Theoretical Computer Science 803 (2020) 105–115

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

# Hardness of and approximate mechanism design for the bike rebalancing problem [☆]

Hongtao Lv [a], Fan Wu [a,*], Tie Luo [b], Xiaofeng Gao [a], Guihai Chen [a]

[a] *Department of Computer Science & Engineering, Shanghai Jiao Tong University, China*
[b] *Institute for Infocomm Research, A*STAR, Singapore*

ARTICLE INFO

*Article history:*
Received 26 December 2018
Received in revised form 7 April 2019
Accepted 18 July 2019
Available online 25 July 2019

*Keywords:*
Approximate mechanism
Sharing economy
Location truthfulness

ABSTRACT

Recently arose in the flourishing sharing economy, the *bike rebalancing* problem is a new challenge that concerns how to incentivize users to park bikes at system-desired locations that better meet bike demands. It can also be generalized to other location-based vehicle or tool sharing problems such as car, truck, drone, and trolley sharing. In this paper, we address this problem using an auction model under a crowdsourcing framework, where users report their original destinations and the bike sharing platform assigns proper relocation tasks to them in order to better balance the bike supply and demand. We first prove two impossibility results: (1) finding an optimal solution to the bike rebalancing problem is NP-hard, and (2) there is no approximate mechanism with bounded approximation ratio that is both truthful and budget-feasible. To overcome this barrier, we introduce two practical constraints and design a two-stage approximate mechanism that satisfies location truthfulness, budget feasibility, individual rationality, and achieves constant approximation ratio. To the best of our knowledge, we are the first to address two dimensional location truthfulness in the regime of mechanism design. In addition, our extensive experiments based on real-world dataset demonstrate that our proposed mechanism can effectively redress the imbalance of bike distribution.


## 1. Introduction

Sharing economy is booming worldwide in recent years. In particular, bike sharing has been deployed in a large number of cities globally and it had reached a million bikes in 2015.[1] While bike sharing offers great convenience and other benefits such as carbon footage reduction, it also gives rise to new challenges, among which a critical one is called the *bike rebalancing* problem. It refers to how to incentivize bike users to park bikes at desired locations so as to minimize the imbalance

☆ This work was supported in part by the National Key R&D Program of China 2018YFB1004703, in part by China NSF grant 61672348 and 61672353, in part by the Open Project Program of the State Key Laboratory of Mathematical Engineering and Advanced Computing 2018A09, and in part by Alibaba Group through Alibaba Innovation Research Program. The opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies or the government.

* Corresponding author.
*E-mail address:* fwu@cs.sjtu.edu.cn (F. Wu).

[1] https://www.citylab.com/city-makers-connections/bike-share/.

https://doi.org/10.1016/j.tcs.2019.07.030

between bike supply and demand. For instance, “hot spots” such as hospitals and shopping malls may have a large demand of bikes, yet meanwhile some other places such as train stations may have parked many unused bikes.[2]

This problem can also be generalized to other location-based vehicle or tool sharing scenarios such as sharing of cars, trucks, drones, and trolleys. In this paper, we seek to tackle this class of problems by designing an incentive mechanism to motivate users to park the vehicles or return tools at system-desired locations that minimize the imbalance between supply and demand. For the ease of description, we refer to this class of problems as the bike rebalancing problem.

We model this problem as a reverse auction where users report their original destinations as their bids and the platform determines a new location for them as a task with associated rewards. There are several challenges in this setting. First, users may misreport their original destinations (e.g., to be further away from “hot spots”) in order to get higher reward. Second, the platform has limited budget and hence wants to use it to the best efficiency, e.g., that maximizes the total revenue of platform. In addition, the mechanism should also satisfy some important and desired properties.

In this paper, we solve the above problem with the challenges. Specifically, our main contributions are summarized as follows:

- We characterize the imbalance between bike demand and supply using the Kullback-Leibler (KL) divergence, and formulate an auction design problem under a crowdsourcing framework.
- We prove two impossibility results: (1) the formulated bike rebalancing problem is NP-hard, and cannot be solved by the classic VCG mechanism; (2) there is no truthful and budget-feasible mechanism for this problem that can achieve a bounded approximation ratio.
- We propose a solution which is a two-stage approximate mechanism under two practical constraints that we introduce. This proposed mechanism achieves location truthfulness, budget feasibility, individual rationality, and constant approximation ratio. To the best of our knowledge, we are the first to study and guarantee 2-D location truthfulness in the regime of mechanism design.
- We conduct experiments using real-world data, and demonstrate the effectiveness of our mechanism as a viable solution.

## 2. Related work

Optimizing bike sharing systems has attracted much research effort [1–3]. A crowdsourcing mechanism that incentivizes users in the bike repositioning process is proposed by Singla et al. [4]. Pan et al. [5] studied a deep reinforcement learning algorithm, which determines the payment by learning from user behaviors. However, all these mechanisms [1–5] do not guarantee location truthfulness.

Singer [6] studied budget feasible mechanism and proved that, in general, the utility of buyers (in our case bike users) can be arbitrarily bad, but it is possible to achieve a bounded approximation for submodular uility functions.

In the field of crowdsourcing [7,8] and crowdsensing [9,10], a large body of works study the allocation and payment of spatial tasks [11–13], and some studies take the quality into consideration [14,15]. In these works, users report their cost for performing tasks, but in reality, users may not know their exact cost. In our work, users report their respective destinations instead, which would be more practical.

There are some works about location truthfulness in the area of facility location games [16–18]. These works are much different from our work because their mechanisms determines the location of a fixed facility for central system, not for respective mobile agents like ours. Moreover, they only study the scenario of one-dimensional locations, while we investigate 2-D locations.

## 3. The model

### 3.1. Preliminary

In the bike rebalancing problem, there is a set of $n$ users $N = \{1, 2, 3, \cdots, n\}$, and a set of $m$ discrete locations $M = \{1, 2, 3, \cdots, m\}$. Each user $i \in N$ who uses a bike needs to indicate or report her intended destination $d_i \in M$ on the map before or during the bike ride. Note that the platform do not consider users who have already reached destination and have just parked their bikes.

As explained in Section 1, it can result in serious imbalance of bike distribution if all the users simply park their bikes at their intended destinations. Hence, the bike sharing platform would like each user $i$ to park her bike at a system-desired location $l_i$ (rather than $d_i$) in order to better match bike demand. In return, since the users will have to travel extra distances, the platform offers each user $i$ an incentive $p_i$ if she accomplishes that task.

We formulate the problem under a crowdsourcing framework as follows. Each location $l \in M$ corresponds to a (relocation) task and each user $i \in N$ reports her intended destination as her bid (not necessarily truthfully). The travel distance between

[2] Such imbalance exists in both docked and dockless bike sharing systems. In the former case, bikes much be parked at prescribed stations. In the latter case, there are no specific stations and bikes can be parked almost freely.

any two points $x$ and $y$ is denoted by $H_{xy}$ and is known to both the platform and the users. The (extra) cost of user $i$ for parking her bike at location $l_i$ rather than her intended destination $d_i$ is denoted by $C_i$ or $C_{d_i l_i} = c \cdot H_{d_i l_i}$, where the constant $c$ is the unit travel cost, which we assume equal for all the users in the system. Hence, the utility of user $i$ who takes a task of location $l_i$ is $u_i = p_i - C_i$, where $p_i$ is the payment to be made to user $i$ and $C_i$ is the cost of user $i$. In addition, we assume that a user does not accept a task if the location $H_{l_i d_i} > h$, where $h$ is a constant.

The platform aims to design a mechanism under a budget $B$ to allocate desirable locations to users for balancing the demand and supply of bikes. The mechanism should satisfy the following properties:

- *Location truthfulness*: the utility of each user who bids truthfully should be no less than the utility if she misreports, i.e., $u_i(d_i, d_{-i}) \geq u_i(d_i', d_{-i})$.
- *Budget feasibility*: the payment to all the users should not exceed the budget, i.e., $\sum_i p_i \leq B$.
- *Individual rationality*: the utility of any user should be nonnegative, i.e., $p_i \geq C_i$.
- *Computational efficiency*: the algorithm used by the mechanism (to compute $l_i$ and $p_i$ for all the users) should terminate in polynomial time.

### 3.2. Problem formulation

We characterize the imbalance of bike distribution using the Kullback-Leibler (KL) divergence, which measures the expected logarithmic difference between two probability distribution $X$ and $Y$, as defined by

$$KL(X||Y) = \sum_i X(i) log \frac{X(i)}{Y(i)}. \tag{1}$$

The smaller the KL divergence is, the smaller the gap between $X$ and $Y$ is, and $KL(X||Y) = 0$ means that $X$ and $Y$ are identical probability distributions. In our case, we denote the bike demand by $Q(l)$, i.e., the proportion of bikes in demand at location $l$ among all the locations $M$. To characterize bike supply, we denote by $A_i$ the set of all the parked bikes before user $i$ parks her bike. Thus, by substituting the bike demand for $X(i)$, and supply for $Y(i)$, we can derive the contribution of a user $i$ who parks at location $l_i$ to be (the derivation is given in Appendix A):

$$\xi_{il_i} = Q(l_i) \log \frac{|A_i(l_i)| + 1}{|A_i(l_i)|} \tag{2}$$

Next, we introduce the notion of submodular function and symmetric submodular function [19].

**Definition 1** *([19]).* A function $V : 2^N \to \mathcal{R}_+$ is submodular if $V(S \cup \{i\}) - V(S) \geq V(T \cup \{i\}) - V(T), \forall S \subseteq T \subseteq N$.

Intuitively, it means that the marginal contribution of a user decreases when the set of winning users becomes larger.

**Definition 2** *([19]).* A function $V : 2^N \to \mathcal{R}_+$ is symmetric submodular if there exist $r_1 \geq \cdots \geq r_n \geq 0$, such that $V(S) = \sum_{i=1}^{|S|} r_i$, $\forall S \subseteq N$.

Symmetric submodularity means that the value of the function is only determined by the cardinality of the set, while the function is submodular at the same time.

Then, we have the following lemma:

**Lemma 1.** *For any given location l, the contribution of a user parking at location l, i.e., Equation (2), is a symmetric submodular function.*

The proof of Lemma 1 is given in Appendix B. Note that the rest of this paper generally applies to any symmetric submodular contribution function and not necessarily the specific form of Equation (2).

Now, we are able to formulate the bike rebalancing problem as

$$\begin{aligned} \mathbf{max} \quad & \xi = \sum_{i \in U, l_i \in M} \xi_{il_i} \\ \mathbf{s.t.} \sum_{i \in U} & p_i \leq B \\ & p_i \geq C_{d_i l_i}, \quad \forall i \in U \end{aligned} \tag{3}$$

where $U \subseteq N$ is the subset of "winning users" who are chosen to perform relocation tasks (i.e., to park bikes at system-specified locations), $p_i$ is the payment to be given to user $i$ and $C_{d_i l_i}$ is user $i$'s parking cost. We aim to design a mechanism (by specifying allocation rule and payment rule) to both solve the optimization problem (3) (i.e., maximize total contribution or revenue of the platform) and satisfy the four properties stated earlier.

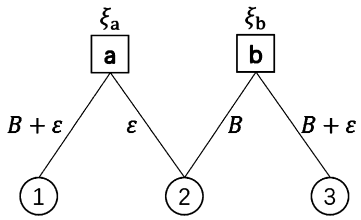


**Fig. 1.** An example that illustrates the impossibility result of Theorem 2, where circles denote users and boxes denote tasks.

## 4. Hardness results

### 4.1. NP-hardness

We prove that the problem (3) is NP-hard.

**Theorem 1.** *The bike rebalancing problem is NP-hard.*

**Proof.** We prove the decision version of the bike rebalancing problem is NP-hard. In the decision version, the question is whether there exists a subset of items $U$ that satisfies both $\sum_{i\in U, l_i\in M} \xi_{il_i} \geq K$ and $\sum_{i\in U} p_i \leq B$ for a given constant $K$.

We use reduction to NP-hardness from the *0-1 knapsack problem* which is a classic NP-complete problem, and is defined as follows.

**Definition 3** *(An Instance of 0-1 Knapsack Problem).* Given a set of $n$ items, each with a positive weight $w_i$ and a positive value $v_i$. Given maximum weight capacity $W$ and a constant $K$, the question is whether there exists a subset of items $U$ that satisfies $\sum_{i\in U} v_i \geq K$ and $\sum_{i\in U} w_i \leq W$.

We simplify the decision version of our problem to an instance where the acceptable range $h$ is small enough such that there is only one choice of $\hat{l_i}$ for each user $i$ and all the $\hat{l_i}$'s are non-overlapping. Thus, the quantities $v_i$, $w_i$, and $W$ in the 0-1 knapsack problem correspond to $\xi_{il_i}$, $p_i$, and $B$ in our case, respectively. Hence, the solution to the instance of the 0-1 knapsack problem is exactly the solution to the instance of our problem. In addition, the above reduction ends in polynomial time, which completes the proof. □

### 4.2. Impossibility of approximate mechanisms

Theorem 1 implies that VCG mechanism is not applicable due to the exponential time complexity of finding an optimal solution. One possible workaround is to make use of the available results in [6,20] where the authors proposed budget feasible approximate mechanisms for submodular utility functions as defined in Section 3.2. However, our problem does not satisfy submodularity: when the set of chosen users expands from $S$ to $T$, the (additional) user $i$'s marginal contribution may *increase* because the user $i$ may have multiple choices of tasks and the task allocated to her (and hence her contribution) may change when $S$ changes to $T$. Therefore, the mechanisms introduced in [6,20,21] cannot be directly used. In fact, we prove that there does not exist an approximate mechanism with bounded approximation ratio for our problem.

**Theorem 2.** *There is no approximate mechanism with a bounded approximation ratio that is truthful, budget feasible and individually rational simultaneously for the bike rebalancing problem.*

**Proof.** Let us consider an example shown in Fig. 1, where the bid of location has been easily converted to the bid of cost by calculating the distance. The bidding profile is $x = \{(B+\epsilon, \infty), (\epsilon, B), (\infty, B+\epsilon)\}$, where $\epsilon$ can be any positive number less than $B$. $\xi_a$ and $\xi_b$ are the contribution of fulfilling task $a$ and $b$ respectively, and $\frac{\xi_b}{\xi_a}$ can be arbitrarily large. In the optimal solution, location $b$ should be allocated to user 2, leading to a total contribution of $\xi_b$. We now show that any truthful, budget feasible and individually rational mechanism can achieve at most a total contribution of $\xi_a$.

Assume for the purpose of contradiction that there exists a mechanism $f$ that satisfies these properties and guarantees a bounded approximation ratio. Let's consider the case of bidding profile $y = \{(B+\epsilon, \infty), (\epsilon, B+\epsilon), (\infty, B+\epsilon)\}$, where user 2 declares $B+\epsilon$ instead for location $b$. In this case, the optimal solution will allocate location $a$ to user 2, so does the mechanism $f$. The reason is that (1) if $f$ allocates location $b$ to user 2, the cost of user 2 is above $B$, so it is neither budget feasible nor individually rational, and (2) if none of the locations $a$ and $b$ is allocated to user 2, then the total contribution is 0, and thus $f$ can not guarantee a bounded approximation ratio. Given this allocation, to achieve truthfulness, the payment to user 2 for parking at location $a$ has to be $B$ because, otherwise, user 2 can misreport $B$ for location $a$. Now, we can compare the bidding profiles $x$ and $y$. In the case of $y$, the utility of user 2 is $B-\epsilon$. In the case of $x$, if mechanism

$f$ allocates location $b$ to user 2, the utility of user 2 is at most 0, so she has incentive to misreport $B+\epsilon$ for location $b$ to change the bidding profile into $y$ to get better utility. Therefore, to ensure truthfulness, mechanism $f$ has two choices in the case of $x$: allocating location $a$ to user 2 or allocating nothing. In either case, the approximation ratio of total contribution is $\frac{OPT}{\xi} \geq \frac{\xi_b}{\xi_a}$ which can be arbitrarily large. Therefore, the mechanism cannot guarantee bounded approximation ratio, which constitutes the contradiction. □

## 5. A two-stage incentive mechanism

Due to the impossibility result of designing an approximate mechanisms in the general case, we introduce two assumptions about practical constraints.

**Assumption 1** *(Large Market Assumption).* We have $n \gg m$ and $C_{d_i l} \ll B$ for each user $i$ and each task location $l$ in the feasible range (i.e., $H_{l_i d_i} > h$ as mentioned in Section 3).

Intuitively, the assumption says that each individual user is negligible as compared to the budget. This assumption is widely used in previous work, e.g., [22,23].

**Assumption 2** *(Similarity Assumption).* For each user $i$ and each task location $l$ in the feasible range, let $\lambda_{il} = \frac{\xi_l}{c_i}$, then $\frac{\max\limits_{i,l} \lambda_{i,l}}{\min\limits_{i,l} \lambda_{i,l}} \leq \eta$.

This assumption says that the value per cost of all tasks do not differ too much (bounded by a known constant).

**Algorithm 1:** The Two-stage Mechanism

**Input:** $N$, $M$, set of bids $d = \{d_1, d_2, \cdots d_n\}$, the contribution of parking the $i$th bike at location $l$ $\xi_{il}$, the conflict network $G = (M, E, W)$, and the neighbor set $\Delta(l)$ (all the adjacent locations) of each location $l$.

**Output:** set of winning allocation $(i, l_i) \in U$, and payment $p_i$, for each winning user $i$.

$U_l \leftarrow \emptyset$, $L \leftarrow \emptyset$, $TC \leftarrow 0$, $N_l \leftarrow 0$;
**for** $l \in V$, $i \in N$ **do**
  **if** $H_{d_i l} \leq h$ **then**
    $N_l \leftarrow N_l \cup i$;
  **end if**
**end for**
**for** $l \in V$ **do**
  $\overline{\xi_l} = \sum_{1 \leq i \leq |N_l|} \xi_{il}$;
**end for**
Sort locations based on $\frac{\overline{\xi_l}}{|\Delta(l)|+1}$ into a list $M'$ in descending order;
**while** $M' \neq \emptyset$ **do**
  Let $l'$ be the head of the list;
  $L \leftarrow L \cup \{l'\}$, $M' \leftarrow M' \backslash \{\Delta(l') \cup l'\}$;
**end while**
**for** $l \in L$ **do**
  $B_l = \frac{\overline{\xi_l} \cdot B}{\sum_{l' \in L} \overline{\xi_l}}$;
  Sort users in set $N_l$ into a list $N_l'$ based on $C_{d_i l}$ in nondecreasing order, and let $j$ be the head of list $N_l'$;
  **while** $C_{d_j l} \leq \frac{B_l}{|U_l|+1}$ **do**
    $U_l \leftarrow U_l \cup (j, l)$, $N_l' \leftarrow N_l' \backslash i$;
    Let $j$ be the new head of $N_l'$;
  **end while**
  Let $i^*$ be the first unselected user in list $N_l'$;
  **for** $(i, l) \in U_l$ **do**
    $p_i = \min\{C_{d_{i^*} l}, \frac{B_l}{|U_l|}\}$;
  **end for**
**end for**

Under these two assumptions, we propose an approximate mechanism as the solution to the bike rebalancing problem. The main idea is to first convert the problem into a submodular problem, by choosing some representative locations and assigning one (or none) of these locations to a user. Note that this is not restrictive because, in reality, not all the locations but only a number of typical locations (e.g., train stations or residential areas) are short of bikes.

However, selecting representative locations is a maximum weighted independent set problem, which is an NP-complete problem [24]. Furthermore, assigning tasks to users to simultaneously achieve truthfulness and budget feasibility remains a challenge to solve.

To this end, we propose a two-stage incentive mechanism. In the first stage, we construct a conflict network $G = (M, E, W)$, where $M$ is the set of all locations, $E$ is the set of edges each connecting two locations if they are accessible to a single user, i.e., $E = \{(l_a, l_b) | \exists i \in N, C_{d_i l_a} \le h \wedge C_{d_i l_b} \le h;\ l_a, l_b \in M\}$, and $W$ is the set of location weights each denoting the maximum possible total contribution of users in the feasible range of a location, i.e., $W = \{w_l | w_l = \sum_{1 \le i \le |N_l|} \xi_{il}, \forall l \in M\}$ where $N_l$ denotes the set of users within the feasible range of location $l$. Then, we use a greedy method to find the maximum weighted independent set of locations.

In the second stage, we divide the budget $B$ among the subset of locations that are selected from $M$, and choose users for each location based on their reported destinations with associated payments determined by a critical price method. The complete procedure is presented in Algorithm 1.

In Algorithm 1, line 2-6 is to determine the candidates that are within the feasible range of each location. Line 7-14 determines a maximum weighted independent set of locations and allocates the budget proportionally to each location based on the number of users. Line 15-26 is to find the optimal set of winners in a greedy manner for each location, and the critical price $p_i = \min\{C_{d_{i^*l}}, \frac{B_l}{|U_l|}\}$ is used as the payment to $i$ where $i^*$ is the first unselected user.

In the following, we prove four important properties of our proposed mechanism: location truthfulness, individual rationality, budget feasibility and computation efficiency. First, to prove location truthfulness, we give the Myerson's Lemma:

**Lemma 2** *([25]). In single parameter auctions, for a normalized mechanism $\mathcal{M} = (f, p)$, where $f$ is the allocation rule and $p$ is the payment rule, $\mathcal{M}$ is truthful iff it satisfies:*

1. ***Monotone allocation:*** *$\forall i \in N$, if $C_i' \le C_i$, then $i \in f(C_i, C_{-i})$ implies $i \in f(C_i', C_{-i})$ for every $C_{-i}$;*
2. ***Threshold payment:*** *payment to each winning bidder is $\inf\{C_i : i \notin f(C_i, C_{-i})\}$.*

**Theorem 3.** *The two-stage incentive mechanism is location truthful.*

**Proof.** In the first stage, it's obvious that users cannot manipulate the selected locations because the sorting of locations only relies on the condition of locations rather than the bids of users. So, let user $i$ be a candidate of location $l$, if she misreports her destination $d_i' \ne d_i$, it must fall into one of following cases:

Case 1: $H_{d_i' l} > h$. In this case, user $i$ either becomes a candidate of another location $l' \ne l$, or fails to be a candidate. In the former scenario, based on non-overlapping characteristic between different selected locations, we have $H_{d_i l'} > h$, so it's beyond the acceptable range of user $i$. In the latter scenario, we easily have that user $i$'s utility $u_i(d_i', d_{-i}) = 0 \le u_i(d_i, d_{-i})$.

Case 2: $H_{d_i' l} \le h$. In this case, each user has at most one choice of task. Thus the problem is a *single parameter* mechanism design problem, in which each bidder has only one private value. So by Lemma 2, we only need to verify the monotonicity and threshold payment rule. The monotonicity is obvious since a user will be considered earlier if reporting $C_i' < C_i$. To verify the threshold payment rule, let $k = |U_l|, C_{i^*} = C_{d_{i^*l}}$, we first consider the case of $B/k \le C_{i^*}$. The user will not be allocated when misreporting a cost $C_i' > B/k$ otherwise it's not budget feasible. If a user declares a lower cost than $B/k$, then it still meets the allocation rule and will be allocated.

In case $C_{i^*} < B/k$, when reporting a cost $C_i' > C_{i^*}$, user $i$ will be considered after user $i^*$. We have $C_i' > C_{i^*} > B/(k+1)$ (otherwise user $i^*$ will be allocated in initial scenario), thus user $i^*$ will win while user $i$ will not be allocated. If declaring $C_i' < C_{i^*}$, $i$ will be a winner obviously. From the above, the threshold payment rule is satisfied, thus the mechanism is truthful. □

**Theorem 4.** *The two-stage incentive mechanism satisfies individual rationality.*

**Proof.** For an unselected user $i$, her payment and cost are both zero, so the utility $u_i = p_i - C_i = 0$. For a winning user $i$, by the line 18 in the algorithm, we have $C_i \le \frac{B_l}{|U_l|}$, and by the nondecreasing order of $N_l'$, we can get $C_i \le C_{i^*}$, where $i^*$ is the first unselected user. Therefore, we have

$$
\begin{aligned}
u_i &= \min\{C_{i^*}, \frac{B_l}{|U_l|}\} - C_i \\
&= \min\{C_{i^*} - C_i, \frac{B_l}{|U_l|} - C_i\} \\
&\ge 0. \quad \square
\end{aligned}
$$

**Theorem 5.** *The algorithm of the two-stage incentive mechanism has a polynomial-time computation complexity.*

**Proof.** The complexity of allocating users to adjacent locations (line 3-7) is $O(|M| \cdot |N|)$. The operation of sorting locations (line 10) is $O(|M| \cdot \log |M|)$. The computation complexity of determining winners for single location (line 17-24) is $O(|N_l| \cdot \log |N_l|)$, so for all selected locations, it's at most $O(|N| \cdot \log |N|)$. Since we have $|N| > \log |M|$, the overall complexity of the two-stage mechanism is $O(|N| \cdot (|M| + \log |N|))$ which is a polynomial-time complexity. □

**Theorem 6.** *The two-stage incentive mechanism is budget feasible.*

**Proof.** In the mechanism, the given budget is divided for each selected location, so we only need to prove the mechanism for each single location is budget feasible. For location $l$ and the set of selected users $A_l$, the price is $\min\{C_{d_{i^*l}}, \frac{B_l}{|U_l|}\}$ where $i^*$ is the first unselected user, so we have

$$\begin{aligned}\sum_{i\in N} p_i &= \sum_{l\in L}\min\{C_{d_{i^*l}}, \frac{B_l}{|U_l|}\}\cdot|U_l| \\ &\le \sum_{l\in L}\frac{B_l}{|U_l|}\cdot|U_l| \\ &= B\end{aligned}$$

which proves the budget feasibility. □

Lastly, to prove the theoretical guarantee of constant approximation ratio, we first introduce a result from the graph theory literature as the following lemma:

**Lemma 3** *([24]). The algorithm of selecting maximum weighted independent set of locations (line 2-14) guarantees that* $\sum_{l\in L}\overline{\xi_l} \ge \frac{\sum_{l\in M}\overline{\xi_l}}{\overline{\Delta}+1}$*, where* $\overline{\Delta}$ *is the maximum number of neighbors.*

**Theorem 7.** *The two-stage incentive mechanism achieves a constant approximation ratio* $2\eta\max\{\eta, \overline{\Delta}+1\}$ *under Assumptions 1 and 2.*

**Proof.** In our mechanism, the weight of locations are denoted as the total contribution containing all feasible users. Thus, we have

$$OPT \le \sum_{l\in M}\overline{\xi_l}. \tag{4}$$

Based on Lemma 3, we can observe that

$$\sum_{l\in M}\overline{\xi_l} \le (\overline{\Delta}+1)\cdot\sum_{l\in L}\overline{\xi_l}. \tag{5}$$

Note that $\overline{\Delta}$ is a constant that is no more than the number of locations within range $2h$ in our problem. Let $OPT_L$ denote the optimal solution with location set $L$, thus we have two cases for $OPT_L$: the budget is fully exhausted or not. If the budget is fully exhausted, based on Assumption 2, we have $OPT_L \ge \frac{1}{\eta}\cdot OPT$. If the budget is not fully exhausted, we can get $OPT_L = \sum_{l\in L}\overline{\xi_l}$. Combining (4), (5), we have that

$$OPT_L \ge \frac{OPT}{\max\{\eta, \overline{\Delta}+1\}}. \tag{6}$$

Moreover, let the optimal solution of a special selected location $l$ be $OPT_l$. It should be noted that in the optimal solution $OPT_l$, the payment for every winning user should be her cost for the task, i.e., we don't need to consider truthfulness here. Similarly, if the budget is fully exhausted for all locations, we get that

$$\sum_{l\in L} OPT_l \ge \frac{1}{\eta}\cdot OPT_L. \tag{7}$$

Otherwise, from Assumption 1, we have that $\sum_{l\in L} OPT_l = OPT_L$.

Next, we show the lower bound of our mechanism:

$$\sum_{i\in U_l}\xi_{il} \ge \frac{1}{2}\cdot OPT_l. \tag{8}$$

Assume for purpose of contradiction that there are $\ell$ users that are selected by $OPT_l$, while our mechanism selects less than $\ell/2$. If we denote the cost of the $i$th user in the initial list $N_l'$ as $C_i$, then we have $C_{\lceil \ell/2\rceil} > 2B_l/l$. However, since we have that $C_{\lceil \ell/2\rceil} \le \cdots \le C_\ell$, and $\sum_{\lceil \ell/2\rceil \le i\le \ell} C_i \le B_l$, we can get that $C_{\lceil \ell/2\rceil} \le 2B/\ell$, which is a contradiction. Note that the contribution function of locations is symmetric submodular, thus we can easily get equation (8).

Combining equations (6), (7) and (8), we get that

$$\sum_{i\in U_l}\xi_{il} \ge \frac{1}{2\eta\max\{\eta, \overline{\Delta}+1\}}\cdot OPT \tag{9}$$

which concludes our proof. □

## 6. Evaluation

We evaluate the effectiveness of our proposed mechanism using a real-world dataset from Mobike,[3] which is a popular bike sharing company in China.

We extract the bike supply and demand information from the dataset which covers $32 \times 32$ regions of Beijing city from 10th to 14th May 2017, with each region being 1.2 km $\times$ 0.6 km. We then build a simulator that generates the number of parking users and the level of demand for each region based on the numeric distribution of this dataset.

### 6.1. Experimental setup

Consistent with our model, we use regions to represent "locations" in our mechanism. We assume that a user who demands a bike only chooses bikes within her region (location). We further divide each region into $8 \times 4$ small regions such that each small region denotes a point on the map. We assume that all the parking users in each region are uniformly distributed over the points of that region. The distance between a user and a location is calculated using their respective centers.

The parameter values are set as follows. The cost of unit distance for each user is $c = 1$ RMB/km, and the maximum acceptable range $h = 2$ km. Unless otherwise specified, the number of existing bikes is 2000, the number of parking users is 3000, and the number of bikes in demand is 4000. We perform each experiment for 50 times and present the average value.

We compare three mechanisms: our proposed two-stage approximate mechanism (TSA), a randomized mechanism (RAN) and a randomized mechanism with selected locations (RAN-SL). In RAN, one user is chosen randomly in each round, and among all her nearby locations with higher demand than her reported location, the platform randomly picks one to allocate to this user and pays her the maximum possible cost $p_i = c * h$ for performing that task. This is repeated until the budget $B$ is used up. RAN-SL is a combination of TSA and RAN, which selects an independent set of locations the same way as in our mechanism, and then, in each round, the platform randomly chooses a location, and allocates it to a randomly chosen candidate user for that location as a task. The payment for each task is also the maximum possible cost $p_i = c * h$, and this process is repeated until the exhaustion of budget $B$. It should be noted that the RAN-SL mechanism is truthful due to the location selection process and the fixed payment, while the RAN mechanism is not truthful.

We use *total contribution* (TC) and *successful service ratio* (SSR) as the evaluation metric. TC is defined as the sum of all the users' contribution, or equivalently the reduction of the KL divergence. Formally,

$$TC = \sum_{l \in M} Q(l) \log \frac{|A_0(l)| + |U_l|}{|A_0(l)|}.$$

SSR is the proportion of demand that is satisfied, which is similar to [4] and, formally,

$$SSR = \frac{\sum_{l \in M} \min\{D(l), |A_0(l)| + |U_l|\}}{\sum_{l \in M} D(l)}.$$

### 6.2. Results

Fig. 2 shows the comparison of total contribution (TC) of the above three mechanisms. Our mechanism TSA outperforms both RAN and RAN-SL, and the advantage is especially large when the budget is lower than 1000, which is of particular interest when budget is tight. Intuitively, we also see that the total contribution increases (equivalently, the imbalance of the bike distribution decreases) when the budget increases, since a larger budget allows for choosing more users to rebalance the bikes.

The comparison of the successful service ratio (SSR) with varying budgets is illustrated in Fig. 3. We observe that SSR of all the three methods increases with the increase of budget. Importantly, our method TSA outperforms the other methods, and the larger difference under tight budget again shows the *budget-saving* advantage of our mechanism.

We also investigate the effect of the number of parking users on TC and SSR. As shown in Fig. 4, the TC of our mechanism clearly surpasses the other mechanisms, and the benefit is more notable when the number of parking users is above 1000. Moreover, the TC of RAN stops increasing when the number of parking users exceeds about 700, which is because it has exhausted the budget. By contrast, this does not happen to RAN-SL and our TSA mechanism, which implies a good scalability of these two mechanisms.

Fig. 5 investigates the effect of the number of parking users on SSR. The performance of our mechanism is superior to RAN-SL and RAN when the number of parking users is greater than 1000, which again demonstrates the effectiveness of our mechanism, because in real life, the scale of bike sharing systems are usually large.

[3] https://mobike.com/global/.

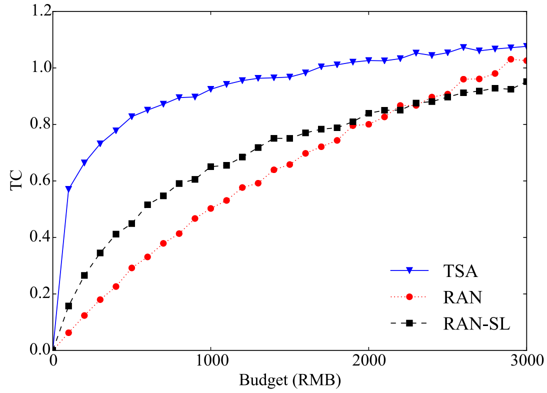


**Fig. 2.** Comparison on TC with varying budget.

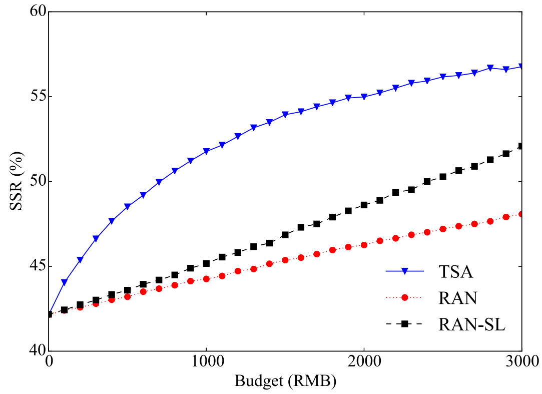


**Fig. 3.** Comparison on SSR with varying budget.

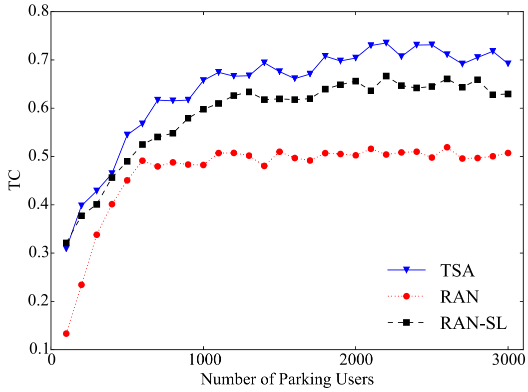


**Fig. 4.** Comparison on TC with varying number of parking users.

## 7. Conclusion

In this paper, we have studied the bike rebalancing problem as a typical case that can be generalized to other location-based vehicle or tool sharing applications in the emerging sharing economy. First, we have proved two impossibility results for optimal and approximate mechanisms. Then, to overcome these barriers, we have proposed a two-stage approximate mechanism by introducing two practical assumptions based on realistic observations. We have proved that this proposed mechanism satisfies desired properties (location truthfulness, budget feasibility, individual rationality) and achieves a constant approximation ratio. Finally, we show that the mechanism is effective and outperforms other choices via our extensive experiments based on a real-world dataset.

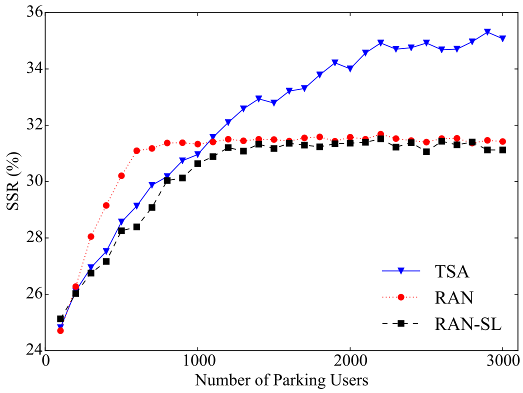


**Fig. 5.** Comparison on SSR with varying number of parking users.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Derivation of user contribution $\xi_{il_i}$

Let $A_0$ be the set of all the (already) parked bikes and $A_0(l)$ the subset of parked bikes at location $l \in M$. We assume that $A_0(l)$ and the demand distribution $Q(l)$ for all the locations $l \in M$ are known to the platform (e.g., through the mobile app and GPS). The destinations of users are continuous in the 2-D area, but locations of tasks $M$ are discrete points, each of which indexes a grid. In our case, we substitute $Q(l)$ for $X(i)$ (demand), and $\frac{|A(l)|}{|A|}$ for $Y(i)$ (supply), where $A(l)$ is the set of all the bikes already parked at location $l$ or will be (due to assignment of relocation task) parked at location $l$, and $A$ is the union of all the $A(l)$. We assume $|A_0(l)| > 0$ for all the locations (which is generally ensured as long as the grid is not too small) to avoid singularity.

From Equation (1) we get that

$$KL(A) = \sum_{l} Q(l) log \frac{Q(l)|A|}{|A(l)|} \tag{A.1}$$

where we omit $Q$ on the left hand side for notational convenience. Then we have

$$KL(A_i \cup (i, l_i)) = \sum_{l \neq l_i} Q(l) log \frac{Q(l)(|A_i| + 1)}{|A_i(l)|} + Q(l_i) log \frac{Q(l_i)(|A_i| + 1)}{|A_i(l_i)| + 1}. \tag{A.2}$$

In this work, our goal is to minimize the imbalance of bike distribution, namely the KL divergence, so we define the contribution of user $i$ as the difference between $KL(A_i)$ and $KL(A_i \cup (i, l_i))$. Based on equation (A.1) and (A.2), we have

$$\begin{aligned} \xi_{il_i} &= KL(A_i) - KL(A_i \cup (i, l_i)) \\ &= log \frac{|A_i|}{|A_i| + 1} + Q(l_i) log \frac{|A_i(l_i)| + 1}{|A_i(l_i)|}. \end{aligned}$$

Denote

$$\xi^1_{il_i} = log \frac{|A_i|}{|A_i| + 1}, \quad \xi^2_{il_i} = Q(l_i) log \frac{|A_i(l_i)| + 1}{|A_i(l_i)|}.$$

We can observe that the sum of the first item only depends on the total number of users $|N|$. Since our objective is to minimize the KL divergence, which is the total contribution of all the users, we can omit the first term $\xi^1_{il_i}$ because the sum of $\xi^1_{il_i}$ is a constant. Thus, we let $\xi_{il_i} = \xi^2_{il_i}$ in the paper.

## Appendix B. Proof of Lemma 1

Let $\xi_l$ denote the total contribution of location $l$, due to the monotonicity of function $\log \frac{x+1}{x}$, if $S_l \subseteq T_l$, we have

$$\xi_l(S_l \cup \{i\}) - \xi_l(S_l) = Q(l) \log \frac{|A_0(l)| + |S_l| + 1}{|A_0(l)| + |S_l|}$$

$$
\begin{aligned}
&\geq Q(l) \log \frac{|A_0(l)| + |S_l| + |T_l \backslash S_l| + 1}{|A_0(l)| + |S_l| + |T_l \backslash S_l|} \\
&= Q(l) \log \frac{|A_0(l)| + |T_l| + 1}{|A_0(l)|} - Q(l) \log \frac{|A_0(l)| + |T_l|}{|A_0(l)|} \\
&= \xi_l(T_l \cup \{i\}) - \xi_l(T_l).
\end{aligned}
$$

Moreover, the function of total contribution $\xi_l = Q(l) \log \frac{|A_0(l)|+|U_l|}{|A_0(l)|}$ depends on cardinality only. Therefore, the contribution of a single location is a symmetric submodular function.

## References


[1] T. Raviv, M. Tzur, I.A. Forma, Static repositioning in a bike-sharing system: models and solution approaches, EURO J. Transp. Logist. 2 (3) (2013) 187–229.
[2] M. Dell'Amico, E. Hadjicostantinou, M. Iori, S. Novellani, The bike sharing rebalancing problem: mathematical formulations and benchmark instances, Omega 45 (2) (2014) 7–19.
[3] S. Ghosh, P. Varakantham, Incentivizing the use of bike trailers for dynamic repositioning in bike sharing systems, in: Proceedings of the Twenty-Seventh International Conference on Automated Planning and Scheduling, ICAPS, 2017, pp. 373–381.
[4] A. Singla, M. Santoni, G. Bartók, P. Mukerji, M. Meenen, A. Krause, Incentivizing users for balancing bike sharing systems, in: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI, 2015, pp. 723–729.
[5] L. Pan, Q. Cai, Z. Fang, P. Tang, L. Huang, Rebalancing dockless bike sharing systems, arXiv preprint, arXiv:1802.04592.
[6] Y. Singer, Budget feasible mechanisms, in: 2010 Annual IEEE Symposium on Foundations of Computer Science, FOCS, IEEE, 2010, pp. 765–774.
[7] D. Zhao, X.-Y. Li, H. Ma, How to crowdsource tasks truthfully without sacrificing utility: online incentive mechanisms with budget constraint, in: IEEE INFOCOM 2014-IEEE Conference on Computer Communications, IEEE, 2014, pp. 1213–1221.
[8] T. Luo, S.S. Kanhere, S.K. Das, T. Hwee-Pink, Incentive mechanism design for heterogeneous crowdsourcing using all-pay contests, IEEE Trans. Mob. Comput. 15 (9) (2016) 2234–2246.
[9] X. Zhang, Z. Yang, W. Sun, Y. Liu, S. Tang, K. Xing, X. Mao, Incentives for mobile crowd sensing: a survey, IEEE Commun. Surv. Tutor. 18 (1) (2016) 54–67.
[10] T. Luo, S.S. Kanhere, J. Huang, S.K. Das, F. Wu, Sustainable incentives for mobile crowdsensing: auctions, lotteries, and trust and reputation systems, IEEE Commun. Mag. 55 (3) (2017) 68–74.
[11] Z. Feng, Y. Zhu, Q. Zhang, L.M. Ni, A.V. Vasilakos, Trac: truthful auction for location-aware collaborative sensing in mobile crowdsourcing, in: 2014 IEEE Conference on Computer Communications, IEEE, 2014, pp. 1231–1239.
[12] Q. Wang, Y. Zhang, X. Lu, Z. Wang, Z. Qin, K. Ren, Real-time and spatio-temporal crowd-sourced social network data publishing with differential privacy, IEEE Trans. Dependable Secure Comput. 15 (4) (2018) 591–606.
[13] P. Cheng, X. Jian, L. Chen, An experimental evaluation of task assignment in spatial crowdsourcing, PVLDB 11 (11) (2018) 1428–1440, https://doi.org/10.14778/3236187.3236196.
[14] D. Peng, F. Wu, G. Chen, Pay as how well you do: a quality based incentive mechanism for crowdsensing, in: Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing, ACM, 2015, pp. 177–186.
[15] S. Yang, F. Wu, S. Tang, X. Gao, B. Yang, G. Chen, On designing data quality-aware truth estimation and surplus sharing method for mobile crowdsensing, IEEE J. Sel. Areas Commun. 35 (4) (2017) 832–847.
[16] A.D. Procaccia, M. Tennenholtz, Approximate mechanism design without money, in: Proceedings of the 10th ACM Conference on Electronic Commerce, ACM, 2009, pp. 177–186.
[17] P. Serafino, C. Ventre, Truthful mechanisms without money for non-utilitarian heterogeneous facility location, in: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI, 2015, pp. 1029–1035.
[18] C.K.K. Fong, M. Li, P. Lu, T. Todo, M. Yokoo, Facility location games with fractional preferences, in: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, AAAI, 2018, pp. 1039–1046.
[19] W. Vickrey, Counterspeculation, auctions, and competitive sealed tenders, J. Finance 16 (1) (1961) 8–37.
[20] N. Chen, N. Gravin, P. Lu, On the approximability of budget feasible mechanisms, in: Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics, 2011, pp. 685–699.
[21] X. Bei, N. Chen, N. Gravin, P. Lu, Budget feasible mechanism design: from prior-free to bayesian, in: Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing, ACM, 2012, pp. 449–458.
[22] R. Vaze, Online knapsack problem and budgeted truthful bipartite matching, in: INFOCOM 2017-IEEE Conference on Computer Communications, IEEE, 2017, pp. 1–9.
[23] R. Gomes, V. Mirrokni, Optimal revenue-sharing double auctions with applications to ad exchanges, in: Proceedings of the 23rd International Conference on World Wide Web, ACM, 2014, pp. 19–28.
[24] S. Sakai, M. Togasaki, K. Yamazaki, A note on greedy algorithms for the maximum weighted independent set problem, Discrete Appl. Math. 126 (2–3) (2003) 313–322.
[25] R.B. Myerson, Optimal auction design, Math. Oper. Res. 6 (1) (1981) 58–73.