

# Comprehensive tempo-spatial data collection in crowd sensing using a heterogeneous sensing vehicle selection method

Yazhi Liu<sup>1</sup> · Jianwei Niu<sup>2</sup> · Xiting Liu<sup>2</sup>

Received: 9 October 2015 / Accepted: 30 March 2016 / Published online: 10 May 2016  
© Springer-Verlag London 2016

**Abstract** The wide distribution of mobile vehicles installed with various sensing devices and wireless communication interfaces has made vehicular mobile crowd sensing possible in practice. However, owing to the heterogeneity of vehicles in terms of sensing interfaces and mobilities, collecting comprehensive tempo-spatial sensing data with only one sensing vehicle is impossible. Moreover, the sensing data collected may expire in the future; as a result, sensing vehicles may have to continuously collect sensing data to ensure the relevance of such data. Although including more sensing vehicles can improve the quality of collected sensing data, this step also requires additional cost. Thus, how to continuously collect comprehensive tempo-spatial sensing data with a limited number of heterogeneous sensing vehicles is a critical issue in vehicular mobile crowd sensing systems. In this work, a heterogeneous sensing vehicle selection (HVS) method for the collection of comprehensive tempo-spatial sensing data is proposed. On the basis of the spatial distribution and sensing interfaces of sensing vehicles and the tempo-spatial coverage of collected sensing data, a utility function is designed in HVS to estimate the sensing capacity of sensing vehicles. Then, according to the utilities of sensing vehicles and the restriction on the number of recruited sensing vehicles, sensing vehicle selection is modeled as a knapsack problem. Finally, a greedy optimal sensing vehicle selection algorithm is designed. Real trace-driven

simulations show that the HVS algorithm can collect sensing data with a higher coverage ratio in a more uniform and continuous manner than existing mobile crowd sensing methods.

**Keywords** Vehicular sensing · Mobile crowd sensing · Comprehensive tempo-spatial data coverage

## 1 Introduction

In mobile crowd sensing systems, ordinary citizens are recruited to collect and share sensing data from their surrounding environments through their mobile devices [1, 2]. On the basis of collected sensing data, mobile crowd sensing systems are able to provide users with various novel applications [3], such as real-time air quality reports [4] and road traffic monitoring [5–7]. An increasing number of vehicles are being installed with various sensing devices and mobile communication interfaces. Therefore, an increasing number of mobile vehicles are able to collect and share various types of sensing data in urban environments. The extensive distribution of these vehicles in urban areas has made vehicular mobile crowd sensing possible in practice.

A vehicular mobile crowd sensing system is composed of sensing vehicles, wireless networks, and sensing data centers. Sensing data are collected by sensing vehicles and then transmitted to a sensing data center via wireless networks. The application of vehicular mobile crowd sensing to transportation can generate numerous advantages. First, it can lower the cost of sensing data collection because vehicles are equipped with several sensors. Second, unlike in single-vehicle sensing systems, sensing data in vehicular crowd sensing systems are collaboratively collected by

✉ Jianwei Niu  
niujianwei@buaa.edu.cn

<sup>1</sup> College of Information Engineering, North China University of Science and Technology, Tangshan 063000, China

<sup>2</sup> State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China

multiple vehicles. As a result, vehicular mobile crowd sensing systems can collect a large amount of comprehensive sensing data. Third, owing to the collaboration between vehicles and sensing data sharing, vehicular crowd sensing systems can provide many novel applications to the field of transportation and other areas. To meet the sensing data requirements of vehicular mobile crowd sensing applications, a large amount of sensing data must be collected. Therefore, vehicular mobile crowd sensing systems are mainly aimed at collecting comprehensive sensing data that encompass all the spatial temporal dimensions of target sensing areas. However, sensing vehicles are heterogeneous. First, the sensing interfaces of vehicles and the types of sensing data that can be collected by vehicles are different because vehicles are manufactured by different companies with no uniform hardware and software standards. Second, vehicles are driven by different people, and the trajectories of the vehicles also vary. Thus, collecting all available sensing data with only one sensing vehicle is impossible.

Sensing data are temporal, and collected sensing data may expire at some time in the future. For example, atmospheric temperature changes with time in any given day; as a result, collected temperature data expire after a few hours. Moreover, different types of sensing data may have different life cycles. For example, the life cycle of noise data is much shorter than that of temperature data. Thus, vehicular crowd sensing systems should continuously collect sensing data in target sensing areas to ensure the relevance of each sensing data type to the target sensing area.

Owing to the heterogeneity of sensing vehicles, one sensing vehicle cannot possibly collect comprehensive tempo-spatial sensing data. Furthermore, as collected sensing data may expire, crowd sensing vehicles should continuously collect sensing data to ensure their relevance. Obviously, in vehicular mobile crowd sensing systems, the quality of collected sensing data increases with the use of many sensing vehicles to collect sensing data. However, a large number of sensing vehicles equates to high cost [8]. Therefore, how to continuously collect comprehensive tempo-spatial sensing data with a limited number of heterogeneous sensing vehicles is a critical issue in vehicular mobile sensing systems.

To address the aforementioned problem, a heterogeneous sensing vehicle selection (HVS) method for the collection of comprehensive tempo-spatial sensing data is proposed in this work. The proposed HVS method considers not only the spatial coverage of collected sensing data but also the temporal coverage of sensing data. A mobility model based on a continuous-time Markov chain is established to forecast the spatial distribution of sensing vehicles. On the basis of the spatial distribution and sensing

interfaces of sensing vehicles and the tempo-spatial coverage of collected sensing data, a utility function is designed in HVS to estimate the sensing capacity of sensing vehicles in the future. According to the utilities of sensing vehicles and the restriction on the number of recruited sensing vehicles, sensing vehicle selection is modeled as a knapsack problem. Finally, a greedy optimal sensing vehicle selection algorithm is designed.

The remainder of this paper is organized as follows. Section 2 reviews the related studies on mobile crowd sensing and vehicle sensing. Section 3 presents the vehicular mobile crowd sensing system model and the construction of tempo-spatial sensing data models. Section 4 presents the vehicular sensing utility formula, which considers the spatial and temporal coverages of sensing data. Section 5 proposes an optimized heterogeneous vehicle selection algorithm based on the utilities of sensing vehicle candidates. Section 6 evaluates the performance of the proposed strategy through real trace-driven simulations. Section 7 provides the conclusions.

## 2 Related work

Mobile crowd sensing [1] is a promising mode of collecting comprehensive sensing data in urban areas. In recent years, a large number of novel applications [3] based on mobile crowd sensing have been deployed. For example, The Common Sense project [9] monitored the contents of pollutants in the air by collecting sensing data from handheld air quality monitors. Yu et al. [10] recovered the noise situation throughout New York City (NYC) based on mobile crowd sensing and inferred the fine-grained noise situation at different times of the day for each region of NYC with the use of data from the city complaint platform, together with social media, road network data, and points of interests. The Mahali project [11] used GPS signals that penetrate the ionosphere for science rather than positioning; a large number of ground-based sensors fed data into a cloud-based processing environment through mobile devices, thus enabling a tomographic analysis of the global ionosphere at an unprecedented resolution and coverage. Zhou et al. [12] presented a novel bus arrival time prediction system based on crowd sensing and predicted bus arrival times by encouraging passengers to collect generally available and energy-efficient sensing resources, including cell tower signals, movement statuses, and audio recordings. Vassilis et al. [13] exploit public transit bus passengers Bluetooth-capable devices to capture and reconstruct micro- and macro-passenger behavior. However, owing to the lack of theoretical research on mobile crowd sensing, existing mobile sensing applications suffer from poor sensing data qualities; thus, the commercial

deployment of mobile crowd sensing systems is restricted. Collecting comprehensive tempo-spatial sensing data is the foremost criterion of mobile crowd sensing systems. Sensing data collected by 85 mobile nodes for two months revealed that people from areas with a small population are willing to collect sensing data [14]. The reverse auction-based dynamic price (RADP) mechanism [15] introduces incentive mechanisms to mobile crowd sensing systems. The most inexpensive sensing data are utilized to increase the total amount of sensing data collected by participants; as a result, the accuracy of sensing data is improved. However, mobile sensing nodes are usually clustered according to time and space. Consequently, the collected data are inhomogeneous in time and space. To collect homogeneous sensing data, the greedy budgeted maximum coverage (GBMC) algorithm [16] maximizes not only the amount of collected sensing data but also the coverage of sensing data. The Information Sampling (ISAM) [17] minimizes the diversity between the collected data and the forecast model to improve the quality of sensing data. Unlike RADP, GBMC and ISAM investigate the homogeneous space distribution of sensing data.

In mobile crowd sensing systems, the mobile crowd participants frequently encounter one another and are thus provided an opportunity to collaborate and provide high-quality sensing services. Collaborative sensing is one of the essential features of participatory sensing systems. Context characteristics are generated from the sensing data collected by multiple nearby participants through the use of a collaborative learning algorithm [18]. MOSDEN (a mobile sensor data engine) [19] is a collaborative mobile sensing framework that can operate on smartphones to capture and share sensing data among multiple distributed applications and users. Similarly, a cloud-assisted collaborative sensing method was proposed in [20] to reduce the energy consumption of mobile phone sensing applications.

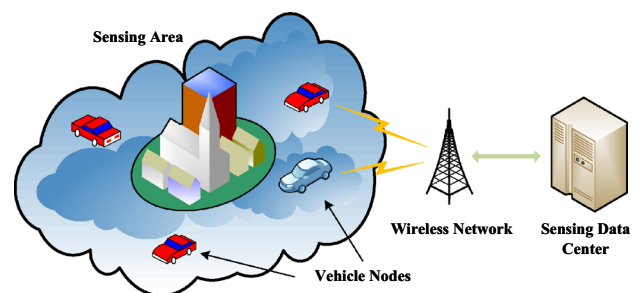
Sensing node selection is a major challenge in mobile crowd sensing systems because of the diversity of the sensing capabilities of mobile devices and the uncontrollable trajectories of mobile nodes [21]. Chien et al. [22] introduced the online task assignment problem in which heterogeneous tasks are assigned to workers with different unknown skill sets. Reddy et al. [23] developed a selection framework to allow organizers to identify well-suited sensing nodes for data collection on the basis of their geographic and temporal availabilities as well as their habits. Tuncay et al. [24] exploited the stability of user behavior and selected sensing nodes according to the fitness of the mobility history profiles of the users. Moreover, modeling the mobility of sensing nodes and forecasting the traces of sensing nodes can reduce the required amount of participating sensing nodes [25, 26]. When selecting sensing nodes, forecasting the traces of the nodes with their

call logs can reduce energy consumption and result in high sensing data coverage ratio [27]. Song et al. [28] proposed a multi task oriented Dynamic participant selection (DPS) algorithm, which selects participants under limited budgets. However, none of the previous studies considered both the heterogeneity of sensing nodes, the limited budget of sensing systems and the properties of the sensing data.

A vehicle is a special type of node in mobile crowd sensing systems. Considering that vehicles are widely distributed in urban areas and equipped with various types of sensors, vehicles provide a natural means to realize mobile sensing. In early vehicle-based environmental sensing applications, only one vehicle is recruited in a system. A novel application is the monitoring of locations through a vehicle radar [29]. Air quality can also be monitored with a CO<sub>2</sub> sensor installed in a vehicle [30]. Similarly, rainy weather can be monitored with an onboard camera [31]. In [32], a laser scanner and an onboard camera were utilized to update a street view map. However, in these applications, no sensing data are shared among vehicles. This condition constrains the extensive deployment of vehicle sensing-based applications. A platform was designed in [33] for large-scale vehicle collaborative sensing. However, the algorithm was designed to plan the trajectories of a robot team to collect sensing data within the shortest time. In [34], multiple probe vehicles were utilized to estimate large-scale traffic in an urban environment. Vehicle sensing can also be utilized for safe driving. A roadway weather information system developed in [35] recruited probe vehicles equipped with GPS, an anti-lock braking system, and acceleration sensors to quickly detect the conditions of road surfaces according to the side slip force of the vehicles in certain road segments.

### 3 System model

A vehicular mobile crowd sensing system is composed of sensing data servers, wireless networks, and vehicles (i.e., sensing nodes), as shown in Fig. 1. Vehicular sensing nodes are installed in multiple types of sensing devices,



**Fig. 1** Architecture of vehicular mobile crowd sensing systems

which can collect multiple types of sensing data. All vehicles have wireless communication interfaces. When a vehicle is selected as a sensing node, the sensors installed in it are turned on to sample the sensing data periodically. All the collected sensing data coupled with the corresponding information of the GPS coordinates are uploaded and stored in the sensing data server via wireless networks. Users of the sensing system can then make a query on the sensing data of interest from the sensing data server.

### 3.1 System architecture

In vehicular mobile crowd sensing systems, a sensing node is a vehicle installed with sensing devices and wireless communication interfaces. All sensing nodes are supposed to be willing to participate in the collection of sensing data. Vehicular sensing nodes are heterogeneous.

On the one hand, vehicles are manufactured by different companies, and the sensing interfaces embedded in them are different. For example, vehicle  $\alpha$  can collect temperature, noise, and humidity data, whereas vehicle  $\beta$  can collect temperature, speed, and vibrations (the notations used in this paper are shown in Table 1). On the other hand, vehicles are driven by different people, and

the trajectories of vehicles are random. Therefore, collecting sensing data that cover an entire tempo-spatial sensing space is difficult for a single vehicle to accomplish. The sensing data center is in charge of a vehicular sensing system. The collected sensing data are stored in the sensing data center. According to the collected sensing data and the sensing interfaces, as well as to the trajectories of the vehicular sensing nodes, the sensing data center selects the proper sensing nodes to collect sensing data.

Different sensing node selection strategies may have different effects on sensing data collection. An example of a sensing node selection scenario is shown in Fig. 2. The right part of the figure shows the collected sensing data in the sensing data center, and the left part of the figure shows the sensing capabilities of the vehicles and the future mobility traces of all available participants. In this sensing scenario,  $\{a, b\}$  is a more preferable set of participants than  $\{a, b\}$  when meeting the task requirements. However, in real-world scenarios with a significantly large number of sensing data requirements and available heterogeneous sensing vehicles, the selection of appropriate sensing vehicles becomes an arduous task. The present work presents a heterogeneous sensing vehicle selection strategy

**Table 1** List of notations

Notation	Description
$I$	The set of subareas
$i, j \in I, 1 \leq i \leq N$	One subarea
$W = \{w_1, w_2, \dots, w_M\}$	M types of sensing data
$w_l$	Life cycle of the $l$ th type of sensing data
$\Omega$	The set of candidate vehicles
$\alpha, \beta$	Sensing vehicles
$Y_\alpha = (y_1^\alpha, y_2^\alpha, \dots, y_M^\alpha)$	Sensing interfaces of vehicle $\alpha$
$\zeta_i^\alpha$	The moment when vehicle $\alpha$ travels into subarea $i$
$\theta_i^\alpha$	The residence time of vehicle $\alpha$ in subarea $i$
$Z(t)$	The valid sensing data in the sensing data center at time $t$
$\zeta_{i,l}(t)$	The valid function of the sensing data of type $l$ in sub area $i$ at time $t$
$X = \{X(t), t \geq 0\}$	Mobility procedure of a vehicle
$X(t)$	The state (i.e., location) of the vehicle at time $t$
$P_{ij}(t)$	The transition probability of a vehicle from sub area $i$ to sub area $j$ after time duration $t$
$\mathbf{Q} = (q_{ij})$	Q Matrix of vehicle's mobility procedure
$u_i^\alpha$	Utility of sensing node $\alpha$ in sub area $i$
$U_\alpha$	The utility of vehicle $\alpha$
$\Phi$	Total budget of the sensing system within one system duty cycle
$\varphi_\alpha$	The cost of sensing vehicle $\alpha$
$S$	Selected sensing vehicles
$\mathbf{H}_\alpha(t)$	The indicate matrix of the sensing data collected by $\alpha$
$\Psi(t)$	Temp indicate matrix of the sensing data in the sensing data center

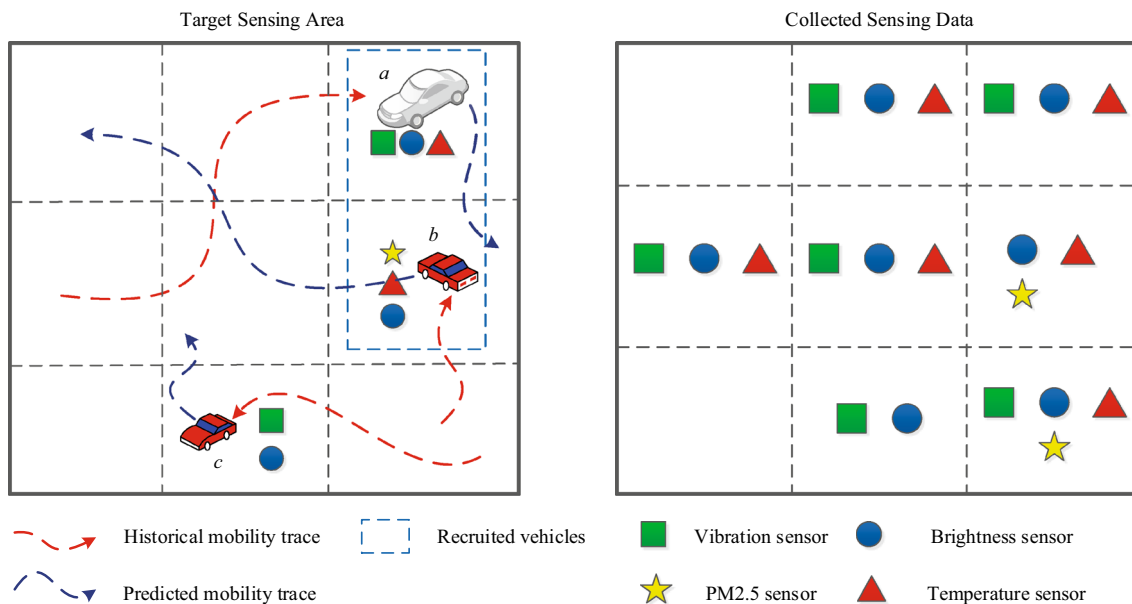


Fig. 2 Data collection by heterogeneous vehicles

that can collect the most comprehensive tempo-spatial sensing data with a limited number of vehicles.

### 3.2 Sensing data model

The geographic area where the vehicular crowd sensing system collects sensing data is defined as the target sensing area or simply the sensing area. The target sensing area is divided into lattice cells according to the geographical location (Fig. 2), and each lattice cell is called a subarea. A subarea is a unit area of the sensing system; the assumption is that the sensing data sampled at any point in the subarea can indicate the sensing value of the subarea. The entire sensing area is divided into  $N$  subareas, as shown in Fig. 2. The set of subareas in the sensing area is denoted as  $I = \{1, 2, \dots, N\}$ . The  $i$ th subarea is denoted as  $i (i \in I, 1 \leq i \leq N)$ .

Given that sensing data are usually sensitive to time (e.g., air temperature data collected an hour ago may no longer be accurate in the succeeding hours), sensing data should be sampled periodically in each unit sensing area. The  $M$  types of sensing interfaces installed in vehicle sensing nodes are denoted as  $W = \{w_1, w_2, \dots, w_M\}$ , and the life cycle of the  $l$ th type of sensing data is denoted as  $w_l$ . If  $Y_\alpha = (y_1^\alpha, y_2^\alpha, \dots, y_M^\alpha)$  denotes the sensing data types that can be collected by sensing node  $\alpha$ , then

$$y_l^\alpha = \begin{cases} 0 \\ 1 \end{cases} \tag{1}$$

where  $y_l^\alpha = 1$  indicates that  $\alpha$  can collect the sensing data of type  $l$  and  $y_l^\alpha = 0$  indicates that  $\alpha$  cannot collect the data of type  $l$ .

Vehicles away collect sensing data of a specific subarea when it is traveling in the subarea. Let  $\zeta_i^\alpha$  denote the moment when vehicle  $\alpha$  travels into subarea  $i$ , and  $\theta_i^\alpha$  denotes the duration vehicle  $\alpha$  stay in subarea  $i$ . Then, the life cycle function of the sensing data of type  $w_l$  collected by  $\alpha$  when it traveling in subarea  $i$  could be denoted as follow.

Vehicles collect the sensing data of a specific subarea when traveling in such subarea. The moment when vehicle  $\alpha$  travels into subarea  $i$  is denoted as  $\zeta_i^\alpha$ , and the time duration within which vehicle  $\alpha$  stays in subarea  $i$  is denoted as  $\theta_i^\alpha$ . Then, the life cycle function of the sensing data of type  $w_l$  collected by  $\alpha$  when traveling in subarea  $i$  can be denoted as follows:

$$h_{ii}^\alpha(t) = \begin{cases} y_l^\alpha & \zeta_i^\alpha \leq t \leq \zeta_i^\alpha + \theta_i^\alpha + w_l \\ 0 & other \end{cases} \tag{2}$$

The valid sensing data in the sensing data center at time  $t$  is denoted as  $Z(t)$ , which is expressed as

$$Z(t) = \begin{bmatrix} \zeta_{1,1}(t) & \zeta_{1,2}(t) & \dots & \zeta_{1,M}(t) \\ \zeta_{2,1}(t) & \zeta_{2,2}(t) & & \zeta_{2,M}(t) \\ \vdots & & & \vdots \\ \zeta_{N,1} & \zeta_{N,2}(t) & \dots & \zeta_{N,M}(t) \end{bmatrix}, \tag{3}$$



where  $\zeta_{i,l}(t)$  is the valid function of the sensing data of type  $l$  in subarea  $i$  at time  $t$ .

$$\zeta_{i,l}(t) \begin{cases} 1 & \text{the sensing data is effective at } t \\ 0 & \text{the sensing data is ineffective at } t \end{cases} \quad (4)$$

### 3.3 Mobility model

A time-continuous life cycle of sensing data is considered in this paper. As a result, to estimate the precise collection time of the sensing data, a time-continuous Markov Chain-based mobility model is used. In vehicular crowd sensing systems, vehicles are traveling in or transferring between subareas. The mobility process is denoted as  $X = \{X(t), t \geq 0\}$ , where  $X(t)$  denotes the state of the vehicle at time  $t$  (i.e., the subarea in which the vehicle stays at time  $t$ ). Normally, the state of a vehicle is only related to its state at the last moment. Consequently, for any  $0 \leq t_0 < t_1 < \dots < t_n < t_{n+1}$ , and  $i_k \in I, 0 \leq k \leq n + 1$ , we have

$$\begin{aligned} P\{X(t_{n+1}) = i_{n+1} | X(t_0) = i_0, X(t_1) = i_1, \dots, X(t_n) = i_n\} \\ = P\{X(t_{n+1}) = i_{n+1} | X(t_n) = i_n\} \end{aligned} \quad (5)$$

Equation (5) means that the mobility process  $X = \{X(t), t \geq 0\}$  is a time-continuous Markov chain. For any  $s, t \geq 0$ , and  $i, j \in I$ , we have

$$\begin{aligned} P\{X(s+t) = j | X(s) = i\} \\ = P\{X(t) = j | X(0) = i\} \\ = P_{ij}(t) \end{aligned} \quad (6)$$

where  $P_{ij}(t)$  denotes the transition probability of a vehicle from subarea  $i$  to subarea  $j$  after time duration  $t$ . The matrix of transition probability is denoted as  $\mathbf{P}(t) = (P_{ij}(t))$ ,  $i, j \in I$ . We define the transfer rate matrix of the mobility process  $X = \{X(t), t \geq 0\}$  as  $\mathbf{Q} = (q_{ij})$ , which is also called the  $Q$  matrix. For any  $i \in I$ , we define

$$q_{ii} = \lim_{\Delta t \rightarrow 0} \frac{1 - P_{ii}(\Delta t)}{\Delta t} \quad (7)$$

For any  $i, j \in I, j \neq i$ , we also define

$$q_{ij} = \lim_{\Delta t \rightarrow 0} \frac{P_{ij}(\Delta t)}{\Delta t} \quad (8)$$

where  $q_{ij}$  denotes the transition intensity of the vehicle from subarea  $i$  to subarea  $j$ . As the number of the subareas in the target sensing area is limited, for  $\forall i \in I, 0 < \sum_{j \neq i} q_{ij} = q_i < \infty$ . The average duration that the vehicle stays in state  $X(0) = i$  is determined by  $q_i$ . Therefore, the value of  $q_i$  can be estimated by the historical mobility traces of the vehicles.

According to the Kolmogorov backward differential equation,

$$\mathbf{P}'(t) = \mathbf{Q}\mathbf{P}(t) \quad (9)$$

If  $\mathbf{P}(0) = \mathbf{I}$ , then the transition probability of the Markov chain  $X$  could be derived as follows:

$$\mathbf{P}(t) = e^{\mathbf{Q}t} \quad (10)$$

## 4 Utility of sensing vehicles

### 4.1 Utility function

The sensing capacity of a sensing vehicle is indicated by the utility of the vehicle. The utility of vehicle  $\alpha$  in the following time interval  $t_u$  is denoted as  $t_\alpha$ . The utility of a sensing vehicle is the metric for its sensing capacity. The amount of sensing data collected by a vehicle depends largely on the duration within which the sensing vehicle collects the sensing data. If the vehicle spends a long time collecting sensing data, its sensing utility will be large. However, the computing cost of the utility will also be high if the sensing time is considerably long. Therefore, a system duty cycle,  $t_u$ , is set to reduce the cost of utility calculation. At the beginning of a system duty cycle, the utilities of the sensing vehicles are updated. Then, the proper sensing nodes are selected according to the utilities of the sensing vehicle candidates. If  $u_i^\alpha$  denotes the utility of sensing node  $\alpha$  in subarea  $i$ , then

$$u_i^\alpha = \sum_{l=1}^M \left\{ y_i^\alpha \cdot \int_0^{t_u} [h_{il}^\alpha(t) - h_{ii}^\alpha(t) \cdot \zeta_{il}(t)] dt \right\} \quad (11)$$

Utility  $u_i^\alpha$  denotes the amount of sensing data that vehicle  $\alpha$  can collect in the next system duty cycle  $t_u$  in subarea  $i$ ;  $h_{il}^\alpha(t)$  denotes the effective duration of the sensing data of type  $l$  collected by  $\alpha$  in subarea  $i$ . If  $\alpha$  arrives in subarea  $i$  at time  $t$  then stays in subarea  $i$  for a duration of  $\tau$  (i.e.,  $\zeta_i^\alpha = t, \theta_i^\alpha = \tau$ ), then

$$\int_{t=0}^{t_u} h_{il}^\alpha(t) dt = \tau + w_l \quad (12)$$

Thus, for any  $\zeta_i^\alpha$  and  $\theta_i^\alpha$ , we have

$$\begin{aligned} \int_{t=0}^{t_u} h_i^l(\alpha, t) dt \\ = \iint_{t \in (0, t_u), \tau \in (0, t_u - t)} \{P(\zeta_i^\alpha = t) \cdot P(\theta_i^\alpha = \tau) \cdot (\tau + w_l)\} \cdot d\tau dt \end{aligned} \quad (13)$$

where  $P(\zeta_i^\alpha = t)$  is the probability density of the time that vehicle  $\alpha$  arrives in subarea  $i$  and  $P(\theta_i^\alpha = \tau)$  is the

probability density of the duration within which vehicle  $\alpha$  stays in subarea  $i$ . Equation (13) is to calculate the cumulative efficient sensing data of type  $l$  that can be collected by node  $\alpha$  in subarea  $i$  in one system duty cycle. If  $\alpha$  arrives subarea  $i$  at time  $t$ , then the effective duration of the collected sensing data in the current system duty cycle is  $t_u - t$ . Therefore, in Equation (13), the region of  $\tau$  is  $(0, t_u - t)$ . Then, the utility of vehicle  $\alpha$  can be calculated as follows:

$$U_\alpha = \sum_{i=1}^N u_i^\alpha \tag{14}$$

### 4.2 Distribution of residence time

**Theorem** If  $\theta_i$  denotes the residence time of a vehicle in subarea  $i$ , then  $\theta_i$  obeys the exponential distribution with parameter  $\lambda_i$ , which is the average residence time of the vehicle in subarea  $i$ .

$$P\{\theta_i \leq t | X(0) = i\} = 1 - \exp(-q_i t) \tag{15}$$

where  $q_i = \lambda_i$ .

*Proof* If the sensing node arrives in subarea  $i$  at time 0, that is,  $X(0) = i$ , and the residence time  $\theta_i$  can be defined as

$$\theta_i = \inf\{t : t > 0, X(t) \neq X(0), X(0) = i\} \tag{16}$$

then

$$P(\theta_i > t | X(0) = i) = P[X(u) = i, 0 \leq u \leq t | X(0) = i] \tag{17}$$

First, the uncountable events are translated into countable events, as shown in Equation (18).

$$B = \{Y : X(u) = i, 0 \leq u \leq t\} = \bigcap_{0 \leq u \leq t} \{Y : X(u) = i\} \tag{18}$$

where  $B$  indicates that the sensing node stay in subarea  $i$  in time duration  $[0, t]$ . Then, the interval  $[0, t]$  is uniformly divided into  $2^n$  subintervals as follows:

$$A_n = \left\{ Y : X\left(\frac{k}{2^n}t\right) = i, k = 0, 1, 2, \dots, 2^n \right\} \\ = \bigcap_{k=0}^{2^n} \left\{ Y : X\left(\frac{k}{2^n}t\right) = i \right\} \tag{19}$$

$A_n$  is the event that sensing node is in subarea  $i$  at time  $\frac{k}{2^n}t$ , where  $k = 0, 1, 2, \dots, 2^n$ . Given that:  $A_{n+1} \subset A_n$ , it could be denoted as follows:

$$A = \bigcap_{n=1}^{\infty} A_n = \lim_{n \rightarrow \infty} A_n \tag{20}$$

$A$  also indicates the event that the sensing node stays in subarea  $i$  from time 0 to time  $t$ . Obviously,  $B = A$ . Therefore,

$$P\{\theta_i > t | X(0) = i\} \\ = P\{X(u) = i, 0 \leq u \leq t | X(0) = i\} \\ = P\{B | X(0) = i\} \\ = P\{A | X(0) = i\} \\ = \lim_{n \rightarrow \infty} P\{A_n | X(0) = i\} \\ = \lim_{n \rightarrow \infty} P\left\{X\left(\frac{k}{2^n}t\right) = i, k = 0, 1, 2, \dots, 2^n | X(0) = i\right\} \\ = \lim_{n \rightarrow \infty} \left\{P_{ii}\left(\frac{t}{2^n}\right)\right\}^{2^n} \\ = \lim_{n \rightarrow \infty} \exp\left\{2^n \ln P_{ii}\left(\frac{t}{2^n}\right)\right\} \\ = \lim_{n \rightarrow \infty} \exp\left\{\frac{\ln[1 - q_i(\frac{t}{2^n}) + o(\frac{t}{2^n})]}{-q_i t / 2^n}(-q_i t)\right\} \\ = \exp(-q_i t) \tag{21}$$

Thus, the residence time of a vehicle in subarea  $i$  is subject to the exponential distribution with parameter  $q_i$ . The average residence time in the subarea is denoted as  $\lambda_i$ , and  $\lambda_i = q_i$ , which means that the parameter of the exponential distribution can be estimated by the average residence time of the sensing vehicles in the subarea.

### 4.3 Distribution of arrival time

The arrival time of a vehicle in subarea  $j$  under the condition that the vehicle is in subarea  $i$  at the initial time (i.e.,  $t = 0$ ) is denoted as  $\xi_{ij}$  and defined as

$$\xi_{ij} = \inf\{t : t > 0, X(t) = j | X(0) = i\} \tag{22}$$

where  $\xi_{ij} = 0$  if  $X(0) = j$ .

$F_{ij}(t)$  is defined as the distribution function of  $\xi_{ij}$  and is expressed as

$$F_{ij}(t) = P\{\xi_{ij} \leq t\}, \quad i, j \in I, i \neq j, t \geq 0 \tag{23}$$

If  $\vartheta_1 = \inf\{t : t > 0, X(t) \neq X(0)\}$  denotes the moment that the vehicle leaves its initial state and  $G_i(x) = P(\vartheta_1 \leq x | X(0) = i) = (1 - \exp(-q_i x))$ , then

$$F_{ij}(t) = q_i^{-1} q_{ij} G_i(t) + q_i^{-1} \sum_{k \neq i, k \neq j, k \in I} q_{ik} \int_0^t F_{kj}(t-u) dG_i(u) \tag{24}$$

If  $\vartheta' = \inf\{t : t > 0, X(\vartheta_1 + t) = j\}$  and  $X'(t) = X(\vartheta_1 + t)$ , then  $\vartheta = \vartheta_1 + \vartheta'$ . Based on the total probability formula and the strong Markov property,  $\vartheta_1$  and  $X(\vartheta_1)$  are conditionally independent when  $X(0) = i$ ; thus, Eq. (24) is true.

## 5 Selection of Heterogeneous Sensing Vehicles

The total cost of the sensing system within a system duty cycle is denoted as  $\Phi$ , and the cost of the sensing vehicle  $\alpha$  is  $\varphi_\alpha$  in a system duty cycle. If  $\Omega$  denotes the set of the sensing vehicle candidates, then the total utility should be maximized under the constraint of the maximum number of vehicles for sensing data collection.

$$\begin{aligned} \text{Max} \sum_{\alpha \in S} U_\alpha \\ \text{s.t.} \sum_{\alpha \in S} \varphi_\alpha \leq \Phi \end{aligned} \quad (25)$$

In Eq. (25),  $S$  denotes the selected sensing vehicles. The target of the optimization problem is to select a set  $S$  of vehicles from  $\Omega$ , and the total cost of the selected sensing vehicles in a system duty cycle is less than  $\Phi$ . Furthermore, the selected vehicles in  $S$  are able to collect the most comprehensive tempo-spatial sensing data (i.e., the sum of the utilities of the vehicles in  $S$  is maximized).

The optimization formula in Eq. (25) is a knapsack problem. Therefore, a greedy algorithm denoted as Algorithm 1 is designed to obtain the solution  $S$  in Eq. (25). The goal of Algorithm 1 is to select a set of vehicles from the candidates so as to collect as much sensing data as possible in the sensing area within the budget at each system duty cycle.

In our algorithm, after a sensing vehicle is selected, the indicate matrix of its expected sensing data are added into a temp indicate matrix of the sensing data in the sensing data center. Then, the utilities of the left sensing candidates are updated according to the new indicate matrix of the sensing data. It will circulate until the total budget of the system in the specific duty cycle is exhausted. The indicate matrix of the sensing data collected by  $\alpha$  is denoted as

$$\mathbf{H}_\alpha(t) = \begin{bmatrix} h_{1,1}(t) & h_{1,2}(t) & \dots & h_{1,M}(t) \\ h_{2,1}(t) & h_{2,2}(t) & & h_{2,M}(t) \\ \vdots & & & \vdots \\ h_{N,1} & h_{N,2}(t) & \dots & h_{N,M}(t) \end{bmatrix} \quad (26)$$

where  $t \in [0, t_u]$ . The temp indicate matrix of the sensing data in the sensing data center is denoted as  $\Psi(t)$ , which is expressed as

$$\Psi(t) = \begin{bmatrix} \psi_{11}(t) & \psi_{12}(t) & \dots & \psi_{1M}(t) \\ \psi_{21}(t) & \psi_{22}(t) & \dots & \psi_{2M}(t) \\ \vdots & \vdots & & \vdots \\ \psi_{N1}(t) & \psi_{N2}(t) & \dots & \psi_{NM}(t) \end{bmatrix} \quad (27)$$

At the beginning,  $\Psi(t) = \mathbf{Z}(t)$ . After the sensing vehicle  $\alpha$  is selected,  $\Psi(t) \leftarrow \Psi(t) + \mathbf{H}_\alpha(t)$ , where

$$\psi_{il}(t) \leftarrow \psi_{il}(t) + \gamma_{il}^\alpha(t) - \psi_{il}(t) \cdot \gamma_{il}^\alpha(t)$$

. Then, the temp utility function of sensing vehicle  $\alpha$  is defined as Eq. (28).

$$U'_\alpha = \left\{ \sum_{l=1}^M \sum_{i=1}^N \int_{t=0}^{t_u} [\gamma_{il}^\alpha(t) - \gamma_{il}^\alpha(t) \cdot \psi_{il}(t)] dt \right\} / \varphi_\alpha \quad (28)$$

---

### Algorithm 1: Optimal heterogeneous vehicle selection algorithm

---

**Input:** Candidate vehicles  $\Omega$ ; system budget in each system duty cycle  $\Phi$ ;  $\varphi_\alpha, \alpha \in \Omega$ ;  $\mathbf{Q} = \{q_{ij}\}$

**Output:** Optimized selected sensing vehicles  $S$

```

1  $\Psi(t) \leftarrow \mathbf{Z}(t)$ ;
2  $\Omega_{tmp} \leftarrow \Omega$ ;
3  $\Phi' \leftarrow \Phi$ ;
4 while  $\Phi' \geq 0$  do
5   for each  $\alpha$  in  $\Omega_{tmp}$  do
6     calculate  $U'_\alpha$ ;
7   end
8    $\alpha \leftarrow \arg \max_{\beta \in \Omega_{tmp}} \{U'_\beta\}$ ;
9    $\Omega_{tmp} \leftarrow \Omega_{tmp} - \{\alpha\}$ ;
10   $\mathbf{S} \leftarrow \mathbf{S} + \{\alpha\}$ ;
11   $\Phi' \leftarrow \Phi' - \varphi_\alpha$ ;
12   $\Psi(t) \leftarrow \Psi(t) + \mathbf{H}_\alpha(t)$ ;
13 end
14 return  $S$ ;
```

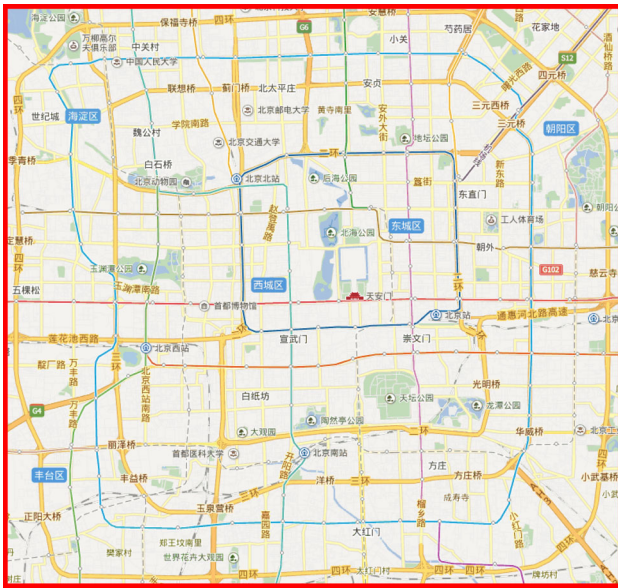
---

## 6 Evaluation

Real vehicle GPS traces from the T-Drive trajectory [36, 37] are utilized to simulate and evaluate the proposed HVS strategy. The T-Drive trajectory data set contains one-week trajectories of 10357 taxis. The total number of GPS points in this data set is approximately 15 million, and the total distance of the trajectories is 9 million kilometers. The time frequency of data sampling is set to 5 seconds. Taxi drivers can usually find an optimal route to a destination based on their experience. Therefore, the T-Drive project attempts to improve the efficiency of the navigation software according to the experience of taxi drivers. The traces of taxis can reflect the characters of passengers. As the mobilities of passengers are not completely random, the mobilities of taxis are also predictable. For example, Huang et al. [38] designed a vehicle mobility model on the basis of the regular patterns derived from the traces of 4000 taxis in Shanghai.

In our simulations, a rectangular region around the Fourth Ring Road of Beijing is adopted as the target sensing area (Fig. 3), whose latitude is between  $(39.840018^\circ \text{ N}, 39.993971^\circ \text{ N})$ , and longitude is between  $(116.276206^\circ \text{ E}, 116.494243^\circ \text{ E})$ . The traces in the rectangular area are used in the simulations. Although the traces are not limited in the target sensing area, we abstract





**Fig. 3** Target sensing area

the area out of the target sensing area as subarea 0 and the subareas in the target sensing area as 1, 2, 3, ... The 1000 vehicles with the longest traces in the rectangular region are employed as sensing vehicle candidates, and the traces are used to calculate the arrival and residence time distribution among the subareas.

The default parameters of the experiments are set as follows. The default number of selected sensing vehicles is 60. The total time of the experiment is 533315 seconds (more than 6 days). The rectangular region is uniformly divided into 100 (10 × 10) subareas. The system duty cycle is set to 1000 s. Ten types of sensors are employed to collect data for the sensing system, and each vehicle has 50% possibility to be equipped with each sensor. The default upper bound of each type of sensing data life cycle is 10 system duty cycles. The cost of the sensing system is assumed to be directly proportional to the number of selected vehicles. Then, the number of selected sensing vehicles is used to denote the cost of the sensing system. The simulation parameters and their default values are shown in Table 2.

We compare the HVS algorithm with three other algorithms, as shown in Table 3. Dynamic participant selection (DPS) [28] is a multitask-oriented mobile crowd sensing system. Although the objective of DPS is to collect sensing data for all sensing tasks instead of comprehensive tempo-spatial sensing data, it proposes a greedy sensing node selection algorithm on the basis of utility calculation. DPS does not consider sensing data life cycle when estimating the sensing capacity of a sensing node; instead, it uses a probability-based mobility model. Random selection (RS) algorithms randomly select sensing vehicles under the

**Table 2** Simulation parameters

Parameter	Default Value
Simulation time	533315 s
Number of selected vehicles	60
Number of vehicle candidates	1000
System duty cycle	1000 s
Sensor types	10
Sensing data life cycle	10 System duty cycles
Area partition	100 (10 × 10)

system budget constraint; specifically, RS1 considers sensing data life cycle, whereas RS2 does not.

During the simulations, the collected sensing data are recorded in a list, and each record is composed as follows:

[node name, time index, location, sensing data type and value( A, B, C, ... ) ]

The sensing data coverage ratio is the ratio of the effective sensing data coverage to the product of the simulation period length and the number of subareas. As the relation between the system budget and the number of selected vehicles is linear, the number of selected vehicles is used in the following simulations instead of the system budget in each system duty cycle.

### 6.1 Impact of the number of selected vehicles on sensing data coverage ratio

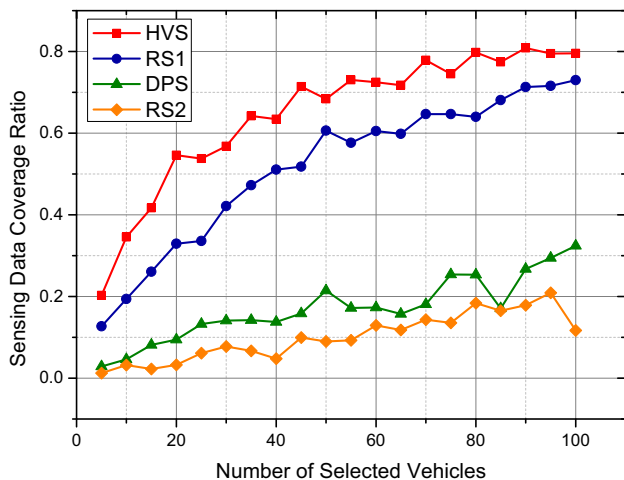
In this subsection, the impact of the number of selected vehicles on the sensing data coverage ratio is evaluated. The number of selected vehicles bears the most significant impact on the sensing data coverage ratio because it determines the amount of data that can be collected. Given the number of sensing vehicles that varies from 5 to 100 in increments of 5 for a system duty cycle, we show the sensing data coverage ratio results in Fig. 4. The collected data cannot match all the data requirements even if 100 available vehicles are selected because the traces are not uniformly distributed in spatial and temporal spaces.

Figure 4 shows the impact of the number of selected vehicles on the coverage ratio of the collected sensing data. The experiment is executed 10 times, and the results in Fig. 4 show the average values from the 10 executions. Figure 4 shows that the coverage ratios of the collected sensing data of the four algorithms increase with an increase in the number of selected vehicles because many vehicles can collect a large amount of sensing data.

Among the four strategies, the HVS algorithm acquires the highest coverage ratio because it considers tempo-spatial sensing data coverage when selecting sensing

**Table 3** Evaluated algorithms

Algorithm	HVS	DPS	RS1	RS2
Node selection	Greedy	Greedy	Random	Random
Mobility model	Time-continuous Markov chain-based	Probability based	No	No
Sensing data life cycle	Yes	No	Yes	No



**Fig. 4** Data coverage ratio under different numbers of selected vehicles

vehicles. Indeed, the life cycle of sensing data is usually much longer than the system duty cycle. Therefore, collecting a certain type of sensing data in each system duty cycle is unnecessary. Sensing data are better collected just before they expire in the sensing data server. As a result, the consideration of the temporal coverage of sensing data allows the HVS algorithm to acquire relatively high sensing data coverage ratios.

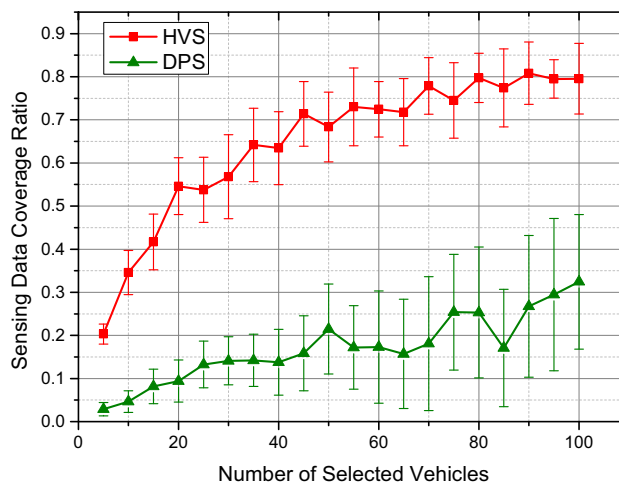
In calculating the coverage ratio, both the HVS algorithm and RS1 consider the sensing data temporal coverage. As a result, their coverage ratios are higher than those of DPS and RS2. Moreover, the coverage ratio of the HVS algorithm is higher than that of RS1 by up to 20%. This result may be ascribed to the fact that the HVS algorithm selects vehicles according to their historical traces, sensing interfaces, and the collected sensing data in the sensing data center, whereas RS1 randomly selects vehicles. The coverage ratio of the HVS algorithm is four times higher than that of DPS because it considers the temporal sensing data coverage when calculating the expected contributions to the existing collected sensing data; furthermore, the HVS algorithm predicts the location of sensing vehicles using a time-continuous Markov chain-based mobility model, which is more accurate than that employed in DPS.

Figure 5 shows the impact of the number of selected vehicles on the standard deviation of the data coverage ratio. Since the relation between the number of selected

vehicles and data coverage ratio of the four algorithms are shown in Fig. 4, only the results of HVS and DPS are shown in Fig. 5 for clarity. In Fig. 5, the standard deviation of DPS increases from 0.015 to 0.17 with an increase in the number of selected sensing vehicles from 5 to 100, whereas the standard deviation fluctuation of the HVS algorithm is small (i.e., from 0.02 to 0.09). Figure 5 also reveals that the standard deviation of the HVS algorithm is lower than that of DPS. The deviation of vehicle traces is high when numerous vehicles are selected, and the high deviation of the vehicle traces leads to a considerable deviation in the coverage ratio of the collected sensing data. The fluctuation of the coverage ratio of the collected sensing data is also significantly reduced with the HVS algorithm because of its adoption of a time-continuous Markov chain-based mobility model and consideration of the uniformity of the temporal coverage of the different types of sensing data. Thus, the sensing data collection of the HVS algorithm is more sustainable and stable than those of the other algorithms.

**6.2 Impact of sensing data life cycle on coverage ratio**

The life cycle of sensing data are usually longer than the system duty cycle. Therefore, collecting a certain type of sensing data in each system duty cycle is unnecessary.



**Fig. 5** Standard deviations of data coverage ratios for different numbers of selected vehicles

Collecting sensing data just before it expires is preferred to keep the sensing data effective along the temporal space. An experiment is conducted in this study to evaluate the impact of the upper bound of the sensing data life cycle on the sensing data coverage ratio. The life cycle of each type of sensing data is randomly chosen at an interval of 0 to the upper bound of the sensing data life cycle.

Owing to the data lifetime are not considered in DPS and RS2, only the coverage ratio of HVS and RS1 are compared in this experiment. In the experiment, the upper bound of the sensing data life cycle changes from 2 to 20 in increments of 2. As shown in Fig. 6, the coverage ratios of the HVS algorithm and RS1 increase with an increase of the upper bound of the sensing data life cycle because the long temporal coverage of sensing time leads to few requirements for sensing data collection. The coverage ratio of the HVS algorithm when the life cycle upper bound is 20 is 55 % higher than that when the upper bound of the sensing data life cycle is 2 system duty cycles. As the HVS algorithm selects sensing vehicles according to their utilities, its coverage ratio is higher than that of RS1 by about 15 % under the same upper bound of the sensing data.

### 6.3 Impact of the sensing area partition on coverage ratio

Another experiment is conducted to test the sensing data coverage ratio for different sensing area partitions. The default number of sensing vehicles is 60. The area partition is selected from  $\{(2, 2), (4, 4), (6, 6), (8, 8), (10, 10)\}$ , as shown in Fig. 7. The HVS algorithm has the highest coverage ratio in each area partition. When the area partition is (6, 6), the HVS algorithm obtains the highest sensing data coverage ratio (i.e., 89 %).

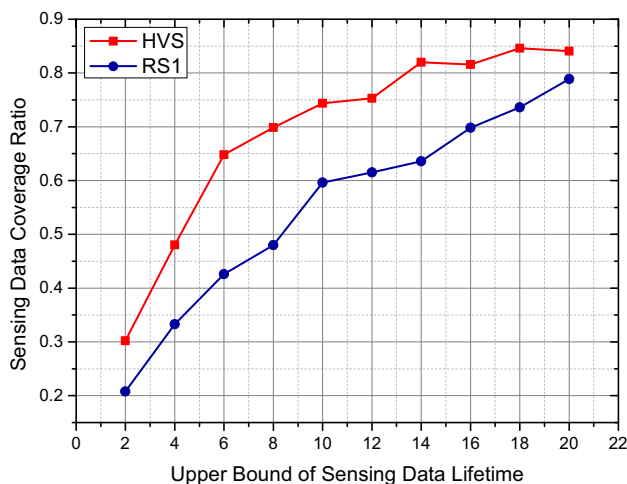


Fig. 6 Impact of the upper bound of the sensing data life cycle on coverage ratio

When the sensing area is fine-grained partitioned into 100 subareas, as a consequence of the system budget limitation in each system duty cycle, the sensing data coverage ratio decreases from 89 to 72 % with an increase in the partition fineness from (6, 6) to (10, 10). As the mobility traces of the vehicles are random, the sensing data coverage ratio is low when the number of divided subareas is small. For the same reason, DPS and RS1 acquire the highest ratios when the area partition is (4, 4). If the number of the divided subareas is small to (2, 2), then the coverage ratio deviation between the HVS algorithm and RS1 is also small to 0.02. The reason is twofold. First, the probability for one vehicle to travel into each subarea is sufficiently high, and the efficiencies of the mobility models are reduced. Second, the requirements for sensing data collection are fewer than those when the sensing area is fine-grained partitioned. However, when the number of subareas increases, the mobility model and the estimation of future sensing capacity show highly significant effects.

### 6.4 Cumulative amount of collected sensing data

In this subsection, the variation trend in the cumulative amount of the collected sensing data during the simulation is evaluated. The simulation time is divided into 30 sub-time intervals. In each subtime interval, the effective coverages of the sensing data collected by the vehicles are computed. Figure 8 shows that the HVS algorithm exhibits a faster data collection speed than the other algorithms. At the end of our simulation, sensing data collected by HVS is 3 times more than that of DPS. As DSP and RS2 do not count the life cycles of the sensing data, the cumulative amounts of the sensing data collected with DSP and RS2 are less than that of the HVS algorithm. Moreover, the

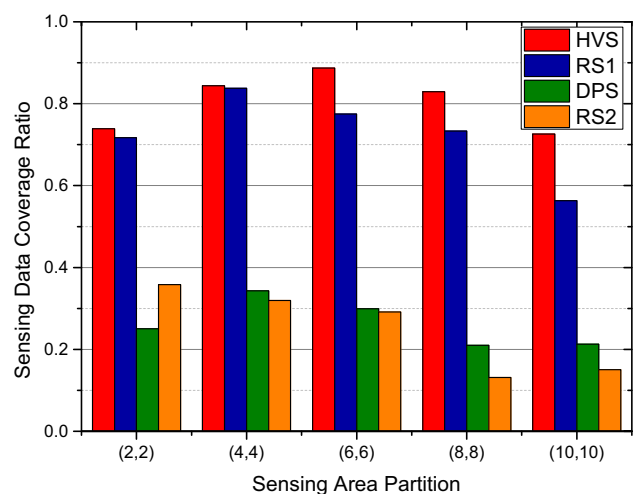


Fig. 7 Impact of sensing area partition on coverage ratio

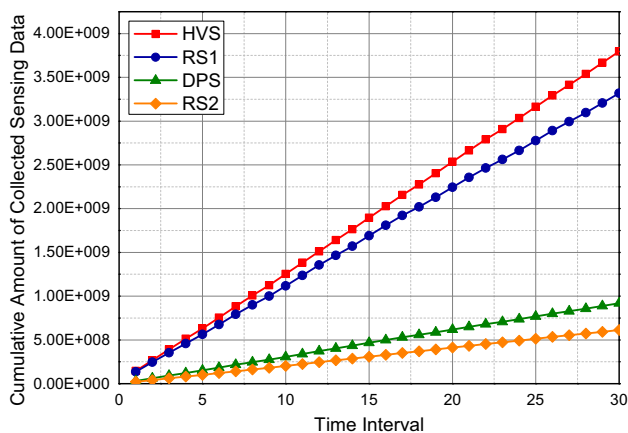


Fig. 8 Cumulative amount of collected sensing data

sensing data collection speeds of DPS and RS2 are much slower than that of the HVS algorithm.

### 6.5 Sustainability of data collection

In this subsection, the sustainability of the data collection of each of the four algorithms is evaluated with the use of the temporal entropy of the collected sensing data. The temporal entropy serves as an indicator of the uniformity of the collected sensing data at different intervals during the simulation process. Therefore, the entropy can evaluate the sustainability of the sensing data collection. A large temporal entropy equates to a highly sustainable sensing data collection algorithm. The method for calculating temporal entropy is shown in Eq. (29).

$$s_t = - \sum_{k=1}^n p_k \log p_k \tag{29}$$

where

$$p_k = \frac{\text{effective data coverage in time interval } k}{\text{total effective data coverage}}$$

In Fig. 9, all temporal entropies increase with the number of selected vehicles. Thus, the HVS algorithm can acquire sensing data sustainably. Furthermore, the HVS algorithm has the highest temporal entropy among the four algorithms when few vehicles are selected to collect sensing data. This result reveals that the HVS algorithm can collect sensing data sustainably with few sensing vehicles because it typically selects vehicles according to their utilities. This feature suggests the capacity of the algorithm to collect scarce sensing data. When numerous sensing vehicles are selected, the four algorithms can acquire uniformly distributed sensing data because the number of vehicles is sufficiently large to generate high deviations among

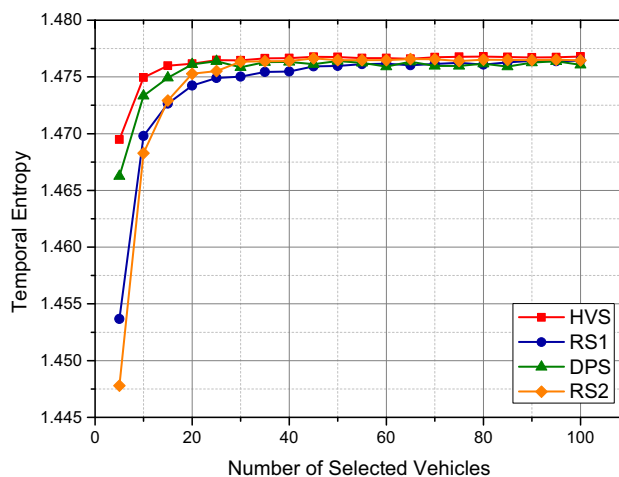


Fig. 9 Temporal entropy at different numbers of selected vehicles

mobilities and heterogeneities and consequently collect various types of sensing data in all subareas.

### 6.6 Spatial uniformity of collected sensing data

In this subsection, the impact of the number of selected sensing vehicles on the spatial entropy of the collected sensing data is evaluated. The spatial entropy of the collected sensing data reveals the spatial uniformity of the collected sensing data. The spatial entropies are derived from the probability distribution of the collected sensing data in all the partitioned subareas, as shown in Eq. (30).

$$s_s = - \sum_{i=1}^N p_i \log p_i \tag{30}$$

where

$$p_i = \frac{\text{data coverage in subarea } i}{\text{coverage of all collected data}}$$

Figure 10 shows the relationship between the spatial entropy and the number of selected vehicles. As shown in Fig. 10, the HVS amount of vehicles has the highest spatial entropy. Thus, this algorithm is able to collect uniformly distributed sensing data with the same number of sensing vehicles. Moreover, the advantage of the HVS algorithm is highlighted when the number of selected vehicles is small for three reasons. First, the mobility model is able to predict the locations of sensing vehicles. Second, when calculating the utility of each sensing vehicle, both the expected collected sensing data and the collected sensing data in the sensing data center are considered. Therefore, the selected sensing vehicles are more likely to collect scarce sensing data; as a result, the spatial uniformity of the collected sensing data is improved. Third, the life cycle of



sensing data is longer than the system duty cycle; consequently, the uniformity of sensing data is further enhanced.

Figure 11 reveals the relationship between the sensing area partition and the spatial entropy of the collected sensing data. Obviously, the spatial uniformity of the collected sensing data is high when several other subareas are partitioned. In Fig. 11, the spatial entropies of the four algorithms almost converge when the area partition is (2, 2) because when the number of subareas is small, the influence of mobility and heterogeneity on sensing capacity is low. Then, the spatial entropy increases with an increase in the number of area partitions. The spatial entropy of the HVS algorithm is slightly higher than those of the other algorithms when the area partition is (10, 10). Nevertheless, the variation tendencies of the four algorithms are almost the same.

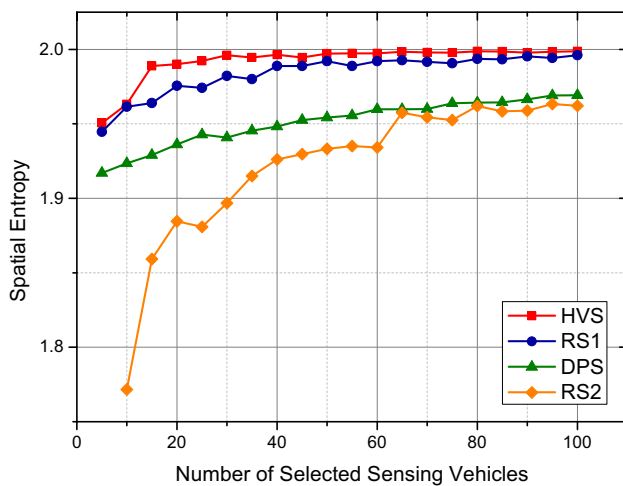


Fig. 10 Spatial entropy at different numbers of selected vehicles

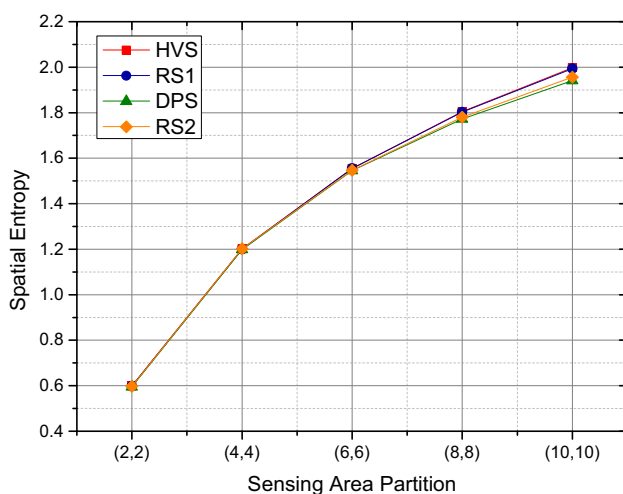


Fig. 11 Impact of area partition on spatial entropy

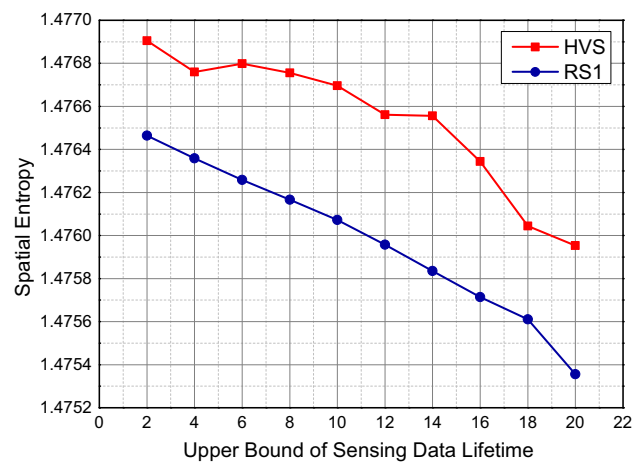


Fig. 12 Impact of the upper bound of the sensing data life cycle on spatial entropy

Figure 12 reveals the impact of the upper bound of the sensing data life cycle on the spatial entropy of the collected sensing data. Since DPS and RS2 do not select participants according to the lifetime of sensing data, the spatial entropies of HVS and RS1 are evaluated in this experiment. The spatial entropies of the HVS algorithm and RS1 decrease with an increase in the upper bound of the sensing data life cycle because the non-uniformity of the life cycles of the different types of sensing data increase with an increase in the upper bound of the sensing data life cycle. Then, the heterogeneity of the collected sensing data will also increase. As a result, the spatial entropy will be small because the upper bound of the sensing data life cycle is high. As the HVS algorithm estimates the types and effective duration of the sensing data that could be collected by vehicles in the following system duty cycle, the spatial entropy of the HVS algorithm is higher and decreases slower than that of RS1. Therefore, the HVS algorithm can mitigate the drawbacks arising from the heterogeneity of the sensing data life cycle and spatial distribution of the collected sensing data.

### 7 Conclusion

The continuous collection of comprehensively covered tempo-spatial sensing data with limited heterogeneous sensing vehicles is a critical issue in vehicular mobile sensing systems. In this work, the HVS algorithm for the collection of comprehensive tempo-spatial sensing data is proposed. First, on the basis of the expected spatial distribution of a vehicle and its sensing interfaces, a utility function is established to estimate the ability of the vehicle to collect sensing data in the next system duty cycle. Then, according to the utility function and system budget



restriction, the sensing vehicle selection problem is formulated as a knapsack problem. Finally, a greedy optimal sensing vehicle selection algorithm is designed. Real trace-driven experiments show that the sensing data collected by the HVS algorithm exhibit higher coverage ratio and more uniform tempo-spatial distribution than those collected by other mobile crowd sensing algorithms.

**Acknowledgments** This work was supported by the National Natural Science Foundation of China (61572060, 61370197, 61170296 and 61190125), the R&D Program (2013BAH35F01) and key technologies R&D program project of Tangshan (15130251a).

## References

- Ganti RK, Ye F, Lei H (2011) Mobile crowdsensing: current state and future challenges. *IEEE Commun Mag* 49:32–39
- Guo B, Yu Z, Zhou X, Zhang D From participatory sensing to mobile crowd sensing. In: *IEEE international conference on pervasive computing and communications workshops (PERCOM 2014 Workshops)*, pp 593–598
- Khan WZ, Xiang Y, Aalsalem MY, Arshad Q (2013) Mobile phone sensing systems: a survey. *IEEE Commun Surv Tutor* 15:402–427
- Devarakonda S, Sevusu P, Liu H, Liu R, Iftode L, Nath B (2013) Real-time air quality monitoring through mobile sensing in metropolitan areas. In: *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing 2013*, pp 1–8
- Zhou P, Cheng Z, Li M Smart traffic monitoring with participatory sensing. In: *Proceedings of the 11th ACM conference on embedded networked sensor systems (SenSys'13)*, pp 1–2
- Banerjee R, Sinha A, Saha A (2013) Participatory sensing based traffic condition monitoring using horn detection. In: *Proceedings of the 28th annual ACM symposium on applied computing*, pp 567–569
- Perttunen M, Kostakos V, Riekkilä J, Ojala T (2015) Urban traffic analysis through multi-modal sensing. *Pers Ubiquitous Comput* 19:709–721
- Liu Y, Li X (2015) Heterogeneous participant recruitment for comprehensive vehicle sensing. *Plos One* 10:1–20
- Dutta P, Aoki PM, Kumar N, Mainwaring A, Myers C, Willet W, Woodruff A 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 (SenSys)*, pp 349–350
- Zheng Y, Liu T, Wang Y, Zhu Y, Liu Y, Chang E Diagnosing new york city's noises with ubiquitous data. In: *Proceedings of the ACM international joint conference on pervasive and ubiquitous computing (UbiComp 2014)*, pp 715–725
- Pankratius V, Lind F, Coster A, Erickson P, Semeter J (2014) Mobile crowd sensing in space weather monitoring: the mahali project. *IEEE Commun Mag* 52:22–28
- Zhou P, Zheng Y, Li M (2014) How long to wait? Predicting bus arrival time with mobile phone based participatory sensing. *IEEE Trans Mob Comput* 13:1228–1241
- Kostakos V, Camacho T, Mantero C (2013) Towards proximity-based passenger sensing on public transport buses. *Pers Ubiquitous Comput* 17:1807–1816
- Chon Y, Lane ND, Kim Y, Zhao F, Cha H, Understanding the coverage and scalability of place-centric crowd sensing. In: *The ACM international joint conference on pervasive and ubiquitous computing (UbiComp'13)*, pp 3–12
- Lee J-S, Hoh B Sell your experiences: a market mechanism based incentive for participatory sensing. In: *IEEE international conference on pervasive computing and communications (PerCom'10)*, pp 60–68
- Jaimes LG, Vergara-Laurens I, Labrador MA A location-based incentive mechanism for participatory sensing systems with budget constraints. In: *IEEE international conference on pervasive computing and communications (PerCom'12)*, pp 103–108
- Ioannis B, Vana K Mobile stream sampling under time constraints. In: *IEEE 14th international conference on mobile data management (MDM'13)*, pp 227–236
- Liu B, Jiang Y, Sha F, Govindan R Cloud-enabled privacy-preserving collaborative learning for mobile sensing. In: *The 10th ACM conference on embedded network sensor systems (SenSys'12)*, pp 57–70
- Jayaraman PP, Perera C, Georgakopoulos D, Zaslavsky A (2013) Efficient opportunistic sensing using mobile collaborative platform mosden. In: *The 9th IEEE international conference on collaborative computing: networking, applications and work-sharing*, pp 1–10
- Sheng X, Tang J, Zhang W Energy-efficient collaborative sensing with mobile phones. In: *The 31rd annual IEEE international conference on computer communications (INFOCOM'12)*, pp 1916–1924
- Chon Y, Lane ND, Kim Y, Zhao F, Cha H, Understanding the coverage and scalability of place-centric crowd sensing. In: *Proceedings of the 2013 ACM international joint conference on pervasive and ubiquitous computing (UbiComp 2013)*, pp 3–12
- Ho C-J, Vaughan JW Online task assignment in crowdsourcing markets. In: *AAAI*, vol 12, pp 45–51
- Reddy S, Estrin D, Srivastava M (2010) Recruitment framework for participatory sensing data collections. *Pervasive Comput* 138–155
- Tuncay G, Benincasa G, Helmy A Autonomous and distributed recruitment and data collection framework for opportunistic sensing. In: *Proceedings of Mobicom'12*, pp 407–410
- Ahmed A, Yasumoto K, Yamauchi Y, Ito M Distance and time based node selection for probabilistic coverage in people-centric sensing. In: *Proceedings of the 8th annual IEEE communications society conference on sensor, mesh and ad hoc communications and networks (SECON 2011)*, pp 134–142
- Hachem S, Pathak A, Issarny V Probabilistic registration for large-scale mobile participatory sensing. In: *Proceedings of the IEEE international conference on pervasive computing and communications (PerCom 2013)*, pp 132–140
- Zhang D, Xiong H, Wang L, Chen G Crowdrecruiter: selecting participants for piggyback crowdsensing under probabilistic coverage constraint. In: *Proceedings of the ACM international joint conference on pervasive and ubiquitous computing (UbiComp 2014)*, pp 703–714
- Song Z, Liu CH, Wu J, Ma J, Wang W (2014) Qoi-aware multitask-oriented dynamic participant selection with budget constraints. *IEEE Trans Veh Technol* 63:4618–4632
- Hull B, Bychkovsky V, Zhang Y, Chen K, Goraczko M, Miu A, Shih E, Balakrishnan H, Madden S Cartel: a distributed mobile sensor computing system. In: *The 4th ACM conference on embedded networked sensor systems (Sensys'06)*, pp 125–138
- Hu S-C, Wang Y-C, Huang C-Y, Tseng Y-C, Kuoz L-C, Chen C-Y Vehicular sensing system for co2 monitoring applications. In: *IEEE VTS Asia pacific wireless communications symposium (APWCS'09)*, pp 168–171
- Görner S, Park S-B, Egbert P (2009) Vision-based rain sensing with an in-vehicle camera. In: *IEEE intelligent vehicles symposium*, pp 279–284
- Hyyppä J, Jaakkola A, Hyyppä H (2009) Map updating and change detection using vehicle-based laser scanning. In: *The IEEE joint urban remote sensing event*, pp 1–6

33. Cortez RA, Luna J-M, Fierro R, Wood J Multi-vehicle testbed for decentralized environmental sensing. In: The IEEE international conference on robotics and automation (ICRA'10), pp 3052–3058
34. Zhu Y, Li Z, Zhu H, Li M, Zhang Q (2013) A compressive sensing approach to urban traffic estimation with probe vehicles. *IEEE Trans Mob Comput* 12:2289–2302
35. Jang J, Baik N (2012) Diagnosing hazardous road surface conditions through probe vehicle as mobile sensing platform. In: The international conference on winter maintenance and surface transportation weather, pp 54–60
36. Yuan J, Zheng Y, Xie X, Sun G Driving with knowledge from the physical world. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining (KDD'11), pp 316–324
37. Yuan J, Zheng Y, Zhang C, Xie W, Xie X, Sun G, Huang Y T-drive: driving directions based on taxi trajectories. In: Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information aystems (GIS'10), pp 99–108
38. Huang H, Zhang D, Zhu Y, Li M, Wu M-Y (2012) A metropolitan taxi mobility model from real gps traces. *J Univ Comput Sci* 18:1072–1092