Pollution spots: a novel method for air pollution monitoring

L. Rodríguez & D. Mendez
Department of Electronic Engineering, Pontificia Universidad Javeriana, Bogota, Colombia

Abstract

The Participatory Sensing paradigm offers the advantages of leveraging the existing communication networks and mobile phones (used as sensing devices), thus reducing the costs of deploying a sensing network. Despite the processing capabilities and the already embedded sensors in smart phones, integrating pollution sensors into cellphones is still an ambitious project. Pollution-Spots proposes a solution by means of using an infrastructure of fixed low-cost sensing devices, and reporting the measurements using Participatory Sensing. The system's name comes from the static sensing units which function as spots the mobile user can hook-up to in order to obtain the environmental data, generating a model that easily escalates, since it does not require users carrying the sensing devices. If the system's administrator requires to enlarge the region of interest, only the installation of an additional Pollution-Spot is required. The device starts measuring and the pedestrian forwards the information, completing the cycle with no extra cost of data transport and/or human resources. Preliminary experiments have been conducted using the platform located in Pontificia Universidad Javeriana's campus. The maps obtained based on the gathered data provide a better spatial resolution of the variables of interest when compared to traditional monitoring stations, proving the effectiveness of the system. These results show the benefits of applying this model in areas with scarce monitoring stations and/or budget limitations. Also, including user participation makes it possible to design applications that improve communities' quality of life, since the maps obtained allow the identification of the areas where the concentration of pollutants might be critical.
Keywords: air pollution, participatory sensing, environmental data visualization.

doi:10.2495/AIR150111

1 Introduction

Participatory Sensing (PS) is a sensing paradigm that takes advantages of the advances and ubiquity of mobile technology, as well as the cellular data network infrastructure already deployed, to create large scale sensing schemes [1]. The PS based schemes cover different data types, mainly information about the environment, weather and traffic congestion, by means of groups of cellular users, that contribute with the measuring and the retrieving of information. The collected data are centralized and are 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, atmospheric particulate matter (PM), and the concentration of different gases) can be used to visualize their behavior over the time-space region of interest, thus helping communities to increase the awareness about pollution, and consequently improving their quality of life.

One concern related to the usage of PS schemes to measure air quality, is the fact that despite the processing capabilities and the embedded sensors in smart phones, these smart phones still do not include the required sensors to acquire the variables of interest, pollutants in this case. Since attaching these sensors to the mobile devices is a burdensome project, it is necessary to come up with a scheme that overcomes this obstacle and, at the same time, exploits the benefits of PS, i.e., massive coverage of a large number of phenomena implied by the mobility of phone carriers, the inclusion of people in the sensing loop, and no costs in terms of data transport. Considering this context, in this paper we propose *Pollution-Spots*, a novel scheme for measuring environmental data, that proposes a solution, by means of an infrastructure of static low-cost air quality sensing devices, inserted into a PS-based scheme, combining user participation 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 escalate.

The name of the system comes from the sensing stations, which are static and therefore can be seen as spots that the user can hook up to in order to obtain the sensing data. In this model, it is easy to enlarge the region of interest, since it only requires the installation of a new Pollution-Spot to acquire the environmental data of that extended region. The data are collected by the pedestrian, who forwards the information using the already deployed wireless data networks (cellular carriers, Wi-Fi, etc.), centralizing this information so that it can be processed and made accessible through a web server.

The rest of this paper is organized as follows: we discuss the related work in Section 2. In Section 3 we detail the proposed scheme, depicting the system's architecture and main functions of the nodes within it. Section 4 describes the challenges associated with the deployment of a PS based scheme and the solutions proposed by the Pollution-Spots model. Section 5 refers to the preliminary results obtained in a case study located in the campus of the Pontificia Universidad Javeriana. Finally, we conclude the paper in Section 6, where we also briefly discuss the guidelines to continue our work.

2 Related work

There are different examples in literature of applications that sense environmental data exploiting the characteristics of PS schemes. 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. 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. This approach also has the disadvantages of requiring multiple sensing tasks (at different time-location points) from the user, a factor that also reduces the number of users willing to collaborate with the system, no matter the involved incentives. Such is the case with Haze Watch [2], which 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.

A similar system, Common Sense [3], has developed hand-held monitors with air quality sensors as a prototype, in order to resolve the absence of the sensors in consumer devices. The monitors can communicate with the database through Bluetooth, 802.15.4, or GPRS radios, depending on either user interest or data fidelity. This project also addresses the challenge of making the monitors as comfortable as possible to collaborators in order to facilitate the measurement task. Another example is P-Sense [1], 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.

Concerning the transport of environmental data to the final user, these delivery techniques commonly use a web server that displays the variables of interest through maps, charts, tables, etc. The work in [4] proposes not only the visualization of raw sensor data, but also contour maps. In this project the sensors are not distributed in a regular fashion, raising the need of clustering the sensors when displaying them in maps, due to the differences in sensor resolution between areas; this is not a concern with our project, since we utilize spatial interpolation techniques that do not require regularly located sensors.

The project in [5] proposes the use of marker maps to display information in specific time-space locations and heat maps with the available measures in a region. They also display tables with information as user request them, by means of exploiting cloud storage services such as Google Fusion Tables for processing and delivering the data. Another interesting idea can be found in [6], where the authors use hierarchical and time-series data to display and compare the corresponding data of each layer. Applying this method with air pollution data helps to get a better perception of the behavior of the variables of interest and discriminate the source of the pollution.

The project Minutely [7], 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. Minutely 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.

3 Proposed scheme

Figure 1 depicts the proposed scheme for Pollution-Spots system's architecture. The model is composed by five layers (from lower to higher level): the Data Collection Layer, the Data Analysis Layer, the Data Network Layer, the Data Storage and Feedback Layer, and the Data Storage, Estimation and Visualization Layer.

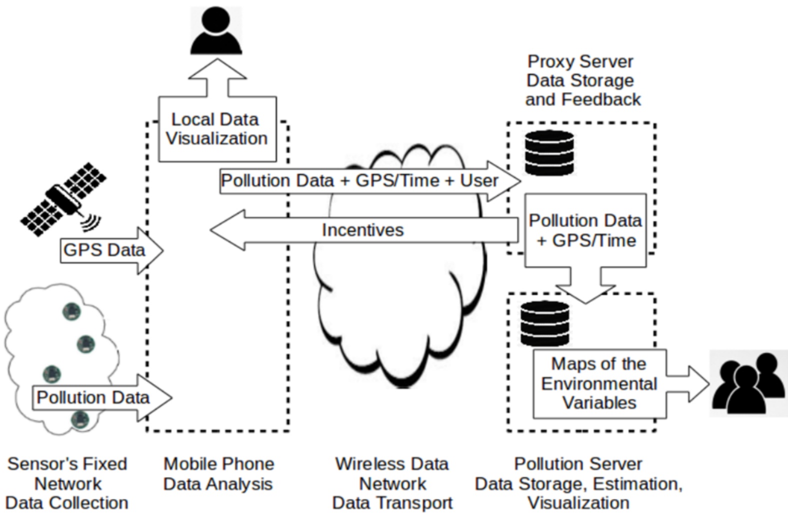

Figure 1: Pollution-Spots system's architecture.

3.1 Data Collection Layer

The first layer is the Data Collection Layer, which is composed of a network of static sensors. Each sensing device has a set of environmental sensors integrated in a board (based on an AVR Atmel processor), so they function as a single unit. The sensing features of the units include measurement of ozone (O_3), relative humidity (RH), temperature (T), particulate matter (PM), and carbon monoxide (CO). Table 1 lists the embedded sensors in the sensing units.

The devices must work as standalone units. For this reason they 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. The collected data, with the corresponding time stamps, are packed in a single frame. The frames are stored in the sensing unit until the mobile user requests them through the mobile sensing application in her cellphone. When requested by the application, the frames are transmitted via Bluetooth to the mobile cellphone and erased from the local storage in the sensing unit.

Table 1: Sensors embedded in the sensing units.

Manufacturer – Sensor	Variable
Membrapor – CO/SF-200	CO
Membrapor – O3/S-5	O_3
Sensirion – SHT15	Temperature and Humidity
Sharp – GP2Y1010AU0	Particulate Matter

3.2 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 (mobile data carriers, Wi-Fi, etc.), the scheme has no extra costs in terms of data transport. The mobile application is developed under Android: the popularity of smart phones based on this operating system increases the possibilities of attracting more users to collaborate with the project.

The mobile application communicates with the Data Storage and Feedback Layer, and the Data Storage, Estimation and Visualization Layer. When the user receives a request for measurements, the application connects to the required Pollution-Spot and receives the data from the sensing device (located in the coordinates specified by the request). The application attaches a GPS-based location stamp to the environmental data and forwards it to the Data Storage and Feedback Layer.

3.3 Data Network Layer

The third layer of the Pollution-Spots system is the wireless network infrastructure based on the IP technology (mobile data carriers, Wi-Fi, etc.) already deployed. By means of this layer, the users forward the sensed data to the higher levels 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. Since this model of data transportation is subject to the perils inherent to a public data network, the development of a privacy preserving mechanism to protect the critical user data is required.

3.4 Data Storage and Feedback Layer

The Proxy Server composes the Data Storage and Feedback Layer, 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 [8, 9], 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.

3.5 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 stored data are used to create the maps of the variables of interest. 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 Challenges of PS based systems

Since the proposed scheme is based on a PS system, it is necessary to address the inherent challenges of this type of sensing paradigm. First, since the system depends on a public wireless network, 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 relays on it. Therefore, 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, at the same time, recognizes her participation. To achieve this, it is necessary to isolate environmental data provided by the user from her 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 [8] and commonly practiced gamification techniques to motivate user collaboration [10]. To simplify the discussion about the algorithm, we first discuss the privacy mechanism and later the incentives mechanisms.

4.1 Privacy mechanisms

After the first collection of measurements, the Pollution Data Server in the Data Storage, Estimation and Visualization Layer interpolates the data points to obtain a data surface of each variable of interest. Each surface is processed to divide the region of interest in areas according to the variability of the data (in terms of spatial coordinates). The cell division algorithm is based on Voronoi decomposition, and it is performed on each environmental data collection as follows:

- At first, the region is divided in cells of uniform size.
- After every iteration, the areas where the variability of the data is low are merged to reduce the amount of data required to reconstruct the data surface since the exact location information is not very important.
- On the other hand, the algorithm splits the areas where the variable of interest has a higher variability, and for these regions the information of the exact location is critical.

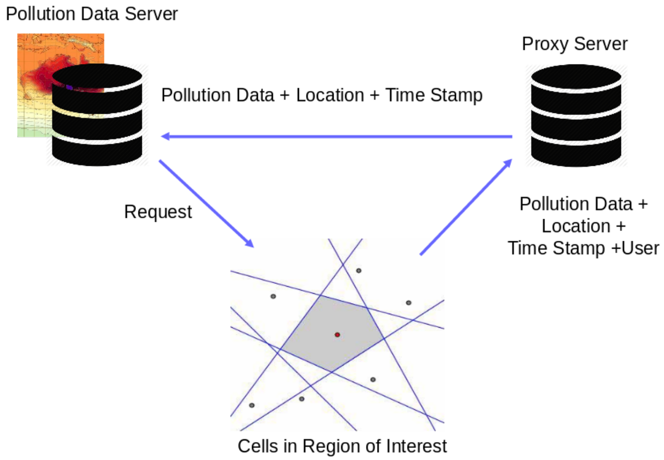

Figure 2: Privacy Mechanism scheme.

With the division of the region of interest in areas of different sizes, the system is able to determine how the information about the data location will be delivered. In large areas (low variability) the location of the data, which is associated to a user, is not so critical and therefore it can be anonymized using points of interest. Spatial coordinates in smaller areas (larger variability), where it is important to know the exact location, must be encrypted to avoid compromising the user's position. This decision of whether to encrypt or anonymize the location data is the main idea of the hybrid privacy scheme proposed in [8] (see Figure 2). Based on a previous cell's division, the Pollution Data Server broadcasts requests to the mobile users, each request specifying the location and time window where the measurements are needed, and the policy to protect the data.

4.2 Incentive mechanism

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 give rewards in form of points, badges and/or recognition. These techniques are a subset of Game Mechanics, the name of the mechanisms used as building blocks for gamifying an application.

There are examples of PS applications that exploit gamification to reward the user's participation [11, 12]. Though the examples illustrate how to use gaming in the context of PS, these works do not address the problem of relating the amount of cooperation supplied from the user and, at the same time, maintain her privacy. The work in [13] is an approach to this issue. The target of this work is applications with no budget constraints, and for this reason, it does not entirely solve the problems addressed here. Also, because of monetary compensations, the rewards are given instantly, so the system does not have memory of user collaboration, as needed to implement gamification.

The proposed solution (depicted roughly in Figure 3) works as follows: When a new measurement task is generated by the Pollution Data Server, the Pollution Data Server assigns a label to it, known as request stamp. This label is calculated through a Hash function, using a key k known to the mobile applications, but not to the Proxy Server. The label is associated to the amount of reward corresponding with the mission, and it is sent only to the Proxy Server. The mobile application receives only the Location-Data information of the request, but can generate the request stamp since it knows the key k.

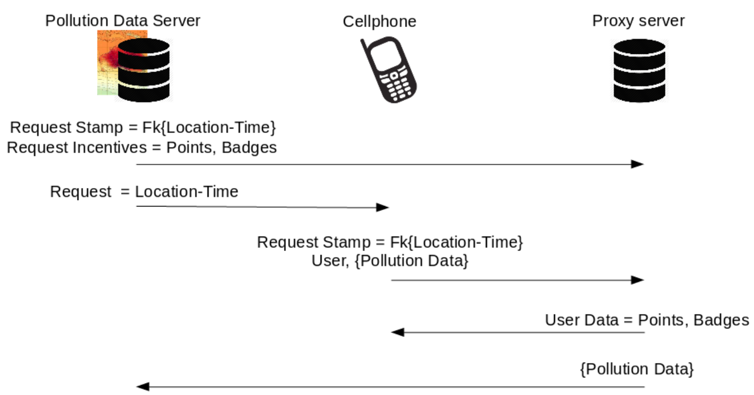

Figure 3: Combined Privacy and Incentive mechanisms in Pollution-Spots.

Once the user performs the measurement task, the application sends the pollution data to the Proxy Server. The application also sends the user ID and the request stamp. If the Proxy Server finds a match between the request stamp sent by a cellphone and the one sent from the Pollution Data Server, it assigns the corresponding reward to the user. The user's category is upgraded in the Proxy Server's database with her new achievement.

The Proxy Server forwards only the environmental data (which do not contain information about the user) to the Pollution Data Server. Since the environmental data (containing the location-time data) have been already treated with the privacy preserving mechanism, it is a hard task for a malicious entity to obtain information about the user while listening to messages between the Data Analysis and the Data Storage and Feedback Layers. Compromising the Pollution Data Server offers no benefits in terms of user data, since the information stored in this server concerns only to the variable of interest.

5 Results and analysis

To verify the performance of the sensors embedded in the modules that will function as Pollution Spots, tests where performed in an area of the city nearby a monitoring station of Bogota's Environmental Agency (SDA – Secretaria Distrital de Ambiente). Due to the lack of specialized equipment and a controlled environment to calibrate the system, it has been a very difficult task to properly validate the level of accuracy of the integrated sensors. However, our experiments show the feasibility of the modules, since the behavior of the measurements is consistent to those provided by the SDA. The differences in the exact values are explained partially by the quality of the sensors in the module and partially by the differences between the location of the measurements.

The data collected from the Pollution-Spots via the smart phones (which receive the measuring tasks that indicate where and when to collect the data), have gaps in time and space, therefore an interpolation technique must be used to create a map to correctly visualize the variable of interest. The spatial data are interpolated using cubic splines, which was selected because cubic interpolation offers a good trade-off between the computational cost and smoothness of the resulting surface. As depicted in Figure 5, the system requires a minimum number of measurements to perform a correct reconstruction of the data surface. The final interpolation generates a 38×38 data points matrix, which can be correctly generated using 50 samples in the area of interest (refer to the right graph in Figure 5). However, with only 10 samples in the area of interest (center graph in Figure 5), the general behavior of the variable can be recognized.

The collected air quality information is available to the users by means of a web server. The main functionality of the web server is the visualization of the data in the region of interest as a partial retribution for the user participation. The air quality data allow the user to get a better perspective of the behavior of the variables in the area of interest.

As presented before, an initial version of Pollution Spots has been deployed in the campus of the Pontificia Universidad Javeriana. Figure 4 shows how the final user can visualize the behavior of the variables of interest in the University campus. The webpage uses color gradient maps and a scale to facilitate the understanding

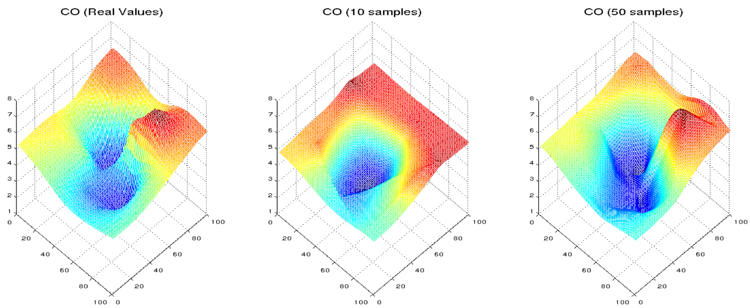

Figure 4: Original data surface for CO (left), reconstruction with 10 samples (center) and with 50 samples (right).

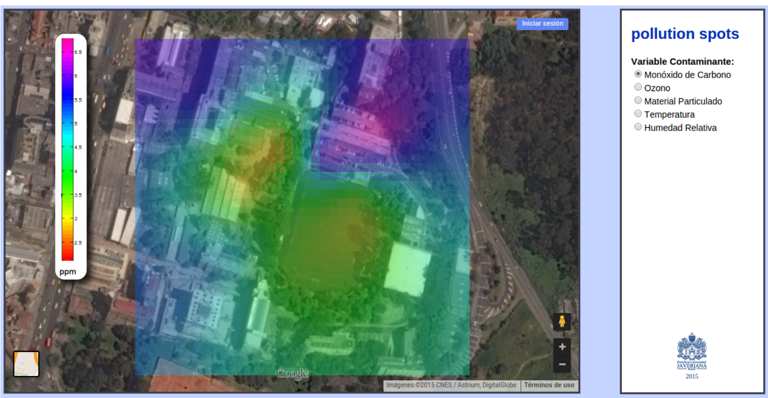

Figure 5: Carbon monoxide concentration (ppm) in the campus of the Pontificia Universidad Javeriana.

of the maps. The figure clearly shows how the lowest concentration levels (red) are obtained in the green areas of the campus, meanwhile the highest levels (magenta) are near the roads and the parking building, which naturally exhibit a greater automobile traffic during the day.

The color data surface is updated with every new iteration (collected data) in the Pollution Data Server, and its visualization represents a very low extra cost in terms of processing. This cost reduction is possible because the algorithm that performs the initial division of the region of interest needs the data surface to work with; taking advantage of this need, a copy of the file with the data representing the variables of interest and their coordinates its made accessible to the web server application, which utilizes it to plot the data surfaces using JavaScript. In the future, the web server will also allow the user to query the database and receive filtered data through tables and charts.

6 Conclusions and future work

Pollution-Spots is a PS-based pollution monitoring system that proposes a cost-effective solution to overcome the absence of sensors to measure air quality in mobile devices, with no extra cost related to data transport. By exploiting the fixed sensor network and user participation, the system requires little user contribution: the user only needs to have the application installed in her cellphone, thus eliminating the downside of carrying the sensing devices. Pollution-Spots also improves the spatial and time resolution when compared to the traditional monitoring systems (owned and managed by county agencies), tackling and solving this important societal problem.

Pollution-Spots is still under development and further experiments and improvement will be done based on the results of the first deployment in the campus of the Pontificia Universidad Javeriana. The complete implementation of the scheme will provide us with more realistic information about its performance, scope and weak points. A critical issue is the trade off between privacy protection and energy consumption that represents the use of Hybrid privacy preserving mechanisms. In order to determine the characteristics and specifications of this mechanism, it is necessary to model the most representative scenarios where the user information could be compromised.

References

[1] Mendez, D., Perez, A., Labrador, M. & Marron, J., P-sense: A participatory sensing system for air pollution monitoring and control. *Proc. of IEEE International Conference on Pervasive Computing and Communications, WiP workshop*, pp. 344–347, 2011.

[2] Carrapetta, J., Youdale, N. & Chow, A., Haze watch, 2010.

[3] Dutta, P., Aoki, P.M., Kumar, N., Mainwaring, A., Myers, C., Willett, W. & Woodruff, A., Demo abstract: Common sense: Participatory urban sensing using a network of handheld air quality monitors. *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, 2009.

[4] Hedgecock, W., Volgyesi, P., Ledeczi, A. & Koutsoukos, X., Dissemination and presentation of high resolution air pollution data from mobile sensor nodes. *Proceedings of the 48th Annual Southeast Regional Conference*, ACM SE'10, pp. 6:1–6:6, 2010.

[5] Devarakonda, S., Sevusu, P., Liu, H., Liu, R., Iftode, L. & Nath, B., Real-time air quality monitoring through mobile sensing in metropolitan areas. *Proceedings of the 2Nd ACM SIGKDD International Workshop on Urban Computing*, UrbComp '13, pp. 15:1–15:8, 2013.

[6] Li, N., Jiang, Z., Liu, Z. & Meng, X., A method of hierarchical time-series data visualization. *Proceedings of the 6th International Symposium on Visual Information Communication and Interaction*, VINCI'13, pp. 113–114, 2013.

[7] Ourcast, Minutely.the minute-by-minute rain forecast at your exact location, 2013.

[8] Vergara-Laurens, I., Mendez, D. & Labrador, M., Privacy, quality of information, and energy consumption in participatory sensing systems. *Pervasive Computing and Communications (PerCom)*, pp. 199–207, 2014.
[9] 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 *Lecture Notes in Computer Science*, pp. 614–620, 2013.
[10] Lands Nathan, B.J., Gamification wiki, 2010.
[11] Han, K., Graham, E., Vassallo, D. & Estrin, D., Enhancing motivation in a mobile participatory sensing project through gaming. *Privacy, Security, Risk and Trust (PASSAT)*, pp. 1443–1448, 2011.
[12] Ueyama, Y., Tamai, M., Arakawa, Y. & Yasumoto, K., Gamification-based incentive mechanism for participatory sensing. *Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pp. 98–103, 2014.
[13] Zhang, J., Ma, J., Wang, W. & Liu, Y., A novel privacy protection scheme for participatory sensing with incentives. *Cloud Computing and Intelligent Systems (CCIS)*, volume 03, pp. 1017–1021, 2012.