Maximizing Spatial-temporal Coverage in Mobile Crowd-Sensing Based on Public Transports 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 intelligent terminal, which collects ubiquitous data easily in metropolis. The existing applications about crowd-sensing based on intelligent terminals mainly consider the current trajectories of the participants, and its quality highly depends on the spatial-temporal coverage which is easily weakened by the trajectories of participants. Nowadays, public transports (PTs) are widely used and affordable in many cities. PTs embedded with substantial sensors can act as participants in crowd-sensing. But distinguishing from intelligent terminals, the trajectories of PTs is schedulable and it is predictable, which sheds an opportunity to achieve high quality crowd-sensing. Therefore, based on the predictable trajectories of PTs, we design a novel system model and formulate the selection of PTs as an optimization problem to maximize the spatial-temporal coverage. In oder to obtain a optimum solution, an approximation algorithm is proposed. The performance of this algorithm can be guaranteed to be close to 1. We evaluate the proposed algorithm with real T-Drive trajectory dataset. The results show that our algorithm achieves a near optimal coverage and outperforms existing alternative algorithms.

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

I. INTRODUCTION

WITH the rapid advance of sensor technology, communication, and mobile computing, mobile crowd-sensing [1] has become a paradigm attracting much attention for collecting distributed data and distributing to general public. With the help of mobile crowd-sensing, the cost of data collection and dissemination over wide range of area can be significantly reduced. Intelligent terminals in different places can easily collect ubiquitous data and share it with potential users in neighborhood [2]-[3]. Equipped with various onboard sensors such as GPS, video cameras, gas sensor and communication module, vehicles become the powerful participants to collect data like intelligent terminals, including traffic monitoring [4]-[5], environment monitoring [6], and urban Wi-Fi characterization [7], etc.

A vehicle-based mobile crowd-sensing system is typically composed of two parts [8]: cloud management platform (CMP) and vehicles embedded with various sensors. An example

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).

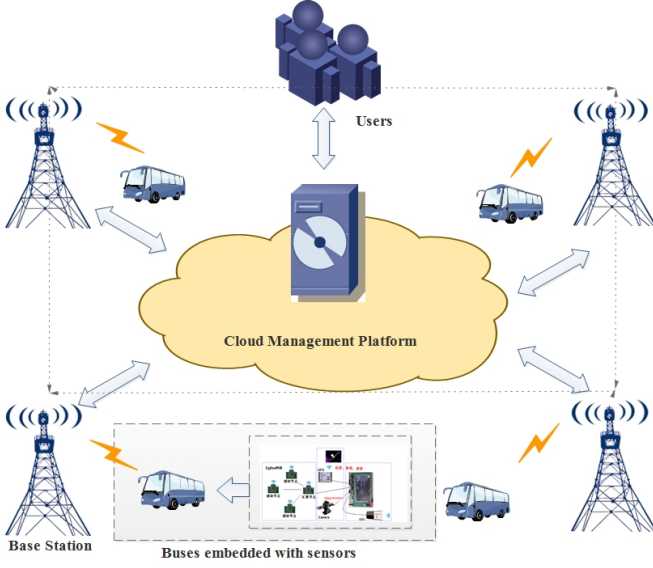Manuscript received XXXX XX, 2018; revised XXXX XX, 2018.

is shown in Fig. 1. The cloud management platform is responsible for selecting a set of vehicles to carry out crowd-sensing tasks and to process data forwarding. The vehicles are sensing nodes distributed in city. In general, it is important to decide which PTs to participate in collaborative sensing. This manifests the success of vehicle-based mobile crowd-sensing. Assuming an extreme case that the CMP selects all vehicles to carry out crowd-sensing task, apparently, it can achieve what is assigned, but multiple vehicles may introduce redundancy since only one is sufficient to conduct the task. Furthermore, modern crowd-sensing application are not only initiated by data centers deployed for large corporations, like Google, Alibaba and public security department, but also from individuals. It is affordable for large corporations to select all participants to carry out task, but the ordinary users can not. Therefore, in order to restrict the numbers of vehicles, the budget of CMP needs to be limited [9]-[11].

The quality of vehicle-based crowd-sensing is easily influenced by the trajectories of vehicles [12]. On the one hand, there no vehicles operating in a specific region at a time, which will lead to the data to be discrete in space. On the another hand, in different time periods, a region is covered at most once, which means the data is discrete in time. So the quality of crowd-sensing is sensitive to space and time. The spatial-temporal coverage (STC) is a fundamental metric of the vehicle-based mobile crowd-sensing quality [1]. Specifically, STC intends to cover as many regions as possible and make sure one region is covered at least once for a period of time. In reality, we are supposed to be aware that the STC is easily weakended by the trajectories of vehicles where vehicles runs randomly. Distinctly, public transports (PTs) strictly follow an schedule, the trajectories of PTs are predictable in spite of the highly dynamic mobility. Considering the future trajectories of PTs, the quality can be effectively improved, instead of only depending on current trajectory as smartphone-based crowd-sensing [13].

In this paper, we investigate how to achieve a high quality of crowd-sensing with the predictable trajectories of PTs and the limited budget of CMP. Analying the relationship between STC and the predictable trajectories of PTs, we design a novel system model by considering the current and future trajectories of PTs and propose an algorithm to select PTs to carry out crowd-sensing tasks. Therefore, a high quality of crowd-sensing can be guaranteed for a period of time. Futhermore, we prove the selection of PTs problem is NP-hard and prove the proposed algorithm can achive a performance

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

guarantee no less than $\left(1 - e^{-1}\right)$.

This paper is organized as follows. Section II reviews the related work. Section III introduces the system model and formulates the selection of PTs as an optimization problem. In Section IV we propose a novel algorithm to solve the selection problem of PTs and analyse the performance guarantee of this algorithm. 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 a significant source of information for smart city. Many researchers are 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. Authors of [11], [18]-[20] proposed a participants recruitment system and formulated the recruitment of participants as a constrained coverage problem but ignored the mobility of vehicle. Authors in [16] constructed a surveillance system based on vehicle with constraint network bandwidth. In [18], Gerla M et al. introduced a crowd-sensing service based on vehicle embedded with cameras to deliver images on demand to users. In [21], K.Han et al. proposed an incentive mechanisms for participant recruitment who interacted with a task requestor in a random order for maximizing the values of finished task. Authors of [22]-[23] studied location-based crowd-sensing systems and major concerned both spatial and temporal coverage based on current location of participants. However, these crowd-sensing systems assumed that the initiators were capable to select participants as many as possible to conduct tasks and the trajectories of vehicles were known hypothetically, which is more suitable for the cases that there were a small number of participants, the unlimited budget of initiators and the participants are

unmovable. Distinguish from the problems above, we make an advance forward. To obtain a high quality of crowd-sensing, we not only consider the current and future trajectories of candidates, but also highlight the limited budget. Then we establish a novel system model and formulate the selection of PTs as an optimization issue solved by a performance guarantee approximation algorithm.

III. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

We divide a region R into a serial of small segments. Let R denotes the set of small segments, $R = \{r_1, r_2, r_3, ..., r_k\}$. The CMP broadcasts a crowd-sensing task to be carried 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 PTs equips with sensor module that we have designed in [23]. Assume there are n PTs can conduct sensing tasks and the set of PTs is denoted by $V = \{v_1, v_2, v_3, ..., v_n\}$. Initially, the CMP obtains the current trajectroy of all PTs according to the schedule and broadcasts the data packet until receives the ACK. If the prediction is not consistent with the actual current trajectory obtained by Global Positioning System (GPS) [24] employed in PTs, it will be updated, respectively. Then we can get the trajectories of a PTs v_i at a specific time t_j, which is denoted by $l_i(t_j) \in R$. Thus the trajectory matrix of PTs can be represented as follows:

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

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

In practice, beacause nearby PTs usually upload similar information, we do not anticipate that all PTs are involved in crowd-sensing. In order to limit the number of PTs, we assume PTs need to be paid a sensing reward from CMP [9]-[11]. The reward can be monetary or non-monetary. Next, we define the sensing reward.

Definition 1: Sensing Reward (SR) a PT is selected to sense often associates with a reward. Let c_i denotes the reward to v_i, which can be acquired through online bidding [30]. The reward vector C is:

$$C = \{c_1, c_2, ..., c_n\} \quad (2)$$

With the limited budget of CMP, not all PTs participate in crowd-sensing. We adopt an indication vector Φ to imply whether a vehicle v_i is selected or not,

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

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

$$C(\Omega) = [C, \Phi] \quad (4)$$

As mentioned above, the quality of crowd-sensing is related to STC. Next, we will introduce the notion of spatial-temporal coverage.

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

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

In [23], we have designed a specific hardware system, which can collect various information such as temprature, flow of traffic, longitude, latitude, and so on. Based on the hardware system, we show an example to explain the implication of STC. In Fig. 2, if users request to collect the flow of traffic at region R where is divided into a serial of segments, that is $R = \{$AB, AD, BC, BE, DE, EF, EH, DH, CF$\}$. The scheduled trajectories 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 of time $\{t_1, t_2, t_3, t_4\}$ the trajectories 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 is capable of selecting two PTs to sense, then we consider two cases as bellows:

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

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

It can be seen that the set of $\{$Bus1, Bus2$\}$ covers five different places in space and the segment of $\{$BC$\}$ is covered three times, so the total STC is 7. On the contrary, the set of $\{$Bus3, Bus4$\}$ only covers four different segments in space and the total STC is 5. So we are more willing to select $\{$Bus1, Bus2$\}$ to sense.

B. Problem Statement

Based on the system model, we are ready to formalize the selection of PTs as a optimization problem for maximizing the STC with limited budget.

Definition 3: Selection of PTs Problem (SPTs) 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 data may have varying importance degrees at different segments and time, such as we are more interested in hotspot with high traffic flow. Therefore, we introduce priority power to indicate a public transport with a higher
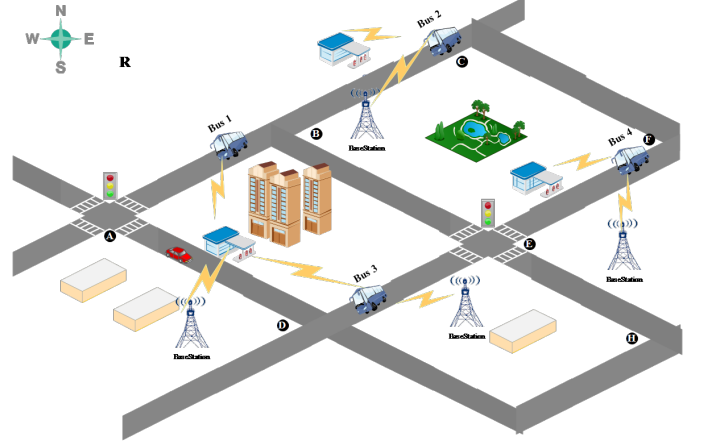

Fig. 2: An example explains the notion of spatial-temporal Coverage.

priority which is more likely to be selected to join crowd-sensing. Analyzing historical data, it is easy to acquire traffic performance index (TPI) [25] which is the congestion level of each segment. Let $D_{t_j}^{l_i}$ denotes the TPI of $l_i(t_j)$ at a specific time t_j, and it is 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 follows:

Definition 4: Priority Power (PP) is the priority of a vehicle to be selected to sense, 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:

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

With the priority power, the STC can be redefined as:

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

and the SPTs can be rewritten as:

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

As a practical matter, it is crucial to select a optimal set of PTs to conduct task with time efficient. Unfortunately, it is NP-hard even though the trajectories of PTs are predictable. In the next section, we prove SPTs is NP-hard and propose an improved approximation algorithm based on greedy algorithm.

IV. SOLUTION TO THE SPTs

A. Complexity Analysis of SPTs

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

Proof: To prove the NP-hard hardness of SPTs, we should demonstrate it belongs to NP firstly. Assuming there is a possible solution Ω', 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 SPTs is NP.

To prove SPTs is NP-hard further, we can construct a reduction in polynomial time from budgeted maximum coverage problem (MCP) as the known NP-hard [26] to SPTs. The MCP is defined as follows.

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

$$x_i \stackrel{mapping}{\mapsto} l_i(t_j), \; S_i \stackrel{mapping}{\mapsto} L(\Omega')$$

$$c_i \stackrel{mapping}{\mapsto} C(\Omega'), \; B \stackrel{mapping}{\mapsto} C_{max}$$

where $\Omega' \subseteq V$. 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 SPTs successfully. So we can obtain a corresponding instance from SPTs for any instance in MCP. As a result, the SPTs is NP-hard.

Consequently, to achieve a truthful and computationally efficient crowd-sensing, it is highly demanded to propose an approximate algorithm to solve SPTs.

B. Approximate Algorithm to Solve SPTs

We have analyzed the NP-hardness of SPTs, it becomes computationally impracticable to select an optimal set of PTs when the total number PTs is large. As for a metropolis like Beijing, the number of PTs under operations is about 30,000 per day by the end of 2016 [25]. To achieve a desired computational efficiency, we propose an approximate algorithm called efficient combination query algorithm (ECQA) to solve SPTs. The ECQA adopts a greedy strategy to solve SPTs. The greedy policy is to select one public transport with most reward efficiency, until the total SR exceeds the limited budget of CMP. Next we define the reward efficiency.

Definition 5: Reward Efficiency (RE) indicates the marginal STC achieved per unit sensing reward. Mathematically, the RE can be computed as follows:

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

The algorithm tries many rounds. A public transport with maximum RE is selected in 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 terminate until exceed the limited budget of CMP. The pseudo-code is listed in table 1.

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 [26]. In this paper, the ECQA can achieve a lower bound ratio of $(1 - e^{-1})$ when the cardinality as q of set S^0 is no less than three, i.e., $q \geqslant 3$. Next, we prove the following theorem about the performance guarantee of ECQA.

The pseudo-code of ECQA 1

Input: set $V = \{v_1, v_2, v_3, ..., v_n\}$ is PTs under operation, set $C = \{c_1, c_2, c_3, ..., c_n\}$ is sensing reward to PTs, C_{max} the limited budget of CMP, an initial set S^0 of cardinality is an integer as 3

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

  $max \leftarrow 0$
  $\Omega \leftarrow \varnothing$
  $S \leftarrow \varnothing$
  for $S^0 \subseteq V, C(S^0) \leqslant C_{max}$ **do**
    $S \leftarrow S^0$
    for $v_i \in V - S$ **do**
      $S' \leftarrow \{S \cup v_i\}$
      $E_i \leftarrow \frac{\sigma(S') - \sigma(S)}{c_i}$
      if $E_i > max$ **and** $C(S') < C_{max}$ **then**
        $S \leftarrow S'$
        $max \leftarrow E_i$
      end if
      if $\sigma(S) > \sigma()\Omega)$ **then**
        $\sigma \leftarrow S$
      end if
    end for
  end for

Theorem 2. The ECQA can achieve a worst performance guarantee to be $(1 - e^{-1})$ of optimum solution when $|q \geqslant 3|$.

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

where Opt is optimum solution.

Proof: Let's redefine $v_i \in V, i = 1, 2, 3, ..., r$ as a public transport 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 [27], After $i - th$, $i = 1, 2, 3, ..., r + 1$ iterations, we can get:

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

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

where c_m denotes the sensing reward to v_m. The detailed proof of inequalities (16), (17) can be found in [26]-[27]. We can easily know that (17) is equivalent to following inequality:

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

where v_{r+1} is selected at $r + 1$ round but not added to Ω because it exceeds the limited budget C_{max}. Applying (16) to (18), we get:

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

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

Assuming $\sigma(\{v_{r+1}\})$ is greater than $\sigma(\{v_i\})$, $i = 1, 2, 3, ..., r$, v_{r+1} should be selected before v_i and included in Ω_r, the opposite is true. Therefore, we can get:

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

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

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

where e is less than three, hence

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

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

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

proves the theorem.

V. VALIDATION

Extensive simulations have been conducted to evaluate 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, we adopt the T-Drive trajectory dataset [28]-[29] which contains a one-week trajectories of 10,357 buses. The dataset contains information about identification, arrival time, longitude, latitude, such as [id: 10002, arival time: 2008-02-03 10:06:48, longitude: 116.41904, latitude: 39.93963]. The total trajectories of buses in this dataset is about 15 million and the total distance of the trajectory reaches 9 million kilometers. We import the data into the Google Global Mapper. As shown in Fig.3, the distribution of the trajectories of vehicles can cover the whole traffic network of Beijing. We extract a subset of buses from the whole dataset for performance research, and the trajectory of the subset spans from February 3, 2008, 6 AM to 10 PM. In our simulation, the size of region is $6km \times 6km$, which is divided into 36 squares with a width of $1km$. The period of time is from $1min$ to $15min$, PTs are associated with SR, which is uniformly distributed in $[0.7, 1.2]$ and the limited budget C_{max} is $[3, 10]$.

B. Algorithm in Comparison

The quality of crowd-sensing is related to STC, we evaluate how the total sensing reward C_{max}, the number of time period m, and the initial size of solution q affect the performance. In this paper, we compare our algorithm with two algorithms and use unpredictable trajectory to vertify the better performance of it. 1) The enumerative algorithm (EA) can always select a optimum set of PTs by exhaustive search. However, the SPTs is NP-hard, when the number of PTs is larger, it becomes infeasible to obtain the optimal solution in polynomial time. Thus, the EA is applied simply when the number of PTs is small. 2) The simulated annealing algorithm (SAA) is often
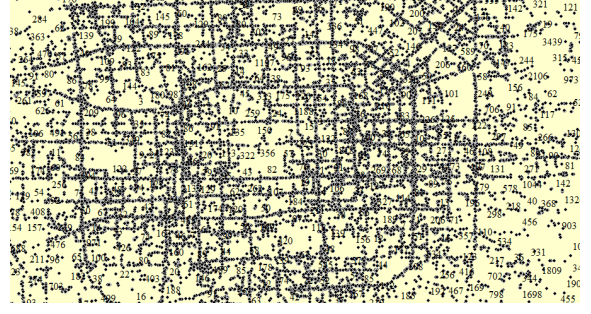

Fig. 3: The vehicle trajectory distribution in Beijing Metropolis according to the dataset

used to solve optimization problems, we improve a SAA to compute the SPTs for maximizing the STC. 3) Using unpredictable trajectory means that we only take the current trajectories of PTs into account as [19]. Furthermore, the results are also compared with the lower bound performance guarantee STC $EA \cdot (1 - e^{-1})$.
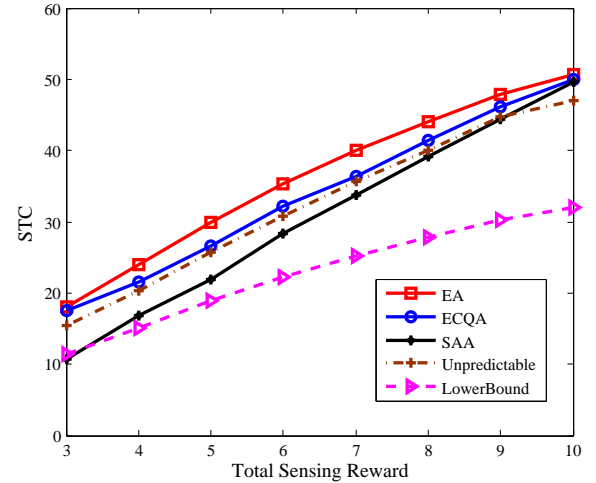

Fig. 4: The total sensing reward C_{max} is $[3,10]$, the initial size of solution is 3, the number of period time is 6.

Fig. 4 to Fig. 9 illustrate 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 SAA, and gets closer to the optimal EA. In Fig. 4, the STC of our algorithm is larger than the SAA. And when the sensing reward is enough, the STC of ECQA and SAA tend to be the optimum. This is because the budget is enough, the CMP may select all PTs to carry out task, the result fits in with the reality. In Fig. 5, the performance guarantee of ECQA fluctuates around 0.9 and still provides a performance guarantee larger than $(1 - e^{-1}) \approx 0.6321$ as we had proved. This result indicates that in real cases, our algorithm is more likely to achieve full-coverage with a high quality of crowd-sensing. Fig. 6 shows that along with the increase of the number of period time m, the STC of ECQA and alternative algorithms have
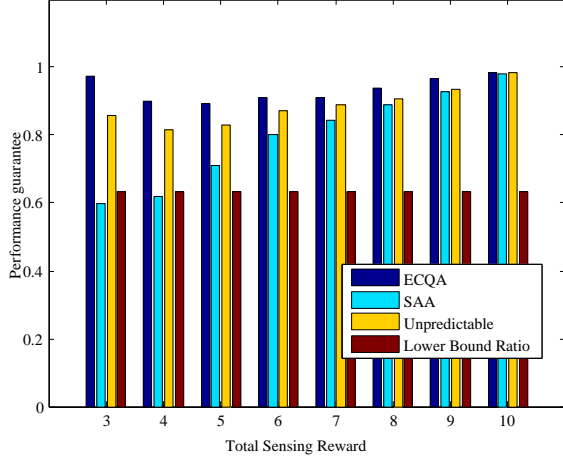
Fig. 5: The total sensing reward C_{max} is [3,10], the initial size of solution is 3, the number of period time is 6.
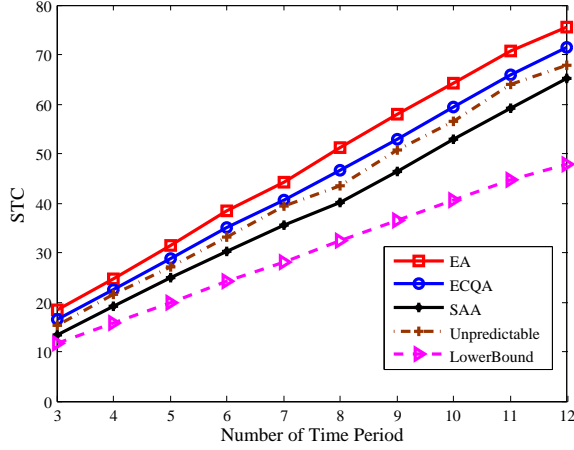

Fig. 7: The number of period is [3, 12], the total sensing reward C_{max} is 6, the initial size of solution is 3.
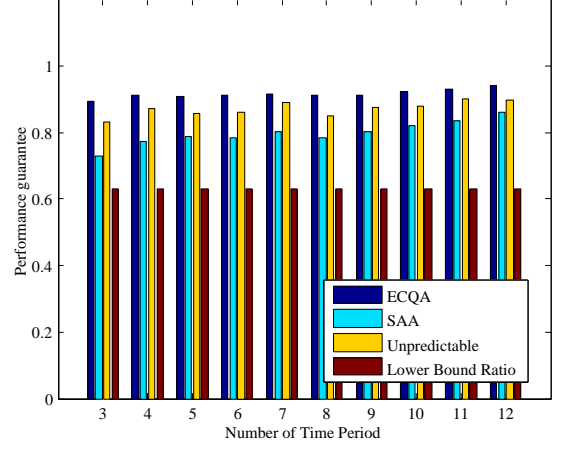

Fig. 6: The number of period is [3, 12], the total sensing reward C_{max} is 6, the initial size of solution is 3.
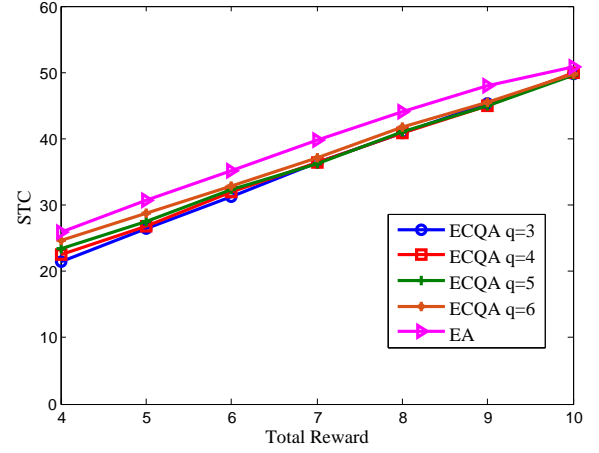

Fig. 8: The initial size of solution is [3, 6], the total sensing reward is [4, 10], the number of period time is 6.

been an rising trend. This is because one segment may be covered many times at difference period times. Additionally, the performance gap between ECQA and EA stayed nearly constant and the performance guarantee is greater than 0.9 consistently as Fig. 7 shown. In Fig. 8 and Fig. 9, we study the influence of the important parameter q on performance of ECQA, the q is varied from 3 to 6. It is easy to see that, the total sensing reward is less than 7, the STC increases with q, otherwise is closer. This results can be understood since the total sensing reward is insufficient and 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. From Fig. 4 to Fig. 7, we can observe the truth that the performance of proposed algorithm using unpredictable trajectory is better than SAA, LowerBound, and gets closer to ECQA. There is an underlying problem can not be ignored, i.e., when the system selects a best solution by using the proposed algorithm with unpredictable trajectory, the PTs covers the same segment at different sensing time,

which means the best solution merely make sure the STC is continuous in time, but discrete in space.

VI. CONCLUSION

In this paper, we have introduced crowd-sensing into vehicular network to construct a vehicle-based crowd-sensing network. The quality of crowd-sensing is sensitive to the trajectories of participants, so we took advantage of scheduled mobility pattern of PTs with predictable trajectory. In this scenario, we need to address how to select PTs to carry out sensing tasks for a high quality of crowd-sensing. We have analyzed the relationship between STC and the predictable trajectory of PTs, and have designed a system model by cosidering the current and future trajectory. Then based on system model, we have formulated the selection of PTs as a optimization problem, which was proved as a NP-hard problem. In oder to maximize STC in polynomial time, we have proposed ECQA to achieve a performance guarantee closer to 1. Finally, the simulations have been performed by using T-Drive trajectory dataset, the results have shown the ECQA outperforms other exsiting algorithms.
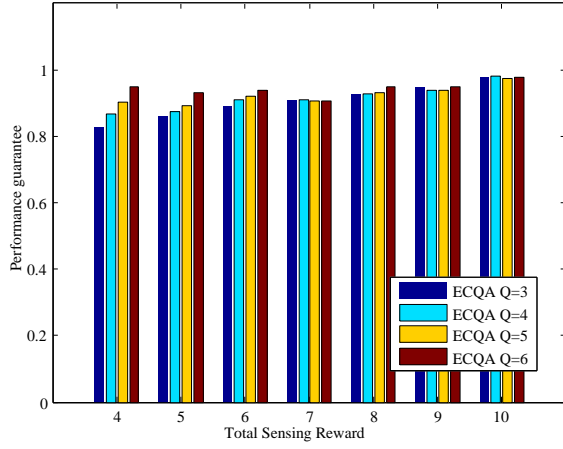
Fig. 9: The initial size of solution is [3, 6], the total sensing reward is [4, 10], the number of period time is 6.

ACKNOWLEDGMENT

REFERENCES

[1] Zhu.Y, Bao.Y, Li.B, "On Maximizing Delay-Constrained Coverage of Urban Vehicular Networks," *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," *ACM MOBICOM*, 2012.

[3] Mohammad Nozari Zarmehri and Ana Aguiar, "Supporting sensing application in vehicular networks," *ACM CHANTS*, 2012.

[4] Koukoumidis, Emmanouil, L. S. Peh, and M. R. Martonosi. "Signal-Guru:leveraging mobile phones for collaborative traffic signal schedule advisory," *International Conference on Mobile Systems, Applications, and Services DBLP*, 2011: 127-140.

[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] Dutta P, Aoki P M, Kumar N, et al. "Common Sense:participatory urban sensing using a network of handheld air quality monitors," *International Conference on Embedded Networked Sensor Systems*, 2009: 349-350.

[7] Farshad, Arsham, M. K. Marina, and F. Garcia. "Urban WiFi characterization via mobile crowdsensing," *Noms IEEE/IFIP Network Operations Management Symposium IEEE*, 2014: 1-9.

[8] Feng Z, Zhu Y, Zhang Q, et al. "TRAC: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing," *INFOCOM 2014 Proceedings IEEE*, , 2014: 1231-1239.

[9] D. Yang, G. Xue, and et al.. "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing," *MobiCom 2012*, pp. 173184.

[10] Reddy, Sasank, et al. "Examining micro-payments for participatory sensing data collections," *ACM International Conference on Ubiquitous Computing ACM*, 2010: 33-36.

[11] He, Zongjian, J. Cao, and X. Liu. "High quality participant recruitment in vehicle-based crowdsourcing using predictable mobility," *Computer Communications IEEE*, 2015: 2542-2550.

[12] Kazemi L, Shahabi C. "GeoCrowd:enabling query answering with spatial crowdsourcing," *International Conference on Advances in Geographic Information Systems*, 2012: 189-198.

[13] Huang, Jihua, and H. S. Tan. "Vehicle future trajectory prediction with a DGPS/INS-based positioning system," *American Control Conference IEEE*, 2006: 6 pp.

[14] Hussain R, Abbas F, Son J, et al. "Vehicle Witnesses as a Service: Leveraging Vehicles as Witnesses on the Road in VANET Clouds," *IEEE International Conference on Cloud Computing Technology and Science*, 2013: 439-444.

[15] Thiagarajan A, Ravindranath L, Lacurts K, et al. "VTrack: accurate, energy-aware road traffic delay estimation using mobile phones," *ACM Conference on Embedded Networked Sensor Systems*, 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," *International Conference on Mobile Ad-Hoc and Sensor Networks*. 2015: 80-87.

[17] Reddy S, Shilton K, Burke J, et al. "Using Context Annotated Mobility Profiles to Recruit Data Collectors in Participatory Sensing," 2009, 5561: 52-69.

[18] Gerla M, Pau G, Weng J T. "Pics-on-wheels: Photo surveillance in the vehicular cloud. International Conference on Computing," *NETWORKING and Communications*, 2013: 1123-1127.

[19] Reddy, Sasank, D. Estrin, and M. Srivastava. "Recruitment Framework for Participatory Sensing Data Collections," *International Conference*, May 17-20, 2010. 2010: 138-155.

[20] 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.

[21] K. Han, C. Zhang, J. Luo, M. Hu, and B. Veeravalli. "Truthful scheduling mechanisms for powering mobile crowdsensing," *IEEE Trans. Comput*, pp. 294307, Jan. 1, 2016.

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

[23] 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*, 2016.

[24] Rai A, Chintalapudi K K, Padmanabhan V N, et al. "Zee:zero-effort crowdsourcing for indoor localization," 2012: 293-304.

[25] https://baike.baidu.com/item.

[26] MAXIM S. "A note on maximizing a sub-modular set function subject to knapsack constraint," *Operations Research Letters*, 200 ,32(5): 41-43.

[27] Khuller S, Moss A, Naor J. "The budgeted maximum coverage problem," Elsevier North-Holland.

[28] 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.

[29] 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.

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