Maximizing Spatial-temporal Coverage in Mobile Crowd-Sensing Based on Public Transport with Predictable Trajectory

Chaowei Wang, *Member, IEEE,* Chensheng Li,Cai Qin,Weidong Wang, *Member, IEEE*,Xiuhua Li

***Abstract*—Mobile crowd-sensing is a prospective paradigm for smartphone, which collects ubiquitous data easily in metropolis. Nowadays, public transports (PTs) are mostly widely used and affordable in many urban. A PT embedded with substantial sensors also can be adopted as participator in crowd-sensing, but distinct from smartphones, the trajectory of PTs are schedulable and its location is predictable, which shed an opportunities to achieve high quality crowd-sensing. However, the existing crowd-sensing application are mainly based on the current locations of the participants and the quality of it highly depends on the spatial-temporal coverage which is easily weakened by the position of participates. Therefore, based on the predictable trajectory of PTs, we design a novel model and present an approximation algorithm to select PTs to participate in urban sensing for maximizing spatiotemporal coverage with constrained sensing reward. We theoretically prove that the selection of PTs problem is NP-hard, the proposed algorithm can achieve a performance guarantee more approximate to 1. The performance of our algorithm is simulated with real T-Drive trajectory dataset. The results show that our algorithm achieves a good coverage closer to optimum and outperform some exiting alternative algorithms.**

***Index Terms*—Mobile crowd-sensing, schedulable trajectory, spatial-temporal coverage, approximation algorithm, performance guarantee.**

I. Introduction

WITH the rapid advance of sensing, communication, and mobile computing, mobile crowd-sensing [1] has become a paradigm attracting much attention for collecting distribute sensory data and distributing to general public. With the help of mobile crowd-sensing the cost of data collection and dissemination tasks over wide range of region can be significantly reduced. Being carried by users locating in different place, smartphones can easily collect ubiquitous data and share data with potential users in neighborhood[2][3]. The vehicle-based mobile crowd-sensing with similar mobility, distributed in large is evolving rapidly. Equipped with various onboard sensors such as GPS, video cameras, gas sensor and also communication module, a vehicle becomes a powerful crowd-sensing application like as a smartphone to collect data and carry on a various of sensing task, including traffic monitoring [4][5], environment monitoring [6],and urban Wi-Fi characterization [7], etc.

Chaowei Wang, Chensheng Li, Cai Qin, Weidong Wang and Xiuhua Li are with School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, China, 100876, (e-mail: {wangchaowei, lichensheng, qincai, wangweidong,lixiuhua}@bupt.edu.cn).

Manuscript received XXXX XX, 2018; revised XXXX XX, 2018.

A vehicle-based mobile crowd-sensing system is typically composed of two parts[8]: cloud management platform (CMP) and mobile vehicles embedded with crowd-sensing application. An example of vehicle-based crowd-sensing is shown in Fig.1. The cloud management platform is responsible for selecting a set of vehicles to complete crowd-sensing task and processing data forwarded to user. Once a vehicle receives authorization of the CMP, it collects the required data and then upload to CMP.

Generally, it is greatly important to select vehicles to participate in collaborative sensing, which manifest the success of vehicle-based mobile crowd-sensing. Assuming an extreme case that the CMP select all vehicles to execute crowd-sensing task, apparently, it can perceive the surroundings and achieve what it is assigned to do, but multiple vehicles in the same region at the same time introduces data redundancy due to a single vehicle is sufficient to cover a geographical region. Therefore, the vehicle usually receive credit or non-monetary reward from CMP [9][11] with constraint budget, and selects a set of vehicle from all vehicles under operation that meet the users requirements for better crowd-sensing performance[11].

Another hand, the location of vehicles have great influence on the quality of vehicle-based mobile crowd-sensing [12], because of the CMP assigns tasks to vehicles that operates in different regions. There is another extreme case in which no vehicle is operating in the region of interest at a specific time that result in blank data. From the above, the quality of crowd sensing is sensitive to location and time, the spatial-temporal coverage (STC) is a fundamental metric of the vehicle-based mobile crowd sensing quality [1]. Particularly, STC intends to cover as many areas of interest as possible or make sure all areas is covered at least once for a period of time.

In reality, we are supposed to be aware that the STC vehicle-based mobile crowd-sensing is more dynamic on account of each vehicle keeps moving persistently as his own schedule. However, public transport, which is distinct from taxi or private car without operating plans, strictly follow an explicit timetable made by operator. Hence the location of each PTs are predictable in spite of the highly dynamic mobility, which shed a chance to achieve high quality crowd-sensing. By taking into consideration of the future location of PTs can effectively prevent the quality of crowd-sensing from affecting for high mobility instead of only depend on current location as smartphone-based crowd-sensing does[13].

In this paper, we concentrate on how to achieve a high quality of crowd-sensing with full use of the predictable

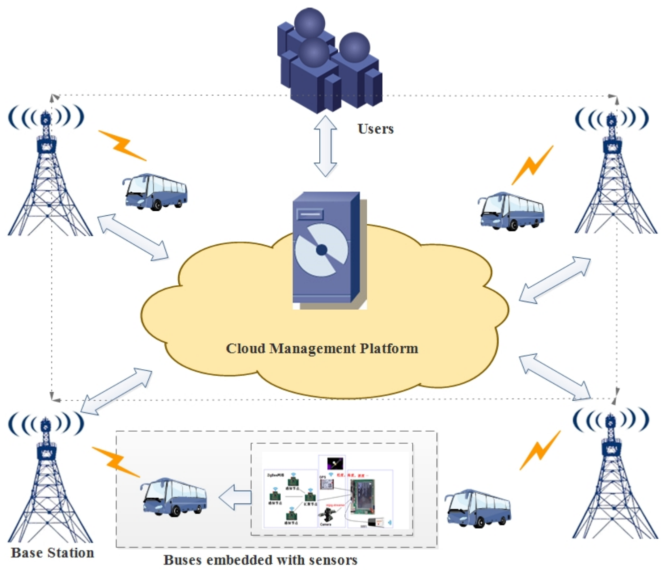

Fig. 1: An example of vehicle-based crowd sensing application.

location of transport vehicles. After analysis on the relations between STC and the location of PTs. To obtain a perfect quality we formulate the problem as maximizing STC with constrained budget. Through thoroughly proof, we find that the problem is NP-hard. And we design a truthful and efficient approximation algorithm, called ECQA, to select a set of PT from all candidates under operation with a high efficiency (minimal budget and maximizing STC), which can approximate the optimal solution within a guarantee performance no less than $\left(1-e^{-1}\right)$, with polynomial-time computation complexity. We also theoretically prove the ECQA guarantee is truthfulness.

This paper is organized as follows.Section II reviews the related work. Section III introduces the system model which considers the relation of the quality of crowd-sensing and STC with budget contrained. Section IV we propose a novel algorithm ECQA to solve the SVP and analyse the performance guarantee of it. Performance evaluation and analysis are provided in Section V. Finally, Section VI draws the conslusions of this paper.

II. RELATED WORK

In recent years, mobile crowd-sensing is now a significant source of information for smart city. Many researchers is dedicated to study the vehicular application of crowd-sensing, e.g., traffic accident evidence collection [14][15], city block monitoring [16] ,bike-net for cyclist experience mapping[17], and many architectures of crowd-sensing were designed. In [11] [18] [19] proposed a participants recruitment system and formulated the problem as a constrained coverage problem but ignore the mobility of vehicle. In [16] constructed a surveillance system based on vehicle with constraint network bandwidth. In [21] introduced a crowd-sensing service based on vehicle embedded with cameras to deliver images on demand to users. In [27] proposed an incentive mechanisms for participant recruitment who interact with a task requestor in a random order for maximizing the values of finished task. In[28][29] studied location-based crowd-sensing systems and major concern both the spatial and temporal coverage based on current location of participants. However, these schemas assumed that the initiator was capable to select candidate vehicles to finish task as much as possible and the location of vehicles are aware hypothetically, which is more suitable for the cases when the number of vehicles is not larger or the budget is unlimited and the vehicles are unmovable. Distinct from the problems above, we make an advance forward,to obtain a high quality of crowd-sensing, not only do we take the current and future location of candidate into account, but also highlight the budget, then we establish system model and formulate the spatial-temporal coverage for high quality of crowd-sensing as a optimization issue solved with a performance guarantee approximation algorithm. In the remainder of this paper, vehicle is mainly to denote PTs.

III. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

In a urban area, we divide region of interest into a serial of small segments. Let R denote the set of all small segments, $R = \{r_1, r_2, r_3, ..., r_k\}$. When the CMP broadcasts a crowd-sensing task to be finished for a period of time, i.e. T. We assume the time is discrete, and we can get $T = \{t_1, t_2, t_3, ..., t_m\}$.The distribution of vehicle is of large-scale, each vehicle equipped with the sensor module that we has designed in [22] is able to take crowd-sensing tasks. We assume there are n vehicles can perform sensing assignments and the set of vehicles is denoted by $V = \{v_1, v_2, v_3, ..., v_n\}$. Initially, the CMP predicts the current position of all vehicles according to the timetable and broadcast the data packet until receive the ACK, if the prediction is not consistent with the actual current location obtained by Global Positioning System (GPS) [23] employed in vehicle, it will be updated, respectively. With the initial location of vehicles and scheduled timetable, we can get the location of a vehicle v_i at a specific time t_j, which is denoted by $l_i(t_j) \in R$. Thus the trajectory of n vehicles can be represented as follows:

$$L(V) = \begin{bmatrix} l_1(t_1) & l_1(t_2) & ... & l_1(t_m) \\ l_2(t_1) & l_2(t_2) & ... & l_2(t_m) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ l_n(t_1) & l_n(t_2) & ... & l_n(t_m) \end{bmatrix} \tag{1}$$

where the size of $L(V)$ is $n \times m$.

In practice, as we are not anticipating that all vehicles are involved in crowd-sensing due to data redundancy. For example, in terms of traffic monitoring, nearby vehicles usually upload the similar information, which should be avoided. Therefore, we regulate that a vehicle selected to take part in crowd-sensing will gain a reward paid by CMP and the budget of CMP is limited to no morn than C_{max}. Next, we define the sensing reward.

Definition 1: Sensing Reward (SR) a vehicle is selected to complete tasks often associated with a reward. Let c_i denote the reward to v_i, which can be acquired through online bidding [29]. The reward vector C for all vehicles is:

$$C = \{c_1, c_2, ..., c_n\} \tag{2}$$

With the constraint of budget of CMP, not all vehicles participate in crowd-sensing, we adopt an indication vector Φ to imply whether a vehicle v_i is selected or not,

$$\Phi_i = \begin{cases} 1 & v_i \in \Omega \\ 0 & otherwise \end{cases} \tag{3}$$

where $\Omega \subseteq V$ is the set of selected vehicles. Let $C(\Omega)$ be the total reward to vehicles in Ω, which can be computed as

$$C(\Omega) = [C, \Phi] \tag{4}$$

As mentioned above, the quality of crowd-sensing is related with STC, which means to cover as many regions of interest as possible and ensure each road segment to be covered once at least within a sensing time T. We also introduce the notion of spatial-temporal coverage.

Definition 2: Spatial-temporal Coverage (STC) determines the quality of crowd-sensing. Formally, which can be defined as

$$\textbf{STC} = \sum_{t_j \in T} \bigcup_{v_i \in \Omega} (l_i(t_j)) \tag{5}$$

In [19], we have designed a hardware system, which can collect various information such as temprature, flow of traffic, longitude, latitude and so on, then we show an example to explain the implication of STC based on it.In Fig.2, if there is a user request to gather the flow of traffic at region R, which is divided into a serials of segments, that is $R =${AB, AD, BC, BE, DE, EF, EH, DH, CF}, the scheduled trajectory of Bus1, Bus2, Bus3, Bus4 is {BC, AB, AD, DE }, {BC, BE, EH },{EH, HD, AD, AB, BE },{EF, BE, AB, AD, DH }, respectively. In a period time $\{t_1, t_2, t_3, t_4\}$ the location of Bus1 to Bus4 is {BC, AD, DE, BC}, {BC, BE, BC, BE },{EH, HD, AB, BE}, {AB, BE, AD ,DH}, respectively. From equality (1), we get

$$L(V) = \begin{bmatrix} BC & AD & DE & BC \\ BC & BE & BC & BE \\ AB & BE & AB & BE \\ AB & BE & AD & DH \end{bmatrix} \tag{6}$$

If the CMP with limited budget is capable of selecting two vehicles to complete the task, then we consider two cases as bellows

$$\begin{aligned} STC(Bus1, Bus2) = &\underbrace{BC}_{t_1} + \underbrace{AD + BE}_{t_2} + \underbrace{DE + BC}_{t_3} \\ &+ \underbrace{BC + BE}_{t_4} \end{aligned} \tag{7}$$

$$STC(Bus3, Bus4) = \underbrace{AB}_{t_1} + \underbrace{BE}_{t_2} + \underbrace{AB + AD}_{t_3} + \underbrace{BE}_{t_4} \tag{8}$$

It can be seen that the set of {Bus1, Bus2} covers five different place in space and the segment of {BC} is covered triple over time. On the contrary, the set of {Bus3, Bus4} simply covers four different place in space, so we are more willing to select {Bus1, Bus2} with more accurate information to participate in crowd-sensing.

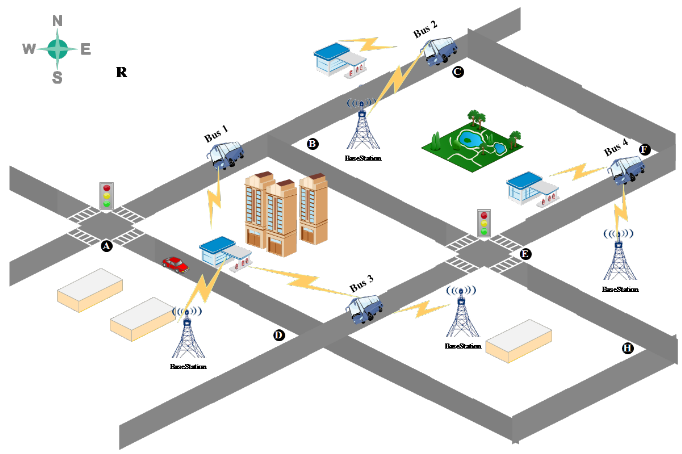

Fig. 2: An example explains the notion of spatial-temporal Coverage.

B. Problem Statement

In a region of interest, each vehicle equipped with amount of sensors which continuously sense the surrounding environment as it passes. However, at a specific moment, if all vehicles on a same road segment are involved in crowd-sensing, it will lead to overlapped coverage. Thus it is highly demanded to select an appropriate set to finish the crowd-sensing tasks and ensure the quality of crowd-sensing. Based on the system model, we are ready to formally define the problem of optimal selection of vehicle (SV) for maximizing the STC with SR budget constraint.

Definition 3: SV Problem (SVP) is to determine a set of vehicle under the budget constraint C_{max} with the objective of maximizing the spatial-temporal coverage.

$$\begin{aligned} &\textbf{max} \ STC(\Omega) \\ &\textbf{s.t.} \quad C(\Omega) \leqslant C_{max} \end{aligned} \tag{9}$$

Actually, the sensing data at different road segments and at a different period time may have varying importance degree, such as we are more interested in hotspot with high traffic flow in the morning rush hour. For this reason, we introduce priority power to indicate that the higher priority of a vehicle, which is more likely to be selected to join in crowd-sensing. Through analyzing historical data, it is easy to acquire traffic performance index (TPI) [24] indicated the congestion level of each road segment. Let $D_{t_j}^{l_i}$ denote the TPI of $l_i(t_j)$ at a specific time t_j, which is assumed known and normalized between 0 and 1, e.g. $D_{t_j}^{l_i} \in (0, 1]$. With the TPI we define priority power as follow:

Definition 4: Priority Power (PP) is the priority of a vehicle to be selected in a crowd-sensing period time, which is a function of $D_{t_j}^{l_i}$ defined as $W_{t_j}^{l_i}(D_{t_j}^{l_i})$. So $D_{t_j}^{l_i} \propto W_{t_j}^{l_i}$, thus the first order derivative of $W_{t_j}^{l_i}(D_{t_j}^{l_i})$ satisfies

$$\frac{\mathrm{d}W_{t_j}^{l_i}}{\mathrm{d}D_{t_j}^{l_i}} > 0 \tag{10}$$

Therefore, priority power is expressed as follows

$$W_{t_j}^{l_i} = \log_2(1 + D_{t_j}^{l_i}) \tag{11}$$

With the priority power, the STC can be redefined as

$$\sigma(\Omega) = \sum_{t_j \in T} \bigcup_{v_i \in \Omega} (l_i(t_j) \cdot W_{t_j}^{l_i}) \tag{12}$$

so the SVP can be rewritten as

$$\begin{aligned} &\textbf{max}\ \sigma(\Omega) \\ &\textbf{s.t.}\quad C(\Omega) \leqslant C_{max} \end{aligned} \tag{13}$$

We hope that the solution of SVP can be found with a time efficient, unfortunately, it is NP-hard even though the track of vehicles is predictable. In the next section, we will prove SVP is NP-hard and propose an improved approximation algorithm based on greedy to solve it.

IV. SOLUTION TO THE SVP

A. Complexity Analysis of SVP

Theorem 1. The SVP is NP-hard even though the trajectory of all vehicles are predictable.

Proof: To prove the NP-hard hardness of SVP, we should demonstrate it belongs to NP firstly,Assuming there is a possible solution $\Omega^{'}$, it is clearly that the correctness of this solution can be certified in polynomial, the time complexity of the checking algorithm is O(n), which means SVP is NP.

To show SVP is NP-hard, we can devise a reduction in polynomial time from budgeted maximum coverage problem (MCP) as the known NP-hard [26] to SVP. The decision version of MCP can be defined as follows.

Given a collection of sets $S = \{S_1, S_2, ..., S_n\}$, each set S_1 with a cost $c_i, 1 \leqslant i \leqslant n$ fetch values from $X = \{x_1, x_2, ..., x_m\}$ with associate with weights $w_i, 1 \leqslant i \leqslant n$. The problem is to find a set $S^{'} \subseteq S$ satisfied with the total cost does not exceed a budget B and the total weight of $S^{'}$ is maximized simultaneously. With all necessary condition of SVP, we make a relational mapping between MCP and SVP as follows:

$$x_i \overset{mapping}{\rightarrow} l_i(t_j),\ S_i \overset{mapping}{\rightarrow} L(\Omega^{'})$$

$$c_i \overset{mapping}{\rightarrow} C(\Omega^{'}),\ B \overset{mapping}{\rightarrow} C_{max}$$

where $\Omega^{'} \subseteq V$. As mentioned before, each vehicle has a priority power which can be mapped to w_i. We have reduced the decision version of MCP to the problem formulation of SVP successfully. So we can obtain a corresponding instance from SVP for any instance in MCP. As a result, the SVP is NP-hard.

Consequently, to achieve a truthful and computationally efficient crowd sensing system, it is highly demanded to propose an approximate algorithm for solving SVP.

B. Approximate Algorithm to Solve SVP

We have analyzed the NP-hardness of SVP, it becomes computationally impracticable to select an optimal set of vehicles from all the candidate vehicles when the number of candidate vehicles is huge. As for a huge metropolis like Beijing, the number of vehicles under operations is about 30,000 per day by the end of 2016 [24]. To achieve the desired computational efficiency, we propose an approximate algorithm called efficient combination query algorithm (ECQA) to solve SVP. While designing the algorithm, not only should we consider to select a vehicle with maximized STC, but also ask for less reward from CMP with budget constraint. Therefore we define the reward efficiency.

Definition 5: Reward Efficiency (RE) indicates the marginal STC achieve per unit reward.

The ECQA adopts a greedy strategy to solve the problem. The greedy policy is to select the next most reward effectiveness vehicle, which means to select a vehicle maximized marginal STC per unit reward, until the total SR exceed the budget of CMP. Mathematically, the reward-efficient for selecting a vehicle v_i can be computed as follows

$$E_i = \frac{\sigma(\Omega^{'} - \sigma(\Omega))}{c_i} \tag{14}$$

The algorithm tries many rounds, a best vehicle with maximum RE is determined as a result of each round. In equation (14), where E_i denotes the reward efficient of vehicle v_i, Ω is solution obtained from $V, \Omega^{'} = \{\Omega \cup v_i\}$ and $v_i \in V - \Omega$. The algorithm will not terminate until the budget constraint is active. The pseudo-code is listed in table 1.

The pseudo-code of ECQA 1

Input: set $V = \{v_1, v_2, v_3, ..., v_n\}$ of vehicle under operation, set SC $= \{c_1, c_2, c_3, ..., c_n\}$ sensing reward of each vehicle, C_{max} the budget constraint of CMP, an initial set S^0 of cardinality is an integer as 3, assume the schedule time of each vehicle is known.

Output: set Ω is the best set of vehicle selected by ECQA.

```
max ← 0
Ω ← ∅
S ← ∅
for S^0 ⊆ V, C(S^0) ⩽ C_max do
    S ← S^0
    for v_i ∈ V − S do
        S' ← {S ∪ v_i}
        E_i ← (σ(S') − σ(S)) / c_i
        if E_i > max  and  C(S') < C_max then
            S ← S'
            max ← E_i
        end if
        if σ(S) > σ()Ω) then
            σ ← S
        end if
    end for
end for
```

The ECQA has a performance guarantee $\rho \leqslant 1$, which indicates we can obtain a solution is ρ times of optimal solution in NP-hard problem [25]. The closer of value of ρ to 1, the more approximation to optimal solution. In this paper, the ECQA can achieve a lower bound ratio of $\left(1 - e^{-1}\right)$ when the cardinality as q of set S^0 is not less than three, i.e. $q \geqslant 3$. Next, we will prove the following theorem about the performance guarantee of ECQA.

Theorem 2. The ECQA can achieve a worst performance guarantee of $\left(1-e^{-1}\right)$ for $|q \geqslant 3|$.

$$\sigma(\Omega) \geqslant \left(1-e^{-1}\right) \cdot \sigma(Opt), q \geqslant 3 \tag{15}$$

where Opt is the set in an optimal solution.

Proof: Lets redefine $v_i \in V, i = 1,2,3,...,r$ as a vehicle added into Ω in i-th iteration, Let Ω_k denote $\bigcup_{i=1}^{k} v_i$, and $\Omega = \Omega_r$. To prove inequality (15), the following two inequalities we can derive from [26], After $i, i = 1,2,3,...,r+1$ iterations, we can get

$$\sigma(\Omega_i) \geqslant \left[1 - \prod_{m=1}^{i}(1-\frac{c_m}{C_{max}})\right] \cdot \sigma(Opt) \tag{16}$$

$$\begin{aligned}\sigma(\Omega_{r+1}) &\geqslant \left[1 - \prod_{m=1}^{r+1}(1-\frac{c_m}{C_{max}})\right] \cdot \sigma(Opt) \\ &\geqslant \left[1 - \left(1-\frac{1}{r+1}\right)^{r+1}\right] \cdot \sigma(Opt) \\ &\geqslant \left(1-e^{-1}\right) \cdot \sigma(Opt)\end{aligned} \tag{17}$$

where c_m denotes the sensing reward to v_m. The detailed proof of inequalities (16), (17)can be found in [25],[26]. We can easily to know that (17) is equivalent to following inequality

$$\sigma(\Omega_{r+1}) = \sigma(\Omega_r) + \sigma(\{v_{r+1}\}) \geqslant \left(1-e^{-1}\right) \cdot \sigma(Opt) \tag{18}$$

where v_{r+1} is selected at $r+1$ round but not added to Ω due to overflow budget constraint C_{max}. Applying (16) to (18), we get

$$\sigma(\Omega_r - S^0) + \sigma(\{v_{r+1}\}) \geqslant \left(1-e^{-1}\right) \cdot \sigma(Opt - S^0) \tag{19}$$

where the set $Opt - S^0$ means that an element belongs to set Opt but not in set S^0.

Assuming $\sigma\left(\{v_{r+1}\}\right)$ is greater than $\sigma\left(\{v_i\}\right), i = 1,2,3,...,r$, if this were the case, v_{r+1} is bound to be selected before v_i and included in Ω_r, so this assumption is invalid. Therefore, we can get

$$q \cdot \sigma\left(\{v_{r+1}\}\right) \leqslant \sigma\left(S^0\right) \tag{20}$$

From (19), (20) the following inequality can be hold

$$\sigma\left(\Omega_r\right) \geqslant \left(1-e^{-1}\right) \cdot \sigma(Opt - S^0) + (1-q^{-1}) \cdot \sigma(S^0) \tag{21}$$

where e is a natural base whose value is less than three, hence

$$\sigma\left(\Omega_r\right) \geqslant \left(1-e^{-1}\right) \cdot \left(\sigma(Opt - S^0) + \sigma(S^0)\right) \tag{22}$$

if and only if $q \geqslant 3$, the inequality (21) makes sense. Clearly, $\sigma(Opt - S^0) + \sigma(S^0) = \sigma(Opt)$, and then

$$\sigma(\Omega_r) \geqslant (1-e^{-1}) \cdot \sigma(Opt),\ for\ q \geqslant 3 \tag{23}$$

Owing to the final output of ECQA as good as Ω_r if not better, this prove the performance guarantee of $(1-e^{-1})$.

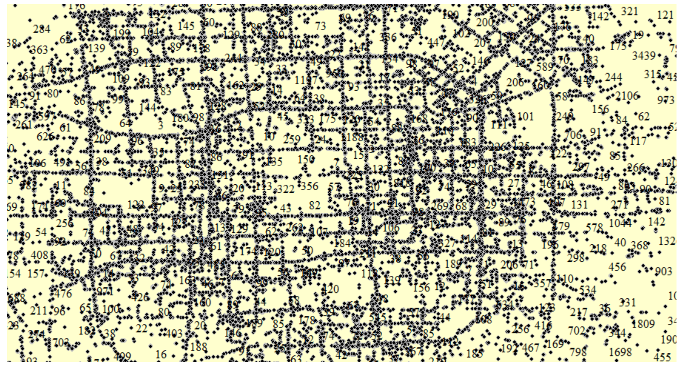

Fig. 3: The distribution of the trajectories of vehicles in the dataset.

V. EVALUATION

Extensive simulations has been conducted to evaluation the performance of our proposed algorithm. The traffic trace dataset we used, the simulation setup, the compared algorithms, and the performance comparison and discussion are presented as follows.

A. *Real Traffic Track Used and Simulation Setup*

In our simulation, to make the evaluation results convincing, the T-Drive trajectory dataset [27], [28] that contains a one-week trajectory of 10,357 buses.The dataset contains information about identification,arrival time,longitude,latitude of vehicles, such as [id:10002, arival time:2008-02-03 10:06:48, longitude:116.41904, latitude:39.93963].The total number of points in this dataset is about 15 million and the total distance of the trajectories reaches 9 million kilometers. We have imported the processed data into the Google Global Mapper, as Fig.3 shows, the distribution of the trajectories of vehicles basically covers the whole traffic network of Beijing. Our simulation is performed on traces extracted from the dataset on February 3, 2008, 6 AM to 10 PM. We randomly extract a small number of vehicles in a region of interest from processed dataset to participate in crowd-sensing, i.e., 10, so that the optimal solution can be found though an enumeration algorithm. Each vehicle is associated with a SR, and the SR of a vehicle is uniformly distributed in [0.7, 1.2].

B. *Algorithm in Comparison*

The quality of crowd-sensing is related to STC, we evaluation how the total sensing reward C_{max}, the number of time period m, and the initial size of solution q impact on the performance. In this paper, we compare the performance of our algorithm with two baseline algorithm. 1) The enumerative algorithm (EA) can always get the optimal vehicles from the candidate vehicles by exhaustive search, however, the SVP is NP-hard, when the number of candidate vehicles is larger, it becomes infeasible to obtain the optimal solution in polynomial time. Thus, the EA is applied simply when the number of vehicles is small. 2) The simulated annealing algorithm (SAA) is often used to solve optimization problems, we improve a SAA to compute the SVP for maximizing the

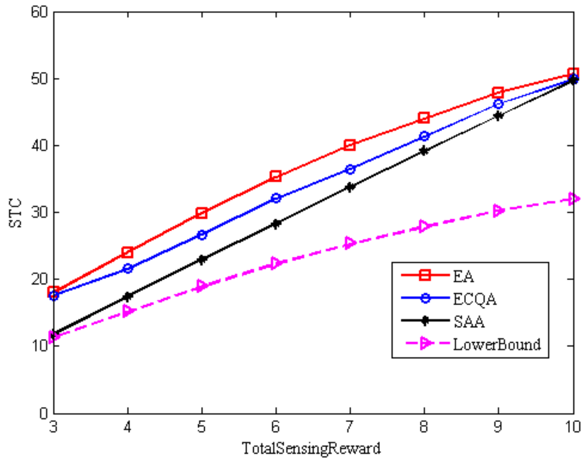

Fig. 4: The total sensing reward C_{max} is [3,10], the initial size of solution is 3, the number of period time is 6.

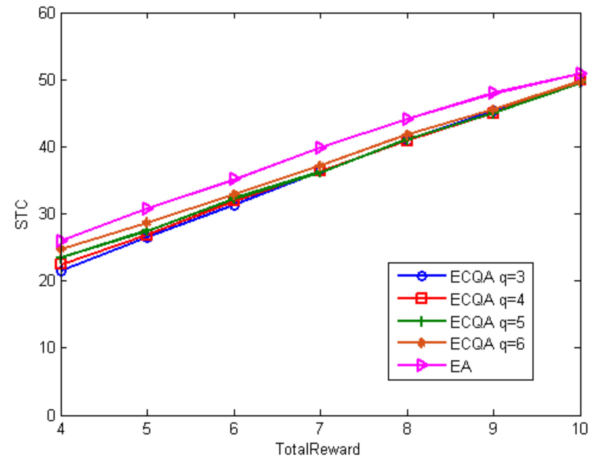

Fig. 6: The initial size of solution is [3, 6], the total sensing reward is [4, 10], the number of period time is 6.

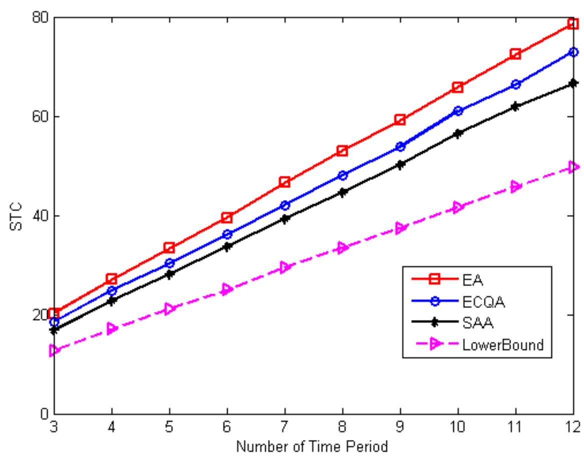

Fig. 5: The number of period is [4, 12], the total sensing reward C_{max} is 6, the initial size of solution is 3.

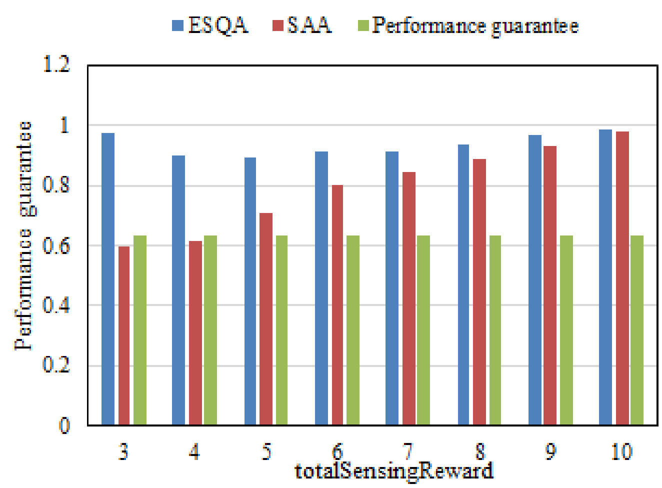

Fig. 7: The total sensing reward C_{max} is [3,10], the initial size of solution is 3, the number of period time is 6.

STC. Furthermore, the results are also compared with the lower bound performance guarantee STC $EA \cdot (1 - e^{-1})$.

Fig.4 to Fig.9 illustrates the performance during the variation of the total sensing reward C_{max}, the number of period time m and the initial size of solution q. In this group of simulations, we extract 10 vehicles from dataset. From Fig.4 to Fig.9 , we can observe that the proposed algorithm in this paper outperforms the one with SAA, and gets closer to the optimal EA. In Fig.4, the STC of our algorithm is larger than the SAA, and if the total sensing reward is enough, the ECQA and EA tends to equal to optimal while the budget is enough, the CMP tends to select all vehicle to finish sensing task . This result fits in with the reality. In Fig.5, the performance guarantee of ECQA fluctuates around 0.9 and still provide a performance guarantee larger than $EA \cdot (1 - e^{-1})$ as we have proved. This result indicates that in real cases, our algorithm is more likely to achieve full-coverage and ensure the integrity of sensing data. Fig.6 shows, along with the increase of the number of period time m, the STC of both

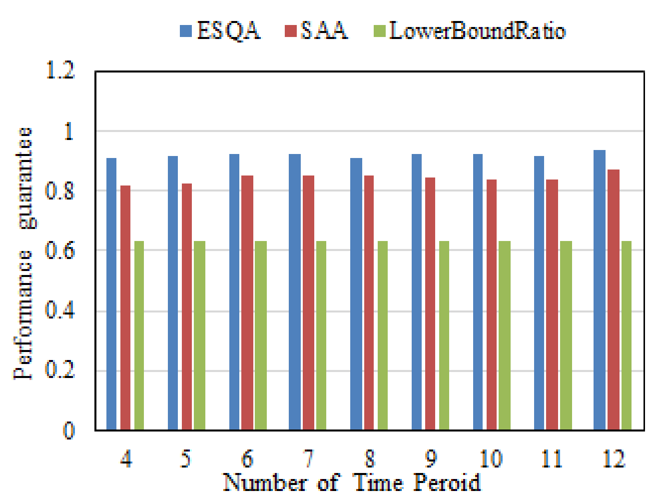

Fig. 8: The number of period is [4, 12], the total sensing reward C_{max} is 6, the initial size of solution is 3.

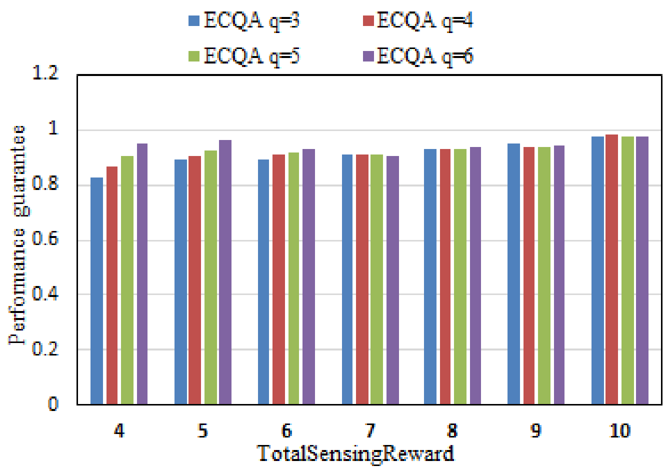

Fig. 9: The initial size of solution is [3, 6], the total sensing reward is [4, 10], the number of period time is 6.

ECQA and competitors become larger and larger. This is due to vehicle's scheduled cyclically movement. The larger number of period time is set, the more covered time in one segment may be as Fig.2 depicted. Additionally, the performance gap between ECQA and EA stayed nearly constant as m and the performance guarantee is greater than 0.9 consistently as Fig.7 shown. In Fig.8 and Fig.9, we study the important parameter q influence on performance of ECQA, the q is varied from 4 to 10. It is easy to see that, when total sensing reward is less than 7, the STC increase with q, otherwise is closer. This result can be understood since the total sensing reward is insufficient, the q is larger, ECQA primarily searches a larger domain for optimal solution, but it takes longer execution time actually. In contract, q has slightly impact on STC, which suggests that we can get a good performance even with a small q and spend less running time simultaneously.

VI. CONSLUTION

In this paper, we introduce mobile crowd-sensing into vehicular network to produce a vehicle-based crowd sensing network. Due to the quality of crowd-sensing is extraordinarily sensitive to the location of participants, so we take full advantage of predictable mobility pattern of public transport buses whose traveling route is scheduled. In this scenario, we need to address a crucial problem of the selection of vehicle to participate in urban sensing for maximizing spatiotemporal coverage. We have proved that the problem of selecting vehicles under a given constraint sensing reward for maximizing spatiotemporal coverage is NP-hard. Then we propose the ECQA which aims to select an optimal vehicles collection for maximizing the spatial-temporal coverage by taking the current and future positions into account of each vehicle. Moreover, through the theoretical analysis and a series of simulation on real T-Drive trajectory dataset, it is demonstrated that the ECQA can achieve a performance guarantee is still greater than $(1-e^{-1})$ of optimum and obtain better performance than alternative algorithm.

ACKNOWLEDGMENT

This paper was supported by National Key R&D Program of China, $NO.2017YFC0804404$, Beijing Municipal Commission of Education (The City's Vehicle Sensing Grid Construction Based on Public Transportation Network).

REFERENCES

[1] Zhu.Y, Bao.Y, Li.B. "On Maximizing Delay-Constrained Coverage of Urban Vehicular Networks[J]." *IEEE Journal on Selected Areas in Communications*, 2012, 30(4):804-817.
[2] Dejun Yang, Guoliang Xue, Xi Fang, and Jian Tang, "Crowdsourcing to smart phones: incentive mechanism design for mobile phone sensing", in*ACM MOBICOM*, 201.
[3] Mohammad Nozari Zarmehri and Ana Aguiar, "Supporting sensing application in vehicular networks", in *ACM CHANTS*, 2012
[4] E.Koukoumidis,L.S.Peh, M.R.Martonosi,"Signalguru:Leveraging mobile phones for collaborative trafc signal schedule advisory", in *Proc. 9th Int. Conf. Mobile Syst., Appl. Serv., ACM*, 2011, pp. 127140.
[5] A. Thiagarajan et al., "Vtrack: Accurate, energy-aware road trafc delay estimation using mobile phones",*7th ACM Conf.Embedded Netw. Sensor Syst., ACM*, 2009, pp. 8598.
[6] P. Dutta, P. M. Aoki, and et al., "Common Sense: participatory urban sensing using a network of handheld air quality monitors," in *SenSys 2009*, pp. 349350.
[7] A.Farshad, M. K. Marina, and F. Garcia, "Urban WiFi characterization via mobile crowdsensing," in *Proc. Netw. Oper. Manage. Symp., IEEE*, 2014, pp. 19.
[8] Feng Z, Zhu Y, Zhang Q, et al. "TRAC: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing[C]" // *INFOCOM, 2014 Proceedings IEEE.* IEEE, 2014:1231-1239.
[9] D. Yang, G. Xue, and et al., "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing," in *MobiCom 2012*, pp. 173184.
[10] S. Reddy, D. Estrin, M. Hansen, and M. Srivastava, "Examining micro-payments for participatory sensing data collections," in *UbiComp 2010*, pp. 3336.
[11] He Z, Cao J, Liu X. "High quality participant recruitment in vehicle-based crowdsourcing using predictable mobility," *Computer Communications. IEEE*, 2015:2542-2550.
[12] L. Kazemi and C. Shahabi, "Geocrowd: Enabling query answering with spatial crowdsourcing," in *ACM SIGSPATIAL 2012*, pp. 189198.
[13] J. Huang and H.-S. Tan, "Vehicle future trajectory prediction with a DGPS/INS-based positioning system," in *American Control Conference*, 2006.
[14] Hussain R, Abbas F, Son J, et al. Vehicle Witnesses as a Service: Leveraging Vehicles as Witnesses on the Road in VANET Clouds[C]// *IEEE International Conference on Cloud Computing Technology and Science. IEEE Computer Society*, 2013:439-444.
[15] Thiagarajan A, Ravindranath L, Lacurts K, et al. VTrack:accurate, energy-aware road traffic delay estimation using mobile phones[C]// *ACM Conference on Embedded Networked Sensor Systems. ACM*, 2009:85-98.
[16] Lee J L, Kim D, Fan L, et al. Barrier-Coverage for City Block Monitoring in Bandwidth Sensitive Vehicular Adhoc Networks[C]// *International Conference on Mobile Ad-Hoc and Sensor Networks. IEEE*, 2015:80-87.
[17] Reddy S, Shilton K, Burke J, et al. Using Context Annotated Mobility Profiles to Recruit Data Collectors in Participatory Sensing[J]. 2009, 5561:52-69.
[18] Gerla M, Pau G, Weng J T. Pics-on-wheels: Photo surveillance in the vehicular cloud[C]// *International Conference on Computing, NETWORKING and Communications. IEEE*, 2013:1123-1127.
[19] Kang L, Poslad S, Wang W, et al. "A Public Transport Bus as a Flexible Mobile Smart Environment Sensing Platform for IoT" *International Conference on Intelligent Environments. IEEE, 2016*.
[20] Rai A, Chintalapudi K K, Padmanabhan V N, et al. "Zee:zero-effort crowdsourcing for indoor localization[C]"// 2012:293-304.
[21] *https://baike.baidu.com/item.*
[22] MAXIM S. "A note on maximizing a sub-modular set function subject to knapsack constraint [J]" . *Operations Research Letters, 2004*,32(5): 41-43.
[23] Khuller S, Moss A, Naor J. "The budgeted maximum coverage problem [M]".*Elsevier North-Holland*.

[24] Jing Yuan, Yu Zheng, Xing Xie, and Guangzhong Sun. "Driving with knowledge from the physical world." *In The 17th ACM SIGKDD international conference on Knowledge Discovery and Data mining, KDD11,* New York, NY, USA, 2011. ACM.

[25] Jing Yuan, Yu Zheng, Chengyang Zhang, Wenlei Xie, Xing Xie, Guangzhong Sun, and Yan Huang. "T-drive: driving directions based on taxi trajectories". *In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems,* GIS 10, pages 99-108, New York, NY, USA,2010. ACM.

[26] Yang D, Xue G, Fang X, et al. "Incentive mechanisms for crowdsensing: crowdsourcing with smartphones[J]". *IEEE/ACM Transactions on Networking*, 2016, 24(3):1732-1744.

[27] X. Zhang, Z. Yang, Y. Liu, J. Li and Z. Ming, Toward efcient mechanisms for mobile crowdsensing, *IEEETrans.Veh.Technol*.,vol.66,no.2, pp. 17601771, Feb. 2017.

[28] K. Han, C. Zhang, J. Luo, M. Hu, and B. Veeravalli, Truthful scheduling mechanisms for powering mobile crowdsensing, *IEEE Trans. Comput.*, vol. 65, no. 1, pp. 294307, Jan. 1, 2016.

[29] K. Han, C. Zhang, and J. Luo, Taming the uncertainty: Budget limited robust crowdsensing through online learning,in *IEEE/ACMTrans.Netw*., vol. 24, no. 3, pp. 14621475, Jun. 2016.