On the Recruitment of Smart Vehicles for Urban Sensing

Sherin Abdel Hamid*, Glen Takahara†, and Hossam S. Hassanein*

*School of Computing, Queen’s University, Kingston, ON, Canada
†Department of Mathematics and Statistics, Queen’s University, Kingston, ON, Canada

Abstract—With the abundant on-board resources in intelligent vehicles, they have become major candidates for providing ubiquitous services, including urban sensing. This paper proposes an efficient recruitment scheme for vehicles in urban sensing applications. Our trajectory-based recruitment (TBR) scheme solves the problem of participant selection by considering spatiotemporal availability of participants. The aim of TBR is choosing the minimum number of vehicles that achieve a required level of coverage for the area of interest. TBR utilizes the easy-to-acquire trajectories of the candidate vehicles as indicators of the availability of participants, and applies a minimal-cover greedy algorithm for selection. The basic greedy algorithm is adapted to handle some practical scenarios, including departing vehicles and varying redundancy requirements. The paper also discusses two data acquisition models for retrieving the sensing data (on-demand and unsolicited). Assessment of TBR shows that it achieves high levels of coverage even when vehicles do not stick to their announced trajectories.

Keywords—Urban sensing, Recruitment, Smart vehicles.

I. INTRODUCTION

Urban sensing is gaining a high interest nowadays with the diversified applications it can provide. Currently, sensors in mobile devices are extensively used to support such applications. Although there is a wide scope of services made feasible with such engagement of mobile devices [1], the use of these devices has challenges, in particular dealing with the relative scarcity of available resources. Concurrently, the plethora of on-board resources in intelligent vehicles is pushing towards utilizing them as mobile providers for ubiquitous services. According to an analysis of the market growth of automotive sensors in North America, the average number of sensors per vehicle is expected to be 70 by end of 2013 [2]. This abundance of various sensors along with other on-board vehicular resources, such as the processing, storage and communication resources, make intelligent vehicles major enablers for many sensing applications and solutions. Furthermore, the mobility of vehicles can be utilized to widen coverage scope and, in turn, the range of applications that can be supported by engaging vehicles in the sensing loop.

We categorize the applications and services that can be provided by vehicular sensing into two categories: 1) instant sensing and 2) on-move sensing applications. Mobile devices typically provide instant sensing. With the high sensing capabilities of vehicles, the scope of such applications can be widened. An example of an instant sensing application is reporting weather conditions such as temperature and ambient barometric pressure. The second category of sensing applications is made feasible by utilizing the movement of vehicles and generating sensing data on the go. Examples include monitoring road conditions, traffic and crowds, and providing estimates of parking availability.

The general architecture of urban sensing consists of three main elements as shown in Fig. 1. These elements are the data contributors/participants, the service provider, and the data consumers/end users. The process involves three main stages as indicated in Fig. 1 as well: 1) The service provider asks data contributors to perform sensing tasks, 2) after collecting the required data, the data contributors send it to the service provider, 3) the service provider, after performing required data analytics, presents meaningful information to the data consumers as part of a subscribed service. Data consumers/end users may also initiate the process asking for specific information.

Although utilizing vehicles as data contributors in urban sensing brings many advantages, it comes with a challenge. In an urban environment there can be many potential participants in an area of interest, especially in a congested area or well-travelled road segment. These participants cannot all be recruited for a sensing task as the recruited participants should be given incentives as a reward for the service they provide and to encourage them to participate in the future. Since monetary incentives have shown to be the most encouraging ones, the service provider paying such rewards would like to minimize the number of recruited participants and the amount paid for each sensing task to handle limited budgets and maximize profit, while providing an acceptable level of service to the end user. Based on the above perspective directing recruitment

![Fig. 1. The architecture of urban sensing.](image-url)
of participants, the main objective of this paper is to introduce a recruitment scheme that selects the minimum number of participants achieving a given level of coverage for the area of interest in a cost-effective manner.

In our recruitment scheme the pool of potential participants is first determined by their spatial and temporal availability to achieve a desired coverage for the area of interest during a given time period. In contrast to some models that consider only instantaneous availability to achieve instantaneous coverage, we consider on-move coverage to support the wide scope of on-move monitoring applications mentioned earlier. With on-move coverage, the number of participants to achieve a desired coverage can be small compared to those achieving coverage without considering mobility of participants. For example, in covering a road to build an estimate of parking availability, we may find that just a few vehicles taking camera shots on the go can provide complete coverage of the road.

As a main part of the on-board vehicular resources, the navigation system is a vital component that provides information to support many of the vehicular applications. Our recruitment scheme is designed to utilize input from such systems that are ubiquitous in intelligent vehicles. By utilizing vehicle trajectories as inputs, we can select the participants in an informed and efficient way. Based on this, we introduce in this paper a trajectory-based recruitment (TBR) scheme to efficiently solve the problem of participant selection in recruiting vehicles to achieve a desired coverage. Assessment of TBR shows that it achieves high levels of coverage even when vehicles do not stick to their announced trajectories. To the best of our knowledge, TBR is the first scheme considering recruitment and selection of participating vehicles for urban sensing.

The remainder of this paper is organized as follows. In Section II, we discuss some related work on utilizing vehicles as sensors, and recruitment for urban sensing. We present our recruitment scheme (TBR) in Section III in its basic case along with two generalized cases that reflect practical situations. In addition, we discuss two data acquisition models that TBR supports. In Section IV, we present a model to assess coverage achieved by our scheme along with assessment results. Finally, we conclude the paper and present our future work in Section V.

II. RELATED WORK

In this section, we touch upon some related work in the area of utilizing vehicles as sources of sensing data, and discuss some available recruitment models for urban sensing.

Many platforms are proposed that utilize the sensory resources of intelligent vehicles. An example is the MobEyes platform [3] that focuses on utilizing vehicular sensors to monitor a vehicle’s surroundings and recognize objects, and utilizing the on-board resources to store the sensed data and share it with other vehicles upon request. Another example is the data-gathering solution proposed in [4] that supports location-aware services utilizing vehicular sensors. In this solution, data requests can be sent to vehicles asking for specific data at specific locations. Vehicles in the area of interest can resolve the request and send the reply back to the requester. Another example is the CarMote system [5] that aims at utilizing vehicular sensors for road surface monitoring.

Although the above mentioned platforms are good examples of using vehicles as mobile sensors, they neglect consideration of the recruitment scheme that chooses the vehicles that will participate in the sensing task. Most of them depend on specific pilot vehicles for evaluation purposes. For use in practical situations, these platforms are in need of some sort of recruitment mechanism for selecting participants.

In the area of participant recruitment for urban sensing, a few models are available in the literature. These models focus mainly on recruiting smart phones to utilize their onboard sensors. In [6], Reddy et al. proposed a recruitment framework in which the selection of participants is dependent on the instantaneous availability of participants along with their participation habits. To maximize the instantaneous coverage of the area of interest within a limited budget, the authors use the budgeted maximum coverage problem [7]. Another mechanism that considers the location and budget constraints is proposed in [8]. In this mechanism, the participants bid for their data, in contrast to the pricing model used in [6] where the participants’ costs are identical. Although these schemes can be effective at selecting mobile devices, they are not efficient for the recruitment of vehicles because they only consider instantaneous sensing and coverage which is not suitable for the wide scope of on-move sensing applications supported by vehicular mobility.

Instead of depending on an efficient recruitment scheme for participant selection, some of the data collection platforms for urban sensing depend on relatively simplistic schemes for collecting data. For example, one scheme uses random selection of data contributors. Another data collection scheme, which we refer to as the ‘naive’ scheme, is one in which the service provider simply asks all the contributors in the area of interest to generate and send data. Although these two schemes are simple, they have drawbacks that hinder their use. The ‘random’ selection scheme is less likely to provide coverage of the area of interest and also more likely to result in collected data which may have undue redundancy, compared to a more targeted recruitment scheme. Although it is the easiest to implement, the ‘naive’ scheme has potentially serious disadvantages. First, by getting data from all participants in an area of interest, the cost of such data to the service provider may be unnecessarily and prohibitively high. Second, with many participants in an area of interest, the data retrieved will have high levels of correlation and redundancy. Such collection of redundant data is considered a waste for both the service provider’s budget and the communication bandwidth.

With these limitations of the available sensing platforms and recruitment models, we are in need of efficient recruitment schemes that ensure the sufficient coverage of the area of interest using the minimum number of participants to minimize the cost, and in a way that utilizes vehicular mobility efficiently to support the on-move sensing applications. These are the main features of our trajectory-based recruitment scheme that we discuss in the next section.

III. THE PROPOSED RECRUITMENT SCHEME AND SUPPORTED DATA ACQUISITION MODELS

As a main component of an intelligent vehicle, the navigation system plays a pivotal role in most of the vehicular applications and services. In addition to providing navigational information to the driver, the output of the navigation system is utilized by a multiplicity of applications including safety, in-
fotainment, and diagnostics. We propose a recruitment scheme that utilizes trajectory information from navigation systems to make vehicle selection decisions. The spatiotemporal information included in vehicle trajectories enables a more informed and reliable selection of a minimum number of participants. As trajectories represent the dynamic availability of participants, we consider them in achieving a desired on-move coverage. Members of the data collection process (drivers registered with the service) will need to enter their destination before starting their trip. This way, the service application can calculate the trajectory and have it stored and ready to be accessed by the service provider when needed. To request a sensing task, a service provider sends sensing requests to the participants. The sensing request defines the sensing task, the area of interest, and the time span of the task. The on-vehicle service application then sends relevant trajectory information (i.e. overlapping parts with the sensing task) to the service provider.

To solve the coverage problem with the minimum number of participants based on availability of trajectories, we consider a related problem in the area of computational geometry. By representing the area of interest and the overlapping parts of participants’ trajectories with the area of interest as intervals, we argue that our problem can be solved with a scheme similar to the minimal-cover problem [9]. Later in this section, we elaborate on the two data acquisition models that are supported by the proposed TBR scheme.

A. Trajectory-Based Recruitment (TBR) Scheme

In this sub-section, we first discuss the minimal cover problem and the greedy algorithm proposed for its solution. We then present our trajectory-based recruitment (TBR) scheme in its basic form by discussing the general formulation of our recruitment problem and how it maps to the minimal cover problem and its greedy solution. Our TBR scheme generalizes the basic algorithm to handle practical situations, including departing vehicles and varying redundancy requirements.

1) Minimal-Cover Scheme:

Problem Definition - The minimal-cover problem is concerned with having a minimum number of overlapping line segments the union of which covers a linear coverage area. It can be described as follows: Given an interval \([a, b]\) and a set \(n\) of intervals \(S = \{[a_i, b_i]\}\), find the smallest subset \(S'\) of \(S\) such that the union of its elements covers the interval \([a, b]\), if such a subset exists; otherwise, report a whole coverage failure.

Solution - The solution is based on a greedy algorithm that works as follows:

(i) Find a member \(m \in S\) that covers \(a\) and has a maximal right endpoint. If such a member does not exist report a failure and exit.

(ii) While failure is not reported and \(m\) does not cover \(b\), find a successor that covers \(b_i\) and has the maximal right endpoint.

(iii) This procedure continues until \(b\) is covered or a successor cannot be found.

Consider the example shown in Fig. 2 where we have an interval \([a, b]\) overlapping intervals. Following the greedy procedure above, we find that the minimum set that covers \([a, b]\) is the one containing segments 2, 4, 6, 7, and 8.

2) Vehicle Selection Scheme:

Based on the information available and the objective of the proposed recruitment scheme, we can define our recruitment problem and the proposed solution as follows.

Inputs
- \(A\) : Area of Interest
- \(S\) : Set of Segments

Output
- \(S' \subseteq S\) : Covering Set of Segments

Problem Definition - Find the minimum number of segments \(S_i \in S\) to form the coverage set \(S'\) such that the coverage function \(F(S', x)\) is defined as

\[
F(S', x) = \begin{cases} 
1 & \text{if } x \text{ is covered by } S' \\
0 & \text{if } x \text{ is not covered by } S'
\end{cases}
\]

Solution - Consider the parts of vehicles’ trajectories that overlap with the area of interest to be the segments and the area of interest itself to be the interval to be covered. We can apply the greedy algorithm above as a solution for the minimal cover problem to achieve our goal and select the minimum number of participating vehicles that can provide coverage for the area of interest. By ensuring that the end of the chosen segment will be covered by a following segment - the one with the maximal right endpoint if many are available - till the end of the area of interest, we can ensure having \(F(S', x) = 1\) \(\forall x \in A\).

In order to handle areas that involve curved roads, such areas can be divided into a series of straight roads that can be dealt with as separate intervals. Generally speaking, any irregular road can be treated as a series of straight roads.

3) Practical Considerations:

The problem defined above assumes complete confidence in vehicle trajectory information and equal importance of road segments. In practical scenarios, such an ideal case is not guaranteed. In the following, we discuss two generalized cases of this basic scheme. These generalized cases reflect practical situations the service provider would face during the recruitment process. These are: i) having a probability that a vehicle will not stick to the trajectory it announced, and ii) having events that require redundancy at some parts of an area of interest.

Case I: TBR with Probability of Leaving:

In realistic scenarios, it is not guaranteed that a vehicle will stick to its announced trajectory. We consider a generalized case of the basic TBR that assigns different probabilities of sticking to the announced trajectory.

For each vehicle, we calculate a degree of confidence \(D_i\) (such that \(0 \leq D_i \leq 1\)) based on the participation history.
of this vehicle assuming that it was involved in earlier tasks, otherwise, \( D_i \) will be set to 1. Based on the computed degree of confidence, a probability of sticking to the announced segment of trajectory \( p(S_i, x) \) \( \forall x \in S_i \) is defined as follows

\[
p(S_i, x) = \begin{cases} 
1 & \text{if } x \leq D_i \\
0 & \text{if } x > D_i 
\end{cases}
\]

where \( x \) is normalized to be in [0, 1] to ease mapping to \( D_i \) values.

Having \( p(S_i, x) \) equal to 1 means that the vehicle will cover this segment and having it equal to 0 means that this part is not covered by this vehicle. Hence, we can define a vehicle’s coverage function to be

\[
f(S_i, x) = p(S_i, x) \forall x \in S_i
\]

where \( S_i \) is the segment announced by vehicle \( i \).

Recall that, as defined earlier, the set covering function \( F(S', x) \) should be equal to 1 \( \forall x \in A \). To compensate for having a part of a vehicle’s trajectory with a probability of being not traversed by the vehicle (not covered), this part should be covered by another vehicle with probability 1. Therefore, we can define the set covering function of a point \( x \in A \) to be the summation of all the vehicles’ coverage functions of \( x \) as follows

\[
F(S', x) = \sum_{S_i \in S'} f(S_i, x) \quad (1)
\]

such that \( F(S', x) \) should be at least 1 to ensure coverage of \( x \).

It may happen that we find more than one segment (vehicle) that compensates the 0 probability of the current segment, in this case, the greedy algorithm chooses the trajectory segment with the maximal right endpoint as defined earlier in the basic greedy algorithm.

The above algorithm can be summarized as follows:

(i) For each segment, calculate \( p(S_i, x) \) based on the computed \( D_i \).

(ii) Map the announced segment of trajectory to a projected one based on \( p(S_i, x) \).

(iii) Apply the greedy algorithm on the projected segments to achieve coverage.

The example in Fig. 3 shows the operation above. We can see that when considering probabilities of leaving, more segments/vehicles are needed to ensure coverage compared to the case with full confidence of sticking to the trajectory.

It may happen that, when considering the probability of leaving, coverage of a certain area may be intermittent if there are no vehicles to compensate the part of the segment with a probability of coverage less than 1. To handle this case, two different approaches can be deployed based on the criticality of the service as follows.

If the service is delay-critical, coverage should be achieved in the exact duration of the event, otherwise, data generated and reported will be obsolete. In this case, the greedy algorithm can be adjusted to provide the maximum coverage possible instead of full coverage. During the selection of vehicles, the parts with no vehicle segments available can be ignored and the algorithm will continue at the next point with available coverage as shown in the example of Fig. 4.

If the service is delay-tolerant, the greedy algorithm can be adjusted such that if a solution with full coverage cannot be achieved, the algorithm will report a failure and it can be re-run at a later time. Re-running the algorithm to achieve full coverage should be accompanied with a maximum threshold of re-runs based on the delay-tolerance of the service.

Intermittent coverage may also happen when a service provider cannot find a sufficient number of vehicles that achieve continuous coverage for the area of interest in the case of monitoring of an environment that is not dense enough, or when the penetration rate of the service and its corresponding application are not high enough in certain areas. These two cases can be handled in the same way discussed above.

**Case II: TBR with Redundancy Requirements:**

The basic case of the proposed scheme assumes that only one vehicle is needed to monitor an area. In practical situations, the service provider may require readings from multiple vehicles monitoring the same area to achieve a certain level of reliability. We adapt the basic case to give a service provider the ability to determine the level of redundancy needed by determining the degree of importance for different parts of the area of interest. For example, as shown in Fig. 5, in the case of a severe accident, a service provider may ask for different degrees of importance relative to the location of the accident with the highest degree at the exact location of the accident, and lower degrees farther from the accident location.

In order to handle this case, we define the degree of importance stated by the service provider for each part of the area of interest to be \( I_r \), where \( \bigcup_r^m I_r = A \), and \( n \) is the number
of parts the area of interest is divided into. I_r will be translated
to the number of vehicles needed to monitor this part. In a
certain part with a certain degree I_r = k, we assign each vehicle/segment in this part a coverage degree C_i such that
\[ C_i = \frac{1}{I_r} \frac{1}{k} \forall S_i \in r \]

To relate to the notations used in the previous and basic cases, we can define a vehicle’s coverage function \( f(S_i, x) \) to be equal to its coverage degree as follows
\[ f(S_i, x) = C_i \forall x \in S_i \]

As aforementioned, the set coverage function of a point \( x \) is defined to be the summation of all the vehicles’ coverage functions of \( x \) as defined in Eqn.1. As mentioned previously, to ensure coverage for a point \( x \), \( F(S^r, x) \) should be at least 1. This implies that to achieve coverage for a point \( x \) in an area with an importance degree \( k \), \( k \) vehicles (each with \( C_i = 1/k \)) are needed to make the value of \( F(S^r, x) \) equal to 1.

The basic greedy algorithm can be adjusted such that for each point \( x \) in a part with \( I_r = k \), the first \( k \) segments covering \( x \) with the maximal right end points will be chosen if \( x \) is not covered at all. Otherwise, if \( x \) is covered by \((k-n)\) segments, \( n \) segments covering \( x \) with the maximal right end points will be chosen.

We remark that the case with redundancy and probability of vehicles leaving is a straightforward extension.

B. Data Acquisition Models

When collecting data, intelligent vehicles can follow different models for data acquisition. In our paper, we consider two models that are enhanced by our recruitment scheme; the on-demand model, and the unsolicited model. These two models differ in when data is generated.

1) On-Demand Model:
In this model, data acquisition is done on-demand and upon request from a service provider. While on the go, vehicles receive requests for sensing tasks. Based on the availability of resources at that moment, the application installed on the on-board unit of the vehicle can decide if the vehicle is ready to participate or not (i.e. accepting the sensing request or declining it). This type of handling sensing tasks without intervention from the driver or any of a vehicle’s occupants falls under the “opportunistic” category of urban sensing.

In addition to being able to handle the services initiated by a service provider, this model is more suited for services initiated by a data consumer - through a service provider - who is in need of specific information.

2) Unsolicited Model:
In this model, vehicles sense their surroundings, collect data, and store it without being tasked. When a service provider needs some information about an area of interest, the provider can check which vehicles have data stored about that area. After selecting the data holders, the service provider informs them to send the stored data. This model involves some sort of advertisements by vehicles about the data they carry. Such advertisements can be handled by metadata that describes the actual data and lists some of its features (e.g. when and where they are generated).

The unsolicited model is only suitable for the delay-tolerant services that allow storing data and reporting it at a later time. An example of such services is using vehicles for monitoring road conditions.

It is worth mentioning that the proposed TBR scheme supports these two data acquisition models. In the on-demand model, the trajectories considered for recruitment are those that vehicles are supposed to follow and can be retrieved from the navigation software. For the unsolicited model, the trajectories are those that vehicles have already traversed and stored sensed data.

IV. COVERAGE ASSESSMENT

We assess the coverage achieved by our proposed recruitment scheme. For clarity of demonstration, we assume that we have a set of available trajectories that represent the solution space, and that vehicles can enter or leave only at intersections. We apply TBR with probability of leaving to this set of trajectories with four different ranges for the degree of confidence \( D_i \), as shown in Table I. For each confidence range, we apply our scheme and find the coverage solution. Then, we assess the obtained solution in terms of the coverage achieved by the chosen trajectories.

To assess a solution, we compute \( p_i \) to be the probability of vehicle \( i \) to leave at an intersection, which is computed based on the equation
\[ q_i = (1 - p_i)^k \]
where \( q_i \) is the probability of vehicle \( i \) to stick to the announced trajectory and is equal to \( D_i \), and \( k \) is the number of intersections vehicle \( i \) may leave at on its trajectory.

We compute the total probability of coverage of a road segment \( R_j \) such that \( 1 \leq j \leq n \) and \( n \) is the number of road segments in an area of interest (as shown in Fig. 6), as follows
\[ P_{R_j} = 1 - \prod p_i \]

where \{\( p_i \)\} included in the multiplication are for the vehicles whose trajectories cover this road segment in the being-assessed solution.

After computing \( P_{R_j} \forall R_j \), we compute the proportional coverage of the area of interest to be
\[ C_{area\_interest} = \frac{\sum P_{R_j}}{n} * 100 \% \]

In Table I, we present the assessment results of solutions
obtained with various ranges of $D_i$. In addition, we consider different densities of vehicles available in the area of interest: 1) a dense environment where each part is covered by redundant vehicles with a minimum of 3 vehicles covering each road segment, and 2) a more sparse environment where each road segment is covered by a maximum of 2 vehicles with some segments are completely not covered. The results show that, in a dense environment, even with low values of $D_i$ (high probabilities of a vehicle not sticking to its trajectory), our scheme achieves high proportional coverage based on the fact that the scheme includes sufficient vehicles in the covering set to compensate for probabilities of leaving which enhances the reliability of our scheme. In a more sparse environment, the achieved proportional coverage is lower because some road segments did not have passing vehicles to cover them. The algorithm worked on achieving the maximum available coverage in this case instead of full coverage.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed the trajectory-based recruitment (TBR) scheme that handles the problem of selection of participating vehicles in urban sensing applications. TBR aims at choosing the minimum number of participants that achieve a certain level of coverage for an area of interest specified in a sensing task. To achieve this target, TBR considers trajectories of vehicles as means for acquiring the spatiotemporal availability of participants that assist in making the selection process more efficient and informative. TBR generalizes the basic recruitment case to some practical cases that a service provider faces during the recruitment process such as probability of a vehicle not sticking to its announced trajectory and having redundancy requirements at certain parts of an area of interest. The assessment results show that our proposed scheme achieves high levels of available coverage even with high probabilities of vehicles not sticking to their trajectories.

We plan to extend our scheme to accommodate budget constraints. We remark that participants would report data with varying quality of information (QoI). We plan to extend TBR to take QoI into consideration, in addition to availability, for directing the selection process. Having varying QoI will lead to having varying costs given as incentives for participants. Development of a dynamic pricing mechanism will be a part of our future work as well.

ACKNOWLEDGEMENT

This research is supported by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

REFERENCES


TABLE I. RESULTS OF COVERAGE ASSESSMENT FOR DIFFERENT RANGES OF THE CONFIDENCE DEGREE

<table>
<thead>
<tr>
<th>Range of Degree of Confidence ($D_i$)</th>
<th>[0.8,1]</th>
<th>[0.5,0.8]</th>
<th>[0.1,0.5]</th>
<th>[0.1,1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{area,interest}$ (Dense Environment)</td>
<td>99 %</td>
<td>89 %</td>
<td>80 %</td>
<td>83 %</td>
</tr>
<tr>
<td>$C_{area,interest}$ (Sparse Environment)</td>
<td>80 %</td>
<td>71 %</td>
<td>55 %</td>
<td>68 %</td>
</tr>
</tbody>
</table>