Review

International Journal of
Distributed
Sensor Networks

International Journal of Distributed Sensor Networks
2018, Vol. 14(8)

DOI: 10.1177/1550147718795351
journals.sagepub.com/home/dsn
SAGE

Maximizing spatial–temporal coverage in mobile crowd-sensing based on public transports with predictable trajectory

Chaowei Wang, Chensheng Li, Cai Qin, Weidong Wang and Xiuhua Li

Abstract
Mobile crowd-sensing is a prospective paradigm especially for intelligent mobile terminals, which collects ubiquitous data efficiently in metropolis. The existing crowd-sensing schemes based on intelligent terminals mainly consider the current trajectory of the participants, and the quality highly depends on the spatial-temporal coverage which is easily weakened by the mobility of participants. Nowadays, public transports are widely used and affordable in many cities around the globe. Public transports embedded with substantial sensors act as participants in crowd-sensing, but different from the intelligent terminals, the trajectory of public transports is schedulable and predictable, which sheds an opportunity to achieve high-quality crowd-sensing. Therefore, based on the predictable trajectory of public transports, we design a novel system model and formulate the selection of public transports as an optimization problem to maximize the spatial–temporal coverage. After proving the public transport selection is non-deterministic polynomial-time hardness, an approximation algorithm is proposed and the coverage is close to 1. We evaluate the proposed algorithm with samples of real T-Drive trajectory data set. The results show that our algorithm achieves a near optimal coverage and outperforms existing algorithms.

Keywords
Mobile crowd-sensing, schedulable trajectory, spatial–temporal coverage, approximation algorithm, performance guarantee

Date received: 31 January 2018; accepted: 18 July 2018

Handling Editor: Christos Anagnostopoulos

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 environmental information and distributing to the 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 sensors, and communication modules, vehicles are considered as intelligent terminals and become powerful participants in data collection. Then public services such as traffic monitoring[4,5] environment monitoring[6] and urban Wi-Fi characterization,[7] etc. are greatly facilitated.

School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, China

Corresponding author:
Chaowei Wang, School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China.
Email: wangchaowei@bupt.edu.cn

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 is shown in Figure 1. The CMP is responsible for selecting a set of vehicles to carry out crowd-sensing tasks and to process data dissemination. The vehicles can be considered as sensing nodes distributed in the city area. In general, it is important to decide which public transports (PTs) to participate in collaborative sensing. Because multiple vehicles may introduce redundancy since only one in a segment is sufficient to conduct the task. Furthermore, modern crowd-sensing application is not only initiated by data centers deployed for large companies, like Google, Alibaba, and Facebook, but also for individuals. It is affordable for large companies to select all participants to carry out task, but the small ones are not able to pay for this. Therefore, the budget of CMP needs to be limited[2,9,10] and it should constrain the number of selected crowd-sensing participants.

The quality of vehicle-based crowd-sensing is easily influenced by the changing trajectory of vehicles.[11] On one hand, if there is no vehicle operating in a specific region at one time, the collected data will be discrete in time. On the other hand, in different time periods, if a region is covered utmost for once, which means the data will be discrete in time. Therefore, the performance of crowd-sensing is sensitive to space and time; then the spatial–temporal coverage (STC) is regarded as a fundamental indicator of the vehicle-based mobile crowd-sensing.[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 weakened by the trajectory of vehicles as they move randomly. However, PTs strictly follow a schedule and route; the

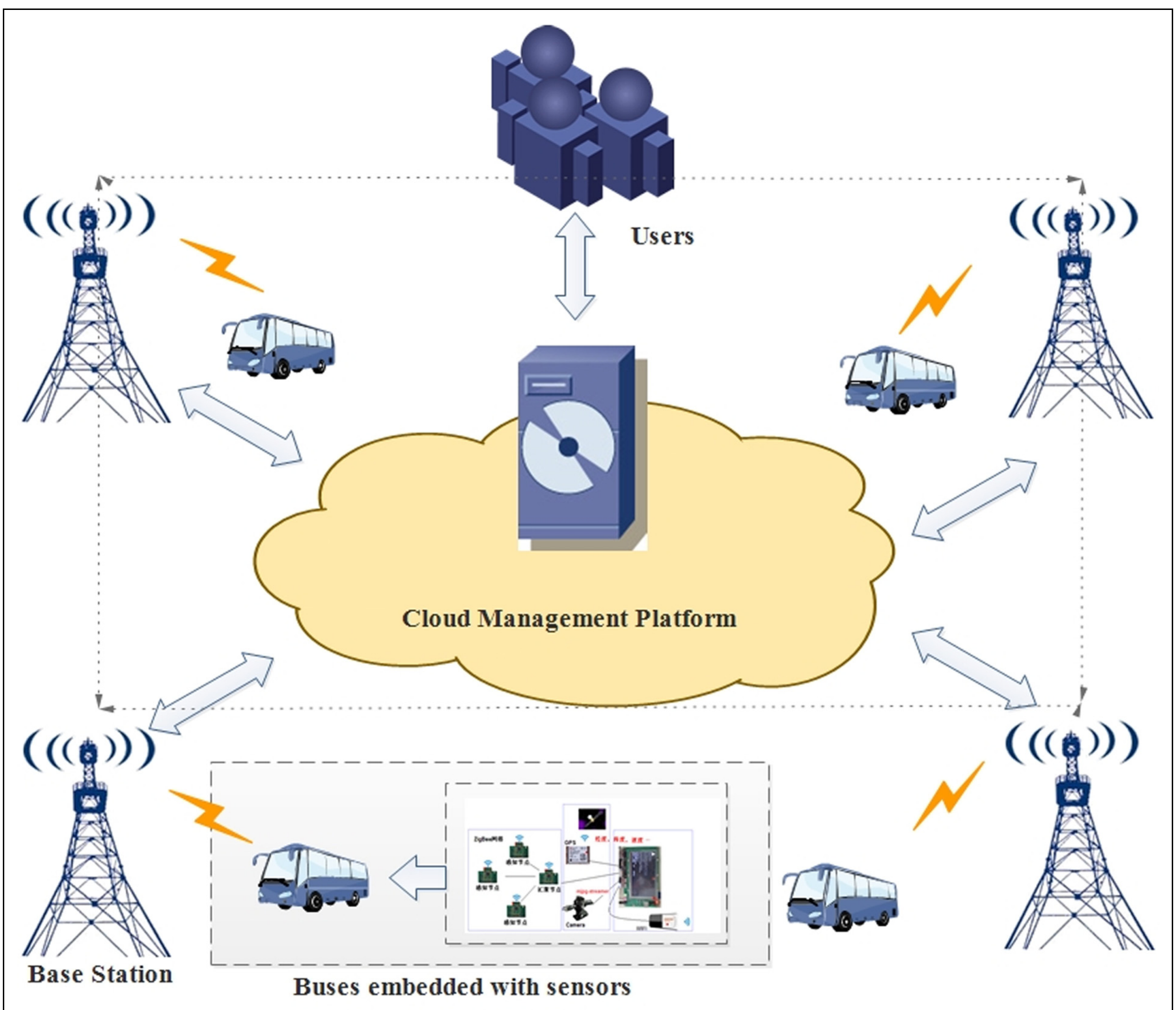

Figure 1. An example of vehicle-based crowd-sensing application.

trajectory of PTs is predictable in spite of the highly dynamic mobility. Considering the future trajectory of PTs, the performance can be effectively improved, which is different from smartphone-based crowd-sensing only considering current trajectory.[12]

In this article, we investigate how to achieve an improved crowd-sensing by PTs with the predictable trajectory and limited budget of CMP. Analyzing the relationship between STC and the predictable trajectory of PTs, we establish a novel system model by jointly considering the current and future trajectory of PTs and propose an algorithm to select PTs to carry out crowd-sensing tasks. Furthermore, we prove the selection of public transport (SPT) problem is non-deterministic polynomial-time hardness (NP-hard) and the proposed algorithm can achieve a performance guarantee not less than $(1 - e^{-1})$.

This article is organized as follows. Section "Related works" reviews the related work. Section "System model and problem formulation" introduces the system model and formulates the SPTs as an optimization problem. In section "Solution to the SPTs," we propose a novel algorithm to solve the selection problem of PTs and analyze the performance guarantee of this algorithm. Performance evaluation and analysis are provided in section "Validations." Finally, section "Conclusion" draws the conclusion of this article.

Related works

In recent years, many researchers focus on studying the vehicular application of crowd-sensing, for example, traffic accident evidence collection,[5,13] city block monitoring,[14] bike-net for cyclist experience mapping,[15] and architectures of crowd-sensing. Authors of the literature[10,16–18] proposed the participants recruitment system and formulated the recruitment of participants as a constrained coverage problem but did not consider the mobility of vehicle. Authors in Lee et al.[14] constructed a surveillance system based on vehicle with constraint network bandwidth. Gerla et al.[16] introduced a crowd-sensing service based on vehicle embedded with cameras to deliver images on demand to users. Han et al.[19] proposed an incentive mechanism for participant recruitment who interacted with a task requestor in a random order for maximizing the values of finished task. Authors of Han et al.[20] and Kang et al.[21] studied location-based crowd-sensing systems and mainly considered both spatial and temporal coverage based on current location of participants. However, these crowd sensing systems assume that the task initiators are capable of selecting all participants to perform the task and the trajectory of the participants is known. In practice, these assumptions only apply to the scenarios with small number of participants, unlimited budget of the initiators, and immovable participants. Different from the problems above, we make a step forward: not only considering the current and future trajectory of candidate, but also highlighting the limited budget. Then we establish a novel system model and formulate the SPTs as an optimization problem solved by a performance guarantee approximation algorithm.

System model and problem formulation

System model

We divide a target region R into a series of small segments. Let R denotes the set of small segments, $R = \{r_1, r_2, r_3, \ldots, r_k\}$. The CMP broadcasts a crowd-sensing task to be carried out for a period of time, that is, T. We assume the time is discrete, and we can get $T = \{t_1, t_2, t_3, \ldots, t_m\}$. The PTs are equipped with sensor modules that we have designed in Kang et al.[21] It is assumed that there are n PTs can conduct the sensing tasks and the set of PTs is denoted by $V = \{v_1, v_2, v_3, \ldots, v_n\}$. Initially, the CMP obtains the current trajectory of all PTs according to the schedule and broadcasts the data packet until it receives the ACK. If the prediction is not consistent with the actual current trajectory obtained by global positioning system (GPS)[22] employed in PTs, it will be updated, respectively. Then we can get the trajectory 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

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

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

In practice, because nearby PTs usually upload overlapped information which brings in redundancy, so we do not anticipate all the PTs are involved in crowd-sensing. In order to limit the number of involved PTs, we assume PTs need to be paid a sensing reward (SR) from CMP.[2,9,10] Next, we define the SR.

Definition 1. SR. A PT is selected to collect data and usually granted a reward correspondingly. Let c_i denotes the reward to v_i, which can be acquired through online bidding.[23] The reward vector C is

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

With 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

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

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

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

As mentioned above, the quality of crowd-sensing is related to STC. Next, we define the notion of spatial–temporal coverage (STC).

Definition 2. STC determines the quality of crowd-sensing. Formally, it can be defined as

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

The union operator in equation (5) is used for obtaining the union set of PTs covering all the target regions at the same period of time and the sums operator in equation (5) is used for calculating the total number of all the target regions covered in all periods of time. These two steps are very important; it is helpful to avoid selecting a set of PTs with the same trajectory.

In Kang et al.,[21] we have designed a specific hardware system, which can collect various information such as temperature, humidity, air quality, flow of traffic, longitude, latitude. Based on the hardware system, we give an example to explain the STC. In Figure 2, if users request to collect traffic information in region R which is divided into a serial of segments as, $R = \{AB, AD, BC, BE, DE, EF, EH, DH, CF\}$. 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 of time $\{t_1, t_2, t_3, t_4\}$, the trajectory of Bus1 to Bus4 is $\{BC, AD, DE, BC\}$, $\{BC, BE, BC, BE\}$, $\{EH, HD, AB, BE\}$, $\{AB, BE, AD, DH\}$, respectively. From equality (equation (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 collect information. Two following cases are considered:

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

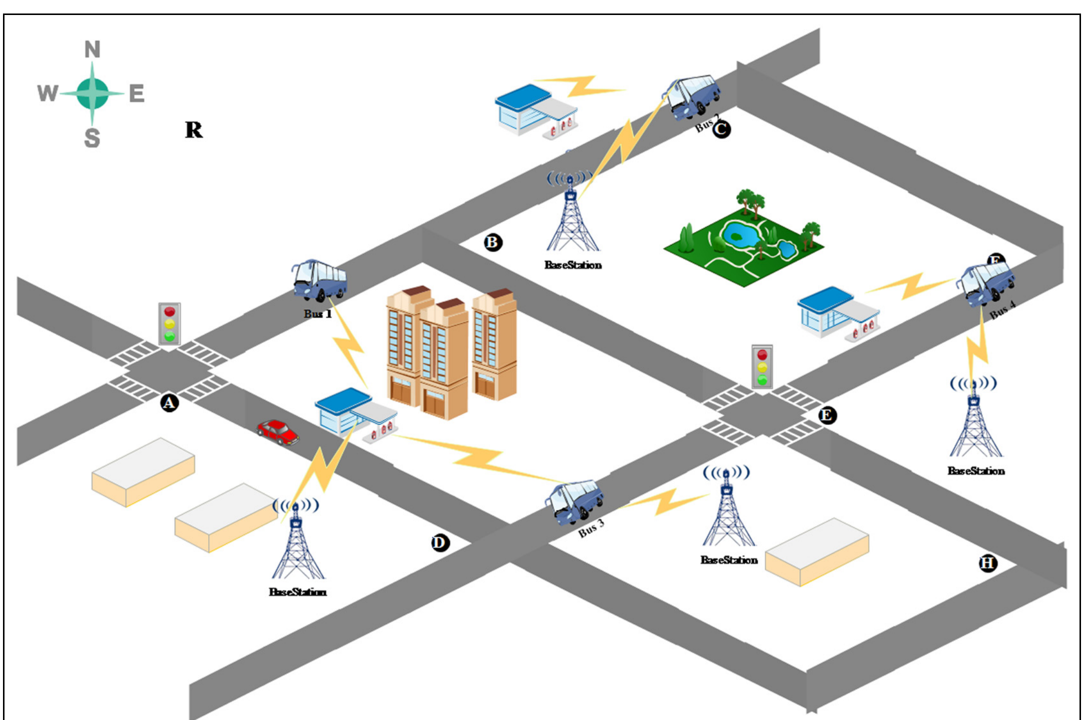

Figure 2. An example explains the notion of spatial–temporal coverage.

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

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. Therefore, we are more willing to select $\{Bus1, Bus2\}$ to sense.

Problem statement

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

Definition 3. SPT problem is to determine a set of vehicle under the budget constraint C_{max} with the objective of maximizing the STC

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

Actually, the collected data may have varying importance degrees at different segments and times, for example, we are more interested in hotspot with high traffic flow. Therefore, we introduce priority power (PP) to indicate a PT with a higher priority which is more likely to be selected to perform crowd-sensing. Analyzing historical data, it is easy to acquire traffic performance index (TPI)[24] 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, for example, $D_{t_j}^{l_i} \in (0, 1]$. With the TPI, we define PP as follows.

Definition 4. 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, PP is expressed as

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

With the PP, the STC can be redefined as

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

and the SPTs can be rewritten as

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

Traditionally, there is supposed to be an optimal solution for SPTs with time efficiency. Unfortunately, it turns to be a NP-hard problem, even though the trajectory of PTs is predictable. In the next section, we will prove SPTs is NP-hard and propose an improved approximation algorithm based on greedy algorithm.

Solution to the SPTs

Complexity analysis of SPTs

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

Proof. To prove SPTs problem is NP-hard, we should demonstrate it falls in NP first. Assuming there is a possible solution Ω', Ω' can be validated in polynomial time and the complexity of checking algorithm is $O(n)$, therefore the SPTs is NP.

To prove SPTs are NP-hard further, we can make a reduction from budgeted maximum coverage problem (MCP), which is a known NP-complete[25] to SPTs. The MCP is defined as follows.

Given a collection of sets $S = \{S_1, S_2, \ldots, S_n\}$, each set S_i with a cost $c_i, 1 \leqslant i \leqslant n$ takes values from $X = \{x_1, x_2, \ldots, x_m\}$ 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

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

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

where $\Omega' \subseteq V$. Eeach vehicle has a PP which can be mapped to w_i. We have reduced the decision version of MCP to the problem formulation of SPTs successfully. Therefore, we can obtain a corresponding instance from SPTs for any instance in MCP. As a result, the SPTs are proved to be NP-hard. Consequently, to achieve a truthful and computationally efficient crowd-sensing, it is highly demanded to propose an approximate algorithm to solve the SPTs.

Approximate algorithm to solve SPTs

As SPTs have been proved to be NP-hard, it becomes computationally impracticable to select an optimal set of PTs when the total number PTs is large. As for a metropolis like the City of Beijing, the number of PTs under operations is over 30,000 per day by the end of

2016.[24] 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 PT with most reward efficiency (RE), until the total SR exceeds the limited budget of CMP. Next we define the RE.

Definition 5. RE is the marginal STC gained per unit SR. Mathematically, the RE for selecting PT i is computed as follows

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

The algorithm goes for many rounds. A PT 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 till the SR exceeds the limited budget of CMP. The pseudo-code is listed in Table 1.

The ECQA has a performance guarantee $\rho \leqslant 1$. In NP-hard problem, we can obtain a solution that is ρ times of the optimal solution.[25] In this article, the ECQA can achieve a lower bound ratio of $(1 - e^{-1})$ when the cardinality as q of set S^0 is not less than 3, that is, $q \geqslant 3$. Next, we prove the following theorem about the performance guarantee of ECQA.

Table 1. The pseudo-code of ECQA 1.

Input: set $V = \{v_1, v_2, v_3, \ldots, v_n\}$ is PTs under operation, set $C = \{c_1, c_2, c_3, \ldots, c_n\}$ 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 ← 0
Ω ← ∅
S ← ∅
for S⁰ ⊆ V, C(S⁰) ⩽ C_max do
  S ← S⁰
  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
```

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

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

where Opt is optimum solution.

Proof. Let us redefine $v_i \in V, i = 1, 2, 3, \ldots, r$ as a PT added into Ω in ith iteration, Let Ω_k denote $\bigcup_{i=1}^{k} v_i$, and $\Omega = \Omega_r$. After i iterations, $i = 1, 2, 3, \ldots, r, r + 1$, we can get the following two inequalities:

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

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

where c_m denotes the SR to v_m. The detailed proof of inequalities (equations (16) and (17)) can be found in Maxim[25] and Khuller et al.[26] We can know that equation (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 the $(r + 1)$th round but not added to Ω because it exceeds the limited budget C_{max}. Applying equations (16)–(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 $Opt - S^0$ means that an element belongs to set Opt but not in S^0.

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

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

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

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

where e is less than 3, hence

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

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

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

proves the theorem.

Validations

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

Real traffic track used and simulation setup

In our simulation, we adopt the T-Drive trajectory data set[27,28] which contains 1-week trajectory of 10,357 buses. The data set contains 4 basic characteristics, i.e., identification, arrival time, longitude and latitude. For instance, it can be denoted as [id: 10002, arrival time: 2008-02-03 10:06:48, longitude: 116.41904, latitude: 39.93963]. The total trajectory number of buses in this data set is about 15 million and the overall length of the trajectory reaches 9 million km. We import the data samples into the Google Global Mapper. As shown in Figure 3, the distribution of the trajectory of vehicles can cover the whole traffic network of Beijing. We extract a subset of buses from the whole data set for validation, and the trajectory of the subset spans from 6 a.m. to 10 p.m. on 3 February 2008. In our simulations, the size of target region is 6 km $\times$ 6 km, which is divided into 36 squares with a width of 1 km. The period of time is from 1 to 15 min, PTs are associated with SR, which is uniformly distributed in $[0.7, 1.2]$ and the limited budget C_{max} is $[3, 10]$.

Algorithm in comparison

As mentioned above, the quality of crowd-sensing is determined by STC; we evaluate how the total SR C_{max}, the number of time period m, and the initial size of solution q affect the performance. In this paper, we validate the proposed ECQA with unpredictable trajectory and compare the proposed algorithm with two other existing algorithms enumerative algorithm (EA) and simulated annealing algorithm (SAA), (1) EA can always select an optimum set of PTs by exhaustive search. However, the SPTs is NP-hard, when the number of PTs is large, 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) SAA is often used to solve optimization problems, we adopt the SAA to compute the SPTs for maximizing the STC. (3) Using unpredictable trajectory means that we only take the current trajectory of PTs into account as Reddy et al.[17] Furthermore, the results are also compared with the lower bound performance guarantee STC $EA \cdot (1 - e^{-1})$.

Figures 4–9 illustrate the performance with the variation of the total SR 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 data set. From Figures 4–9, we can observe that the proposed algorithm outperforms the SAA and gets closer to the optimal EA. In Figure 4, the STC of our algorithm is better than the SAA. And when the SR is enough, the

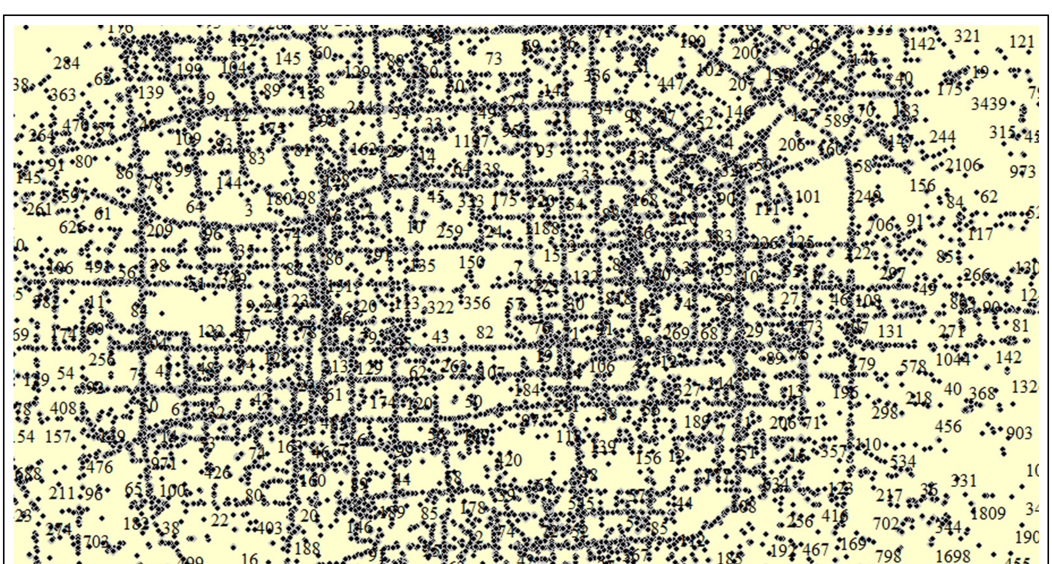

Figure 3. The vehicle trajectory distribution in Beijing Metropolis according to the data set.

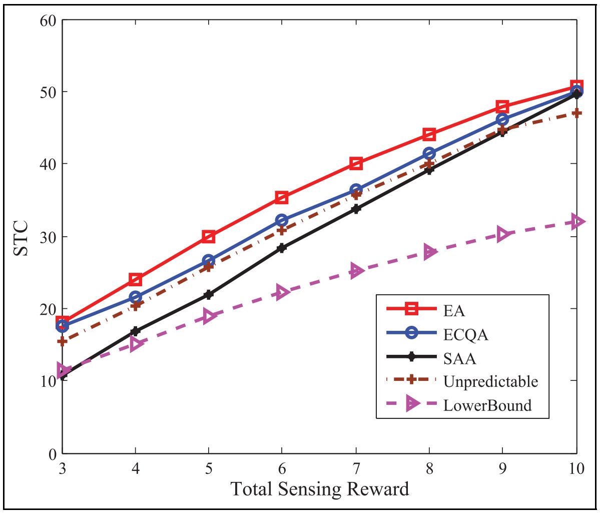

Figure 4. The STC versus total sensing reward C_{max}.

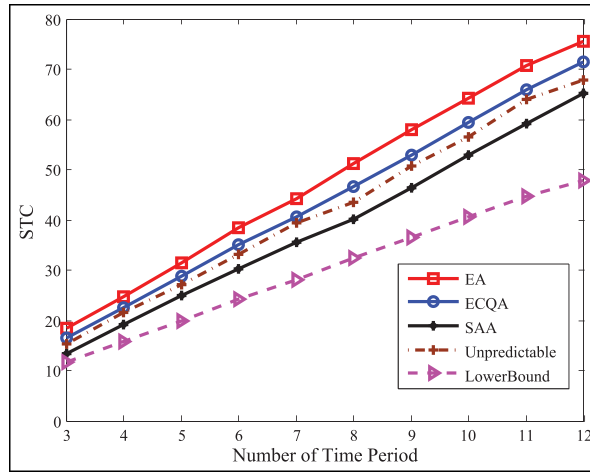

Figure 6. The STC versus the number of period *T*.

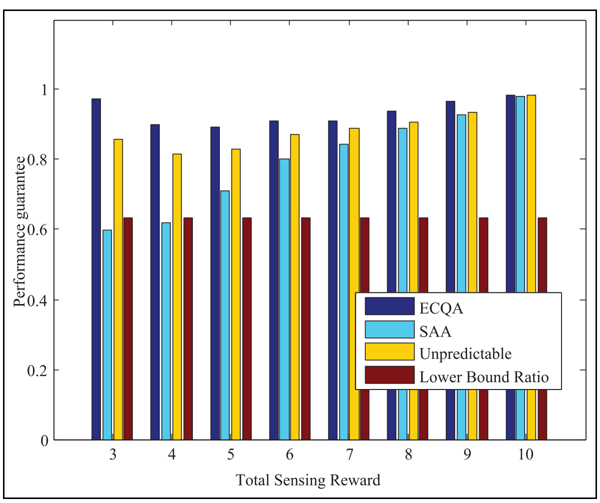

Figure 5. The performance guarantee versus total sensing reward C_{max}.

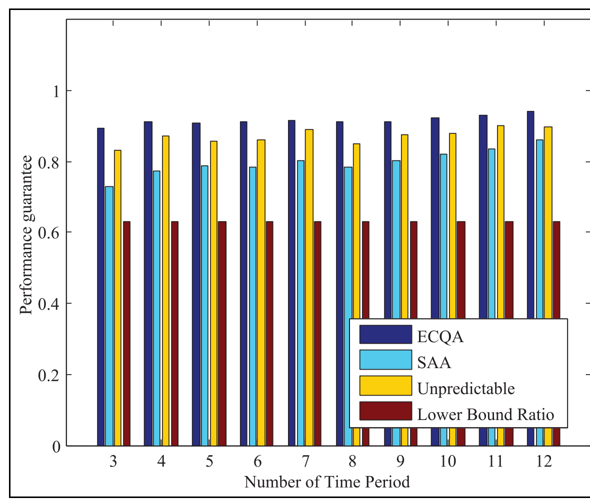

Figure 7. The performance guarantee versus the number of period *T*.

STC of ECQA and SAA tends to be the optimal. This is because the budget is adequate, the CMP may select all PTs to carry out sensing task, the results fit in with the reality. In Figure 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. Figure 6 shows that the STC of ECQA and alternative algorithms increases with the number of period time m. This is because one segment may be covered many times at different time periods. In addition, the performance gap between ECQA and EA stayed nearly constant and the performance guarantee is greater than 0.9 consistently

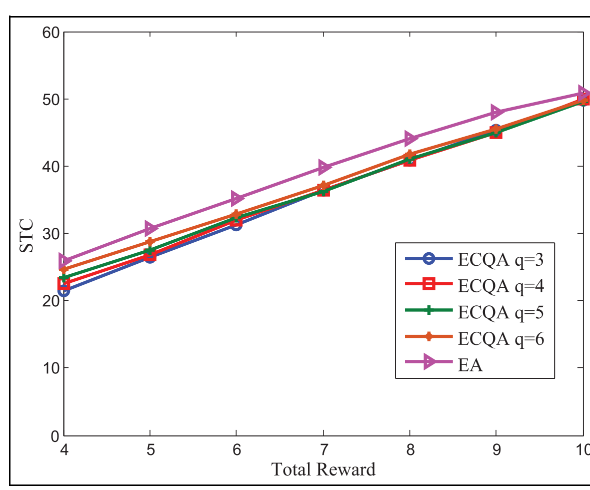

Figure 8. The STC versus the initial size of solution *q*.

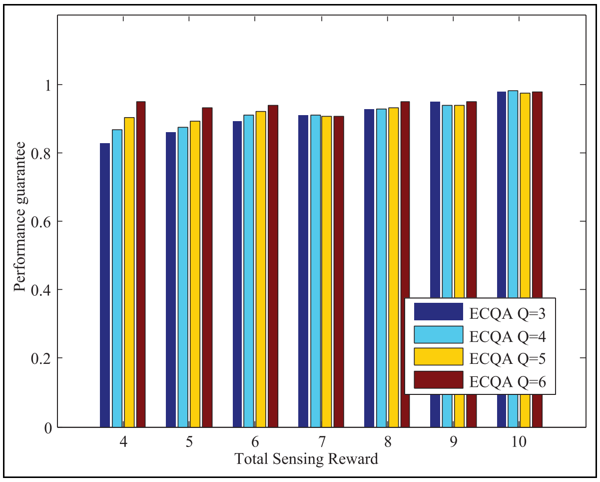

Figure 9. The performance guarantee versus the initial size of solution *q*.

with Figure 7. In Figures 8 and 9, we show 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 SR is less than 7; the STC increases with q, otherwise is closer. These results can be understood since the total SR is insufficient and the q is larger. ECQA primarily searches for 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 Figure 4–7, we can observe that the proposed algorithm with unpredictable trajectory outperforms the SAA and lower bound ratio algorithms, and gets closer to ECQA in terms of STC and performance guarantee. There is an underlying problem cannot be ignored, that is, when the system selects a best solution using the proposed algorithm with unpredictable trajectory, the PTs cover the same segment at different sensing times, which means the best solution merely makes sure the STC is continuous in time, but discrete in space.

Conclusion

In this article, 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 trajectory of participants, so we took advantage of scheduled mobility pattern of PTs with predictable trajectory. We have analyzed the relationship between STC and the predictable trajectory of PTs and designed a system model by considering both the current and future trajectory. Then we have formulated the SPTs as a optimization problem, which is proved to be a NP-hard problem. In order to maximize STC in polynomial time, we have proposed ECQA to achieve a performance guarantee close to 1. Finally, the simulations have been performed by adopting the T-Drive trajectory data set, and the results show the proposed ECQA outperforms other existing algorithms.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This article was supported by the National Science Foundation of China, no. 61271186 and National Key R&D Program of China, no. 2017YFC0804404, and Beijing Municipal Commission of Education (The City's Vehicle Sensing Grid Construction Based on Public Transportation Network).

References

1. Zhu Y, Bao Y and Li B. On maximizing delay-constrained coverage of urban vehicular networks. *IEEE J Sel Area Commun* 2012; 30(4): 804–817.
2. Yang D, Xue G, Fang X, et al. Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing. In: *Proceedings of the 18th annual international conference on mobile computing and networking*, Istanbul, 22–26 August 2012. New York: ACM.
3. Zarmehri MN and Aguiar A. Supporting sensing application in vehicular networks. In: *ACM MobiCom workshop on challenged networks*, Istanbul, 22 August 2012. New York: ACM.
4. Koukoumidis E, Peh LS and Martonosi MR. SignalGuru: leveraging mobile phones for collaborative traffic signal schedule advisory. In: *International conference on mobile systems, applications, and services*, Bethesda, MD, 28 June–1 July 2011, pp.127–140. New York: ACM.
5. Thiagarajan A, Ravindranath L, LaCurts K, et al. VTrack: accurate, energy-aware road traffic delay estimation using mobile phones. In: *Proceedings of the 7th ACM conference on embedded networked sensor systems*, Berkeley, CA, 4–6 November 2009, pp.85–98. New York: ACM.
6. Dutta P, Aoki PM, Kumar N, et al. Common Sense: participatory urban sensing using a network of handheld air quality monitors. In: *Proceedings of the 7th ACM conference on embedded networked sensor systems*, Berkeley, CA, 4–6 November 2009, pp.349–350. New York: ACM.
7. Farshad A, Marina MK and Garcia F. Urban WiFi characterization via mobile crowdsensing. In: *IEEE/IFIP network operations management symposium*, Krakow, 5–9 May 2014, pp.1–9. New York: IEEE.
8. Feng Z, Zhu Y, Zhang Q, et al. TRAC: truthful auction for location-aware collaborative sensing in mobile

crowdsourcing. In: *INFOCOM: IEEE conference on computer communications*, Toronto, ON, Canada, 27 April–2 May 2014, pp.1231–1239. New York: IEEE.
9. Reddy S, Estrin D, Hansen M, et al. Examining micropayments for participatory sensing data collections. In: *Proceedings of the 12th ACM international conference on ubiquitous computing*, Copenhagen, 26–29 September 2010, pp.33–36. New York: IEEE.
10. He Z, Cao J and Liu X. High quality participant recruitment in vehicle-based crowdsourcing using predictable mobility. In: *IEEE conference on computer communications*, Kowloon, Hong Kong, 26 April–1 May 2015, pp.2542–2550. New York: IEEE.
11. Kazemi L and Shahabi C. GeoCrowd: enabling query answering with spatial crowdsourcing. In: *Proceedings of the 20th international conference on advances in geographic information systems*, Redondo Beach, CA, 6–9 November 2012, pp.189–198. New York: IEEE.
12. Huang J and Tan HS. Vehicle future trajectory prediction with a DGPS/INS-based positioning system. In: *American control conference*, Minneapolis, MN, 14–16 June 2006, pp.1–6. New York: IEEE.
13. Hussain R, Abbas F, Son J, et al. Vehicle witnesses as a service: leveraging vehicles as witnesses on the road in VANET clouds. In: *IEEE 5th international conference on cloud computing technology and science*, Bristol, 2–5 December 2013, pp.439–444. New York: IEEE.
14. Lee JL, Kim D, Fan L, et al. Barrier-coverage for city block monitoring in bandwidth sensitive vehicular adhoc networks. In: *10th international conference on mobile adhoc and sensor networks*, Maui, HI, 19–21 December 2014, pp.80–87. New York: IEEE.
15. Reddy S, Shilton K, Burke J, et al. Using context annotated mobility profiles to recruit data collectors in participatory sensing. In: Choudhury T, Quigley A, Strang T, et al. (eds) *Location and context awareness*, vol. 5561. Berlin: Springer, 2009, pp.52–69.
16. Gerla M, Pau G and Weng JT. Pics-on-wheels: photo surveillance in the vehicular cloud. In: *International conference on computing, networking and communications*, San Diego, CA, 28–31 January 2013, pp.1123–1127. New York: IEEE.
17. Reddy S, Estrin D and Srivastava M. Recruitment framework for participatory sensing data collections. In: Floréen P, Krüger A and Spasojevic M (eds) *Pervasive computing: 8th international conference, pervasive 2010, Helsinki, Finland, May 17–20, 2010 proceedings*. Berlin: Springer, 2010, pp.138–155.
18. Zhang X, Yang Z, Liu Y, et al. Toward efficient mechanisms for mobile crowdsensing. *IEEE T Veh Technol* 2017; 66(2): 1760–1771.
19. Han K, Zhang C, Luo J, et al. Truthful scheduling mechanisms for powering mobile crowdsensing. *IEEE T Comput* 2016; 65: 294–307.
20. Han K, Zhang C and Luo J. Taming the uncertainty: budget limited robust crowdsensing through online learning. *IEEE/ACM T Netw* 2016; 24(3): 1462–1475.
21. Kang L, Poslad S, Wang W, et al. A public transport bus as a flexible mobile smart environment sensing platform for IoT. In: *12th international conference on intelligent environments*, London, 14–16 September 2016. New York: IEEE.
22. Rai A, Chintalapudi KK, Padmanabhan VN, et al. Zee: zero-effort crowdsourcing for indoor localization. In: *Proceedings of the 18th annual international conference on mobile computing and networking*, Istanbul, 22–26 August 2012, pp.293–304. New York: ACM.
23. Yang D, Xue G, Fang X, et al. Incentive mechanisms for crowdsensing: crowdsourcing with smartphones. *IEEE/ACM T Netw* 2016; 24(3): 1732–1744.
24. https://baike.baidu.com/item
25. Maxim S. A note on maximizing a submodular set function subject to knapsack constraint. *Oper Res Lett* 2004; 32(1): 41–43.
26. Khuller S, Moss A and Naor J. The budgeted maximum coverage problem. *Inf Process Lett* 1999; 70: 39–45.
27. Yuan J, Zheng Y, Xie X, et al. Driving with knowledge from the physical world. In: *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining (KDD'11)*, San Diego, CA, 21–24 August 2011. New York: ACM.
28. Yuan J, Zheng Y, Zhang C, et al. T-drive: driving directions based on taxi trajectories. In: *Proceedings of the 18th international conference on advances in geographic information systems*, San Jose, CA, 2–5 November 2010, pp.99–108. New York: ACM.