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
{liuqing, tangruiming, steph}@nus.edu.sg
†Institute for Infocomm Research, A*STAR, Singapore
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 the profitability, individually rationality, computationally efficiency, and truthfulness for each of the four mechanisms. Our study analytically shows that one of the mechanisms, called TruTeam, is superior to the others particularly due to its computationally efficiency and truthfulness. Furthermore, 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 [10]. 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 a price for their participation. The problem differs from the usual team formation problem (e.g. [2], [3], [11]) as it considers not only the skills required by the task and possessed by the workers but also economic incentives and several criteria that guarantee profitability for requesters and candidate workers, social welfare, and truthfulness [13]. A large body of research on crowdsourcing markets have confirmed the intuition that financial incentives increase workers' interest [8] and effort level [6] as well as the attractiveness to experienced workers [12]. However, to the best of our knowledge, existing pricing models (e.g. [15], [16], [9]) for crowdsourcing platforms only consider individual workers without taking into account teamwork which is crucial in many collaborative environments [10].

In this paper, we design efficient and effective task allocation and pricing mechanisms for the selection of a team of workers. We start by presenting two baseline mechanisms which pay the selected workers the same as their bids. The first baseline mechanism looks for the cheapest team by enumerating all possible teams in a brute force manner. It is *profitable* for the requester and for the selected workers but is not efficient. The mechanism does not ensure that workers have incentive to bid their true costs, and hence they could cheat in order to gain higher payment. A mechanism that encourages workers to bid their true costs is called a *truthful* or *incentive-compatible* mechanism. We refer to this first 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 strategy to form a team of workers with lower bids and higher 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 overbidding. Next, We design two truthful mechanisms in which workers will bid their true costs.

We first adapt the Vickrey-Clarke-Groves auction mechanism which is an extension of the Vickrey second-price auction to multiple goods. We refer to this mechanism as VCG, and show that it is profitable, individually rational, truthful but not efficient.

Finally, we design a mechanism that combines the greedy selection rule and a payment scheme calculated using the other candidates' bids. A selected worker receives a payment that is the highest bid she could have placed and still be selected. We refer to this mechanism as TruTeam. It possesses all the properties we desire, i.e., it is efficient, profitable, individually rational and truthful.

Using synthetic scenarios, we evaluate the efficiency and effectiveness of the above 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 the 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, individually rationality, prove or disprove truthfulness for each of the four mechanisms. In addition, we also provide their computational complexity.
- We extensively evaluate the efficiency and effectiveness of the four mechanisms and show that TruTeam is an efficient, profitable, individually rational and truthful mechanism for team formation in crowdsourcing mar-

kets.

II. Related Work

A. Pricing Mechanisms

Various models and techniques have been proposed for pricing in crowdsourcing platforms. Budget-feasible mechanisms maximize a requester's profit under a budget constraint while satisfying other properties such as truthful bidding. In recent work [15], the authors designed a framework for tasks pricing and allocation in an online setting. 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 [16] 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 [9], 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 design a payment mechanism that pays the selected workers the same amount while ensuring budget feasibility and one-way-truthfulness and achieving near-optimal utility.

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 is studied in [4], where each job needs to be performed on a set of machines. This extended problem is proved to be NP-hard [4].

The authors of [7] proposed a new variant of the task allocation problem. In their model, the workers are connected in a social network and they only have local knowledge about resources so that the tasks can only be assigned to its neighbour workers. This problem was proved to be NP-hard and a max-flow network model was proposed to solve the problem.

In [5], [11], the authors model the team formation problem under the assumption that workers need to communicate when performing a task. The authors consider both individual skills and communication cost. The selected team is able to complete the given task and has minimum communication cost. The authors of [2] studied a different metric, the workload of workers, when selecting a team to complete a given task. The authors of [3] consider both workload and communication costs when selecting a team to perform the task.

In this paper, we consider a pricing mechanism 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 when selecting a team, but rather focus on the truthfulness of a pricing mechanism. 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* $W = \{w_1, ..., w_n\}$, and the task needs a set 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 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. It is *not* necessary that $b_i = c_i$.

We assume that a worker cannot lie about the skills that she has. Some existing crowdsourcing platforms such as MTurk [1] 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 less the payment to the selected workers.

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

where S is the set of selected workers. 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 less her cost of performing the task.

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

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

B. Task, Skills, Teams

The skill profile of the given task is an l-dimension vector $s_* = (s_*[1], ..., s_*[l])$, where $s_*[i] = \{0, 1\}$ represents 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 $s_{w_i} = (s_{w_i}[1], ..., s_{w_i}[l])$, similarly defined but representing the skills w_i possesses.

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

$$s_T[j] = \bigvee_{w_i \in T} s_{w_i}[j], \ j = 1, ..., l \tag{3}$$

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

$$s_T[j] \geq s_*[j], \ j = 1, ..., l \tag{4}$$

TABLE I
NOTATIONS

notations	description
n	number of workers
l	number of skills
W, w_i	the set of all the workers, a worker
c_i, b_i, u_i	cost, bid, utility of worker w_i
T	a team of workers
s_{w_i}, s_T	the skill profile of worker w_i, and of team T
v	value of the task
s_*	the skill profile of the task
S	the set of the selected workers
p_i	payment to worker w_i

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. Each selected worker's payment should be no less than her true cost.

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

Incentive Compatibility (Truthfulness). Bidding her true cost is each worker's dominant strategy. Formally, if u_i , u_i' are the respective utilities of worker w_i when she bids truthfully and untruthfully, 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 [14]:*

- *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 an 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.

$$S = \arg\min_{T} \sum_{w_i \in T} p_i \tag{5}$$

$$s.t.\ s_T[j] \geq s_*[j], \quad j = 1, ..., l \tag{6}$$

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

IV. MECHANISMS

A. An Optimal Mechanism

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

In Line 1, for every possible team, Mechanism 1 checks whether it is able to complete the task, by checking whether Equation 4 holds. All the teams which are able to complete the task are collected in D. Line 2 to line 3 initialize necessary variables. The cheapest team is selected using line 4 to line 6, by enumerating all the teams in D. If the total bid of this selected team is larger than the value of the task, the task is abandoned (line 7 to line 8); otherwise, the task is performed and each selected worker is paid with her bid (line 10 to line 11).

Mechanism 1: Optimal Allocation and Payment (OPT)

```
   Input: b_{1~n}, s_{w_1~w_n}, v, s_*
   Output: S, p_{1~n}
1  D ← all possible teams that can cover s_*;
2  S ← ∅; p_{1~n} ← 0;
3  T ← ∅; MIN = ∞;
4  foreach T ∈ D do
5      if Σ_{w_i∈T} b_i < MIN then
6          S ← T; MIN = Σ_{w_i∈T} b_i
7  if v − MIN < 0 then
8      S ← ∅
9  else
10     foreach w_i ∈ S do
11         p_i ← b_i
12 return {S; p_1, ..., p_n}
```

Now, let us check whether Mechanism 1 satisfies the pre-defined properties (namely, computationally efficiency, individually rational, profitable and truthful) in the previous section.

Lemma 1. *OPT is not Computationally Efficient.*

Proof. The complexity of OPT is $O(2^n)$ (where n is the number of workers) since it considers every possible team (namely, every possible subset of W). □

Lemma 2. *OPT is Individually Rational.*

Proof. A rational worker bids the task with a value no less than her cost to perform this task, i.e., $b_i \geq c_i$. Therefore, the utility for each selected worker is $u_i = p_i - c_i \geq p_i - b_i = 0$. □

Lemma 3. *OPT is Profitable.*

Proof. This is obvious, since if the the payment to the selected team is larger than the value of the task, this task will not be performed. When performing the task, the utility of the requester is $U_R = v - \sum_{w_i \in S} p_i = v - \sum_{w_i \in S} b_i \geq 0$. □

Lemma 4. *OPT is not Truthful.*

Proof. OPT pays each selected worker with exactly the value of her bid, therefore she has incentive to bid a higher value than her true cost.

We will use an example to show that OPT is not truthful. There is a task with $v = 50$ and $s_* = (1)$ (the task requires only one skill). w_1 and w_2 sign up to do this and $s_{w_1} = s_{w_2} = (1)$. Their true costs are $c_1 = 1$ and $c_2 = 5$, respectively. If they bid their true cost ($b_1 = 1$ and $b_2 = 5$), obviously, w_1 is selected and her payment is $p_1 = 1$. In this case, $u_1 = p_1 - c_1 = 0$. However, if w_1 bids a higher value than her true cost, e.g., $b'_1 = 3$, she will still be selected, and her payment is $p'_1 = b'_1 = 3$ for this time. Then, her utility becomes $u'_1 = p'_1 - c_1 = 3 - 1 = 2 > u_1$. Therefore, under this mechanism, in order to increase her utility, w_1 may overbid instead of bidding her true cost. ☐

B. A Greedy Mechanism

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

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

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

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

Based on this intuition, we devise Mechanism 2 (we refer to it as GREEDY). Line 2 to line 9 select a team or traverse all the workers based on the heuristics described above. Every time, when we find such a worker with the minimum cost, we have to check if the value of the task can cover the payment to this worker (we will not select a worker if the task is not profitable after selecting this worker), in order to guarantee that the task is profitable (line 4 to line 8). We repeatedly selecting workers until a satisfactory team is found or all the workers have been considered. If a satisfactory team is found, we pay each worker her bid (line 10 to line 12); otherwise the task is abandoned (line 14).

Lemma 5. *GREEDY is Computationally Efficient.*

Proof. Every time, selecting the worker with minimal value of $\frac{b_i}{\Delta_i(S)}$ takes $O(n)$ time. We have to consider at most n workers. Therefore, the complexity of this mechanism is $O(n^2)$. ☐

Lemma 6. *GREEDY is Individually Rational.*

Proof. A rational worker bids the task with a value no less than her cost to perform this task, i.e., $b_i \geq c_i$. Therefore, the utility for each selected worker is $u_i = p_i - c_i \geq p_i - b_i = 0$. ☐

Lemma 7. *GREEDY is Profitable.*

Proof. Every time we select a work, we check if the remaining value of the task can cover the payment of this worker. If not, the worker will not be selected. This process guarantees that the value of this task can cover the payment to the whole team. ☐

Mechanism 2: Small Bid First and Pay As You Bid (GREEDY)

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

```
1  S ← ∅; p_{1~n} ← 0;
2  repeat
3      w_i ← argmin_{w_i ∈ W\S} b_i / Δ_i(S);
4      if v > b_i then
5          S ← S ∪ {w_i};
6          v ← v − b_i;
7      else
8          W ← W\{w_i};
9  until s_S cover s_* or W\S = ∅;
10 if s_S cover s_* then
11     foreach w_i ∈ S do
12         p_i ← b_i;
13 else
14     S ← ∅
15 return (S; p_1, ..., p_n)
```

Lemma 8. *GREEDY is Not Truthful.*

Proof. Since this mechanism pays each selected worker exactly her bid. Every worker has incentive to bid a higher value than her true cost (the counter example in OPT shows that this mechanism is not truthful). ☐

C. A VCG based Mechanism

We adapt the traditional Vickrey-Clarke-Groves (VCG) mechanism [13] to our problem, and design a mechanism called VCG.

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

$$S = \arg\min_{T} \sum_{w_i \in T} c_i \tag{8}$$
$$s.t. \ \ s_S[j] \geq s_*[j], \quad j = 1, ..., l$$

Payment rule. VCG pays each selected worker w_i the d-ifference 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:

$$p_i = (\min_{T} \sum_{w_j \in T \wedge w_i \notin T} c_j) - \sum_{w_j \in S \wedge j \neq i} c_j \tag{9}$$
$$s.t. \ \ s_T[j] \geq s_*[j], \quad j = 1, ..., l$$

where S is defined in (8).

We devise a mechanism (in Mechanism 3) based on the VCG. We refer to this mechanism as VCG. In line 1, Mechanism 3 selects all the teams which are able to complete the task (the same as Mechanism 1 does). Line 2 to line 3 initialize necessary variables. Line 4 to line 6 select the team of workers according to the allocation rule of VCG. Line 7 to line 8 compute the payment to each worker according to the payment rule of VCG. The task is abandoned if it is profitable (line 9 to line 10).

Mechanism 3: VCG mechanism

Input: $b_{1\sim n}$, $s_{w_1 \sim w_n}$, v, s_*
Output: S, $p_{1\sim n}$
1 $D \leftarrow$ all possible teams that can cover s_*;
2 $S \leftarrow \emptyset$; $p_{1\sim n} \leftarrow 0$;
3 $T \leftarrow \emptyset$;MIN $= \infty$;
4 **foreach** $T \in D$ **do**
5   **if** $\sum_{w_i \in T} b_i < MIN$ **then**
6         $S \leftarrow T$; MIN $= \sum_{w_i \in T} b_i$
7 **foreach** $w_i \in S$ **do**
8     $p_i \leftarrow \min_{T \in D} \sum_{w_j \in T \wedge w_i \notin T} b_j - \sum_{w_j \in S \wedge j \neq i} b_j$
9 **if** $v - \sum_{w_i \in S} p_i < 0$ **then**
10     $S \leftarrow \emptyset$; $p_{1\sim n} \leftarrow 0$
11 **return** $\{S; p_1, ..., p_n\}$

Now, let us checking whether this mechanism satisfies the pre-defined properties.

Lemma 9. *VCG is not Computationally Efficient.*

Proof. Running time of VCG is $O(n2^n)$, where n is number of workers. There are $O(2^n)$ teams to consider, therefore it takes $O(2^n)$ to select the cheapest team. To decide the payment of each selected worker w_i, it takes again $O(2^n)$ time to find the cheapest team that excludes w_i. The total time complexity is $O(n2^n)$. ☐

Lemma 10. *VCG is Individually Rational.*

Proof. If worker w_i is selected, then

$$\begin{aligned} u_i &= p_i - c_i \\ &= \min_T \sum_{w_j \in T \wedge w_i \notin T} c_j - \sum_{w_j \in S \wedge j \neq i} c_j - c_i \\ &= \min_T \sum_{w_j \in T \wedge w_i \notin T} c_j - \sum_{w_j \in S} c_j \\ &\geq 0 \end{aligned} \tag{10}$$

Note that $b_i = c_i$ because this is a truthful mechanism (we will prove it later). The last inequality holds because team S is overall the cheapest team. ☐

Lemma 11. *VCG is Profitable.*

Proof. This is obvious since the task is abandoned if it is not profitable. ☐

Lemma 12. *VCG is Truthful.*

Proof.
- Assume that worker w_i is selected by bidding her true cost and she is also selected by bidding untruthfully. We want to compare her utilities when she bids truthfully and untruthfully, respectively. u_i, u_i', p_i, p_i' are the corresponding utilities, payments of w_i, when she bids truthfully and untruthfully, respectively. S' is the team of selected workers when w_i bids untruthfully.

$$\begin{aligned} u_i' &= p_i' - c_i \\ &= \min_T \sum_{w_j \in T \wedge w_i \notin T} c_j - \sum_{w_j \in S' \wedge j \neq i} c_j - c_i \\ &= \min_T \sum_{w_j \in T \wedge w_i \notin T} c_j - \sum_{w_j \in S'} c_j \end{aligned} \tag{11}$$

$u_i' \leq u_i$, since $\sum_{w_j \in S'} c_j \geq \sum_{w_j \in S} c_j$.
- When w_i is not selected when she bids truthfully (i.e., $u_i = 0$) and she is selected by underbidding.

$$\begin{aligned} u_i' &= p_i' - c_i \\ &= \min_T \sum_{w_j \in T \wedge w_i \notin T} c_j - \sum_{w_j \in S'} c_j \\ &= \sum_{w_j \in S} c_j - \sum_{w_j \in S'} c_j \leq 0 \end{aligned} \tag{12}$$

Therefore. a worker gets her maximum utility by bidding her true cost. ☐

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 TruTeam (Mechanism 4) which is a mechanism that satisfies all the four properties; i.e., it is computationally efficient, individually rational, profitable, and truthful.

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.

Payment rule. The intuition of the payment rule is to pay each selected worker the highest cost she can report while still being selected [9]. 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 $W \backslash \{T \cup \{w_i\}\}$ (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

$$b_i \leq \frac{b_{j_1}}{\Delta_{j_1}(T)} \times \Delta_i(T) \tag{13}$$

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:

$$p_i' = \frac{b_{j_1}}{\Delta_{j_1}(T)} \times \Delta_i(T) \tag{14}$$

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

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 $T \cup \{w_{j_1}, w_{j_2}, ..., w_{j_k}\}$ (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.

$$p_i = \max_{j \in \{j_1, j_2, ..., j_k\}} \{\frac{b_j}{\Delta_j(T)} \times \Delta_i(T)\} \tag{15}$$

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

Example 1. *The task has value $V = 50$ and $s_* = (1, 1, 1)$. Four workers $W = \{w_1, w_2, w_3, w_4\}$ want to do this task. Their bids are $b_1 = 4, b_2 = 12, b_3 = 6, b_4 = 15$. $s_{w_1} = (1, 0, 0)$, $s_{w_2} = (0, 1, 1)$, $s_{w_3} = (1, 1, 0)$, and $s_{w_4} = (1, 1, 1)$.*

We select the first worker to perform the task: $\frac{b_1}{\Delta_1(\emptyset)} = 4/1$, $\frac{b_2}{\Delta_2(\emptyset)} = 12/2$, $\frac{b_3}{\Delta_3(\emptyset)} = 6/2$, $\frac{b_4}{\Delta_4(\emptyset)} = 15/3$. w_3 minimizes the value $\frac{b_i}{\Delta_i(S)}$ ($S = \emptyset$ for now). Therefore, w_3 is selected.

Now, we need to decide the payment for w_3. The mechanism works as follows.

$W' = \{w_1, w_2, w_4\}$. $\frac{b_1}{\Delta_1(\emptyset)} = 4/1$, $\frac{b_2}{\Delta_2(\emptyset)} = 12/2$, $\frac{b_4}{\Delta_4(\emptyset)} = 15/3$. Therefore, w_1 minimize the value $\frac{b_i}{\Delta_i(T)}$ ($w_i \in W', T = \emptyset$ for now). If w_3 wants to be selected, her payment is $p_3 = \max\{4/1 \times 2, 0\} = 8$.

$W' = \{w_2, w_4\}$. We compare the value $\frac{b_i}{\Delta_i(T)}$ between w_2 and w_4: $\frac{b_2}{\Delta_2(\{w_1\})} = 12/2$, $\frac{b_4}{\Delta_4(\{w_1\})} = 15/2$. w_2 minimizes the value $\frac{b_i}{\Delta_i(T)}$ ($w_i \in W', T = \{w_1\}$ for now). If w_3 wants to be selected, her payment is $p_3 = \max\{12/2 \times 1, 8\} = 8$.

The remaining value of the task, namely 50, can cover the payment to w_3, namely, $p_3 = 8$. Therefore, w_3 is the first selected worker, $S = \{w_3\}$.

We select other workers in the same way. Finally, the selected workers is $S = \{w_2, w_3\}$, and the payments are $p_2 = 15$, $p_3 = 8$.

Lemma 13. *TruTeam is computationally efficient, with a time complexity of $O(n^2 l)$.*

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, the complexity of this mechanism is $O(n^2 l)$. □

Lemma 14. *TruTeam is individually rational.*

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

$$\begin{aligned} b_i &\leq \frac{b_{j_1}}{\Delta_{j_1}(S)} \times \Delta_i(S) \\ &\leq max_{j \in \{j_1, j_2, ..., j_k\}} \{\frac{b_j}{\Delta_j(T)} \times \Delta_i(T)\} \\ &= p_i \end{aligned} \tag{16}$$

We assume a worker will not bid bellow her true cost, i.e. $b_i \geq c_i$ (In fact, we will show in Lemma 16 that $b_i = c_i$). Therefore, w_i's utility is $u_i = p_i - c_i \geq p_i - b_i \geq 0$. □

Lemma 15. *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. □

Lemma 16. *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 (15). 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.

In the following detailed proof, b'_i is an untruthful bid, p'_i and u'_i are the corresponding payment and utility of worker w_i when she bid b'_i.

- If w_i is selected by biding her true cost c_i:
  - If w_i overbids, i.e., $b'_i > c_i$
    * If w_i is not selected by overbidding, her utility drops to 0.
    * If w_i is still selected by overbidding. Suppose $\{w_1, w_2, ..., w_{i-1}, w_i, w_{i+1}, ..., w_q\}$ is the original selected team when w_i bids c_i, $\{w_1, w_2, ..., w_{i-1}, w_{j_1}, w_{j_2} ..., w_{j_x}, w_{j_y}, ..., w_{j_q}\}$ is the selected team if w_i did not participate. Then $p_i = \max_{j \in \{j_1,...,j_q\}} \{\frac{b_j}{\Delta_j(T)} \times \Delta_i(T)\}$. Let $\{w_1, w_2, ..., w_{i-1}, w_{j_1}, w_{j_2}, ..., w_{j_x}, w_i, ..., w_{q'}\}$ be the selected team when w_i overbids. $\{w_1, w_2, ..., w_{i-1}, w_{j_1}, w_{j_2} ..., w_{j_x}, w_{j_y}, ..., w_{j_q}\}$ is still the selected team if w_i did not participate. So, $p'_i = \max_{j \in \{j_y,...,j_q\}} \{\frac{b_j}{\Delta_j(T)} \times \Delta_i(T)\}$. We have $p_i \geq p'_i$ since $\{j_y, ..., j_q\} \in \{j_1, ..., j_q\}$.
  - If w_i underbids, w_i will be selected for sure. Suppose $\{w_1, w_2, ..., w_{i-x}, ..., w_i, w_{i+1}, ..., w_q\}$ is the original selected team when w_i bids c_i, $\{w_1, ..., w_{i-x}, ...w_{i-1}, w_{j_1}, w_{j_2}, ..., w_{j_q}\}$ is the selected team if w_i did not participate. Then $p_i = \max_{j \in \{j_1,...,j_q\}} \{\frac{b_j}{\Delta_j(T)} \times \Delta_i(T)\}$. Let $\{w_1, , ..., w_{i-x}, w_i, ..., w_{q'}\}$ be the selected team when w_i underbid, $\{w_1, ..., w_{i-x}, ...w_{i-1}, w_{j_1}, w_{j_2}, ..., w_{j_q}\}$ is still the selected team if w_i did not participate. So, $p'_i = \max_{j \in \{i-x+1,...,i-1,j_1,...,j_q\}} \{\frac{b_j}{\Delta_j(T)} \times \Delta_i(T)\}$. Obviously, $p_i \leq p'_i$. Worker w_i is the i^{th} selected worker when bidding c_i, because $\frac{b_j}{\Delta_j(T)} \leq \frac{c_i}{\Delta_i(T)}$ ($j \in \{1, 2, ..., i-1\}$). Therefore, $p'_i = \max_{j \in \{i-x,...,i-1,j_1,...,j_q\}} \{\frac{b_j}{\Delta_j(T)} \times \Delta_i(T)\} \leq \max\{c_i, max_{j \in \{j_1,...,j_q\}} \{\frac{b_j}{\Delta_j(T)} \times \Delta_i(T)\}\} = \max\{c_i, p_i\} = p_i$. Therefore, $p'_i = p_i$
- If w_i is not selected by biding its true cost c_i:
  - If w_i overbids, w_i is still not selected for the task.
  - If w_i underbids, i.e., $b'_i < c_i$
    * If w_i is still not selected, her utility keep unchanged.
    * If w_i is luckily selected by underbidding. Suppose $\{w_1, w_2, ..., w_{j_0}, w_{j_1}, ..., w_{j_q}\}$ is the selected team when w_i bids c_i. Originally, w_i is not selected since $\frac{b_j}{\Delta_j(T)} < \frac{c_i}{\Delta_i(T)}$ ($j \in \{1, 2, ..., j_q\}$). Suppose $\{w_1, w_2, ..., w_{j_0}, w_i, ..., w_{q'}\}$ is the selected team if w_i underbids. Set $\{w_1, w_2, ..., w_{j_0}, w_{j_1}, ..., w_{j_q}\}$ determines the payment to worker w_i, as $p'_i = \max_{j \in \{j_1,...,j_q\}} \{\frac{b_j}{\Delta_j(T)} \times \Delta_i(T)\} \leq \max_{j \in \{1,2...,j_q\}} \{\frac{b_j}{\Delta_j(T)} \times \Delta_i(T)\} < c_i$. The utility of worker w_i is $u'_i = p'_i - c_i < 0$.

TABLE II
PARAMETERS OF OUR TWO DATASETS

	n: no. of workers	l: no. of skills	(u, σ)	$[C_1, C_2]$
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 $n = 1000$	$(l/5, 0.4)$	$[1, v]$

To conclude, w_i has no incentive to bid a value other than her true cost. $\square$

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).
- **Social Welfare**: defined as the sum of all the players' utilities, including the requester and the workers in our case. If the task can not be completed, the social welfare is normalized as 0. Otherwise, the social welfare is

$$\begin{aligned} \pi &= U_R + \sum_{w_i \in W} u_i \\ &= v - \sum_{w_i \in S} p_i + \sum_{w_i \in S} (p_i - c_i) \\ &= v - \sum_{w_i \in S} c_i \end{aligned} \tag{17}$$

- **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 mechanisms that are computationally efficient (namely, GREEDY and TruTeam). The parameters of the two datasets are listed in Table II and explained bellow.

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, $b_i = c_i$. 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 in these four mechanisms.

To generate each worker's skill profile, we first generate the number of skills she has, using the normal distribution (u, σ). Suppose w_i has x skills, then we randomly choose x different skills out of all the l skills.

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 (i.e. when the number of required skill is 5, number of worker only reaches 25, it tasks VCG about 1 minute to get the results). However, all the four mechanisms achieve strictly positive requester's utility (e,f,g,h) and social welfare (i,j,k,l). And we also observed that GREEDY and OPT perform similarly as each other, so are TruTeam and VCG. 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, the 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.

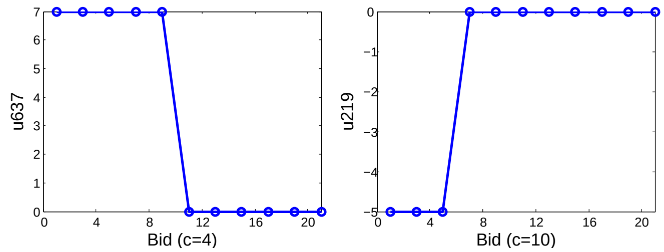
In Fig. 2(a) and 2(b), the running time of TruTeam contains a small peak when the number of workers is between 100 and 300. This happens when the task is complex (requiring skills as many as $l = 50$) and the number of workers is relatively small ($n < 300$). Thus the team size is fairly large and TruTeam 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 to perform a task. 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) yield 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 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) Social Welfare: Fig. 2 (i,j,k,l) presents the results on social welfare in different settings. When workers bid truthfully (i,k), GREEDY and TruTeam will select the same set of workers, and hence the social welfare of both mechanisms are supposed to be the same. Some trivial differences are caused by the randomization used in our Monte Carlo simulations, or in some instances, a valid team cannot be formed in one mechanism due to lack of competent and cheap workers but can be formed in the other.



(a) Utility of Worker w_{637} (b) Utility of Worker w_{219}

Fig. 3. Workers's Utility

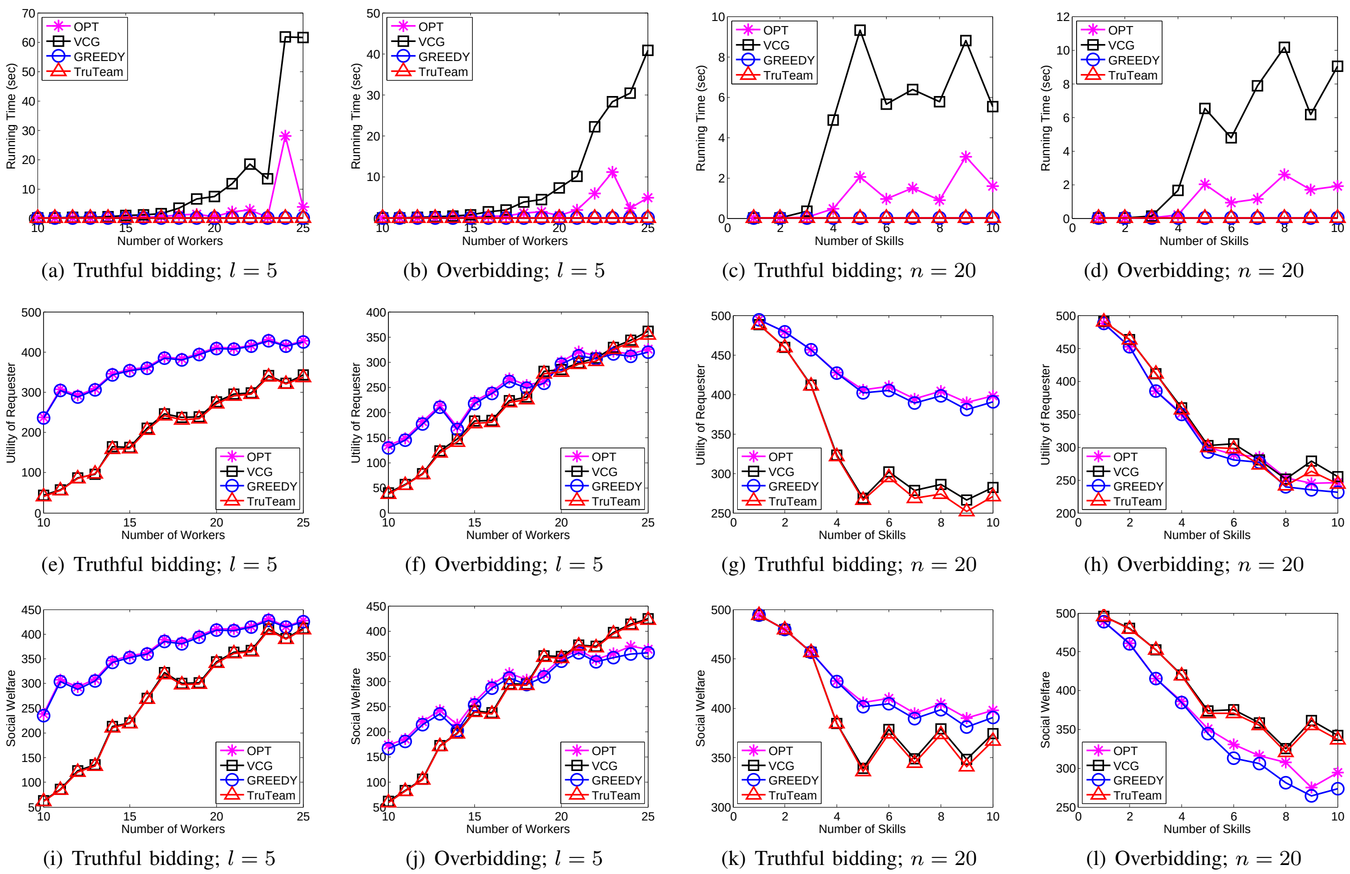
In the case that workers are strategic and thus overbid (under GREEDY), we observe in Fig. 2(j) 2(l) that TruTeam outperforms GREEDY, as a result of GREEDY forming a team of higher-cost workers. This demonstrates another benefit (in addition to requester's profit) of ensuring truthfulness (as in TruTeam).

4) 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 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 *dominate strategy* to bid her true cost in order to maximize her utility.

VI. Conclusion

In this paper, we formulate the problem of team formation in crowdsourcing markets, and solve it by providing four candidate incentive mechanisms: OPT, GREEDY, VCG, and TruTeam. Then, we prove that all the four mechanisms satisfy profitability and individually rationality. On the other hand, we disprove truthfulness for OPT and GREEDY, and prove that for VCG and TruTeam. Furthermore, by calculating the time complexity for the four mechanisms, we find that OPT and VCG are not efficient whereas GREEDY and TruTeam are. To derive quantitative intuitions, we carried out extensive simulations. The results demonstrate that among the four mechanisms, TruTeam is an efficient, profitable, individually rational and truthful pricing mechanism for team formation in crowdsourcing markets.



(a) Truthful bidding; $l = 5$ (b) Overbidding; $l = 5$ (c) Truthful bidding; $n = 20$ (d) Overbidding; $n = 20$

(e) Truthful bidding; $l = 5$ (f) Overbidding; $l = 5$ (g) Truthful bidding; $n = 20$ (h) Overbidding; $n = 20$

(i) Truthful bidding; $l = 5$ (j) Overbidding; $l = 5$ (k) Truthful bidding; $n = 20$ (l) Overbidding; $n = 20$

Fig. 1. Simulation results on the Small Dataset

REFERENCES

[1] Http://www.mturk.com.

[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] Y. Azar, "On-line load balancing," in *Online Algorithms*. Springer, 1998, pp. 178–195.

[5] 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.

[6] 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.

[7] M. De Weerdt, 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.

[8] D. DiPalantino and M. Vojnovic, "Crowdsourcing and all-pay auctions," in *ACM Conference on Electronic Commerce (EC)*. ACM, 2009, pp. 119–128.

[9] 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.

[10] 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.

[11] 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.

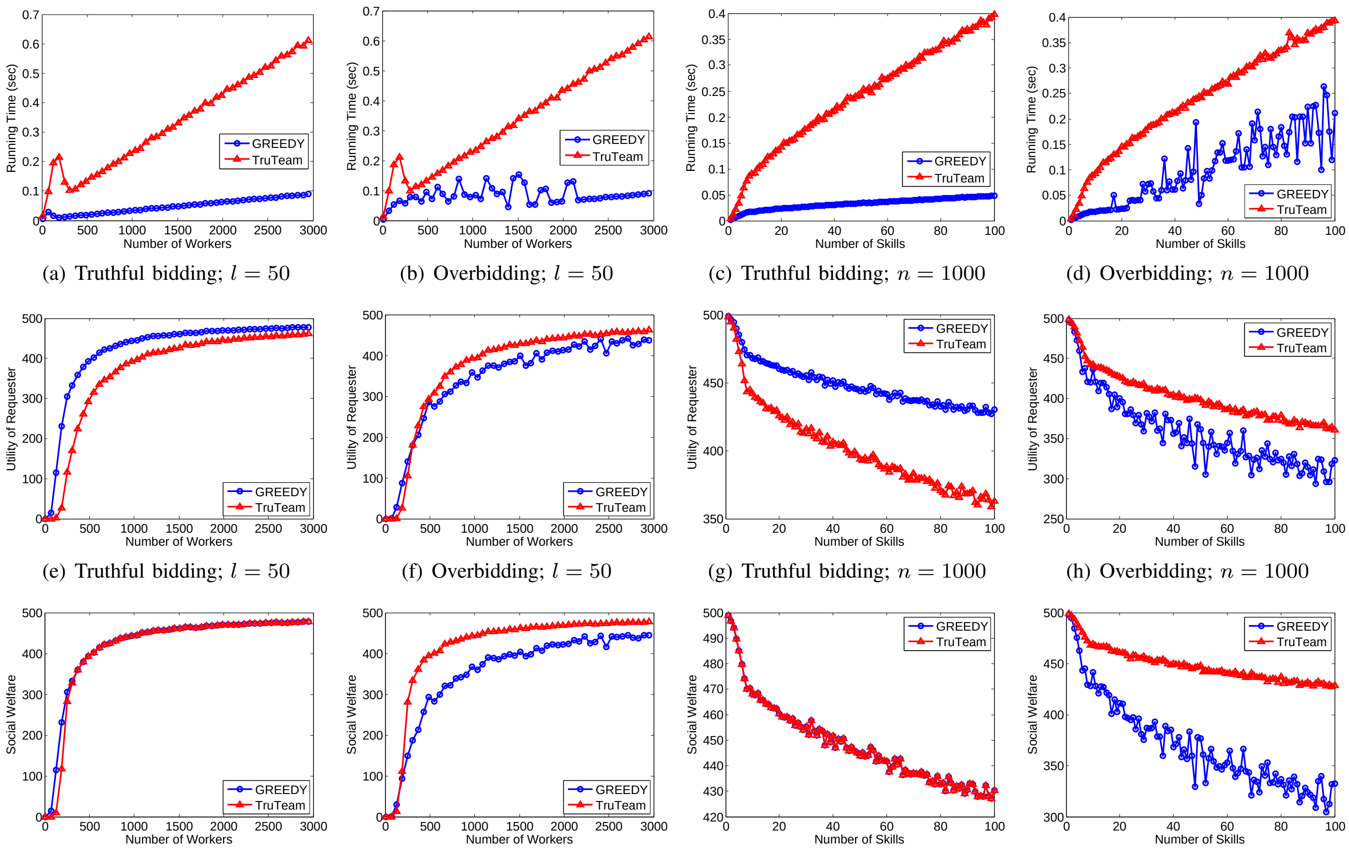
[12] W. Mason and D. J. Watts, "Financial incentives and the performance of crowds," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 2, pp. 100–108, 2010.

[13] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic game theory*. Cambridge University Press, 2007.

[14] Y. Singer, "Budget feasible mechanisms," in *51st Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2010*. IEEE, 2010, pp. 765–774.

[15] Y. Singer and M. Mittal, "Pricing mechanisms for crowdsourcing markets," in *Proceedings of the 22st International Conference on World Wide Web (WWW)*, 2013, pp. 1157–1166.

[16] A. Singla and A. Krause, "Truthful incentives in crowdsourcing tasks using regret minimization mechanisms," in *Proceedings of the 22st International Conference on World Wide Web (WWW)*, 2013, pp. 1167–1178.



(a) Truthful bidding; $l = 50$ (b) Overbidding; $l = 50$ (c) Truthful bidding; $n = 1000$ (d) Overbidding; $n = 1000$

(e) Truthful bidding; $l = 50$ (f) Overbidding; $l = 50$ (g) Truthful bidding; $n = 1000$ (h) Overbidding; $n = 1000$

(i) Truthful bidding; $l = 50$ (j) Overbidding; $l = 50$ (k) Truthful bidding; $n = 1000$ (l) Overbidding; $n = 1000$

Fig. 2. Simulation results on the Large Dataset