

# MONITORING AIR POLLUTION BY COMBINING A STATIC INFRASTRUCTURE WITH A PARTICIPATORY SENSING APPROACH: DESIGN AND PERFORMANCE EVALUATION

DIEGO MENDEZ, JULIAN COLORADO, LAURA RODRIGUEZ,  
ANDRES CHACON & MATEO HERNANDEZ

Department of Electronics Engineering - Pontificia Universidad Javeriana, Bogota, Colombia.

## ABSTRACT

Participatory Sensing (PS) is a new fast-growing sensing approach that involves the participation of mobile phone users, and the corresponding communication infrastructure, to create a large-scale monitoring system. Using PS-based system makes it possible to measure and detect variables and events with an improvement in spatial and time resolution over traditional monitoring system. *Pollution-Spots* proposes an air pollution monitoring solution by means of using an infrastructure of fixed low-cost sensing devices, and reporting the measurements using a PS approach. The sensing devices acquire the variables and the pedestrian forwards this information, completing the cycle with no extra cost of data transport and/or human resources. However, including humans to the sensing loop, rises new challenges, such as protecting user private data, motivating user's participation, and reducing mobile phone's power consumption, all while maintaining the quality of the collected data. *Pollution-Spots* proposes a combined algorithm that protects the participant's private information and also implements a gamification technique to encourage the participation without any monetary reward. The proposed system has proven to be energy efficient when compared to similar approaches, with the additional benefit of considering the quality of the collected information, which is normally affected by privacy protection algorithms.

*Keywords:* air pollution, incentives, participatory sensing, privacy protection.

## 1 INTRODUCTION

Participatory Sensing (PS) is a new fast-growing sensing approach that involves the participation of mobile phone users, and the corresponding installed infrastructure, to create a large-scale monitoring system [1, 2]. Using a PS-based system makes it possible to measure and detect variables and events with an improved spatial and time resolution, compared to a traditional monitoring system, constituted mainly by a small set of stations in the city. The collected data are centralized and generally used to increase collective knowledge. In the case of environmental data, and more specifically the data concerning air quality, the sensing variables (such as temperature, particulate matter (PM), and the concentration of different pollutants) can be used to visualize their behavior over the time-space region of interest, thus helping communities to increase awareness about pollution, and consequently improving their quality of life.

One of the main obstacles when trying to implement an air pollution monitoring system using a PS approach is the fact that no mobile phone includes the required embedded sensors to acquire the variables of interest. Additionally, attaching these sensors to our current smartphones would be a burdensome project and most participants will reject the idea of carrying additional equipment on a daily basis. Considering this context, this paper presents *Pollution-Spots*, a novel scheme to monitor air pollution by combining an infrastructure of already deployed low-cost pollution stations (the *Pollution-Spots*) and the participation of mobile phone users to transmit and centralize the data, so as to avoid the extra cost of data transport and/or human resources. The result is a cost-effective air quality sensing system that is easy

to deploy and scale. In this model, it is simple to enlarge the region of interest, since it only requires the installation of a new *Pollution-Spot* to acquire the environmental data of that new region. This paper is an extended version of the work presented in [3].

However manageable the deployment of *Pollution-Spots* is, when implementing a PS-based monitoring system, the designer needs to face a set of challenges concerning the roll of humans in the sensing loop. First, how is the system going to encourage the participation of the mobile phone users? The users' participation is crucial since they are the only mechanism to collect and centralize the environmental information, nevertheless, why would a user sacrifice time to participate in such a system? Second, the installation of a pollution monitoring app in the user's phone implies that it will require additional battery usage, bandwidth and computational resources. Therefore, the application and the whole system need to be designed in such a way that utilizes these limited resources in a smart way. And third, since the participants have to share personal information, such as ID and location, the system should properly protect the user's private data while maintaining a good quality of the collected information. *Pollution-Spots* considers these issues and proposes a solution for each one of them.

The rest of this paper is organized as follows: related work is discussed in Section 2. Section 3 details the proposed scheme, depicting the system's architecture and main functions of the nodes. Section 4 describes the combined algorithm that includes the benefits of privacy protection while also encouraging the participation of mobile phone users. Section 5 refers to the results obtained for different scenarios. Section 6 concludes the paper with the main contributions of this work.

## 2 RELATED WORK

There are different examples in literature of applications that sense environmental data exploiting the characteristics of PS schemes, especially when considering an urban scenario where we can encourage the participation of many mobile phone users [4]. Besides air pollution monitoring, similar approaches have also used this approach to measure other variables of interest, such as noise pollution [5]. The work depicted in these examples has a common denominator: the concept of the implied mobility of the sensors, such that they are at all times attached to the mobile devices, as it is presented in [6], for instance. This notion has the disadvantage of requiring trust of the users with the sensing devices, reducing the possibilities of obtaining more granulated data, since there is a limited number of trusted users, and an even more limited number of sensing devices. A similar system to the one proposed here, Common Sense [7], has developed handheld monitors with air quality sensors as a prototype, in order to resolve the absence of the sensors in the consumer devices. The monitors communicate with the database through Bluetooth, 802.15.4, or GPRS radios. The project Haze Watch [8, 9] relies on air pollution sensors attached to motor vehicles to perform the readings. The sensing data are sent to the database using iPhones, which also have to be attached to the motor vehicles. Project Minutely [10], though working with forecast data (instead of raw data as intended in this project), provides a good example of how to integrate people in the sensing loop: it combines the weather radar from the Bureau of Meteorology in the United States and Australian areas, with personal reports from the participants, to deliver real-time weather forecasts.

A very interesting document is presented in [11] where the reader can find a compilation of community-led implementations of environmental monitoring initiatives that are based on urban PS, including both handheld (portable). The document elaborates on the different challenges and opportunities that such a system might have. Another example is P-Sense [2], a

system where the environmental data are collected by a set of individual external sensors integrated in a board. The sensing devices can be carried along with the cellphone or function as standalone devices due to their independent power supply. The proposed architecture of P-Sense is used as a reference point for *Pollution-Spots*.

As it is shown in many PS systems [12–18], the simplest way to encourage the participation of the users requires offering a monetary reward, but for some applications such a budget is not possible to have. Some incentive mechanisms do not rely on a monetary reward [8, 10, 19, 20] but on the participant’s benefit when they take part in such a system, like obtaining information about the variable under study or just philanthropic satisfaction. Since our project does not include a monetary budget for compensation, we rely on these latter techniques. There are also hybrid algorithms that simultaneously use different privacy protection techniques to get the best of each. Examples of such a mechanism are the work presented in [21], which uses an extension of  $l$ -diversity applied to the V-MDAV algorithm (tessellation), and the work in [22] which uses a decision scheme to select an encryption or an anonymization approach. An improved version of this system is presented in [23] where the decision process depends on the dynamics of the variable of interest in small cells in which the area under study has been subdivided. There are also examples where the set of points sent to each user according to their individual mobility patterns [24, 25], we recognize the positive outcome of using these agent-based methods, and place them in consideration as future stages in our project.

### 3 PROPOSED SCHEME: POLLUTION-SPOTS

Figure 1a depicts the proposed scheme for the *Pollution-Spots* system’s architecture. The model is composed by five layers: (i) Data Collection, (ii) Data Analysis, (iii) Data Network, (iv) Data Storage and Feedback, and (v) Data Storage, Estimation and Visualization.

- **Data Collection Layer:** The first layer is composed of a network of static sensing units, each one has a set of environmental sensors integrated in a board (based on an Atmel AVR processor), therefore working as a single unit. The sensing features of the units include measurement of Ozone (Membrapor O3/S-5), Relative Humidity and Tempera-

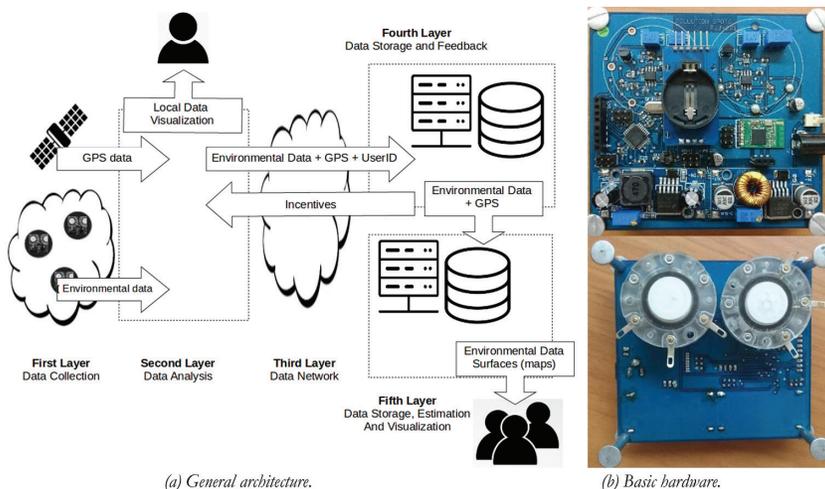


Figure 1: Pollution-Spots PS system.

ture (Sensirion SHT15), Particulate Matter (Sharp GP2Y1010AU0F), and Carbon Monoxide (Membrapor CO/SF-200). Since they must work as standalone units, the devices include an independent power supply based on an energy harvesting module and a solar panel, making the solution autonomous and sustainable. The set of sensors collects measurements at regular time intervals, and when requested by the application, the data are transmitted to the mobile cellphone and erased from the local storage. Each one of these sensing units is called a *Pollution-Spot* and its basic hardware can be seen in Figure 1b.

Bluetooth 2.0 is the communication interface between the Pollution Spots and user's smartphones. The reason for choosing Bluetooth over Wi-Fi is power consumption in the standalone device, which rely only on their autonomous power supply, thus making Wi-Fi a burdensome choice, energy wise. Since there's no direct link of communication between the Data Collection Layer and Data Network Layer, the system depends on mobile users, as the Data Analysis Layer, to get the gathered data from the units using Bluetooth, and forward it to higher levels of the scheme, through the wireless data network.

- **Data Analysis Layer:** Mobile phones compose the second layer of the system. In hands of the participants, they act as a point of gathering, analysis, and forwarding of the acquired data. Since the mobile devices transmit the data using the existing wireless data networks, the scheme has no extra costs in terms of data transport. The mobile application (developed for Android devices) communicates with the Data Storage and Feedback Layer, and the Data Storage, Estimation and Visualization Layer. When the application server in the Data Storage, Estimation and Visualization Layer broadcasts the locations where data are needed, the application receives this request for measurements, and informs the user of the nearest *Pollution-Spot*. If the user agrees to perform the measurements (accepts the task), the application will receive the data from the sensing device (located in the coordinates specified by the request), as soon as the device is in range. The application attaches a GPS-based location stamp to the environmental data and forwards it to the Data Storage and Feedback Layer.
- **Data Network Layer:** The third layer of the *Pollution-Spots* system is the wireless network infrastructure based on the Internet Provider technology already deployed (mobile data carriers, Wi-Fi, etc.). By means of this layer, the users forward the sensed data to the higher layers of the architecture. The user is allowed to decide whether to use the data cellular network or Wi-Fi, and also when to send the data to the centralized server.
- **Data Storage and Feedback Layer:** This layer is composed by the Proxy Server and acts as an intermediary between the Data Storage, Estimation and Visualization Layer and the mobile phones in the Data Analysis Layer. This mediation is necessary to guarantee an extra layer of protection for the participants' private information [23, 26], which will be discussed later on. The server also manages a database (based on MySQL) with the data from the participants. The user data quantify the participation of each participant, which can be used to provide incentive mechanisms to maintain the required amount of users in the system. The Proxy Server forwards only the environmental data with the time-location stamp to the Pollution Data Server, disengaging user data from environmental data.
- **Data Storage, Estimation and Visualization Layer:** The Pollution Data Server, located in the Data Storage, Estimation and Visualization Layer, is the core of the *Pollution-Spots* network, and it implements environmental data storage and management. The management of the pollution data includes algorithms for data interpolation, privacy protection, incentives, as well as verification of the measured data quality. These algorithms will be discussed in the following section.

4 COMBINED ALGORITHM FOR INCENTIVES AND PRIVACY PROTECTION

When implementing a PS-based system, it is necessary to address the inherent challenges of this type of sensing paradigm. First, the system depends on a public wireless network, hence, user location data could be compromised. This problem rises the question: how to implement techniques to protect user privacy? Another obstacle is the reduced budget for encouraging the crucial user collaboration, since centralizing the collected data relies on it. Considering this, which mechanisms are adequate for an air pollution monitoring PS system that has no funds to compensate user participation?

In order to address these questions, *Pollution-Spots* implements a combined algorithm that protects user privacy and, simultaneously, recognizes his participation. To achieve this, it is necessary to isolate environmental data provided by the user from his identity. This is the reason of the existence of the Data Storage and Feedback Layer. The proposed algorithm is based on the Hybrid Privacy Mechanism presented in [23], and commonly practiced gamification techniques to motivate user collaboration [27, 28]. The privacy mechanisms are discussed first, and later the incentives mechanisms are presented.

4.1 Hybrid Privacy Mechanism

The proposed Hybrid Privacy Mechanism is described by Figure 2. The input data for each iteration  $R_i$  are: the surface data obtained by interpolating the measurements ( $M_{R_i}$ ), the initial number of clusters ( $k$ ), the maximum number of cell divisions ( $n$ ) and the minimum size of the cell ( $S_{min}$ ) [23].

The first stage of the algorithm calculates the center for the  $k$ -clusters on the surface  $M_{R_i}$ . Using each center  $C_{xy}$ , the algorithm computes their corresponding Voronoi space  $V_i$  and the gradient  $g_i$  of the variable represented by the surface  $M_{R_i}$  within the space  $V_i$ . The interval  $I$  between the minimum  $g_{min}$  and maximum  $g_{max}$  gradients is quantified in  $n$  parts and classified according to each  $g_i$  into an interval  $I_j$ , and the area is divided in  $j$  centers  $P_{xy}$ . The regions with a low gradient (null) are regrouped to generate larger cells with center  $Q_{xy}$ . Depending on the size of each new cell ( $P_{xy}$  or  $Q_{xy}$ ), the algorithm decides to encrypt or to anonymize

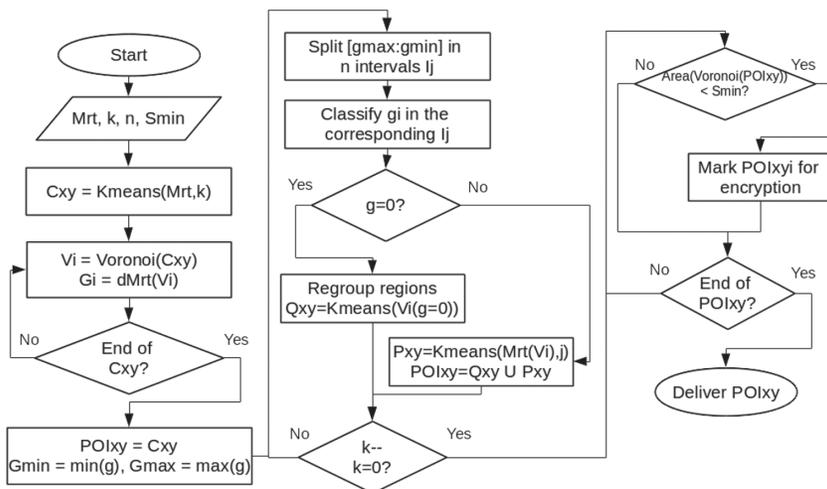


Figure 2: Decision algorithm for the Hybrid Privacy Mechanism at each iteration  $R_i$ .

(Points-of-Interest - *POI*) to report the locations in that area. Finally, the algorithm returns the *POI* of iteration  $R_{t+1}$ , which are defined by  $P_{xy} \cup Q_{xy}$ .

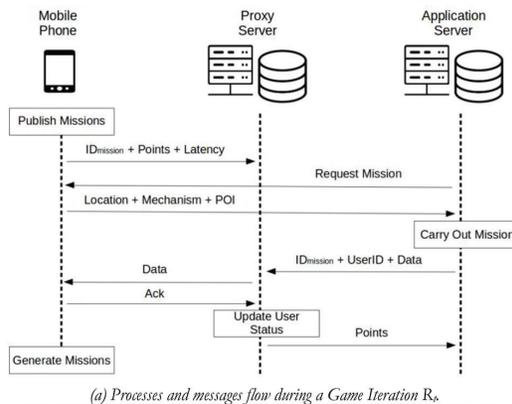
Combining the privacy protection techniques with the incentives enables to reward the user's participation without compromising the critical data. We use the labels  $ID_{mission}$  in order to identify the information that has been published in the network at each iteration  $R_t$  for a mission (a task to collect data from a *Pollution-Spot*).

#### 4.2 Game Iteration $R_t$

As previously mentioned, the database located in the Data Storage and Feedback Layer holds the information that quantifies the user participation in the system. This information is essential when applying the incentives mechanism, because it is used to assign rewards in form of points, badges and/or recognition. These techniques are a subset of Game Mechanics.

Each iteration of the system is defined as an interval called Game Iteration  $R_t$ . This interval defines the time between the missions associated to a specific  $R_t$ , the time involved during the execution of a mission by the user, the collection of information and the data processing according to the next Game Iteration ( $R_{t+1}$ ). Figure 3a details the flow of processes and messages between the applications running in the cellphone, the Proxy Server and the application server, during a Game Iteration  $R_t$  (assuming the user has already installed the *P-Spots* App on the smartphone, as depicted in Figure 3b [left]).

The iteration  $R_t$  starts when the application server publishes new missions to the Proxy Server. The data associated to the missions which are required by the Proxy Server are: the



(b) Different screens of the *P-Spots* App. On the left, the welcome and register screen; on the center, the main menu allows the user to access the rules statement and the missions; on the right, the user can select a mission based on its location in the map.

Figure 3: Details of how the application and the system interact with the user.

$ID_{mission}$  label (calculated through a hash function), the points corresponding to each mission and the latency (mission valid time). After the application requests a mission (GUI depicted in Figure 3b [center]), the server replies with the missions generated for the iteration  $R_i$ . The application needs the following mission's data: latitude and longitude, mechanism (whether or not to encrypt data) and the *POI* generated by the algorithm. Once the application receives the aforementioned data, it displays the closest three missions to the user on a map, including the user's position and the missions (blue dot marker and red flags depicted on Figure 3b [right]). The user can now go to the closest *Pollution-Spot* and the application can download the pollution data (connecting via Bluetooth, as explained before), process the information and send it using the selected privacy mechanism.

## 5 PERFORMANCE EVALUATION OF THE COMBINED ALGORITHM

This section presents the performance evaluation of the combined algorithm. The results and analysis of the privacy and incentive mechanisms, and power consumption are presented next.

We carried out a quantitative analysis for the main privacy protection mechanisms introduced herein. We used Matlab for the simulation and a region of interest with a surface of 100 *units*<sup>2</sup>. The average quantity of people in the area was set to 60, taking into account random initial positions and random people mobility along the surface. In order to model the mobility of the people, we used a function to generate random paths. Those paths only modify the probability of the direction that one person takes while moving in the area, assuming an average speed of 1.5 *m/s* [29], and the Euclidian distance to calculate the position of the people. To compare the privacy mechanisms, we used the same algorithms proposed in [23] for the best Quality of Information: Tessellation, *POI*, Random Perturbation and the Hybrid Mechanism (*POI* and Encryption).

The parameters and assumptions made to simulate each mechanism were:

- **Tessellation:** The cells were divided randomly, using the location of the three pedestrians closest to the center of the cell. Also, at first, the centers of the cells are regularly distributed along the area of interest.
- **POI:** The area of interest is divided in a regular grid of 5×5 cells, the *POI* are the centers of the cells.
- **Random Perturbation:** To add random noise to the data we used a normal distribution with mean  $\mu=0$ , and variance  $\sigma^2=5$  units.
- **Hybrid Mechanism:** To simulate the scenario where combined *POI* and encryption mechanism are used, we chose an initial cluster number of  $k_0 = 25$  and size minimum size of the cell  $S_{min} = 200$ .

The initial number of cells and clusters were chosen using try and failure method, selecting the most cost effective (energy and performance wise), using the metrics that we'll discuss in next section.  $S_{min} = 200$  was chosen using the same metrics.

### 5.1 Privacy Mechanisms

In order to quantify the performance of the privacy mechanisms, first the different proposed metrics are presented, and later the results are analyzed.

### 5.1.1 Quality of Information

The determination coefficient ( $R^2$ ) is used in order to assess the quality of the information of each mechanism. The coefficient measures the difference between the real observations compared to those affected by the privacy mechanism [22] and it is defined as shown in eqn (1):

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \tag{1}$$

where  $y_i$  are the surface data obtained by interpolating the measurements without being affected with any protection mechanism,  $f_i$  are the surface data obtained by interpolating those measurements modified by the privacy mechanism and  $\bar{y}_i$  is the mean value of the observations.

### 5.1.2 Probability to Locate the User

To evaluate the behavior of each mechanism regarding the privacy protection issue, the *Probability to locate the user* ( $P_{Loc}$ ) is used, (see eqn [2]).  $P_{Loc}$  is the probability to know the user's location within the area  $S_A$  given an overall area  $S_T$  (for each iteration  $R_i$ ) [22].

$$P_{Loc} (M_{R_i}, D_i) = 1 - \frac{\bar{S}_A}{S_T} \tag{2}$$

where  $M_{R_i}$  is the surface data obtained by interpolating the observations during the iteration  $R_i$ ,  $D_i$  is the data package reported by the user,  $S_T$  is the overall data surface area  $M_{R_i}$ , and  $\bar{S}_A$  is the average area that enables locating the user. The location is estimated by using eqn (3):

$$\bar{S}_A = \sum_{j=1}^k (S_{A_j} * P_{A_j}) \tag{3}$$

where  $S_{A_j}$  corresponds to the areas for estimating the probability of the user's location  $P_{A_j}$ . Each  $S_{A_j}$  is estimated by using a percentile  $Perc_j$ , the map of values  $M_{R_i}$  and the package  $D_i$ .

The estimation function returns a set of areas  $S_{A_j}$ , with a probability  $P_{A_j}$  of finding the user at the iteration  $R_i$ . First, the algorithm calculates the distribution of the differences between the data surface  $M_{R_i}$  and the report  $D_i$ . Second, the algorithm selects the areas  $S_{A_j}$  according with a set of percentiles that ensure the proper quality during the estimation. For the application, percentiles 5%, 10%, 15%, 20% and 25% were used. Two levels of privacy management are considered: (i) Proxy Server which has access to the user data and the obfuscated observations by the *POI* algorithm; and (ii) Environmental Data Server which has access to both encrypted and obfuscated data, but no to the user data.

We propose the assessment of the probability of locating the user in three possible scenarios: The adversary has access to the Proxy Server ( $P_{Loc,S,P_r}$ ); the adversary has access to the Environmental Data Server ( $P_{Loc,S,P_d}$ ); or the adversary has access to both servers ( $P_{Loc,S,P_r,P_d}$ ).

### 5.1.3 Results and Analysis

Taking into account the conditions introduced before, we now present the results and corresponding analysis for the different scenarios.

5.1.3.1 Quality of Information

Figure 4 shows the performance of each mechanism according to the Quality of Information. One can note that the Hybrid Mechanism presents accurate results whereas Tessellation presents a lower performance. The low performance for the latter is due to the random mobility of the users in the region of interest, causing irregular cell divisions, which consequently causes that the center of certain cells is far away from the real positions.

5.1.3.2 Probability of locating the user

The general results of evaluation of the probability of locating the participant under different scenarios are shown in Table 1.

When the adversary has access to the Proxy Server  $P_{Loc,S,P_r}$ , the best results are achieved by using the Hybrid Mechanism, for which the system has access to the fraction of the observations that use the data disturbance given by *POI* as the privacy mechanism. The probability of locating the user is 0.9882, however, since only 89.60% of the users make use of *POI* for reporting the observations, the resultant probability is  $P_{Loc,S,P_r} = 0.8855$ .

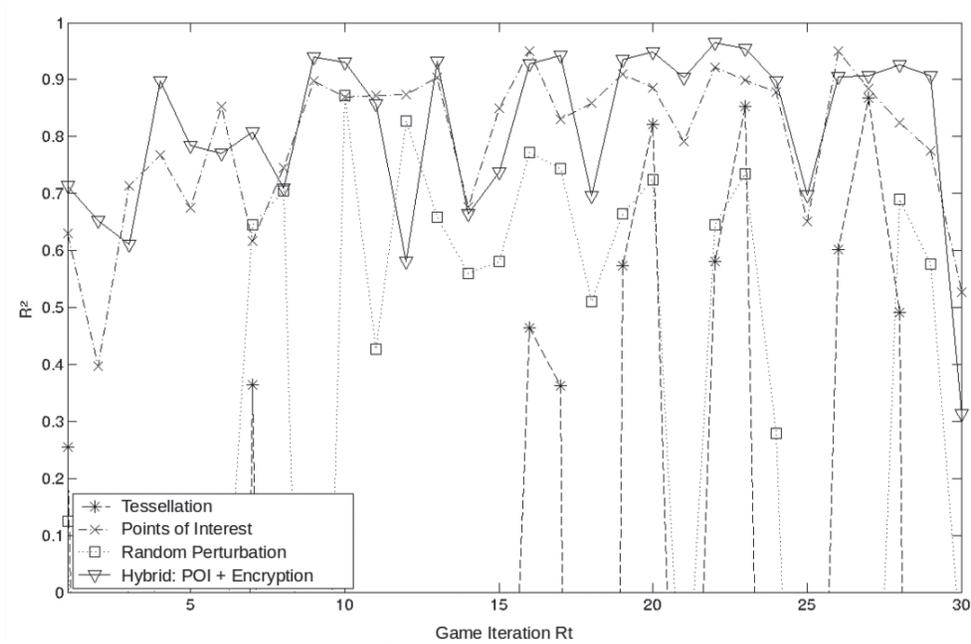


Figure 4: Coefficient of determination ( $R^2$ ) for every privacy mechanism.

Table 1: Probability to locate a participant.

Mechanism	$P_{Loc,S,P_r}$	$P_{Loc,S,P_d}$	$P_{Loc,S,P_r,P_d}$
POI	0.983	0.400	0.983
Random perturbation	0.989	0.989	0.989
Tessellation	1.000	0.317	1.000
Hybrid	0.885	0.988	0.989

When the adversary has access to the Environmental Data Server  $P_{Loc.S,P_d}$ , the best performance is achieved by the Tessellation case, for which the adversary has access to the obfuscated data but not to the real locations, and since  $k=3$  and we have 60 participants, the probability to locate the user is as low as  $P_{Loc.S,P_d} = 0.317$ .

Finally, when the adversary has access to both servers  $P_{Loc.S,P_r,P_d}$ , the best performance is achieved by the *POI* mechanism, having the same probability obtained in the first scenario. However, the Hybrid Mechanism has a very similar performance.

### 5.2 Incentive Mechanisms

In order to analyze the performance for the incentive mechanisms, we developed a testbed protocol in Matlab, as shown in Figure 5. It is worth noting that the simulation of pedestrians within this PS system is scalable, therefore, adding new parameters to the simulation requires just including the new variable to the sum and recalculate the weights for each module.

- **Ramdon Path:** This module calculates the pedestrian’s  $X$  and  $Y$  displacements by using the weights assigned to each possible change of direction of the participant.
- **Path with Target:** This module calculates the pedestrian’s  $X$  and  $Y$  displacements by considering the direct path that takes the user to the target destination.
- **Most Probable Path:** This module simulates the most probable paths for the user’s displacement on the surface. The system assigns more weight to the most probable transited paths instead to those paths less transited (i.e. it is probable that the users transit by the internal streets of the university campus rather than crossing grass paths).
- **Path with Incentives:** This is the most relevant module since it allows to verify the effectiveness of applying the incentive mechanisms to the PS system. It works by the same principle of the ‘Path with Target’ module, however in this case, the target position is the nearest *Pollution-Spot*. In addition, the *Pollution-Spot* should not drastically change the

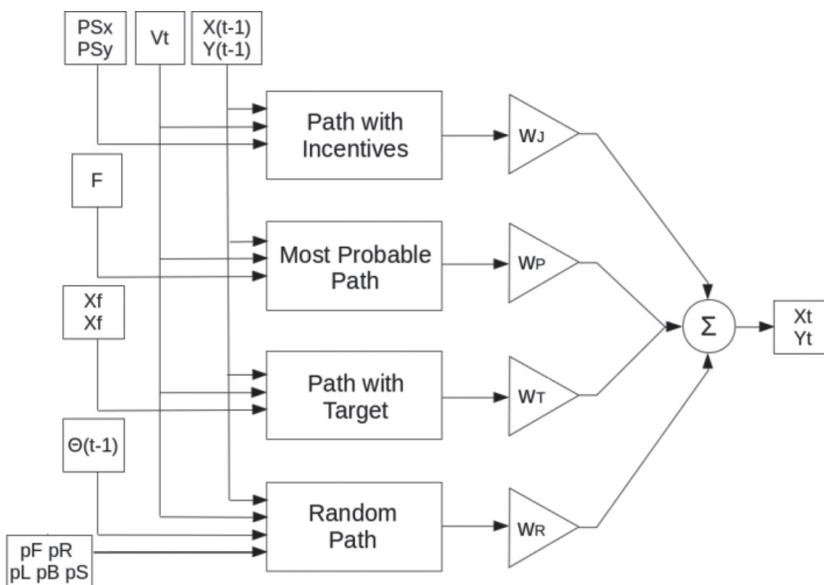


Figure 5: Testbed structure to simulate different incentive mechanisms.

normal direction of the user. Equation (4) describes the behavior of the participants due to the incentive mechanisms:

$$dX_{It} = V_t * \cos(\theta_{It}) * w_{Ji} \quad dY_{It} = V_t * \sin(\theta_{It}) * w_{Ji} \quad (4)$$

where  $w_{Ji}$  is the instantaneous step that the incentive mechanism contributes to the participant, which depends on three parameters: current participant's score  $S_t$ , the environmental coefficient  $G_t$  (how likely the participant will change his path to collect pollution data from a *Pollution Spot*) and the score received from measuring from that particular spot  $w_{PS}$ .

The output of each module is multiplied by the weight that this type of walk has in the participants behavior, allowing us to model different scenarios. The final value of the weights are as follows: Random Path  $w_R = 0.01$ , Path with Target  $w_T = 0.30$ , Most Probable Path  $w_P = 0.34$ , and Path with Incentives  $w_J = 0.35$ . Two different experiments were carried out to measure the contribution of the incentives mechanisms to the quality of information of the collected data.

### 5.2.1 Changing the Number of Participants N

These experiments measured the Quality of Information through the coefficient of determination ( $R^2$ ) when the number of participants is modified. With these simulations, it is possible to detect the minimum number of participants necessary to maintain a certain level of quality on the collected data. The coefficients used were  $G_t = 0.8$  and  $w_J = 0.35$ . Figure 6a presents the results of this experiment.

### 5.2.2 Variation of the Environmental Coefficient $G_t$

These tests were carried out to simulate the behavior of the participants depending on how willingly they would change their current path to collect data from a nearby *Pollution-Spot* (variation of  $G_t$ ). For these experiments  $w_J = 0.35$  for the incentives mechanisms and  $N=20$ . Figure 6b presents the results of this experiment.

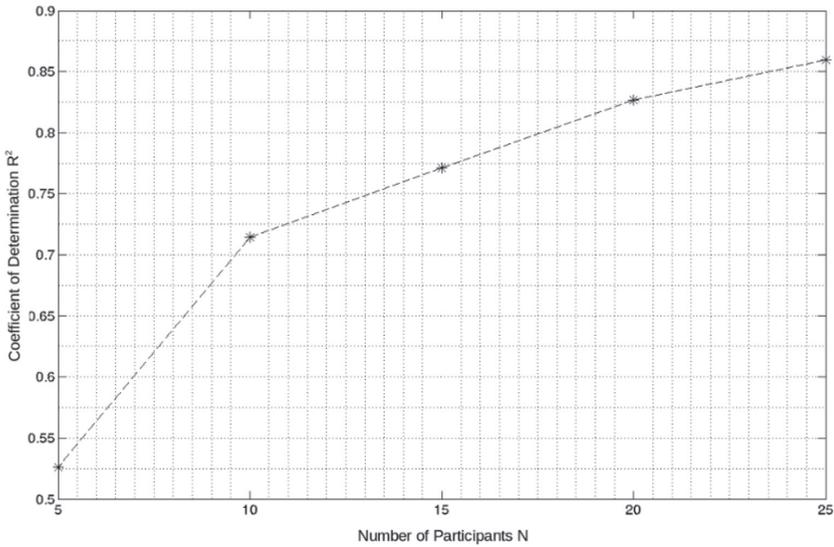
From the reported results, it is clear that the number of participants in the system affects the quality of information in a greater way, than the behavior of these participants once they have joined the game and installed the application. This result proves again how important it is to encourage the participation of more mobile users in order to maintain the system in a stable state.

## 5.3 Energy Consumption

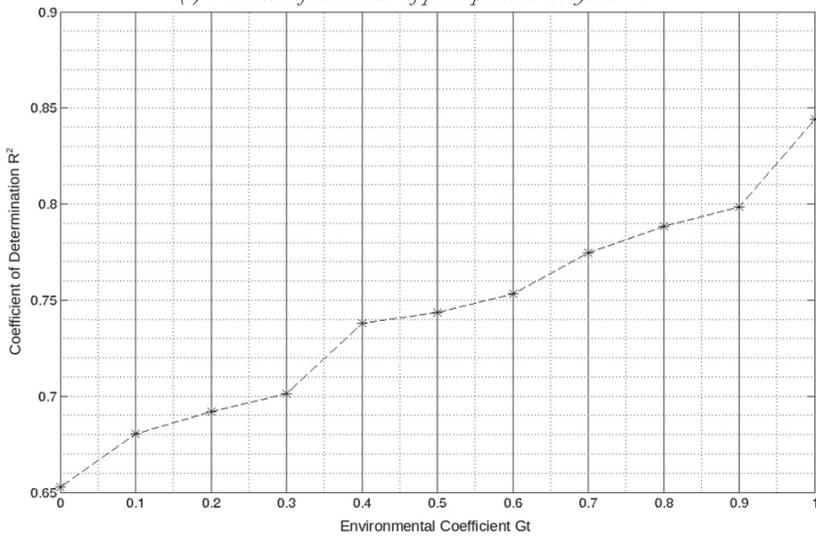
For each privacy mechanism, we carried out 20 tests for processing and transmitting the corresponding user profile registry. We measured three processes:

- Plain-text registry transmission using *TCP/IP sockets* for Tessellation and Random Perturbation.
- Registry encryption using an AES symmetrical key of 128 bits in size for double encryption schemes.
- Registry transmission via SSL connection for encryption techniques.

It was assumed that the registry of 1MB in size was already stored with the corresponding structure for each privacy protection technique. The testbed was composed by a Samsung



(a) Variation of the number of participants in the system N.



(b) Variation of the environmental coefficient  $G_t$ .

Figure 6: Contribution of the incentive mechanisms in the quality of information.

S3-mini phone (GT-I9195L) running Android 4.4.2 and the university campus Wi-Fi as the default communication channel. In order to calculate the energy used by each privacy protection mechanism, the phone was connected to a Tektronix PS280 power supply ( $V_{DC} \approx 4.2V$ ) and we measured the electrical current  $A_i$  consumed at time intervals of  $\delta_i = 0.25s$  [23]:

$$E = \sum_i (V_{DC} * A_i * \delta_i)$$

The energy consumption on the handset for each of the mechanisms can be summarized as follows: for the *POI*, Random Perturbation and Tessellation mechanisms, each consumes  $1.914J$ ; the Double Encryption mechanism consumes  $8.886J$ ; and the proposed Hybrid

mechanism consumes 8.161J. Experiments were carried out by taking into account that all registries are stored within the phone's memory, therefore, the energy consumption is similar for those techniques that only use TCP sockets for sending messages. The methods that use encryption present higher levels of energy consumption, since these methods use AES and SSL every time that the data are sent to the server. As a result from the experiments, the hybrid mechanism achieves a better performance by encrypting just one fraction of the messages while still protecting the user's private information.

## 6 CONCLUSIONS

*Pollution-Spots* is an air monitoring novel scheme that combines an infrastructure of static monitoring stations and the participation of users to collect the information. A custom hardware was designed, reducing the total cost per sensing unit and allowing us to scale the network to many spots, possibly covering large areas with a better spatial resolution when compared to traditional systems.

The proposed system not only collects pollution data, but also encourages the user participation, protecting his private information while still maintaining a good quality of the collected information. The system also has to consider a low power consumption and low computational load in order to reduce the effect of the implemented mechanisms into the user's mobile phone. With this in mind, the hybrid mechanism that decides whether to encrypt or anonymize the user's information achieves a better tradeoff between quality of information and low power consumption.

Game theory has proven to be an interesting mechanism to encourage the participation of mobile users when there is no monetary budget to compensate them. This result proves that the privacy protection and incentives mechanisms have to be carried out at the same time, for which a combined algorithm was proposed and developed, maintaining the benefits of the hybrid mechanism and at the same time implementing a 'green game' to encourage the mobile users to participate.

Our current and future work focuses on the implementation of *Pollution-Spots* as part of a Smart Campus project. One of the most challenging tasks involves the calibration and deployment of these sensors in order to validate the acquired readings when compared to the values reported by the official environmental monitoring system of the city. Another challenge would be considering and including agent based techniques as mentioned in Section 2.

## ACKNOWLEDGMENT

The results presented here are partial results of the **Diseño de una Red de Monitoreo de la Calidad del Aire** project, funded by the Pontificia Universidad Javeriana (Project ID: 00005835), through the internal R&D fund.

## REFERENCES

- [1] Mendez, D., Perez, A., Labrador, M. & Marron, J., *P-sense: A participatory sensing system for air pollution monitoring and control*. PerCom, WiP workshop, pp. 344–347, 2011.
- [2] Guo, B., Yu, Z., Zhou, X. & Zhang, D., *From participatory sensing to mobile crowd sensing*. 2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS), pp. 593–598, 2014.

- [3] Rodriguez, L. & Mendez, D., *Pollution spots: A novel method for air pollution monitoring*. WIT Transactions on Ecology and the Environment (Air Pollution XXIII), WIT Press, London, 2015.
- [4] Mendez, D., Diaz, S. & Kraemer, R., *Wireless technologies for pollution monitoring in large cities and rural areas*. 2016 24th Telecommunications Forum (TELFOR), pp. 1–6, 2016.
- [5] Hachem, S., Mallet, V., Ventura, R., Pathak, A., Issarny, V., Raverdy, P.G. & Bhatia, R., *Monitoring noise pollution using the urban civics middleware*. 2015 IEEE First International Conference on Big Data Computing Service and Applications, pp. 52–61, 2015.
- [6] Hasenfratz, D., Saukh, O., Sturzenegger, S. & Thiele, L., *Participatory air pollution monitoring using smartphones*. *Mobile Sensing*, **1**, pp. 1–5, 2012.
- [7] Dutta, P., Aoki, P.M., Kumar, N., Mainwaring, A., Myers, C., Willett, W. & Woodruff, A., *Demo: Common sense: Participatory urban sensing using a network of handheld air quality monitors*. Proceedings of the Conference on Embedded Networked Sensor Systems, Berkeley, CA, 2009.
- [8] Carrapetta, J., Youdale, N. & Chow, A., *Haze watch*, 2010.
- [9] Hu, K., Sivaraman, V., Luxan, B.G. & Rahman, A., *Design and evaluation of a metropolitan air pollution sensing system*. *IEEE Sensors Journal*, **16**(5), pp. 1448–1459, 2016.  
<https://doi.org/10.1109/jsen.2015.2499308>
- [10] Ourcast, *Minutely.the minute-by-minute rain forecast at your exact location*, 2013.
- [11] Balestrini, M., Ladera, T.D., Polvora, A. & Nascimiento, S., *Mapping Participatory Sensing and Community-led Environmental Monitoring initiatives*. Making Sense: European Commission, 2016.
- [12] Koutsopoulos, I., *Optimal incentive-driven design of participatory sensing systems*. *INFOCOM, 2013 Proceedings IEEE*, pp. 1402–1410, 2013.
- [13] Lee, J.S. & Hoh, B., *Sell your experiences: a market mechanism based incentive for participatory sensing*. PerCom. IEEE, New York, pp. 60–68, 2010.
- [14] Wang, H., Zhao, Y., Li, Y., Zhang, K., Thepvilojanapong, N. & Tobe, Y., *An optimized directional distribution strategy of the incentive mechanism in sense util-based participatory sensing environment*. *Mobile Ad-hoc and Sensor Networks*, pp. 67–71, 2013.
- [15] Jaimes, L., Vergara-Laurens, I. & Labrador, M., *A location-based incentive mechanism for participatory sensing systems with budget constraints*. *Pervasive Computing and Communications (PerCom)*, 2012 IEEE International Conference, pp. 103–108, 2012.
- [16] Zhang, J., Ma, J., Wang, W. & Liu, Y., *A novel privacy protection scheme for participatory sensing with incentives*. *CCIS*, **3**, pp. 1017–21, 2012.
- [17] Ueyama, Y., Tamai, M., Arakawa, Y. & Yasumoto, K., *Gamification-based incentive mechanism for participatory sensing*. PERCOM Workshops, pp. 98–103, 2014.
- [18] Zhang, Y. & van der Schaar, M., *Reputation-based incentive protocols in crowdsourcing applications*. *INFOCOM, 2012 Proceedings IEEE*, IEEE, pp. 2140–2148, 2012.
- [19] Burke, J.A., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S. & Srivastava, M.B., *Participatory sensing*. Center for Embedded Network Sensing, 2006.
- [20] Anawar, S. & Yahya, S., *Empowering health behaviour intervention through computational approach for intrinsic incentives in participatory sensing application*. *ICRIIS*, pp. 281–5, 2013.
- [21] Machanavajjhala, A., Kifer, D., Gehrke, J. & Venkitasubramaniam, M., *L-diversity: Privacy beyond k-anonymity*. IN ICDE, 2006.

- [22] Vergara-Laurens, I. & Labrador, M., *Preserving privacy while reducing power consumption and information loss in LBS and participatory sensing applications*. GLOBECOM Workshops, pp. 1247–1252, 2011.
- [23] Vergara-Laurens, I., Mendez, D. & Labrador, M., Privacy, quality of information, and energy consumption in participatory sensing systems. *PerCom*, pp. 199–207, 2014.
- [24] Zenonos, A., Stein, S. & Jennings, N., *Coordinating measurements for air pollution monitoring in participatory sensing settings*. Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems, pp. 493–501, 2015.
- [25] Chen, C., Cheng, S., Gunawan, A., Misra, A., Dasgupta, K. & Chander, D., *Traccs: Trajectory-aware coordinated urban crowd-sourcing*. Proceedings of the Second AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2014), pp. 30–40, 2014.
- [26] Vergara-Laurens, I., Mendez-Chaves, D. & Labrador, M., *On the interactions between privacy-preserving, incentive, and inference mechanisms in participatory sensing systems*. Network and System Security, Springer Berlin Heidelberg, volume 7873 of LNCS, pp. 614–620, 2013.
- [27] Han, K., Graham, E., Vassallo, D. & Estrin, D., Enhancing motivation in a mobile participatory sensing project through gaming. *Privacy, Security, Risk and Trust*, pp. 1443–1448, 2011.
- [28] Ueyama, Y., Tamai, M., Arakawa, Y. & Yasumoto, K., *Gamification-based incentive mechanism for participatory sensing*. PERCOM Workshops, pp. 98–103, 2014.
- [29] Richard, L., Knoblauch, M.T.P. & Nitzburg, M., *Study compares older and younger pedestrian walking speeds*, 1997.