Reputation-Aware, Trajectory-Based Recruitment of Smart Vehicles for Public Sensing

Sherin Abdelhamid, *Member, IEEE*, Hossam S. Hassanein, *Fellow, IEEE*, and Glen Takahara, *Member, IEEE*

***Abstract*— With the abundant on-board resources in smart vehicles, they have become major candidates for providing ubiquitous services, including public sensing. One of the challenges facing such ubiquitous utilization is the recruitment of the participating vehicles. In this paper, we present the reputation-aware, trajectory-based recruitment (RTR) framework that handles recruitment of vehicles for public sensing. The framework considers the spatiotemporal availability of participants along with their reputation to select vehicles that achieve desired coverage of an area of interest within a budget cap. The framework consists of a reputation assessment scheme, a pricing model, and a selection scheme collaborating for a main recruitment objective; maximizing coverage with minimum cost. We propose greedy heuristic solutions targeting the selection problem in real-time. The RTR framework generalizes the basic selection problem to handle some practical scenarios, including departing vehicles and varying redundancy requirements. We also propose a reputation assessment scheme and a pricing model as parts of the framework. Extensive performance evaluation of the proposed framework is conducted and the evaluation shows that the proposed greedy heuristics are able to achieve results close to previously obtained optimal benchmarks under different scenarios, and that the framework succeeds in achieving high levels of coverage even when vehicles do not stick to their announced trajectories.**

***Index Terms*— Public sensing, recruitment, smart vehicles, reputation systems.**

I. Introduction

PUBLIC sensing is gaining a high interest nowadays with the diversified applications it can provide. Currently, sensors in mobile devices are extensively used to support public sensing services [1], [2]. However, the use of these devices has challenges, in particular dealing with the relative scarcity of available resources. Concurrently, there is a plethora of on-board resources in smart vehicles pushing towards utilizing them as mobile providers for ubiquitous services [3]. According to an analysis of the market growth of automotive sensors in North America, the average number of sensors per vehicle reached 70 in 2013 [4]. The abundant sensors along with other on-board vehicular resources, such as processing, storage and communication resources, make smart 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.

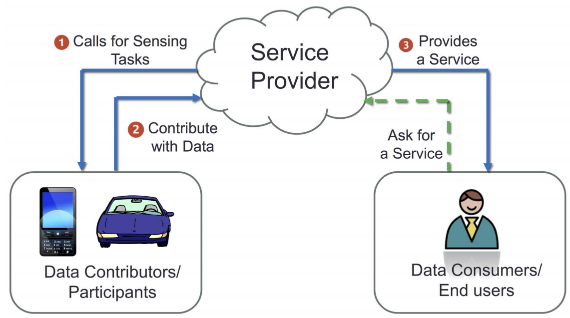


Fig. 1. The architecture of public sensing.

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

Although utilizing vehicles as data contributors in public sensing brings many advantages, it comes with challenges. There can be many potential participants in an area of interest, especially in a congested/well-travelled area. These participants cannot all be recruited for a sensing task as the recruited participants should be given incentives as a reward for the

Manuscript received September 3, 2016; revised February 14, 2017 and June 24, 2017; accepted July 2, 2017. Date of publication August 16, 2017; date of current version May 2, 2018. This work was supported by the Natural Sciences and Engineering Research Council of Canada under Grant STPGP 479248. The Associate Editor for this paper was J. Li.
(Corresponding author: Sherin Abdelhamid.)

S. Abdelhamid is with the Faculty of Computers and Information, Ain Shams University, Cairo 11566, Egypt (e-mail: shereen@cis.asu.edu.eg).

H. S. Hassanein is with the School of Computing, Queen's University, Kingston, ON K7L2N8, Canada (e-mail: hossam@cs.queensu.ca).

G. Takahara is with the Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L2N8, Canada (e-mail: takahara@mast.queensu.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2017.2727053

service they provide, and to encourage them to participate in the future. Hence, effective recruitment schemes are needed to ensure the selection of the right number of "trusted" participants achieving a required level of coverage for an area of interest in a cost effective manner. Based on the above perspective directing recruitment of participants, the main objective of this paper is to introduce a recruitment framework that handles the aforementioned recruitment requirements.

In our recruitment framework, the pool of potential participants is first determined by their spatial and temporal availability to consider those spatially available in the area of interest during the task 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. 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 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 many vehicular applications. Our recruitment framework is designed to utilize input from such systems represented in vehicle trajectories as indicators of the on-move availability of those vehicles.

In practice, with the diversity of participants' behavior and the various capabilities of the candidate vehicles, depending solely on availability for recruitment would not be efficient at selecting the best candidates out of the available selection pool. The reputation of participants must be considered along with availability to maximize the benefit out of the chosen participants by selecting those who are more likely to contribute with high quality data. Taking reputation into consideration also promotes selecting trusted participants and discarding those with malicious behavior. Furthermore, since in real scenarios SPs will have a budget cap that the total recruitment cost should not exceed, introducing a budget constraint while selecting the participants is called for. Consequently, in this paper, we present a **r**eputation-aware, **t**rajectory-based **r**ecruitment (RTR) framework that accommodates the consideration of participants' reputation and the budget constraint while building on the availability of participants.

The proposed RTR framework consists of three main modules: 1) a reputation assessment scheme, 2) a pricing model, and 3) a selection scheme. The outputs of the first two modules are fed as inputs to the third module. Keeping in mind the recruitment perspective discussed above, the main objective of our selection problem targets maximizing the available coverage, while minimizing the total recruitment cost in a reputation-aware manner with budget consideration. In our previous work in [5], we formulated the selection problem as an integer linear programming (ILP) optimization problem. The purpose of the optimization formulation is to present performance benchmarks for the recruitment objective giving some performance bounds. In this paper, greedy heuristic solutions are presented. Such heuristic solutions are needed to cope with practical real-time selection requirements. Another major part of the framework is the proposed reputation assessment scheme that feeds the selection scheme with the reputation scores of the candidate participants. We consider different metrics under the assessment scheme proposing a mechanism for assessing each of these metrics under two different data acquisition models. As part of the framework, a dynamic pricing model is also proposed for computing the participants' recruitment cost, to be used by the selection scheme.

The performance of the framework is extensively evaluated comparing the optimal and greedy selection solutions. The results show that the proposed heuristics are able to achieve results close to the optimal benchmarks. In addition, extensive evaluation is conducted to show the quality of the collected sensing data and the performance of the selection scheme under some practical cases.

To the best of our knowledge, the proposed RTR framework is the first work that considers the reputation of participants and the budget constraint in recruiting vehicles for public sensing services.

The remainder of this paper is organized as follows. In Section II, we discuss some related work in the areas of utilizing vehicles as a sensing resource, recruitment for public sensing, and reputation assessment. The proposed RTR framework is introduced in Section III, focusing on its selection part and the proposed greedy heuristic solutions. In Section IV, we present the proposed reputation assessment scheme and pricing model as parts of the RTR framework. In Section V, we discuss the performance of the proposed greedy heuristics comparing them to the benchmark performance results of the ILP formulations. Finally, we conclude the paper and present our future work in Section VI.

II. Related Work

A. Vehicle as a Mobile Sensor

Many platforms are proposed for utilizing the sensory resources of smart vehicles under the public sensing domain, which is also referred to as Mobile Crowd Sensing (MCS). An example is the MobEyes platform [6] 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 [7] 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. Other platforms are enhanced by use of the Internet which they use for sending the sensed data to remote servers. Examples of such platforms are discussed in [8]. Many systems are also proposed to utilize in-vehicle sensors for specifically enhancing intelligent transportation systems (ITS). Wang *et al.* [9], investigate a plethora of ITS services enabled by mobile sensing. Other examples under this category include the work in [10]–[12]. Tang *et al.* [10], propose the Collecting Lane-based Road Information via Crowdsourcing (CLRIC) method. CLRIC automatically extracts detailed lane structure of roads by using data collected by vehicles. With the same target of building accurate lane-level information, the work presented in [11] enhances coarse, inaccurate maps through local sensor information from a 3D lidar and a positioning system. For a different ITS target, Ruhhammer *et al.* [12] present an approach for extracting multiple intersection parameters through analyzing logged data from a test fleet. Currently, there is a focus on utilizing vehicular sensors for road condition monitoring services. The CS-Monitoring system in [13] and the work presented in [14] are examples of this category of applications. Further insights into the use of vehicles as mobile sensors are presented in [15].

Although the above mentioned platforms are good examples of using vehicles as mobile sensors in public sensing/MCS services, they neglect the consideration of a recruitment scheme to choose which vehicles 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 a recruitment mechanism for selecting participants.

B. Recruitment for Public Sensing

In the area of participant recruitment for public sensing/MCS, a few models are available in the literature. These models focus mainly on recruiting smartphones to utilize their on-board sensors. Reddy *et al.* [16], propose a recruitment framework that considers a participant's availability and participation habits for selection. To maximize the coverage of the area of interest within a limited budget, Khullera *et al.* [17] use the greedy solution of the budgeted maximum coverage problem. Their recruitment framework differs from the proposed framework in that it does not consider on-move availability as it is limited to smartphone use. In addition, it does not consider availability and reputation simultaneously; it supports selection by any of the metrics independently based on user choice. Similar to [16], the recruitment framework proposed in [18] does not consider on-move availability of participants. The main objective of this recruitment framework is to maximize the total tempo-spatial behavior similarity for participants subject to reputation and budget constraints. Some other models are proposed for recruitment purposes that do not pay attention to the reputation of participants. A mechanism that considers the location and budget constraints is proposed in [19]. This mechanism depends on the Reverse Auction pricing model [20] in which the participants bid for their data, in contrast to the pricing model used in [16] where the participants' costs are identical, and to the dynamic approach considered in our proposed pricing model. Although these schemes can be effective at selecting smartphones, 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.

A few schemes have been proposed in the literature for the purpose of recruiting vehicles for crowdsourcing applications, including public sensing. The solutions proposed in [21] depend on mobility prediction for acquiring vehicle trajectories as indicators of the availability of participants. Their main recruitment target is selecting participants to maximize coverage with limited budget. Two solutions are proposed to target different crowdsourcing application requirements. One solution depends on a greedy heuristic algorithm, while the other one is based on a genetic algorithm. Han *et al.* [22], propose two algorithms for trajectory-based node selection in vehicular crowdsensing. The algorithms differ in when they acquire the vehicle trajectories and run the selection process. An offline algorithm is proposed that assumes knowing a priori which vehicles will be at the sensing area in addition to knowing the vehicle arrival times. The second algorithm handles the selection process in an online manner to handle vehicle dynamics. In this algorithm, the selection decision is made on the fly once a vehicle arrives to the sensing area. These two algorithms do not consider paying rewards/incentives back to participants, therefore, they do not take the recruitment budget into account. In our previous work in [23], we propose the trajectory-based recruitment (TBR) scheme for recruiting vehicles for urban sensing. The TBR scheme shares with the proposed RTR framework consideration of the on-move availability of vehicles, but it only considers vehicle availability in the selection process, whereas here we consider the reputation and budget constraints. Although these vehicular recruitment schemes improve on the smartphone-based schemes for use in vehicular environments, they suffer from limitations and the lack of trust in the quality of the reported data since they only depend on participant availability for selecting participants and ignore the important factor of participant reputation/behavior.

With these limitations of the available sensing platforms and recruitment models, we are in need of efficient recruitment frameworks that ensure the required coverage of the area of interest with both reputation and budget considerations, and in a way that utilizes vehicular mobility efficiently to support the on-move sensing applications. These are the main features of our recruitment framework presented in this paper.

C. Reputation Assessment

In the area of reputation/trust assessment, we classify the assessment schemes into two main categories; redundancy-dependent and redundancy-independent. In the former category, the SP/truster responsible for computing a reputation score for each participant depends on correlated readings reported from other participants who are asked to do the same task. The outlier detection technique [24] is commonly used under this category. To compute a reputation score/trust level of a reporter, outlier detection can be used to measure the distance of the reported data value to a common value (e.g., average of the correlated data) such that the shorter the distance is, the higher the reputation/trust of the reporter. The system presented in [25] is an example that uses outlier detection for building a reputation system for smartphone-based sensing applications.

In the second category, redundancy-independent, the truster does not require redundant data to assess the reputation of a specific participant. The truster depends on assessment metrics that take the performance history of the trustee into consideration, and/or depend on some current features associated with the trustee's device/data. The aforementioned recruitment

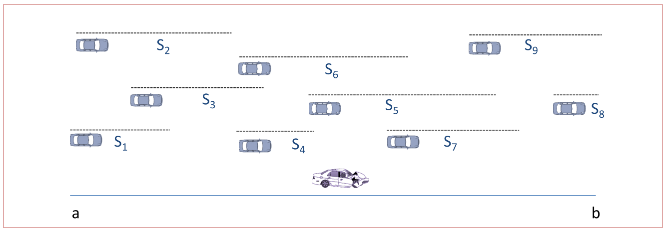


Fig. 2. An example showing trajectory segments of vehicles in an event area.

framework presented in [16] is an example of a system within this category. Since in our framework we target recruiting participants in a cost-effective manner, we do not consider redundant coverage, unless required, as discussed later. Therefore, we adopt the redundancy-independent category in our reputation assessment scheme.

III. The Reputation-Aware, Trajectory-Based Recruitment Framework

As a main component of a smart 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, infotainment, and diagnostics. We remark that with the assistance of this system, the trajectories of vehicles can be easily acquired and utilized as a precise indicator of vehicles' availability. As mentioned earlier, we consider the participants' spatiotemporal availability as a main criterion for recruiting participants and achieving a required coverage. By noting that vehicles' trajectories overlap with sensing parameters (the sensing area and duration) defined in the sensing request, we can tighten our solution space to those that are spatiotemporally available in the area of interest. In addition, as trajectories represent on-move availability, they are suited for handling recruitment for the wide scope of on-move sensing applications.

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 SP when needed. To request a sensing task, an SP 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 SP to start the selection process. Fig. 2 shows an example of a trajectory segment solution space existing in the targeted area of an event.

With the diversity of drivers' behavior and vehicles' capabilities, considering reputation of participants and their reported data is an important criterion that aids in distinguishing among participants and picking those that ensure an adequate level of quality. As in practice an SP responsible for the recruitment process will have a budget cap that cannot be exceeded, it is necessary to include a budget constraint in the selection process. Bearing in mind this perspective, we present the reputation-aware, trajectory-based recruitment (RTR) framework that considers both the spatiotemporal availability and reputation of participants while accommodating budget constraints in recruiting vehicles for public sensing services. The RTR framework consists of three main modules: 1) a reputation assessment scheme used for computing a reputation score for each candidate participant, 2) a dynamic pricing model for computing a recruitment cost for each participant based on his/her reputation score and the distance traversed, and 3) a selection scheme for choosing the participants to be recruited to achieve a required level of coverage for the area of interest in a reputation-aware manner with cost consideration. The outputs of the first two modules; the participants' reputation scores and recruitment costs, are fed as inputs to the selection scheme along with the vehicles' trajectories to start the selection process. In this section, we focus on the selection part of the framework. The reputation assessment scheme and the pricing model are detailed in the next section.

For the selection part, our earlier work [5] presented an ILP formulation for the selection problem as a benchmark for the sake of providing the upper bounds of the recruitment solutions. In the following, we present greedy heuristic solutions that follow the objective of the optimization formulation. Such heuristic solutions are needed to handle real-time services. In addition, we generalize the basic solution for the selection problem to handle practical situations. Later in this section, we elaborate on data acquisition models that are supported by the proposed RTR framework.

A. System Model and Problem Formulation

We consider an area of interest A with a trajectory segment set $\mathcal{S}$ of S segments spatiotemporally available in this area. An arbitrary segment is denoted by $i \in \mathcal{S}$. Each segment $i \in \mathcal{S}$ is associated with a reputation score r_i and a recruitment cost c_i computed using the reputation scheme and the pricing model, respectively, as discussed in the next section. A budget cap B and a reputation threshold R_{Th} will be determined by the SP interested in the recruitment process.

Based on the information available and the main recruitment target, we can define our recruitment problem as follows.

Inputs:

$$\begin{aligned} A &: \text{ Area of interest} \\ \mathcal{S} &: \text{ Set of trajectory segments} \\ B &: \text{ Budget cap} \\ R_{Th} &: \text{ Reputation threshold} \end{aligned}$$

Output:

$$\mathcal{S}' \subseteq \mathcal{S} : \text{Covering Set of Segments}$$

Problem Definition: Find a segment set $\mathcal{S}' \subseteq \mathcal{S}$ that achieves maximum coverage to the area A, satisfying a recruitment objective while considering the reputation and budget constraints, $r_i \geq R_{Th} \ \forall \ i \in \mathcal{S}'$ and $\sum_{i \in \mathcal{S}'} c_i \leq B$, respectively.

Since SPs would usually favor getting the covering solution with the minimum cost, our recruitment objective targets minimizing the total recruitment cost while achieving the maximum available coverage.

It is worth mentioning that although Fig. 2 shows a straight road, our model is not restricted to this road topology.

The proposed model is generic and can support a multiplicity of roads based on the fact that curved/non-straight roads can be treated as a series of straight roads.

B. The Proposed Greedy Heuristic Solutions

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 variations to the set cover problem (SCP) [26] altering it to an interval cover one.

One of the variations of the SCP that considers a budget cap in the selection process is known as the budgeted maximum coverage problem (BMCP) [17]. The BMCP does not consider the reputation-awareness part and its objective does not match with our recruitment objective. Therefore, the solution for the BMCP cannot be applied to our work. In the following, we discuss our basic reputation-aware budgeted maximum coverage (RBMC) solution in addition to an extended version, RBMC-MC, that handles our recruitment objective.

1) The Reputation-Aware Budgeted Maximum Coverage (RBMC) Greedy Solution: In the basic SCP, the input is a set of elements $\mathcal{E} = \{E_1, E_2, ..., E_m\}$ and a collection of subsets $\mathcal{H} = \{H_1, H_2, ..., H_n\}$ s.t. $\bigcup_i H_i = \mathcal{E}$. The goal is to find a solution that covers the set $\mathcal{E}$ with the minimum number of subsets from $\mathcal{H}$. The greedy approximation solution proposed for solving the SCP works through iterations to find the subset that covers the maximum number of uncovered elements in each iteration until all the elements in the set $\mathcal{E}$ are covered.

In our proposed RBMC solution, we alter the SCP to consider covering an interval instead of a set of elements and generalize it to accommodate the reputation and budget constraints. The solution maps the collection of subsets to the collection of trajectory segments and the domain of elements to the area of interest to be covered. The basic RBMC solution targets maximizing the coverage, which corresponds to the first stage of the optimization formulation.

Algorithm 1 shows our basic RBMC solution. Let $\mathcal{S}$ be the set of trajectory segments, $\mathcal{G} \subseteq \mathcal{S}$ be the collection of segments forming the selected covering set, and W_i be the length of the interval (area) covered by segment i but not covered by any segment in $\mathcal{G}$. The algorithm consists of two main procedures. The procedure *REPUTATION_FILTER($\mathcal{S}$)* is a pre-processing step to filter out the segments with a reputation score r_i below the threshold R_{Th}, and store those segments satisfying the reputation constraint in $\mathcal{U}$ to apply the selection process on. The procedure *GREEDY_BUDGETED_COVERAGE($\mathcal{U}$)* aims at selecting the collection of segments $\mathcal{G}$ that maximizes W_i without exceeding the given budget cap B. The output set $\mathcal{G}$ holds the minimum number of segments achieving the maximum available coverage with satisfying both the budget and reputation constraints.

Algorithm 1 The Reputation-Aware Budgeted Maximum Coverage (RBMC) Solution

```
REPUTATION_FILTER(S)
begin
U ← φ
for all  S_i ∈ S do
    if r_i ≥ R_Th then
        U ← U ∪ S_i
return U
end

GREEDY_BUDGETED_COVERAGE(U)
begin
G ← φ and C ← 0
while U ≠ φ do
    select S_i ∈ U that maximizes W_i
    if C + c_i ≤ B then
        G ← G ∪ S_i
        C ← C + c_i
    U ← U \ S_i
    update W_j for each S_j ∈ U
    for all  S_j ∈ U do
        if W_j = 0 then
            U ← U \ S_j
return G
end
```

2) The Reputation-Aware Budgeted Maximum Coverage With Minimum Cost (RBMC-MC) Greedy Solution: An extended version of RBMC is the RBMC-MC solution that handles the main recruitment objective of maximizing the coverage while minimizing the total cost.

Algorithm 2 shows the common approach followed in our RBMC-MC solution. After the pre-processing *REPUTATION_FILTER($\mathcal{S}$)* procedure, the procedure *GREEDY_BUDGETED_COVCOST($\mathcal{U}$)* is used to handle the selection process. It works similarly to the *GREEDY_BUDGETED_COVERAGE($\mathcal{U}$)* procedure in Algorithm 1, aiming at selecting the collection of segments $\mathcal{G}$ that maximizes W_i without exceeding B, with an added condition that handles the case when there is more than one trajectory having the same coverage weight W_i. In this case, the procedure selects the one with the minimum recruitment cost c_i. The output is the covering set $\mathcal{G}$. A final post-processing step is included in the algorithm to improve the output towards the recruitment objective. In this step, handled through the *POST_PROCESSING($\mathcal{G}, \mathcal{U}'$)* procedure, the algorithm works on reducing the redundant coverage in the covering set through replacing each selected trajectory in $\mathcal{G}$ with another trajectory from the reputation-filtered set $\mathcal{U}'$ that has the same unique coverage and shorter length than the replaced one, if there is any.

C. Practical Considerations

Our main recruitment problem assumes complete confidence in vehicle trajectory information and equal importance of road parts. In practical scenarios, such an ideal case is not guaranteed. In the following, we discuss two generalized cases of the basic problem. These generalized cases reflect practical situations the SP 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.

Algorithm 2 The Reputation-Aware Budgeted Maximum Coverage With Minimum Cost (RBMC-MC) Solution

$\mathcal{U} \leftarrow REPUTATION\_FILTER(\mathcal{S})$ //as in Algorithm 1

GREEDY_BUDGETED_COVCOST($\mathcal{U}$)
begin
$\mathcal{G} \leftarrow \phi$, $C \leftarrow 0$, and $\mathcal{U}' \leftarrow \mathcal{U}$
while $\mathcal{U} \neq \phi$ **do**
  select $S_i \in \mathcal{U}$ that maximizes W_i
  if there are many candidates with the same maximum W_i **then**
    select $S_i \in \mathcal{U}$ that maximizes W_i and minimizes c_i
  if $C + c_i \leq B$ **then**
    $\mathcal{G} \leftarrow \mathcal{G} \cup S_i$
    $C \leftarrow C + c_i$
  $\mathcal{U} \leftarrow \mathcal{U} \setminus S_i$
  update W_j for each $S_j \in \mathcal{U}$
  for all $S_j \in \mathcal{U}$ **do**
    if $W_j = 0$ **then**
      $\mathcal{U} \leftarrow \mathcal{U} \setminus S_j$
$\mathcal{G} \leftarrow POST\_PROCESSING(\mathcal{G}, \mathcal{U}')$
return $\mathcal{G}$
end

POST_PROCESSING($\mathcal{G}, \mathcal{U}'$)
begin
for all $S_k \in \mathcal{G}$ **do**
  if there is a segment $S_i \in \mathcal{U}'$ with the same unique coverage of S_k **and** $length(S_i) < length(S_k)$ **then**
    replace S_k with S_i
    update the unique coverage of all $S_k \in \mathcal{G}$
return $\mathcal{G}$
end

1) Case I: RTR 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 selection problem 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. D_i is computed as the average of traversed portions of the segments in the past interactions. For first-time participants, D_i is set to 0.5 as an average of the potential D_i values. Based on the calculated degree of confidence, a probability of sticking to the announced segment of trajectory i, $p(i, y)$, $\forall y \in i$, is defined as follows

$$p(i, y) = \begin{cases} 1 & \text{if } y \leq D_i \\ 0 & \text{if } y > D_i \end{cases} \tag{1}$$

where y is a location point on trajectory i and is normalized to be in $[0, 1]$ to ease mapping to D_i values.

Having $p(i, y)$ 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. 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. To handle this case, the use of the proposed greedy algorithms can be adjusted as follows

(i) For each trajectory i, calculate $p(i, y)$ based on the computed D_i.
(ii) Map the announced segment of trajectory i to a projected one based on $p(i, y)$.
(iii) Apply the greedy algorithms on the projected segments to achieve coverage.

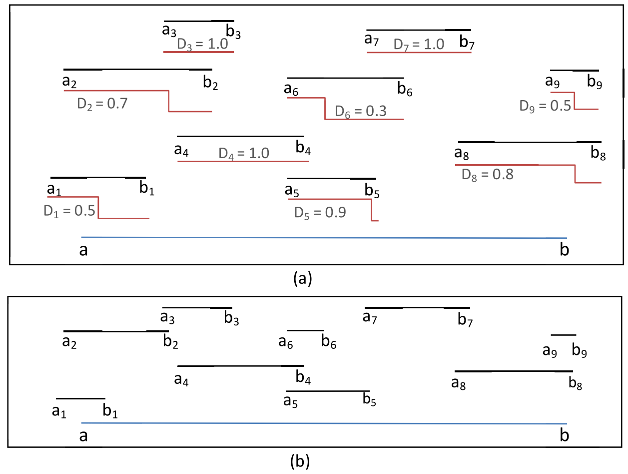


Fig. 3. An example of RTR with probability of leaving. In part (a), each segment is mapped to a probability distribution based on its corresponding D_i. Part (b) shows the projected segments.

The example in Fig. 3 shows the first two steps of the procedure above. We highlight that with probability of leaving, more segments are needed to ensure coverage compared to the case with full confidence of sticking to the trajectory.

It may happen, when considering the probability of leaving, that coverage of a specific area may be intermittent if there are no vehicles satisfying the constraints 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 algorithms can be used to provide the maximum coverage possible at that time. If the service is *delay-tolerant*, the greedy algorithms can be adjusted such that if a solution with the required coverage cannot be achieved, the algorithms will report a failure and they can be re-run at a later time. Re-running the algorithm should be accompanied with a maximum threshold of re-runs based on the delay-tolerance of the service.

Intermittent coverage may also happen in sparse environments, or when the penetration rate of the service and its application are not high enough in some areas. These two cases can be handled in the same way discussed above.

2) Case II: RTR With Redundancy Requirements: The basic problem assumes that only one vehicle is needed to monitor a point of interest. In practical situations, the SP may require readings from multiple vehicles monitoring the same area to

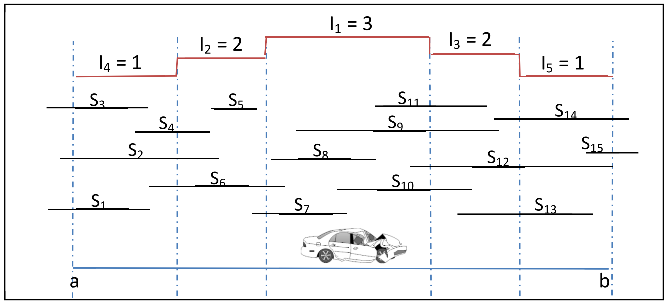


Fig. 4. An example of RTR with redundancy requirements. The area of interest is divided into 5 main parts each with a degree of importance based on its proximity to the event.

achieve a certain level of reliability. We adapt the basic case to give an SP 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. 4, in the case of a severe accident, an SP 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.

In order to handle this case, we define the degree of importance stated by the SP for each part of the area of interest to be I_{rp}, where $\bigcup_{1}^{m} rp = A$, and m is the number of parts the area of interest A is divided into. I_{rp} is translated to the number of vehicles needed to monitor part rp in the area. The RBMC and RBMC-MC algorithms can be adjusted to handle this case. Each part can be divided into sectors and the algorithms can be adjusted to ensure that each sector in a part with $I_{rp} = g$ has g vehicles covering it through selecting more segments using the same selection procedures until satisfying each sector's redundancy requirement, as long as the budget constraint allows for it.

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

D. Data Acquisition Models

When collecting data, smart vehicles can follow different models for data acquisition. We consider two models: solicited and unsolicited. These models differ in when data is generated.

1) The Solicited Model: In this model, data acquisition is done on-demand upon a request from an SP. 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 the vehicle's occupants falls under the "opportunistic" category of public sensing [27].

In addition to handling service requests initiated by an SP, this model can also handle requests initiated by a data consumer through an SP.

2) The Unsolicited Model: In this model, vehicles sense their surroundings, collect data, and store it without being tasked. When an SP needs some information about an area of interest, the provider can check which vehicles having data stored about that area. After selecting the data holders, the SP informs them to send the stored data. This model can involve 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.

The proposed RTR framework supports these two data acquisition models. In the solicited 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. Reputation Assessment and Recruitment Pricing

In this section, we present our reputation assessment scheme and pricing model that are responsible for computing a reputation score and a recruitment cost for each participating vehicle, respectively. These parameters are used by the selection module as discussed in the previous section.

A. Reputation Assessment

In assessing the reputation of a participant, the data acquisition model controls the metrics used for assessment and computing a reputation score. In the following, we delineate how the reputation score can be computed for each participant according to the data acquisition model to be used.

1) Reputation Assessment in the Solicited Model: Since in the solicited model data will be collected after the participant gets tasked, the computed reputation score will help in anticipating the behavior of the participant and the quality of the reported data. The score will be used in the selection process as the expected reputation of the participant.

a) Computing the reputation score: We adopt the Beta reputation system [28] for computing a reputation score, r, for each candidate participant. The Beta system is used in reputation and trust management systems for computing reputation/trust scores for a set of trustees by a centralized truster interacting with them. In our reputation assessment scheme, we consider the truster to be the SP responsible for the recruitment process and the trustees to be the candidate participants.

The use of the Beta system for computing a participant's reputation score involves computing two variables x and y based on the past interactions with the participant. The system parameters α and β are computed based on the x and y values according to Eq. 2. The reputation score is computed as the expectation of the Beta distribution of the computed α and β, as in Eq. 3.

$$\alpha = x + 1 \text{ and } \beta = y + 1, \text{where } x \,\&\, y \geq 0 \tag{2}$$

$$E(p) = \frac{\alpha}{\alpha + \beta} \tag{3}$$

One of the approaches used in the Beta system for computing its x and y variables involves the assessment of the participant after an interaction with the SP and providing the assessment as a single value υ in the [0,1] range. Then, the x and y values can be computed according to Eq. 4.

$$x = w\frac{(1+\upsilon)}{2} \text{ and } y = w\frac{(1-\upsilon)}{2} \tag{4}$$

We adopt the approach above in computing the reputation score in the solicited model. After an interaction with a participant i, the SP computes a per-interaction assessment, as discussed later, and plugs the assessment value υ_i into Eq. 4 for computing the x_i and y_i parameters with considering the weight w to be 1. These parameters are used for computing α_i and β_i according to Eq. 2. The expectation of the score, $E(p)_i$, can be computed according to Eq. 3. Then, the expected score is used as the participant's reputation score after normalizing it to be in the [0,1] range as shown in Eq. 5.

$$r_i' = [E(p)_i]_{norm} = \frac{E(p)_i - min(E(p))}{max(E(p)) - min(E(p))} \tag{5}$$

In the aforementioned procedure, only the last interaction with a participant is used for computing the reputation score. Ignoring the historical information (i.e., the past contributions) may be misleading if the participant has an uncommon performance in the latest interaction. In our scheme, we consider the last n interactions with a participant for computing his/her reputation score. If the SP has encountered more than n interactions with that participant, a sliding window of length n will be considered when the score has to be updated after an interaction with that participant. For each interaction, the assessment value υ is computed followed by the computation of the corresponding x and y parameters. Aggregated x and y values of the n interactions are computed as in Eq. 6.

$$x_{ag} = \sum_{j=1}^{n} \lambda^{(n-j)} x_j \text{ and } y_{ag} = \sum_{j=1}^{n} \lambda^{(n-j)} y_j \tag{6}$$

where x_{ag} and y_{ag} are the aggregated x and y values, respectively. The parameter λ is an aging factor used to give lower weight to the old contributions than the recent ones such that $0 \leq \lambda \leq 1$ [28]. Then, the x_{ag} and y_{ag} are used for computing α_{ag} and β_{ag} according to Eq. 2, which are used for computing $E(p)_{ag}$ as in Eq. 3. The final reputation score to be considered in the recruitment process is computed as follows for participant i

$$reputation\ score(r_i) = [E(p)_{ag_i}]_{norm} \tag{7}$$

where $[E(p)_{ag_i}]_{norm}$ is $E(p)_{ag}$ of participant i normalized to the [0,1] range using Eq. 5.

b) Per-interaction assessment: A participant's reputation score in the solicited model is computed based on per-interaction assessments as discussed above. In this part, we discuss how the per-interaction assessment is handled and define the different metrics that are used for computing the assessment value (υ).

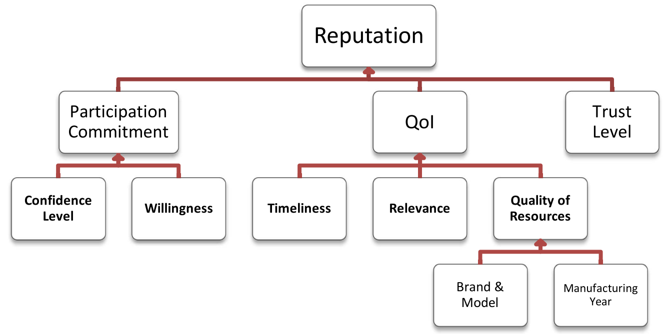


Fig. 5. The reputation metrics used in the solicited model.

The per-interaction assessment value is computed based on three main metrics: *1) Participation Commitment*, *2) Quality of Information (QoI)*, and *3) Trust Level*. The first two main metrics involve underlying sub-metrics as delineated below and shown in Fig. 5.

(i) Participation commitment: For assessing the commitment of a participant, two metrics can be considered as follows.

1) *The confidence of sticking to the announced trajectory (CoT)*: It may happen that a participant does not stick to the trajectory announced to the SP either intentionally or for a sudden detour. For a contribution j by participant i, this metric can be assessed as a binary variable as shown below:

$$CoT_i^j = \begin{cases} 1 & \text{if } i \text{ traversed the whole announced} \\ & \text{trajectory} \\ 0 & \text{if } i \text{ did not stick to the whole} \\ & \text{announcedtrajectory} \end{cases} \tag{8}$$

2) *The willingness to participate (WP)*: We represent a participant's willingness as the number of times that participant has participated in an evaluation period, e_prd, ending with that last interaction (e.g., e_prd can be a 1 month period). This number is then normalized to be in the [0,1] range taking into consideration the values achieved by the other candidate participants. Let the date of an interaction j to be Dt_j, the value of the willingness metric of participant i performing j can be computed as

$$WP_i^j = \Big[no.\ of\ \ participations_i \,\big|_{Dt_j - e\_prd}^{Dt_j}\Big]_{norm} \tag{9}$$

The values of these two metrics are combined using a weighted additive utility function to compute the participation commitment, P, of a participant i after a contribution j, as follows

$$P_i^j = w_1^P \times CoT_i^j + w_2^P \times WP_i^j \tag{10}$$

where w_l^P is the weight of each metric such that $\sum_{l=1}^{2} w_l^P = 1$.

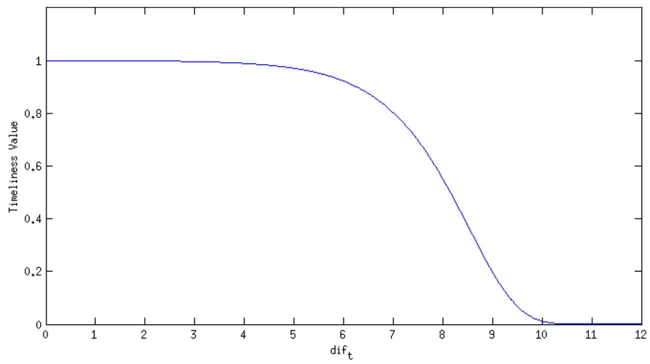


Fig. 6. An example of inverse Gompertz function with $a = c = 1$ and $b = 2 \times 10^{-4}$ for $t\_dr = 10$.

(ii) Quality of information (QoI): We consider three main metrics that can give a valuation of the quality of the reported information.

Timeliness (TM): Timeliness represents how promptly a participant sends the required information after getting assigned a sensing task. The timeliness value is the highest when the reply is promptly sent after the task assignment. Let the reply time be (T_r), the task assignment time be (T_a), and the task duration be (t_dr). The timeliness value is at its maximum when the difference between T_r and T_a (dif_t) is almost 0 and it decreases exponentially as that time difference increases. We use the inverse Gompertz function, shown in Fig. 6, to evaluate timeliness as its shape is compatible with the timeliness evolution discussed above. The lower asymptote of the general inverse Gompertz function is 0 and the function approaches it in infinity. In our case, we need the function to reach 0 when dif_t exceeds t_dr as that means the reported information is useless. In short, we can measure the timeliness value of the information reported by participant i in interaction j using the inverse Gompertz function if dif_t does not exceed t_dr, and assign it 0 otherwise. This is summarized in Eq. 11.

$$TM_i^j(dif_t) = \begin{cases} ae^{-be^{c(dif_t)}} & \text{if } dif_t \leq t\_dr \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

The parameter a of the inverse Gompertz function represents the higher asymptote which is equal to 1 in our case since the maximum of the Timeliness value is 1. The parameters b and c represent the displacement on the x-axis and the decay rate, respectively. We need the function to be as close to 0 as possible at $dif_t = t\_dr$. This fact directs the assignment of the parameters b and c. By assigning c the value 1, the value of b that lets $TM_i^j(dif_t)$ close to 0 (let it be 0.01) at $dif_t = t\_dr$ can be computed as follows

$$b = \frac{-ln(0.01)}{e^{t\_dr}} \tag{12}$$

Relevance (RL): Relevance of the reported information to the sensing task can be spatial, temporal, or both. The spatial relevance (S_RL) measures the portion of the reported information that spatially fits in the area of interest declared in the sensing task. Per a participation report, it can be computed as the length of the covered area overlapping with the area of interest ($l\_overlap$) to the total length of the covered area (l_total). This computation is shown in Eq. 13 for information reported by participant i in interaction j. The temporal relevance (T_RL) measures the portion of the reported information that timely fits in the task interval (t_inter). Considering the start and end times of the reported information, it can be computed as the length of the reporting duration overlapping with t_inter ($d\_overlap$) to the total length of the reporting duration (d_total). We show this computation in Eq. 14 for information reported by participant i in interaction j. For computing the spatiotemporal relevance, which is the value that represents our RL metric, we multiply the S_RL and T_RL values, as shown in Eq. 15.

$$S\_RL_i^j = \frac{l\_overlap_i^j}{l\_total_i^j} \tag{13}$$

$$T\_RL_i^j = \frac{d\_overlap_i^j}{d\_total_i^j} \tag{14}$$

$$RL_i^j = S\_RL_i^j \times T\_RL_i^j \tag{15}$$

Quality of Resources (QR): For assessing this metric, a hashtable can be created with a record for each brand model and a corresponding weight based on its features and manufacturing year.

The values of these three metrics are also combined using a utility function as shown in Eq. 16 to compute the QoI, Q, reported from a participant i after a contribution j, per

$$Q_i^j = w_1^{QS} \times TM_i^j + w_2^{QS} \times RL_i^j + w_3^{QS} \times QR_i^j$$
$$\text{where } \sum_{l=1}^{3} w_l^{QS} = 1 \tag{16}$$

(iii) Trust level: The trust level (TL) metric measures how much a candidate participant is reliable and can be trusted to perform a task. Some participants may behave deceitfully seeking to maximize their own gains. Other malicious participants may provide incorrect data with the purpose of degrading/misleading the service to be provided. This metric is used to detect and disqualify such malicious/untrusted participants. Many techniques can be used to measure the TL value of a participant after an interaction. These techniques vary based on the nature of the task that participant has performed and is being assessed after, and the type of sensors used. For example, if the task involved using the on-board camera for capturing on-road images or videos, the viability and correctness of the reported images/videos can be evaluated using image processing algorithms to measure the degree of feature matching between past images of the road/object defined in the sensing task and the images/videos of this road/object reported by the participant. This degree of matching is an indicator of the participant's reliability and can be used as his/her TL. Other sensing tasks may involve reporting raw sensing data such as traffic volumes and pollution levels. In such cases, the reported value can be compared to the range of expected values computed based on the average of past values. The degree of closeness to the expected range

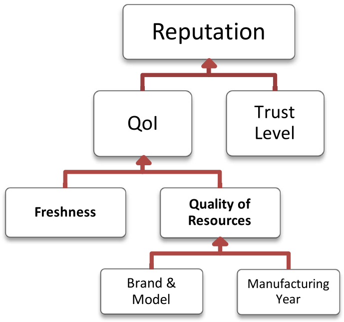


Fig. 7. The reputation metrics used in the unsolicited model.

can represent the TL of the reporter. The computed TL value should be normalized to the [0,1] range before plugging it into Eq. 17 of our assessment scheme.

Finally, the per-interaction assessment, v, of a participant i after a contribution j is computed by combining the participation commitment (P_i^j), the QoI (Q_i^j), and the trust level (TL_i^j) using the following additive utility function

$$v_i^j = w_1^v \times P_i^j + w_2^v \times Q_i^j + w_3^v \times TL_i^j \text{ where } \sum_{l=1}^{3} w_l^v = 1 \tag{17}$$

2) Reputation Assessment in the Unsolicited Model: Computing reputation scores in the unsolicited model primarily depends on the advertised metadata that gives an insight into the quality of the to-be-retrieved data.

a) Computing the reputation score: Two main metrics can be used for assessing the reputation in the unsolicited model: *1) Quality of Information (QoI)* and *2) Trust Level (TL)*. The QoI metric involves underlying sub-metrics as shown in Fig. 7. Since the being-assessed data has been already captured, the participation commitment metric is not needed in this model.

The assessment needed for computing a reputation score for a participant involves two parts; pre-interaction and post-interaction. Since the data to be collected has been already measured, the evaluation of the QoI metric and its sub-metrics is handled using a pre-interaction assessment before selecting the participants based on the features of the data they hold and the features of the data-holding vehicles themselves.

The post-interaction assessment is done by the SP after each interaction with a participant to compute his/her TL. The results of the post interaction assessments are stored for future use to be utilized for computing an expected TL value for a participant. This expected TL is combined with the computed value of the QoI metric corresponding to the data advertised by that participant, resulting in a reputation score to be considered in the selection of that participant. The aggregation of the past TL post-interaction assessments for computing the expected TL value is handled using the Beta system following the same steps detailed in the system use in the solicited model.

After computing the expected TL value of participant i (TL_i) and the QoI of the data advertised by that participant (Q_i), these two values are combined using an additive utility function to compute the reputation score (r_i) to be considered in the selection process. This computation is shown below.

$$r_i = w_1^r \times Q_i + w_2^r \times TL_i \text{ where } \sum_{l=1}^{2} w_l^r = 1 \tag{18}$$

In the following, we detail how the pre-interaction and post-interaction assessments are handled.

b) Pre-interaction assessment: This assessment is used for computing the QoI value of the data advertised by a participant right before considering the participant in the selection process. The two metrics we use for computing the QoI value are the freshness of the advertised data and the holding vehicle's quality of resources.

Freshness (FR): Let the information acquisition time be T_q and the time of information collection by the SP be T_c. Freshness evaluates how recent this measured information is such that the minimal the difference between T_c and T_q ($diff_f$) is, the higher the freshness value is. The freshness value decreases as $diff_f$ increases. We also use the inverse Gompertz function to represent the freshness evolution and evaluate its value. The information to be considered is bounded by a time window with length (w_len) such that if $diff_f$ exceeds w_len, the information is useless and its freshness value is assigned 0. Eq. 19 summarizes how the freshness value of the information hold by participant i is computed.

$$FR_i(diff_f) = \begin{cases} ae^{-be^{c(diff_f)}} & \text{if } diff_f \leq w\_len \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

with having $a = c = 1$ and b computed as in Eq. 12 replacing t_dr with w_len.

The ***Quality of Resources (QR)*** can be computed as discussed under the solicited model.

The values of these two metrics are combined using a utility function as shown in Eq. 20 to compute the QoI (Q) of the data advertised by participant i.

$$Q_i = w_1^{QU} \times FR_i + w_2^{QU} \times QR_i \text{ where } \sum_{l=1}^{2} w_l^{QU} = 1 \tag{20}$$

c) Post-interaction assessment: This assessment is used for computing a TL value of a participant after an interaction. The TL metric is used for the same purposes aforementioned under its use in the solicited model; yet, different techniques of measuring the TL value can be used. Since the unsolicited model involves advertisements of the carried data through metadata, a matching evaluation can be used that involves comparing the advertisement metadata (adv_meta) received by the SP and corresponding metadata extracted from the received data (rec_meta). A similarity function $f_{sm}(adv\_meta, rec\_meta) \rightarrow s$ can be used for such a comparison resulting in a similarity score s, in the [0,1] range, representing the matching degree between adv_meta and rec_meta. The similarity score s is an indicator of the reporter reliability and can be used as his/her TL as summarized in Eq. 21 for participant i after an interaction j. Many similarity functions are proposed for data matching in semantic web and data management applications, with the simple string matching being the basic form [29].

$$TL_i^j = s = f_{sm}(adv\_meta, rec\_meta) \tag{21}$$

The TL_i^j is mapped to the v_i^j value to be plugged into Eq. 4 of the Beta system to initiate its operation targeting computing the expected TL_i value.

B. Pricing Model

Taking reputation into consideration along with participant availability, a dynamic pricing model that computes participants' rewards based on their reputation score can be adopted. Reward/price assigned to each participant is proportional to distance traversed (a measure of availability) as well. We compute a participant price pr_i which, in turn, is the cost c_i incurred by the SP for recruiting participant i, as follows

$$c_i = C_{\text{init}} + (C_m * d_i * r_i) \tag{22}$$

where C_{init} is a constant initial reward paid to incentivize participants; C_m is a constant cost per meter determined by the SP; d_i and r_i are the covering distance (in meters) and reputation score of participant i, respectively, $\forall\ i \in \mathcal{S}$.

For flexibility of implementation, the rewards and costs are represented as a number of tokens that can be mapped to any form of incentives by the SP.

It is worth mentioning that the operation of the recruitment framework is generic and is not restricted to the use of the assessment scheme and pricing model presented above.

V. Performance Evaluation

In this section, we present numerical results of the proposed RTR selection scheme comparing the ILP optimization and greedy heuristic solutions of the main recruitment objective targeting maximizing the coverage while minimizing the total recruitment cost (which we refer to as MinCost in the ILP and RBMC-MC in the heuristics). The results of the ILP optimization represent upper bounds of the budgeted reputation-aware recruitment that can be achieved. In addition, to show the gain achieved through minimizing the recruitment cost, we compare the solutions to solutions targeting only the maximum coverage (which we refer to as MaxCov in the ILP and RBMC in the heuristics). The solutions are compared in terms of the *ratio of achieved coverage* and the *total recruitment cost*. In addition, we consider two more metrics that show the quality of the sensed data collected by the selected participants. The first metric is the *amortized quality index (AQI)* indicating the overall quality of sensing averaged over the selected group of participating vehicles. The AQI is computed as the average QoI, avg(Q_i), of the selected participants. The other metric is the *trustworthiness index (TI)* that indicates the level of trust in the viability and correctness of the collected data over the selected vehicles. The TI metric is computed as the average of $[QR_i \times TL_i]$ of the selected participants. We also study the effect of changing the reputation threshold R_{Th}, while keeping the density of vehicles fixed through comparing the two heuristic solutions considering different values of R_{Th}.

Considering the practical challenges, we also assess the coverage achieved by the selection scheme with probability of vehicles leaving their announced trajectories. We evaluate the RBMC solution considering different ranges of the degree of confidence D_i.

All the shown results represent the average results of running the algorithms for 1000 rounds per comparison.

A. Implementation Setup

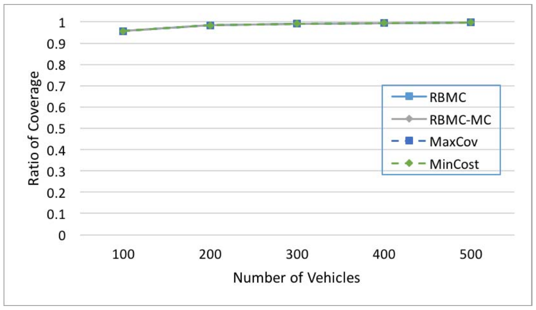
We use Gurobi 5.6.3 [30] to solve the ILP optimization formulation with Matlab as a simulation environment. The greedy heuristic algorithms are implemented in C++. We simulate an area of interest of $5Km$ divided into 100 road sectors, each is 50 meters long. The vehicular trajectories are randomly generated within that area of interest. C_{init} is set to 1, C_m is set to 0.01, and R_{Th} is set to 0.5.

We consider the reputation assessment under the solicited model since the assessment under the unsolicited one is much simpler. Our reputation assessment scheme is applied to compute a reputation score $r_i\ \forall\ i \in \mathcal{S}$. For each participant i, the CoT_i metric is assigned the value 1 as we do not assume a probability of leaving in the basic comparisons. The WP_i, TM_i, RL_i, and TL_i values are randomly generated in the [0,1] range. The QoR hashtable has 10 records each associated with a value in the [0.5,1] range. The values of the participation commitment (P_i) and QoI (Q_i) metrics are then computed based on Eqs. 10 & 16, respectively. The per-interaction assessment (v_i) is computed based on Eq. 17. Equal weights are given to the underlying metrics in each utility function. The final reputation score (r_i) is computed using the Beta reputation system as discussed in Section IV-A-1. We consider only one interaction in computing the reputation score ($n = 1$) for simplicity, since considering more past interactions will not affect the overall performance comparison.

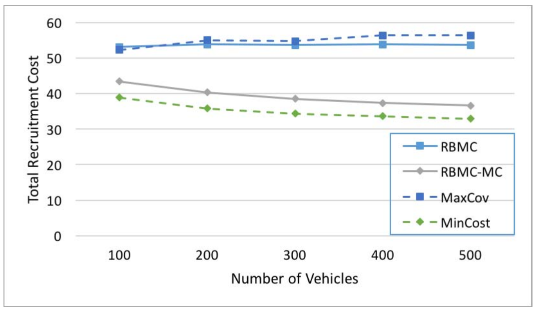
We apply our pricing model to compute c_i according to Eq. 22 $\forall\ i \in \mathcal{S}$.

B. Numerical Results and Analysis

For the comparison between the ILP and heuristic solutions, we first compare them in terms of the first two metrics with a budget cap that allows for achieving full coverage to the area of interest (B is set to 100). Fig. 8 shows the results of this comparison for various densities of vehicles (number of vehicles per area of interest). In terms of the ratio of achieved coverage, Fig. 8(a) shows that all the solutions achieve the same coverage as they all work on achieving the maximum coverage that can be provided by the available trajectories. No restrictions on the number of chosen trajectories are encountered since the budget cap allows for that. With increasing the vehicle density, the solutions succeed in achieving better coverage since the opportunity that more road sectors have vehicles passing by increases. In Fig. 8(b), the solutions are compared in terms of the total recruitment cost. The results show that RBMC-MC achieves better performance compared to the other heuristic solution, and with a slight increase to the lower bound achieved by the ILP MinCost.

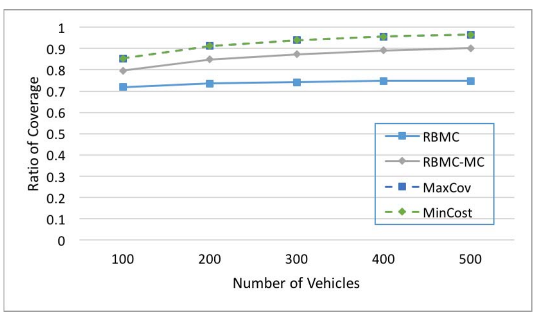


(a)

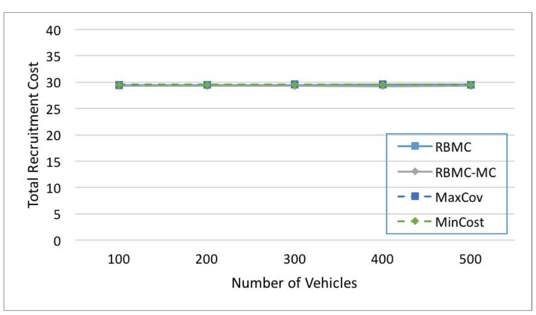


(b)

Fig. 8. Performance results with B = 100 and full coverage can be achieved. (a) Ratio of coverage with varying densities. (b) Total recruitment cost with varying densities.



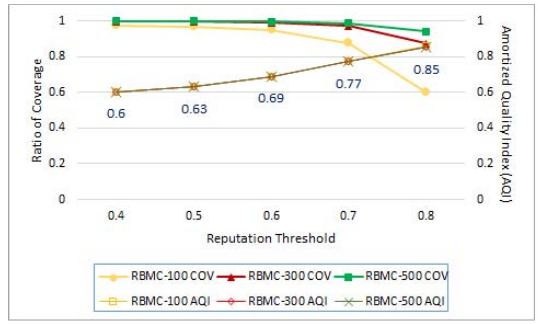
(a)



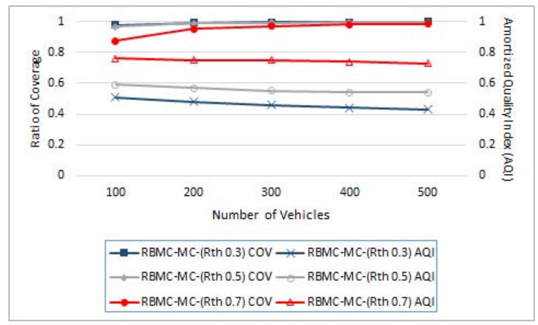
(b)

Fig. 9. Performance results with B = 30 and only partial coverage can be achieved. (a) Ratio of coverage with varying densities. (b) Total recruitment cost with varying densities.

Second, we perform the same comparison but with a strict budget cap that only allows for achieving partial coverage (B is set to 30). Results are shown in Fig. 9. In terms of the achieved coverage ratio, Fig. 9(a) shows that the RBMC-MC solution succeeds in achieving higher coverage



(a)

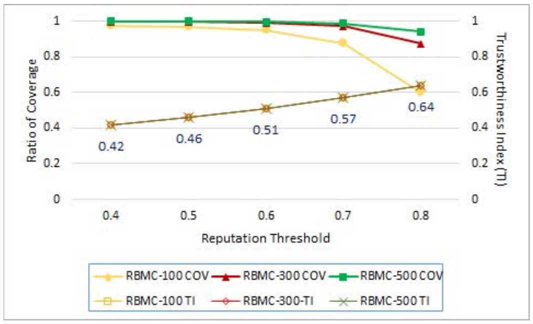


(b)

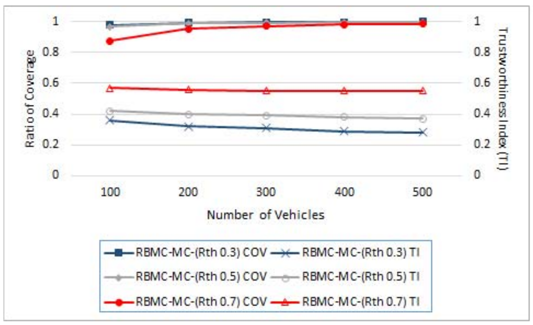
Fig. 10. The tradeoff between the AQI and ratio of coverage. (a) The tradeoff between the AQI and coverage with varying reputation thresholds. (b) The tradeoff between the AQI and coverage in RBMC-MC with varying densities.

than the basic RBMC. The reason is that while RBMC-MC selects a trajectory/vehicle, it tries to minimize the recruitment cost. Therefore, it manages to recruit more vehicles, hence, achieving higher coverage using the same budget cap. The optimal coverage achieved by the ILP solutions is slightly higher than the best coverage achieved by the heuristics through RBMC-MC. In terms of the recruitment cost, Fig. 9(b) shows that all the solutions incur almost the same total recruitment cost. The reason is that all the solutions try to achieve their objectives limited by the budget cap, therefore, their total recruitment cost will always be close to the limited budget cap.

In terms of evaluating the reputation and quality of sensed data, Fig. 10 shows the results related to the AQI metric. In Fig. 10(a), we show the effect of increasing the reputation threshold on both the ratio of coverage and the AQI considering the RBMC algorithm with different densities (100, 300, 500 vehicles per the area of interest). The results show that with increasing the reputation threshold, the ratio of coverage decreases due to the decrease in the number of candidate participants, while the AQI increases due to raising the reputation bar, resulting in a tradeoff. As expected, in RBMC, changing the vehicular density for the same reputation threshold has no effect on the AQI value. The reason is that even with increasing the number of candidate vehicles, the average reputation over the selected participants stays the same since the RBMC solution only focuses on the unique coverage of the candidate participants, given that they are all having reputation within the defined range. On the contrary, we can see in Fig. 10(b) that with using the RBMC-MC algorithm,



(a)



(b)

Fig. 11. The tradeoff between the TI and ratio of coverage. (a) The tradeoff between the TI and coverage with varying reputation thresholds. (b) The tradeoff between the TI and coverage in RBMC-MC with varying densities.

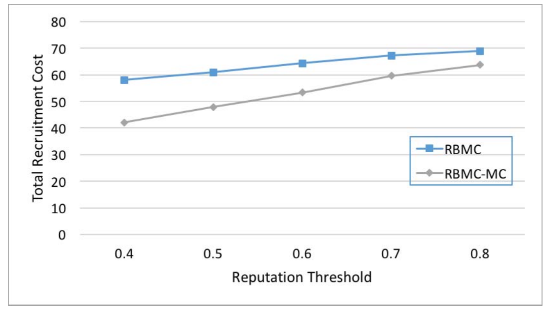


Fig. 12. Total recruitment cost with varying reputation thresholds.

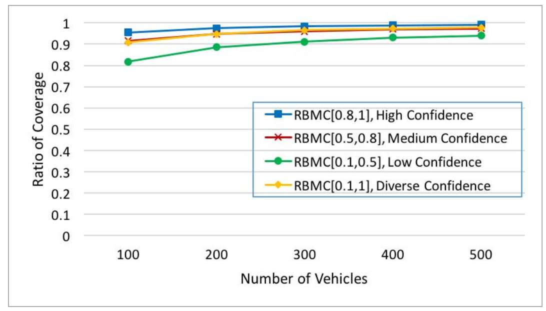


Fig. 13. Coverage assessment for different ranges of the confidence degree.

the AQI value decreases with increasing the vehicular density while maintaining the same reputation threshold. The reason is that since the RBMC-MC algorithm works on minimizing the cost, with increasing the vehicular density the algorithm may select less expensive options which are linked to participants having lower reputation resulting in lower AQI. Apparently, with increasing the density, the ratio of coverage increases resulting in a tradeoff between the AQI and coverage in this case as well. Fig. 10(b) shows these results for different reputation thresholds (0.3, 0.5, and 0.7).

The results involving the TI metric are shown in Fig. 11. The same above discussion related to the AQI metric applies to the TI metric. Fig. 11(a) shows that there is a tradeoff between the ratio of achieved coverage and the TI value of the covering set with changing the reputation threshold. Fig. 11(b) shows that the same tradeoff is maintained under the RBMC-MC algorithm while changing the vehicular density.

We study the effect of changing the reputation threshold R_{Th} on the total recruitment cost while keeping the density of vehicles fixed (200 vehicles per the 5Km area). We compare the performance of the two heuristic solutions with different values of R_{Th}. As expected, Fig. 12 shows that with increasing R_{Th}, the total recruitment cost increases and the two solutions converge because they all get restricted to very limited options which are the vehicles with r_i above/equals to the threshold.

Finally, we assess the coverage achieved by the RTR framework with probability of vehicles leaving their announced trajectories. We compare the RBMC solution considering this probability with four different ranges of the degree of confidence D_i. In Fig. 13, we present the assessment results obtained with these four ranges of D_i considering different densities of vehicles available in the area of interest. The results show that, in a dense environment (400-500 vehicles in the 5Km area), even with low values of D_i (high probabilities of a vehicle not sticking to its trajectory), our framework achieves a high coverage ratio. This is attributed to the fact that the selection scheme includes sufficient vehicles in the covering set to compensate for probabilities of leaving, which enhances the reliability of our framework. In a sparse environment (100 vehicular densities), the achieved coverage ratio is lower because of the non-sufficient availability of covering vehicles. We remark that RBMC-MC with probability of leaving, and RBMC with redundancy and probability of leaving are straightforward extensions.

VI. Conclusions and Future Work

In this paper, we proposed the reputation-aware, trajectory-based RTR framework for recruiting vehicles for public sensing services. The framework utilizes the spatiotemporal availability of participants and their reputation to select a set of vehicles that achieves coverage of an area of interest with a budget cap. We proposed a reputation assessment scheme and a pricing model as parts of the framework that are used to feed the third module, the selection scheme, with a reputation score and a recruitment cost of each candidate participant to start the selection process. In a previous work, we formulated the selection problem as an ILP optimization problem for a main recruitment objective; maximizing coverage with minimum cost. In this paper, we presented greedy heuristic solutions that handle the aforementioned recruitment objective for real-time services. The RTR framework generalizes the basic selection case to some practical cases that an SP faces during the recruitment process such as probability of a vehicle not sticking to its announced trajectory and having redundancy requirements. The performance evaluation results showed that the proposed greedy heuristics achieve results close to the optimal benchmarks and succeed in efficiently handling cases with high probabilities of vehicles not sticking to their trajectories.

Our future work includes investigating techniques from the computational geometry field that can be used for solving the trajectory-based selection process and comparing them to our greedy heuristics. We plan also to study the use of auction-based pricing in our framework for having the participants bid for their prices instead of computing them by the SPs.

REFERENCES

[1] W. Z. Khan, Y. Xiang, M. Y. Aalsalem, and Q. Arshad, "Mobile phone sensing systems: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 402–427, 1st Quart., 2013.

[2] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, "A survey of mobile phone sensing," *IEEE Commun. Mag.*, vol. 48, no. 9, pp. 140–150, Sep. 2010.

[3] S. Abdelhamid, H. Hassanein, and G. Takahara, "Vehicle as a resource (VaaR)," *IEEE Netw.*, vol. 29, no. 1, pp. 12–17, Jan./Feb. 2015.

[4] W. J. Fleming, "New automotive sensors—A review," *IEEE Sensors J.*, vol. 8, no. 11, pp. 1900–1921, Nov. 2008.

[5] S. A. Hamid, H. Abouzeid, H. S. Hassanein, and G. Takahara, "Optimal recruitment of smart vehicles for reputation-aware public sensing," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2014, pp. 3160–3165.

[6] U. Lee, E. Magistretti, M. Gerla, P. Bellavista, and A. Corradi, "Dissemination and harvesting of urban data using vehicular sensing platforms," *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 882–901, Feb. 2009.

[7] C. E. Palazzi, F. Pezzoni, and P. M. Ruiz, "Delay-bounded data gathering in urban vehicular sensor networks," *Pervasive Mobile Comput.*, vol. 8, no. 2, pp. 180–193, 2012.

[8] U. Lee and M. Gerla, "A survey of urban vehicular sensing platforms," *Comput. Netw.*, vol. 54, no. 4, pp. 527–544, 2010.

[9] X. Wang, X. Zheng, Q. Zhang, T. Wang, and D. Shen, "Crowdsourcing in ITS: The state of the work and the networking," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 6, pp. 1596–1605, Jun. 2016.

[10] L. Tang, X. Yang, Z. Dong, and Q. Li, "CLRIC: Collecting lane-based road information via crowdsourcing," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 9, pp. 2552–2562, Sep. 2016.

[11] A. Joshi and M. R. James, "Generation of accurate lane-level maps from coarse prior maps and lidar," *IEEE Intell. Transp. Syst. Mag.*, vol. 7, no. 1, pp. 19–29, Mar. 2015.

[12] C. Ruhhammer, N. Hirsenkorn, F. Klanner, and C. Stiller, "Crowdsourced intersection parameters: A generic approach for extraction and confidence estimation," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2014, pp. 581–587.

[13] K. Xie, W. Luo, X. Wang, L. Wang, and J. Wen, "Road condition gathering with vehicular DTN," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr./May 2015, pp. 576–581.

[14] M. Aki *et al.*, "Road surface recognition using laser radar for automatic platooning," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 10, pp. 2800–2810, Oct. 2016.

[15] S. Abdelhamid, H. S. Hassanein, and G. Takahara, "Vehicle as a mobile sensor," *Procedia Comput. Sci.*, vol. 34, pp. 286–295, Jan. 2014.

[16] S. Reddy, D. Estrin, and M. Srivastava, "Recruitment framework for participatory sensing data collections," in *Proc. 8th Int. Conf. Pervasive Comput.*, vol. 6030. 2010, pp. 138–155.

[17] S. Khullera, A. Mossb, and J. Naor, "The budgeted maximum coverage problem," *Inf. Process. Lett.*, vol. 70, no. 1, pp. 39–45, 1999.

[18] Y. Zeng and D. Li, "A self-adaptive behavior-aware recruitment scheme for participatory sensing," *Sensors*, vol. 15, no. 9, pp. 23361–23375, 2015.

[19] L. G. Jaimes, I. Vergara-Laurens, and M. A. Labrador, "A location-based incentive mechanism for participatory sensing systems with budget constraints," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Mar. 2012, pp. 103–108.

[20] J.-S. Lee and B. Hoh, "Dynamic pricing incentive for participatory sensing," *Pervasive Mobile Comput.*, vol. 6, no. 6, pp. 693–708, 2010.

[21] Z. He, J. Cao, and X. Liu, "High quality participant recruitment in vehicle-based crowdsourcing using predictable mobility," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr./May 2015, pp. 2542–2550.

[22] K. Han, C. Chen, Q. Zhao, and X. Guan, "Trajectory-based node selection scheme in vehicular crowdsensing," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Nov. 2015, pp. 1–6.

[23] S. A. Hamid, G. Takahara, and H. S. Hassanein, "On the recruitment of smart vehicles for urban sensing," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 58–63.

[24] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artif. Intell. Rev.*, vol. 22, no. 2, pp. 85–126, 2004.

[25] K. L. Huang, S. S. Kanhere, and W. Hu, "On the need for a reputation system in mobile phone based sensing," *Ad Hoc Netw.*, vol. 12, pp. 130–149, Jan. 2014.

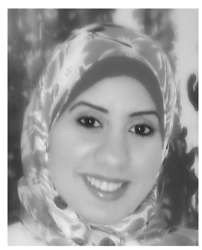
[26] U. Feige, "A threshold of ln n for approximating set cover," *J. ACM*, vol. 45, no. 4, pp. 634–652, 1998.

[27] B. Guo, Z. Yu, X. Zhou, and D. Zhang, "From participatory sensing to mobile crowd sensing," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PERCOM Workshops)*, Mar. 2014, pp. 593–598.

[28] A. Jsang and R. Ismail, "The beta reputation system," in *Proc. 15th Bled Electron. Commerce Conf.*, 2002, pp. 41–55.

[29] R. da Silva, R. Stasiu, V. M. Orengo, and C. A. Heuser, "Measuring quality of similarity functions in approximate data matching," *J. Informetrics*, vol. 1, no. 1, pp. 35–46, 2007.

[30] *Gurobi Optimization*, accessed on Jun. 23, 2017. [Online]. Available: http://www.gurobi.com/



Sherin Abdelhamid (S'12–M'15) received the B.Sc. and M.Sc. degrees in computer science from Ain Shams University, Egypt, in 2005 and 2009, respectively, and the Ph.D. degree from the School of Computing, Queen's University, Canada, in 2014. She is currently an Assistant Professor with the Faculty of Computers and Information, Ain Shams University. Her research interests include intelligent transportation systems, vehicular and wireless multi-hop networks, vehicular cloud computing, information-centric networks, crowdsourcing, and Internet of Things. She has several publications in top venues and has received many awards and recognitions, including the best paper award of the AHSN Symposium of the IEEE GLOBECOM 2013.

Hossam S. Hassanein (S'86–M'90–SM'05–F'17) is the Founder and the Director of the Telecommunications Research Laboratory, Queen's University School of Computing. He is a leading authority in the areas of broadband, wireless and mobile networks architecture, and protocols and performance evaluation. His record spans over 500 publications, in addition to numerous keynotes and plenary talks in flagship venues. He has received several recognitions and best paper awards at top international conferences. He is a former Chair of the IEEE Communication Society Technical Committee on Ad hoc and Sensor Networks. He is an IEEE Communications Society Distinguished Speaker.

Glen Takahara (M'97) received the B.A. degree in mathematics from The University of British Columbia, Vancouver, BC, Canada, in 1988, and the M.S. and Ph.D. degrees in statistics from Carnegie Mellon University in 1990 and 1994, respectively. He is currently a Professor in mathematics and engineering with the Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada. His research interests are in communications theory, networking, applied probability, and statistics.