in T \wedge w_i \notin T} c_j \right) - \sum_{w_j \in S} c_j \\ &= u_i \end{aligned} \quad (10)$$

which proves the truthfulness. The proof of other properties are deferred to [19] due to space constraint.  $\square$

##### D. Efficient and Truthful Mechanism (TruTeam)

OPT is an optimal allocation mechanism if every worker bids truthfully. GREEDY is computationally efficient but not a truthful mechanism. VCG is a truthful mechanism but is not computationally efficient.

In this section, we present the mechanism TruTeam (Mechanism 1) which satisfies all the four properties (i.e., TruTeam is computationally efficient, individually rational, profitable, and truthful).

---

#### Mechanism 1: Efficient and Truthful mechanism for Team formation (TruTeam)

---

**Input:**  $b_{1 \sim n}, s_{w_1 \sim w_n}, v, s_*$

**Output:**  $S, p_{1 \sim n}$

```

1  $S \leftarrow \emptyset; p_{1 \sim n} \leftarrow 0;$ 
2 repeat
3    $w_i \leftarrow \arg \min_{w_i \in W \setminus S} \frac{b_i}{\Delta_i(S)};$ 
4    $W' \leftarrow W \setminus \{S \cup \{w_i\}\};$ 
5    $T \leftarrow S;$ 
6   repeat
7      $w_j \leftarrow \arg \min_{w_j \in W' \setminus T} \frac{b_j}{\Delta_j(T)};$ 
8      $p_i \leftarrow \max\{\frac{b_j}{\Delta_j(T)} \times \Delta_i(T), p_i\};$ 
9      $T \leftarrow T \cup \{j\};$ 
10  until  $\Delta_i(T) = 0$  or  $v - p_i < 0;$ 
11  if  $v \geq p_i$  then
12     $S \leftarrow S \cup \{w_i\};$ 
13     $v \leftarrow v - p_i;$ 
14  else
15     $W \leftarrow W \setminus \{w_i\};$ 
16 until  $s_S$  cover  $s_*$  or  $W \setminus S = \emptyset;$ 
17 if  $s_S$  not cover  $s_*$  then
18    $S \leftarrow \emptyset; p_{1 \sim n} \leftarrow 0$ 
19 return  $(S; p_1, \dots, p_n)$ 

```

---

**Allocation rule.** In each iteration, it selects the worker who has the smallest cost per marginal skill contribution, i.e., the lowest  $\frac{b_i}{\Delta_i(S)}$ . This is the same as GREEDY.

TABLE I  
PARAMETERS OF OUR TWO DATASETS

	$n$ : no. of workers	$l$ : no. of skills	$(u, \sigma)$	$[C_1, C_2]$
Small	$n = 10 \sim 25$ , fixing $l = 5$	$l = 1 \sim 10$ , fixing $n = 20$	$(l/3, 0.4)$	$[1, v/5]$
Large	$n = 10 \sim 3000$ , fixing $l = 50$	$l = 1 \sim 100$ , fixing $n = 1000$	$(l/5, 0.4)$	$[1, v]$

**Payment rule.** The intuition of the payment rule is to pay each selected worker the highest cost she can report while still being selected [11]. This is the “threshold price” stated in Theorem 1, and we will show that overbidding under TruTeam does no good to improve a worker’s utility.

Now, we explain in detail how to determine the payment to each selected worker. When computing the payment to worker  $w_i$ , let’s see how this mechanism selects a team without  $w_i$ ’s participation. It selects from set  $W \setminus \{T \cup \{w_i\}\}$  ( $T$  is the selected worker set before  $w_i$ ) the worker ( $w_{j_1}$ ) who minimizes the value  $\frac{b_j}{\Delta_j(S)}$ . Therefore

$$b_i \leq \frac{b_{j_1}}{\Delta_{j_1}(T)} \times \Delta_i(T) \quad (11)$$

Otherwise, if  $\frac{b_i}{\Delta_i(T)} > \frac{b_{j_1}}{\Delta_{j_1}(T)}$ , we would have selected  $w_{j_1}$  instead of  $w_i$  according to the allocation rule of this mechanism. Therefore, we set the payment to worker  $w_i$  equal to this value:

$$p'_i = \frac{b_{j_1}}{\Delta_{j_1}(T)} \times \Delta_i(T) \quad (12)$$

However,  $p'_i$  may not be the highest bid that  $w_i$  can report while still being able to be selected, because  $T \cup \{w_{j_1}\}$  may not cover all the skills required by the task. Suppose  $T \cup \{w_{j_1}, w_{j_2}, \dots, w_{j_k}\}$  (without  $w_i$ ) is the set of workers selected according to this mechanism. We set the payment equal to the following threshold price as described in lines 6-10.

$$p_i = \max_{j \in \{j_1, j_2, \dots, j_k\}} \left\{ \frac{b_j}{\Delta_j(T)} \times \Delta_i(T) \right\} \quad (13)$$

Note that,  $T$  is updated every time by including a new worker  $w_{j_x}$  ( $x = 1, 2, \dots, k$ ), i.e.,  $T = T \cup \{w_{j_x}\}$ .

In order to be profitable,  $w_i$  is selected to perform this task only if the task’s remaining value is not less than  $p_i$  (we also update the set of selected workers as  $S = S \cup \{w_i\}$ ), as shown in lines 11-13. Otherwise we skip  $w_i$  and consider next candidate worker as described in lines 14-15.

Repeat the above process until the task can be completed by the set of workers  $S$  or all the workers have been considered.

**Lemma 4.** *TruTeam is computationally efficient, with a time complexity of  $O(n^2l)$ .*

*Proof.* Selecting the worker who has the minimal value  $\frac{b_i}{\Delta_i(S)}$  takes  $O(n)$  time. Deciding the payment for the selected worker takes  $O(nl)$ . Since there are  $n$  workers, time complexity of this mechanism is  $O(n^2l)$ .  $\square$

**Lemma 5.** *TruTeam is individually rational.*

*Proof.* From the above payment rule, we can see that for any selected worker  $w_i$ .

$$\begin{aligned} b_i &\leq \frac{b_{j_1}}{\Delta_{j_1}(S)} \times \Delta_i(S) \\ &\leq \max_{j \in \{j_1, j_2, \dots, j_k\}} \left\{ \frac{b_j}{\Delta_j(T)} \times \Delta_i(T) \right\} \\ &= p_i \end{aligned} \quad (14)$$

We assume a worker will not bid below her true cost, i.e.  $b_i \geq c_i$  (In fact, we will show in Lemma 7 that  $b_i = c_i$ ). Therefore,  $w_i$ ’s utility is  $u_i = p_i - c_i \geq p_i - b_i \geq 0$ .  $\square$

**Lemma 6.** *TruTeam is profitable.*

*Proof.* Every time when a worker is considered, we check if the remaining value of the task can cover the payment to this worker. If not, the worker will not be selected. This process guarantees that the value of this task is more than the total payment to the whole team.  $\square$

**Lemma 7.** *TruTeam is truthful.*

*Proof.* According to Theorem 1, we need to prove that (1) the allocation rule is monotone and (2) the payment to each selected worker  $w_i$  is the threshold price  $p_i$ .

The monotonicity of the allocation rule is obvious, since if  $w_i$  is selected, she will also be selected by bidding a smaller value which leads to a smaller cost per marginal skill contribution.

For the threshold price, recall  $p_i$  from equation (13). If  $b_i > p_i$ ,  $w_i$  will be placed after the last selected worker, thus she will not be selected to perform the task. Therefore,  $p_i$  is the threshold price.<sup>1</sup>  $\square$

**Theorem 2.** *TruTeam is computationally efficient, individually rational, profitable and truthful.*

## V. EVALUATION

In this section, we compare the performance of the four mechanisms, namely OPT, VCG, GREEDY and TruTeam in terms of the following metrics:

- **Running Time:** the actual CPU time on a computer.
- **Requester’s Utility:** defined in Eqn. (1).
- **Truthfulness:** We verify the truthfulness of TruTeam by evaluating workers’ utility if they overbid or underbid.

### A. Simulation Setup

We generate two different datasets to evaluate our mechanisms. A **Small** dataset is used to evaluate all the four mechanisms and a **Large** dataset is used to evaluate the two computationally efficient mechanisms (i.e., GREEDY and TruTeam). The parameters of the two datasets are listed in Table I and explained below.

We set the value of the task  $v = 500$  which is unknown to workers. Each worker’s true cost  $c_i$  is uniformly drawn from  $[C_1, C_2]$ . In the case that all workers are truthful, we set  $b_i = c_i$ . In the case of overbidding, we randomly select  $k \in [1, n]$  workers and let each of them overbid a random value  $r_i$  (i.e.,  $b_i = c_i + r_i$ ) where  $r_i \in [1, v]$ . We do not consider underbidding, since no rational worker will underbid under these four mechanisms.

To generate each worker’s skill profile, we first generate the number of skills she has, using the normal distribution  $(u, \sigma)$ . Suppose  $w_i$  has  $x$  skills, then we randomly assign  $x$  different skills out of all the  $l$  skills to her.

<sup>1</sup>Bidders will not underbid, either. The intuition is that if  $w_i$  underbids and is selected, her payment will not cover her cost. See [19] for the details.

All the simulations were run on a Windows PC with a 3.40GHz CPU and 8 GB memory. Each data point is averaged over 100 measurements.

## B. Results

Fig. 1 shows the comparison of the four mechanisms conducted over the Small dataset. We observe in Fig. 1 (a,b,c,d) that OPT and VCG do not scale well when the number of workers or skills becomes large. However, all the four mechanisms achieve strictly positive requester’s utility (e,f,g,h). Therefore, in this section, we focus on the computationally efficient mechanisms, GREEDY and TruTeam, and explain in detail the results pertaining to them collected on the Large dataset.

1) *Running Time*: The results are presented in Fig. 2 (a,b,c,d). Generally, as the number of workers or the number of skills increases, running time of both mechanisms increases. Although GREEDY outperforms TruTeam, since the payment determination process of TruTeam is more complicated than that of GREEDY, TruTeam maintains a high efficiency. For example, it completes in about 0.6 second even when the number of worker reaches 3000. More importantly, TruTeam ensures truthfulness which is crucial to incentive mechanisms to counteract possible cheating behaviors in practice.

In Fig. 2(a) and 2(b), running time of TruTeam contains a small peak when the number of workers is between 100 and 300. This happens when the task is complex (requiring skills as many as  $l = 50$ ) and the number of workers is relatively small ( $n < 300$ ). Thus the team size is fairly large and TruTeam needs to check every other worker’s bid and skills when deciding the payment to each selected worker, resulting in higher running time.

2) *Requester’s Utility*: Fig. 2 (e,f,g,h) present the requester’s utility in different settings. Generally, the requester’s utility increases as the number of worker increases (e,f), since the requester has more “cheaper” workers to select from. On the other hand, the requester’s utility drops as the number of required skills becomes larger (g,h), which is a natural result of the increase of complexity of the task.

In the case of truthful bidding, as is shown in Fig. 2(e) 2(g), the untruthful mechanism (GREEDY) yields higher utility than the truthful mechanism (TruTeam) because TruTeam pays each selected worker no less than her bid.

However, in the case of overbidding which is a more realistic setting, Fig. 2(f) 2(h) show that TruTeam outperforms GREEDY. This demonstrates that in real crowdsourcing markets where workers are strategic and speculating higher payment (e.g., by trying to overbid), TruTeam generates higher profit for the requester by ensuring truthful bidding.

3) *Truthfulness*: Lastly, we verify the truthfulness of TruTeam by examining the utilities of two randomly chosen workers,  $w_{637}$  and  $w_{219}$ . We set  $n = 1000$  and  $l = 50$ , and their true costs are  $c_{637} = 4$  and  $c_{219} = 10$ , respectively. In Fig. 3(a), we observe that  $w_{637}$  is selected to perform the task if she bids her true cost  $c_{637} = 4$ , and her utility reaches the optimal value 7. If she overbids a value no less than 11, she is not selected and therefore her utility drops to 0. In Fig. 3(b), it is observed that  $w_{219}$  is not selected to do the task if she bids her true cost  $c_{219} = 10$ , and hence her utility is 0. This is the optimal utility she can get because even though she can be selected to

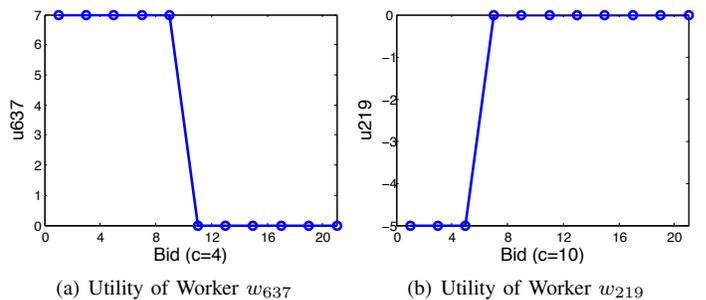


Fig. 3. Workers’ Utilities.

do the task, which only happens if she under bids (below 7), her payment will not be able to cover her true cost and hence she will receive a negative utility, as indicated in Fig. 3(b).

TruTeam ensures that it is every worker’s *dominant strategy* to bid her true cost in order to maximize her utility.

## VI. CONCLUSION

In this paper, we formulate team formation in crowdsourcing as a task allocation and pricing problem, and provide four candidate mechanisms as solutions: OPT, GREEDY, VCG, and TruTeam. We prove that although all the four mechanisms satisfy profitability and individual rationality, only VCG and TruTeam satisfy truthfulness, and only GREEDY and TruTeam are efficient. Simulation also demonstrate that TruTeam is the only one, among the four candidates, that is efficient, profitable, individually rational and truthful.

To the best of our knowledge, this is the first study on team formation in collaborative crowdsourcing markets. Future research could be in the direction of taking into account the quality of contribution [20] and trustworthiness of workers [21] when selecting and rewarding workers, or considering previous collaborations among workers and inter-dependency among multiple tasks when forming multiple teams.

## ACKNOWLEDGMENT

This work is supported by the French Ministry of European and Foreign Affairs under the STIC-Asia program, CCIPIX project.

## REFERENCES

- [1] A. Kittur, J. V. Nickerson, M. Bernstein, E. Gerber, A. Shaw, J. Zimmerman, M. Lease, and J. Horton, “The future of crowd work,” in *Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work (CSCW)*. ACM, 2013, pp. 1301–1318.
- [2] A. Anagnostopoulos, L. Becchetti, C. Castillo, A. Gionis, and S. Leonardi, “Power in unity: forming teams in large-scale community systems,” in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM)*. ACM, 2010, pp. 599–608.
- [3] —, “Online team formation in social networks,” in *Proceedings of the 21st International Conference on World Wide Web (WWW)*. ACM, 2012, pp. 839–848.
- [4] T. Lappas, K. Liu, and E. Terzi, “Finding a team of experts in social networks,” in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, 2009, pp. 467–476.
- [5] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic game theory*. Cambridge University Press, 2007.
- [6] D. DiPalantino and M. Vojnovic, “Crowdsourcing and all-pay auctions,” in *ACM Conference on Electronic Commerce (EC)*. ACM, 2009, pp. 119–128.
- [7] Y. Chen, T.-H. Ho, and Y.-M. Kim, “Knowledge market design: A field experiment at google answers,” *Journal of Public Economic Theory*, vol. 12, no. 4, pp. 641–664, 2010.

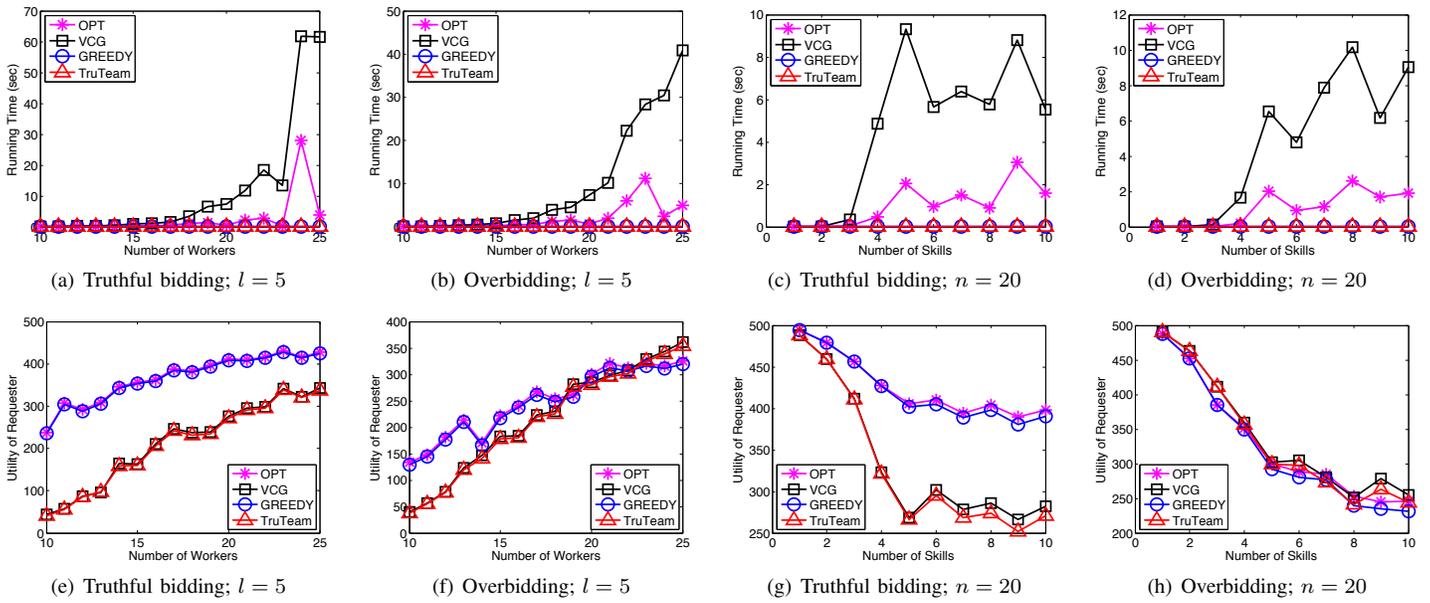


Fig. 1. Simulation results on the Small Dataset.

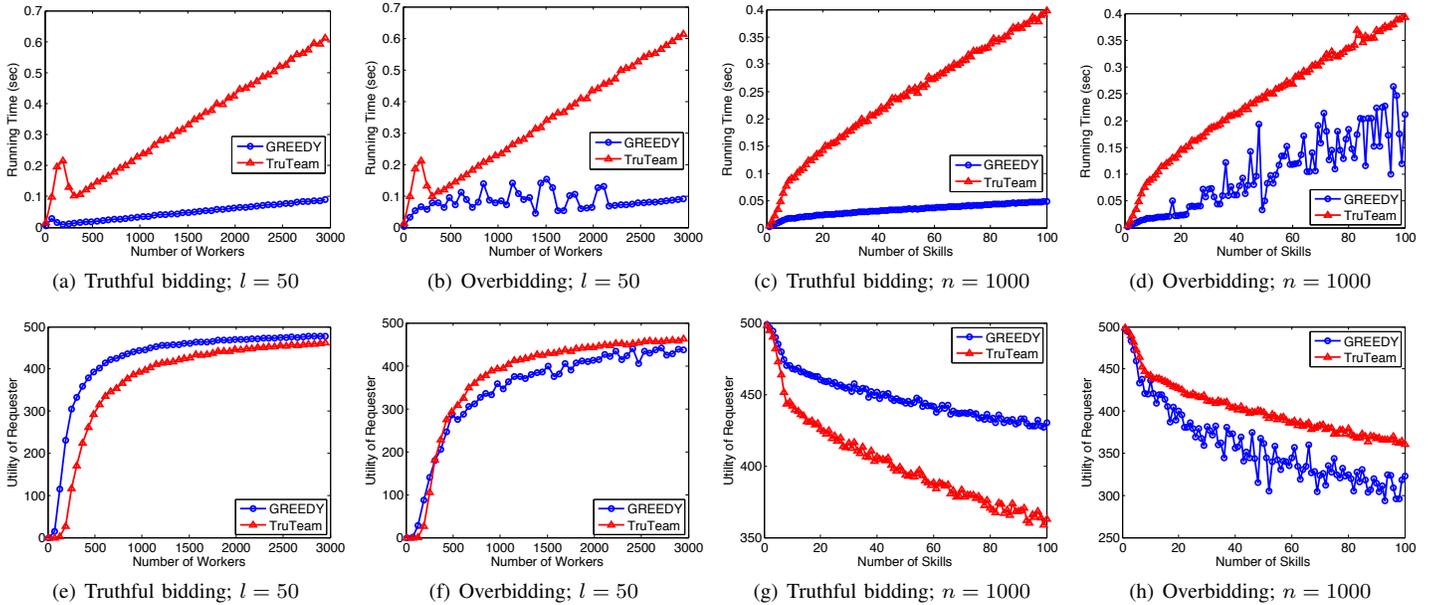


Fig. 2. Simulation results on the Large Dataset.

- [8] W. Mason and D. J. Watts, "Financial incentives and the performance of crowds," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 2, pp. 100–108, 2010.
- [9] Y. Singer and M. Mittal, "Pricing mechanisms for crowdsourcing markets," in *Proceedings of the 22nd International Conference on World Wide Web (WWW)*, 2013, pp. 1157–1166.
- [10] A. Singla and A. Krause, "Truthful incentives in crowdsourcing tasks using regret minimization mechanisms," in *Proceedings of the 22nd International Conference on World Wide Web (WWW)*, 2013, pp. 1167–1178.
- [11] G. Goel, A. Nikzad, and A. Singla, "Allocating tasks to workers with matching constraints: truthful mechanisms for crowdsourcing markets," in *Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion*, 2014, pp. 279–280.
- [12] Y. Singer, "Budget feasible mechanisms," in *51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2010. IEEE, 2010, pp. 765–774.
- [13] T. Luo, H.-P. Tan, and L. Xia, "Profit-maximizing incentive for participatory sensing," in *IEEE INFOCOM*, 2014, pp. 127–135.
- [14] T. Luo, S. S. Kanhere, S. K. Das, and H.-P. Tan, "Optimal prizes for all-pay contests in heterogeneous crowdsourcing," in *IEEE MASS*, 2014.
- [15] Y. Azar, "On-line load balancing," in *Online Algorithms*. Springer, 1998, pp. 178–195.
- [16] M. De Weerd, Y. Zhang, and T. Klos, "Distributed task allocation in social networks," in *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. ACM, 2007, p. 76.
- [17] K. Balog and M. De Rijke, "Determining expert profiles (with an application to expert finding)," in *International Joint Conferences on Artificial Intelligence (IJCAI)*, vol. 7, 2007, pp. 2657–2662.
- [18] Amazon Mechanical Turk, <http://www.mturk.com>.
- [19] Institute for Infocomm Research, Tech. Rep. [Online]. Available: <http://www1.i2r.a-star.edu.sg/~luot/TR1409-TruTeam.pdf>
- [20] C.-K. Tham and T. Luo, "Quality of Contributed Service and market equilibrium for participatory sensing," in *IEEE DCOSS*, 2013, pp. 133–140.
- [21] T. Luo, S. S. Kanhere, and H.-P. Tan, "SEW-ing a simple endorsement web to incentivize trustworthy participatory sensing," in *IEEE SECON*, 2014, pp. 636–644.