Evaluating Mobility Models in Participatory Sensing

Ricardo Lent, Marcelino Minero, Robin North, and Javier Barria
Imperial College London
London SW7 2BT, United Kingdom
{r.lent, mm9609, robin.north, j.barria}@imperial.ac.uk

ABSTRACT
Pervasive wireless sensor networks offer many mission-oriented opportunities. In this paper we evaluate the potential benefits of combining fixed and mobile sensor nodes as part of a mission-oriented, participatory sensor network deployed on an urban environment that has the objective of detecting the extent of the emission of a substance of interest (e.g., pollutant). We evaluate the system both by assuming simple mobility models for sensors and more realistic models that seek to reduce operating costs.

Categories and Subject Descriptors
I.6.3 [Computing Methodologies]: Simulation and Modeling—applications

General Terms
Design, Measurement, Performance

Keywords
Pervasive Computing, Mobile Computing, Wireless Sensor Networks, Simulation

1. INTRODUCTION
Environmental quality in urban spaces is a top priority towards reducing public health issues [1]. This is (partially) addressed by using monitoring networks that estimate the amount of pollutants in the atmosphere through measurement procedures. The accuracy and complexity of the data acquired strongly depends on the methodology used to perform these measurements and the mitigation actions being considered. For example, the service requirements in terms of spatial coverage, time to alarm, and acceptable error limits may vary for recurrent pollution (e.g., traffic-related air pollution) as opposed to one-off accidental releases (e.g., Spillage of a hazardous chemical). In this context, the performance of the monitoring network also depends on the cost and resources assigned for the deployment of such network.

A common approach to implement these monitoring networks is to use Wireless Sensor Networks (WSNs) due to its versatility. As a consequence of this characteristic, the number of monitoring networks that use WSNs is large and spans across dissimilar areas since the procedures for estimating a phenomenon of interest would depend on the application or field of interest. For example, a WSN used to estimate a phenomenon in the ground (e.g., a toxic cloud) may greatly differ from those used in buildings (e.g., a fire) or at sea.

In particular, the estimation of a phenomenon in urban scenarios might be accomplished by gathering a large number of readings from mobile sensors carried by people or vehicles using an approach recently referred as participatory sensing [2], [3]. The data values obtained from these readings, describing some properties of the phenomenon, can be used to either represent the phenomenon as a layered distribution of data values (level curves) [4] or to obtain a binary view of the phenomenon (i.e., the data values are greater and lesser than certain threshold).

The aim of this paper is to assess the impact of different participatory sensing mobility models in the estimation of an environmental phenomenon of interest. Our work will focus on the estimation of a particular boundary of the phenomenon (contour) using a WSN that includes fixed and mobile units. This approach will be performed following a four-step procedure. First, an initial estimation of the phenomenon is obtained using fixed sensor units. Second, a virtual sampling path is formed with the streets adjacent to this initial estimation. Third, a set of mobile sensor units is assigned and/or negotiated a mission to cover all or part of this virtual path to perform additional sampling towards improving the quality of the initial estimation. It is in this step where the impact of several mobility models on the quality of the estimation is investigated. Last, the previous step is repeated according to the values obtained by fixed and mobile units and relevant sampling paths are iteratively determined according to the dynamics of the event.

This paper is organised as follows. Related work is presented in Section 2. Section 3 describes the models and assumption used in this study. Section 4 presents our results and finally, Section 5 provides final remarks.

2. LITERATURE REVIEW
Within the context of this study, the goal of a WSN is to estimate a phenomenon occurring in the physical environment where they are deployed. Such phenomenon would normally be anomalous concentrations of chemicals, pollutants or temperature. The data values obtained by the WSN...
are then used to describe relevant properties of the phenomenon, such as extent and location over time.

An important technique consists in generating a layered distribution using data values obtained from sensor networks [4]. This technique is an ad-hoc solution that can be targeted to specific scenarios such as static (time-invariant) [6] or dynamic events [7], [8]. These scenarios range from the implementation of solutions using static or mobile sensing units to scenarios where both types of sensing units coexist.

Considering that a phenomenon can be estimated using the measurements from a number of mobile sensors another approach is to take advantage of the mobile sensors carried by people (e.g., [9]) or from vehicles with predetermined mobility [10], [3]. This approach is known as participatory sensing [2], social sensing or crowdsensing [5] among other terms.

Participatory sensing has the potential to estimate a phenomenon without using dedicated sensing networks lowering overall costs. However, some drawbacks exist such as calibration, privacy concerns [11] and increased processing mainly due to the number of readings [3]. A characteristic of this approach is that the estimation of the phenomenon is obtained from a number of sensors in which their mobility might not be fully controlled [12]. Two of such scenarios are presented in [10] and [3]. For example, in [3] the use of street sweepers to collect environmental data implies that the mobility of these vehicles is limited to pre-established routes. In this regard, the use of this approach involves only mobile sensors that upload their measurements to a central location in which estimation procedures are performed. Some of these approaches present the results to the user using Web interfaces [12], [13], [14]. In the framework here developed we set up the mechanisms so that user participation could be enhanced by using an incentive mechanism.

Despite the advantages of the participatory sensing, the uncontrolled mobility of sensors has an adverse effect on the efficiency of the monitoring task [4]. As an example, after an initial estimation of a phenomenon is obtained, we might be interested in directing mobile sensing units towards points closer to the estimated boundary instead of gathering additional readings of areas far from this estimation. In this paper we will investigate the effect of controlled and uncontrolled mobility models.

### 3. MODELS

A goal of this work is to develop a realistic simulation of sensors deployed on an urban setting. For this, we consider a specific area of the world—the South Kensington district in London (Imperial College location). The selected area comprises latitudes between 51.48701 and 51.51106 and longitudes between -0.20697 and -0.15514. A digital map of the area was obtained from the OpenStreetMap project and processed to extract a road graph suitable for mobile sensors (Fig. 1).

In a real-world scenario, the location of static sensors could be restricted by logistic or topographic reasons. In the evaluation study, we considered bus stops as possible locations for the static sensors, which have the advantage of a permanent electrical power supply and network connectivity. The latter is normally used by Transport for London to update electronic information boards about waiting times. At the beginning of a simulation run, a number of sensors are randomly placed on any of the 150 bus stops that exist in the area. Since these sensors are intended to remain static, their location is known by the control server.

Mobile sensors on the other hand are assumed to be carried by pedestrians or cyclists, who are assigned a random moving speed in the range 1–10 m/s. Mobile sensors are assumed to be embedded in or connected to a smart phone. With probability $P_p$, a mobile will pause at an intersection for a random time $T_p$. Mobile sensors will require the assistance of a GPS receiver or comparable system to correlate sensor readings to locations.

A control server or detector is responsible for receiving sensor readings and for computing the possible contour of the event. We assume that the server and the sensors can communicate through a network (e.g., the Internet). The server may in addition issue control messages to some mobiles to influence their mobility towards specific locations as we will discuss shortly.

#### 3.1 Pollutant Emission and Dispersion

When the emission of a substance of interest (i.e., pollutant) occurs, the wind will disperse it throughout the environment. The well-known Gaussian plume dispersion model by a point source allows to estimate the pollutant concentration at ground-level as it would be detected by a real sensor at $(x,y)$ relative to the origin of emission:

$$C(x, y, 0) = \frac{Q}{2\pi K} \exp\left(-\frac{u(y^2 + H^2)}{4Kx}\right)$$

Location $(x,y)$ is also relative to the direction of wind, so vector $(x,y)$ is rotated an angle $\alpha$. Both wind direction $(\alpha)$ and intensity $(U)$ evolve as a random walk with random steps $\alpha_w$ and $U_w$ respectively at each simulation iteration. The emission origin is randomly selected and does not change during the course of one simulation. This relatively simple case provides a suitable basis for the assessment of alternative sensor network mobility models.

#### 3.2 Sensing

An exact estimation of a phenomenon in an urban scenario would require a large number of sensors along with detailed meteorological and topographical conditions before the sampling procedure is performed. In general, however, such exact estimation is infeasible and not necessary. As an example, let us consider a factory plume whose emissions,
due to meteorological conditions, are directed towards an urban area. In this scenario, we might be interested in estimating the extent of pollution caused by a deposition process (e.g. dry deposition) in that particular area. This might be a particular threshold of contamination, beyond which the pollution is not considered dangerous. Depending on the dynamics of the scenario of interest, two approaches may be used: random sampling and transect sampling.

In a more general scenario, when the extent and dynamics of the phenomenon is unknown a transect sampling approach might be a better choice. This approach requires to trace lines (not necessarily parallel) across an area and to sample at regular intervals along these lines. These intervals are determined by the length of the transect line and the number of samples required. In this particular application we must consider that the position of the transect lines is not only determined by the extent and dynamics of the phenomenon but also by the topology of streets and other urban elements (e.g. green areas).

In this study, we assume that sensors operate in binary mode—can only report the presence of absence of a given substance after comparing readings to a threshold \( \theta \). Sensing and reporting occurs in cycles of \( \Delta_S \) s each. Given that the location of static sensors does not change, static sensors only need to report changes in their status (i.e., when going from positive to negative detection and vice-versa). Mobile sensors on the other hand always submit both their location and sensing state at the end of each reporting cycle.

### 3.3 Detection Model

For the sake of simplicity, we assume a convex hull-based contour detection mechanism by forming the minimal convex set \( P \) containing all points within the detection area that are known to be above the detection threshold. Better results are to be expected from more elaborated models. However, for comparison purposes our assumption is adequate given that we maintain the exact testing assumptions (i.e., binary sensors and same detection model) for all test cases.

Detection points are continuously reported by sensor nodes to a central server. Those points are determined by the location of static sensors and the location from which mobile sensors reported readings. In the latter case, a history of reports is assumed to be relevant only within a predefined time \( \tau \). Static nodes without an associated state report are assumed to be non-detecting. Otherwise, their associated reports would be valid but only within a predefined time \( \tau \).

Note that this system does not require any synchronization among sensors. The detector simply computes the contour asynchronously using whatever information available at the time.

### 3.4 Tour Estimation for Mobiles

The region between sensors detecting values below and above the threshold is the area with highest uncertainty and the most likely location of the real contour. We define set \( T_t \) as the desired tour (sequence of locations to visit) for mobiles at time \( t \), which we estimate by expanding \( P_t \) by a given factor. \( T_t \) is a scaled up version of \( P_t \) by a growth factor that we define as half the average distance between detecting and non-detecting sensing locations, which the server can calculate from the available data. \( T_t \) is periodically calculated by the control server every \( \Delta_T \) seconds.

### 3.5 Error Estimation

Let \( E_t \) be the area affected by the pollutant at time \( t \), i.e., \( E_t \) contains the set of points \((x, y)\) s.t. \( C_t(x, y, 0) > \theta \). Also, convex set \( P_t \) represents all the control server’s knowledge about pollution in the area (i.e., the detected area), which is periodically estimated from sensor reports as previously explained.

Two metrics are of main interest for evaluating the accuracy of \( P_t \) and therefore, of the contour detection mechanism. The first metric is the most relevant to this study and evaluates the error caused by not including polluted areas into the detected area and can be formulated as the conditional probability: \( P(\bar{P}_t/E_t) \). The second metric evaluates the error of including non-polluted areas into the detected area: \( P(P_t/E_t) \). Given that we have assumed a convex hull-based detection of the affected areas, \( P(P_t/E_t) \) tends to be very small, so we will focus on evaluating only the former metric.

A pragmatic way of calculating \( P(\bar{P}_t/E_t) \) is by comparing the real pollution status and the detection status for a set of sampling points. These points could be either uniformly covering the area of interest or better yet, covering only the roads in that area given that buildings could restrict pollution propagation. Given a non-empty set of sampling points \( S, s = (s_x, s_y, 0) \in S \) (assuming ground level), \( P(\bar{P}_t/E_t) \) could be approximated by:

\[
P(\bar{P}_t/E_t) = \frac{1}{|S|} \sum_{s \in S} 1\left[ C_t(s) > \theta \right] \cap (s \notin P_t)
\]

where \( 1[x] = 1 \) when \( x \) true and 0 otherwise. Note that \( P(\bar{P}_t/E_t) \) is equivalent to the mean square error (MSE) of locations with a pollution above the threshold level but not detected by the control server.

In the simulations, \( E_t \) and \( P_t \) were not evaluated continuously but at rate \( \Delta_E^{-1} \) and \( \Delta_P^{-1} \) respectively.

### 3.6 Mobility Models

We consider a number of mobility models to evaluate the possible improvements that mobile sensors could bring to the system. Three of these models are tour oblivious: In the Random Walk model, mobiles randomly decide which street to follow at each intersection. In the Random Waypoint model, mobiles randomly choose a destination intersection and move towards there on the shortest path through the streets. The third model, the Random Segment, is similar to the Random Waypoint, but mobiles acquire a new location once they arrive at their destinations—this models a node exiting the system and another entering at a different location. Three additional mobility models are tour aware. In the first one—the Touring Model, nodes start at a random location on the area and move to the closest vertex of the tour assuming that a tour has already been determined by the server. Once at that vertex, the mobile follows the tour either following the right or left direction (randomly selected). The second model—Detour, is a tradeoff between Random Segment and Touring Model. Mobiles start at a random location and move towards another random location. However, they detour to visit a fraction of the tour, which is a parameter of the system.

The last model, the Budget model, is equivalent to the Detour model with an additional property that prevents the control server from assigning certain detours to mobiles. We
assume that an incentive mechanism is in place that motivates mobiles to detrour and sense areas of interest to the control server. A budget places a limit to the maximum detour that the control server could incur.

For this, each detour is associated with a cost that is proportional to the difference between the total detour length and the direct route length from the initial location to the final destination of the mobile. Every time a new mobile becomes available for covering a possible detour, the server compares its available budget with the estimated detour cost. A detour is then assigned only if the server can "pay" the mobile’s detour costs at that time. The server’s budget increases over time at a predefined rate.

To implement this idea, we have modeled the server’s budget $B$ as a counter managed by a token bucket algorithm. $B$ increases overtime at a rate $b$ and a detour assignment will be granted if and only if $B \geq c$, where $c$ is the cost of the detour. If the assignment is successful, $B$ decrements $c$ units. In the typical application of this system to air pollution management, the token bucket algorithm can be used to implement temporal budget constraints.

4. SIMULATIONS

We will present Monte Carlo simulation results. Between 500 and 1000 samples were used to estimate each average value with each sample representing a single simulation run. The simulations were executed on a computing cluster of 100 cores. Table 1 shows the simulation parameters that were used. The study consisted of two hours of simulated time.

4.1 Random Models and Touring Model

It is expected that system detection accuracy would depend on the number of sensors available to track emission dispersions on the target area. Figure 2 depicts the resulting average mean square detection error as a function of the number of static sensors deployed on the scenario assuming no mobile sensors. Clearly, MSE could be reduced by increasing the number of static sensors. However, the location constraint of using only bus stop locations for the static sensors limits the sensing coverage of the system to the main roads. As a result, the MSE was unable to drop below 0.45 even when almost every bus stop was hosting a sensor node.

Mobile sensors could improve the detection accuracy of the system given that they can expand the coverage area. Figures 3 and 4 depict the ensemble MSE (over time) resulting from mobiles using the Random Walk and Random Waypoint mobility models respectively. It is interesting to note that even with the least appealing model (i.e., Random Walk), it was possible to improve the MSE with the help of the mobiles. Random Waypoint offered slightly better performance than Random Walk as it can be observed.

Table 1: Simulation Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pause probability $P_p$</td>
<td>0.1</td>
</tr>
<tr>
<td>Pause time $T_p$</td>
<td>uniform [0, 10]</td>
</tr>
<tr>
<td>Wind angle $\alpha_w$</td>
<td>uniform [-0.01, 0.01]</td>
</tr>
<tr>
<td>Wind intensity $U_w$</td>
<td>uniform [-0.02, 0.02]</td>
</tr>
<tr>
<td>Plume initial height $H$</td>
<td>20 m</td>
</tr>
<tr>
<td>Plume emission rate $Q$</td>
<td>20 Kg/s</td>
</tr>
<tr>
<td>Eddy diffusion coef. $K$</td>
<td>1 m²/s</td>
</tr>
<tr>
<td>Sensing interval $\Delta S$</td>
<td>20 s</td>
</tr>
<tr>
<td>Emission calc. interval $\Delta P$</td>
<td>20 s</td>
</tr>
<tr>
<td>Tour calc. interval $\Delta T$</td>
<td>20 s</td>
</tr>
<tr>
<td>Error calc. interval $\Delta E$</td>
<td>10 s</td>
</tr>
<tr>
<td>Detection threshold $\Theta$</td>
<td>0.001</td>
</tr>
<tr>
<td>Data time-to-live $\tau$</td>
<td>1000 s</td>
</tr>
</tbody>
</table>

Mobile sensors: 0

Figure 2: Detection mean square error (MSE) with only static sensors.

Figure 3: Detection MSE with the Random Walk model during the first two hours of the pollutant emission event.

Figure 4: Detection MSE with the Random Waypoint model with mobile sensors.

Once a potential tour is estimated, dedicated mobile nodes could be assigned to follow the tour, continuously generating readings from points of interest. Undeniably, following a designated tour of points of interest can provide better results than simply using mobiles moving randomly (Figure 5). Using multiple mobiles produced the best results.
4.2 Partial Touring

However impractical, making a significant number of mobiles follow the complete sensing tour can provide the lowest MSE. We turn now our attention to more practical cases where mobiles are asked to detour from their normal path to cover a portion of the tour. Figures 6 and 7 depict this case for scenarios having 50 static sensors and either 5 or 10 mobile nodes. The plots show cases for a target detour (i.e., tour distance covered) of 0, 0.5, 1, and 3 Km. The first case corresponds to nodes that ignore the detour request and move directly to their destination, which produces a MSE comparable to the Random Waypoint as it would be expected.

There is a relationship between the Tour Coverage Ratio (TCR) and the lowest MSE that could be achieved by the detection system. The TCR is the ratio between the tour length that is covered by the mobile route over the total tour length. Measured values for this relationship can be found in Figure 8 (left) for different number of mobiles and 50 static sensors. The MSE value shown in that figure corresponds to the average level observed during the second simulated hour (i.e., approximately after the system has reached its steady state). Results strongly suggest that larger TCRs produce little additional improvement and that higher benefits could be achieved by making use of larger number of mobiles covering shorter distances.

This observation can be further appreciated in Figure 8 (right). The MSE obtained by making mobiles cover about half of the tour (Detour 3000 model) produced comparable results to those from the Touring Model where mobiles were dedicated to go around the tour. Moreover, the simulations also suggest that the mobility behavior of mobiles becomes less significant for scenarios with very large number of mobiles. Figure 8 (right) indicates that the MSE of the different mobility models used in the simulations tend to converge. If the number of mobiles is not large, then significant accuracy improvements could be gained by appropriately guiding the mobility path of mobile sensors.

4.3 Budget Model

The last simulation set corresponds to various cases with a budget model. Figure 9 (left) depicts the steady-state MSE observed for increasing budget values and for different number of mobiles.

The achievable TCR was assumed to be linearly proportional to the detour cost, which is the difference between the total routing distance with and without a detour. This re-
6. REFERENCES


