On the Interactions between Privacy-Preserving, Incentive, and Inference Mechanisms in Participatory Sensing Systems

Idalides J. Vergara-Laurens[1], Diego Mendez-Chaves[2], and Miguel A. Labrador[1]

[1] University of South Florida, Tampa, FL, 33620
ijvergara@mail.usf.edu, mlabrador@usf.edu
[2] Pontificia Universidad Javeriana, Bogota, Colombia
diego-mendez@javeriana.edu.co

Abstract. In Participatory Sensing (PS) systems people agree to utilize their cellular phone resources to sense and transmit the data of interest. Although PS systems have the potential to collect enormous amounts of data to discover and solve new collective problems, they have not been very successful in practice, mainly because of lack of incentives for participation and privacy concerns. Therefore, several incentive and privacy-preserving mechanisms have been proposed. However, these mechanisms have been traditionally studied in isolation overseeing the interaction between them. In this paper we include a model and implement several of these mechanisms to study the interactions and effects that they may have on one another and, more importantly, on the quality of the information that the system provides to the final user. Our experiments show that privacy-preserving mechanisms and incentive mechanisms may in fact affect each other's performance and, more importantly, the quality of the information to the final user.

Keywords: Participatory sensing, privacy-preserving, incentive mechanisms, inference mechanisms, P-sense.

1 Introduction

Participatory Sensing (PS) is a new data collection paradigm based on the availability of millions of cellular users equipped with smart applications, a large diversity of sensors, and Internet connectivity at all times. The availability of such a large number of mobile nodes opens the possibility to collect very large amounts of data and from places not possible or economically feasible before. For example, P-Sense [8] is an application that requires users to sense the level of pollution as they travel to build accurate pollution maps that can be used by the community and governmental organizations for many different purposes. However, users might no be willing to participate in this system if they also have to spend their data plans and batteries without any direct benefit in return. Therefore, for some PS systems, incentive mechanisms need to be included to guarantee a minimum level of participation for the system to be able to actually work. Similarly, most users will not be willing to participate if as a result of their data reporting, their privacy is not guaranteed. Therefore, privacy-preserving mechanisms need to be in place for these PS systems. Finally, inference and data analysis mechanisms are

J. Lopez, X. Huang, and R. Sandhu (Eds.): NSS 2013, LNCS 7873, pp. 614–620, 2013.

also usually included as part of a PS system to make estimations of the variables of interest in places where no data have been collected from, to make predictions, or to make any other type of analysis that will bring additional information to the final users.

However, one important problem is that these mechanisms have been devised and studied in an isolated or independent manner, as if they were the only mechanisms working in the system. Therefore, this paper presents a model to study the interactions between privacy-preserving, incentive, and inference mechanisms that have not studied before. In particular, this paper answers to the following questions:

- What effect do privacy-preserving mechanisms have in the quality of the information that the system provides to the final user?
- What effect do incentive mechanisms have in the quality of the information that the system provides to the final user?
- What effect do privacy-preserving and incentive mechanisms working together have in the quality of the information that the system provides to the final user?

The rest of the paper is organized as follows. Section 2 includes a brief description of the privacy-preserving, incentive, and inference mechanisms available in the literature and the ones used in this paper. Section 3 describes the model and performance metrics utilized in this work to study the effects produced by these mechanisms. Section 4 presents the performance evaluation of available privacy-preserving and incentive mechanisms. Finally, Section 5 presents the most important conclusions and provides directions for additional research.

2 Related Work

This section provides a brief literature review on privacy-preserving, incentive, and inference mechanisms, as they related to the work in this paper.

Privacy-Preserving Mechanisms: The main idea of **anonimization** is to *generalize* the users' data to a group of users in such a way that the user cannot be distinguishable from the group [2]. On the other hand, **obfuscation** techniques assume that the identity of the participant is or could be known [10]. Differently from anonymization techniques, the key idea is to modify the real location of the participants without considering the location of other participants. Finally, **encryption-based** techniques rely on cryptographic methods to guarantee the privacy of the participants with no modification of the actual data [4].

Incentive Mechanisms: Most of these mechanisms are based on *reverse auction* techniques. For instance, in the *Reverse auction based dynamic price scheme (RADP-VPC-RC)* presented in [6], each user makes a bid offering her sensed data and the system buys the k cheapest ones. Further, RADP-VPC-RC tackles the problem of cost explosion and avoids users from dropping out of the system. The work presented in [5] extends this approach with the *Greedy Incentive Algorithm (GIA)*, which uses not only the price but also the locations of the users. The key idea is to buy the k cheapest samples that maximize the covered area avoiding to buy samples that are closely located.

Inference Mechanisms: They aim to estimate the variables of interest in those places where data are not available. In this area, Kriging is one of the most widely used techniques in geostatistics (a branch of statistics that focuses on spatio-temporal datasets) [7]. All the different variations of the kriging estimator are modified versions of the best linear regression estimator [3, 9].

3 System Model and Performance Metrics

The system model consists of four components: *Sensed Data, Privacy Mechanism, Incentive Mechanism* and *Inference Engine* (Figure 1). The sense data component corresponds to the data reported by the participants of the PS application and is used as the input to the other components of the system. The privacy mechanism, receives sensed data and produces modified data according to the selected privacy mechanism. The incentive mechanism, selects a subset of the input data according to the incentive mechanism implemented in the system. The last component, the inference engine, is used to produce estimations of the selected variables in the areas of interest.

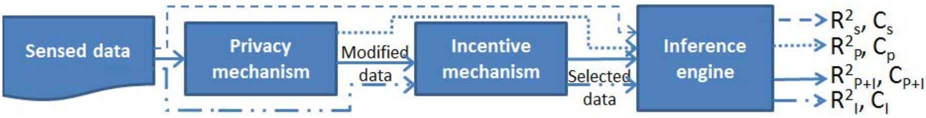

Fig. 1. System model and performance metrics

In addition, the proposed model includes several *data paths*. The first data path applies to those systems that implement neither privacy-preserving nor incentive mechanisms. The second data path considers a system that implements a privacy-preserving mechanism but it does not consider an incentive mechanism. The third path considers a system that does not implement a privacy-preserving mechanism but it does include an incentive mechanism. The fourth path considers a system that implements all components. Finally, the performance metrics for the model are: 1-**The quality of the estimation** R^2 measures how different the real data and the estimations are. 2-**The average displacement** as a result of changing the original location of the participant by privacy-preserving mechanisms. 3-**The coverage** of the area of interest C after the incentive mechanism is applied.

4 Performance Evaluation

In order to collect real environmental, we utilized the P-Sense system for measuring CO (ppm), CO_2 (ppm), combustible gases (ppm), air quality (4 discrete levels), temperature (F), and relative humidity (%) on the air [8]. Moreover, during three months data were collected 3 times a day for one hour, 3 to 4 days a week at the campus of the University of South Florida, in Tampa, Florida. The campus area is approximately 3.9 km^2, which is represented using a grid of 105x105 units in this project, i.e. each unit is equivalent to 20 meters. On the other hand, the implemented privacy-preserving mechanisms are: 1-**Tessellation** [2] (anonimization technique) varying the

k parameter from 3 to 9; 2-**Points of Interests** [10] (obfuscation technique) using a grid from 4x4 up to 11x11 cells; and 3-**Random Perturbation** [1] (obfuscation technique) using uniform distribution {[1,5],[5,15] and [10,20]}, normal distribution $\{[\mu = 5, 10, 10, 15, 15, 15, 20, 20, 20, \sigma = 1, 1, 3, 1, 3, 5, 1, 3, 5\}$, and exponential distribution, $\{\lambda = 5, 10, 15, 20\}$. Finally, the incentive mechanism utilized in the experiments is the GIA mechanism proposed in [5], as this is the only one in the literature that studies coverage considering the location of the users. The parameters used in the evaluation are the ones included in the GIA paper [5]: coverage radius=5, true valuation=uniform [0,10], and budget=[20,260].

4.1 Experiments and Results for Each Data Path

First Data Path: Neither Privacy nor Incentive Mechanisms. Table 1 presents the obtained R_S^2 and C_S when we apply the inference mechanism to the original data, i.e., without privacy nor incentive mechanism. Note the good quality of the estimations (R_S^2 is very close to 1), i.e., the estimated values are very close to the real values collected by the P-Sense system, and the coverage (33%) achieved by the system.

Table 1. The quality of the estimations (R_S^2) and coverage (C_S) for the first data path

Variable	Temperature	Relative humidity	Air Quality	CO_2	CO	Combustion gases	Coverage
R_S^2	0.92	0.89	0.91	0.92	0.88	0.96	33.14%

Second Data Path: Privacy but Not Incentive Mechanism. Figure 2 shows the quality of the estimations (R_P^2) produced by several privacy-preserving mechanisms for three of the environmental variables measured by the P-Sense system according to the average displacement. As it can be seen from the figure, all variables present a similar behavior regardless of the privacy-preserving mechanism: different privacy mechanisms with similar average displacement produce similar quality of estimation. This is an important conclusion because it means that none of the privacy-preserving schemes studied here is actually a good scheme in systems that do not tolerate low quality estimations. Moreover, Table 2 presents the relationship between the quality of the estimation (R_P^2), coverage (C_P), and average displacement for the temperature variable using the implemented privacy-preserving mechanisms. For example, *Points of Interest 6x6* and *Random Perturbation (all cases)* have similar average displacement and different coverage but very similar R_P^2. In conclusion, the impact of the average displacement over the quality is significantly greater than the impact of the coverage on the quality of estimation.

Third Data Path: Incentive but not Privacy Mechanism. Figure 3 shows the quality of the estimation (R_I^2) when the GIA mechanism is applied with different radii as a function of the budget. As the figure shows, the quality of estimation and number of selected participants decrease when the coverage radius increases. However, a very low radius implies the possibility of selecting participants that are very close to each other, which spend the budget but provide similar (redundant) information [5]. In conclusion,

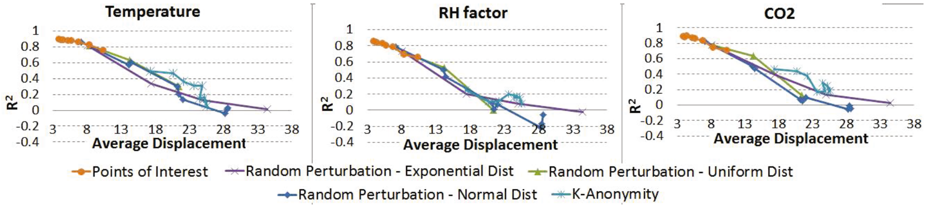

Fig. 2. R_P^2 as a function of the average displacement for several privacy-preserving mechanisms applied to each environmental variable

Table 2. Relationship between quality of estimation (R_P^2), coverage (C_P) and average displacement for the temperature

–	Tessellation			Points of Interest				Random Perturbation		
	$k = 3$	$k = 5$	$k = 7$	$4x4$	$6x6$	$8x8$	$10x10$	Uniform $[1; 10]$	Normal $(\mu = 5, \sigma = 1)$	Exponential $(1/\lambda = 5)$
R_P^2	0.47	0.31	0.16	0.76	0.86	0.88	0.89	0.82	0.86	0.82
C_P (%)	12.57	8.166	5.67	11.11	20.73	28.21	30.76	33.65	33.67	33.83
Avg Disp	20.66	23.66	24.45	10.45	6.8	5.25	4.21	8.42	7.18	8.12

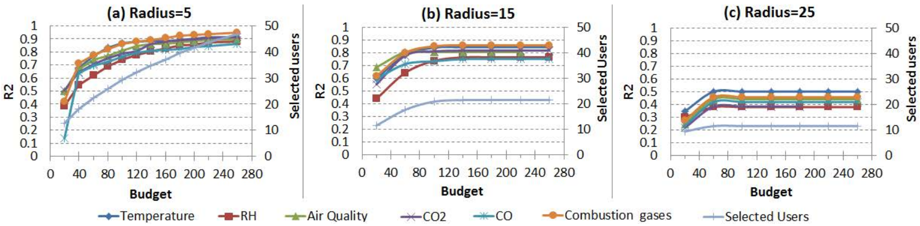

Fig. 3. R_I^2 as a function of the budget available to the GIA algorithm using different radii

selecting the appropriate radius should be function of the needed quality of estimation as well as the variability of the data being measured.

Figure 4-a shows the coverage (C_I) achieved by the GIA incentive mechanism. Note that, as expected, the coverage increases with the budget since a higher budget means more selected participants. However, although the effect of budget increments on the coverage is constant (figure 4-a), the effect of budget increments on quality of estimation is not significant after some value (figure 3-a). Therefore, the optimal budget depends mostly on the desired quality of estimation instead of the coverage.

Fourth Path: Privacy and Incentive Mechanisms. Figure 4 shows the coverage (C_{P+I}) achieved by the system as a function of the budget when the GIA mechanism is used along with the different privacy-preserving mechanism (three graphs on the right). Basically, the coverage increases with the budget. However, it can be seen how some privacy-preserving mechanisms limit the area of coverage of the GIA algorithm, providing a flat coverage value regardless of the budget available (Tessellation and Points of Interest). This is due to the number of selected points since a larger k as well as a smaller number of points of interest means fewer reporting points.

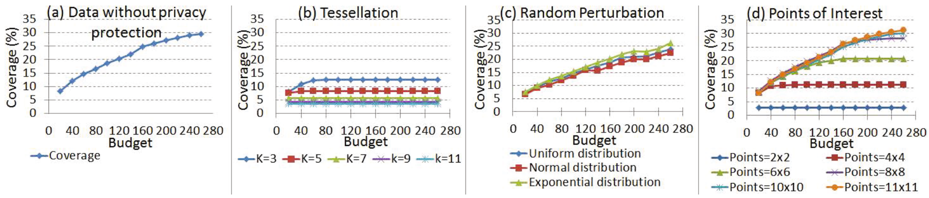

Fig. 4. Coverage achieved by the GIA mechanism as a function of the budget without privacy protection (C_I) (left) and when privacy-preserving mechanisms are present (C_{P+I})

Additionally, Figure 5 shows the quality of estimation (R^2_{P+I}) achieved by the system for different budgets and privacy-preserving mechanisms. In the case of Tessellation, the budget has no major effect on the quality of estimation because the incentive mechanism buys the anonymized locations and, when it buys one of the k users, the others are discarded. In the case of Points of Interest, the situation is similar: the budget affects the coverage but it does not affect the quality of estimation much because the number of point of interest defines the number of reporting locations. In conclusion, when privacy and incentive mechanisms work together, the most affecting factor is the average displacement produced by the privacy mechanisms.

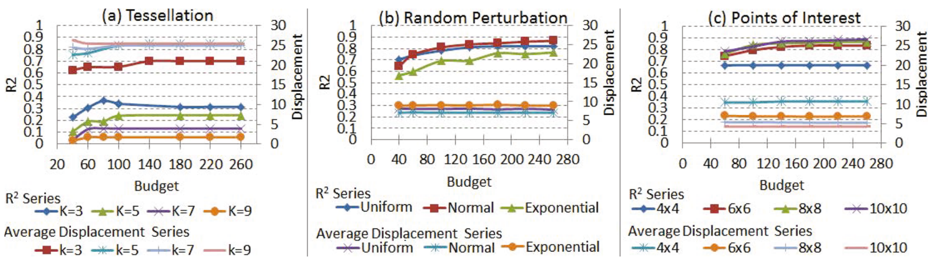

Fig. 5. Quality of estimation achieved by each privacy-preserving mechanism as a function of the budget used in the GIA mechanism (R^2_{P+I}) for temperature

5 Conclusions and Research Directions

This paper presents a model to study the interactions between privacy-preserving, incentive, and inference mechanisms in participatory sensing systems. A performance evaluation is carried out to evaluate the impact of these mechanisms on the quality of estimation (R^2), as provided by the inference system to the final user, as well as the area of coverage (C) achieved by the incentive mechanism. In the case where no incentive mechanism is included in the system, the impact of privacy-preserving mechanisms on the quality of the information to the final user depends on the average displacement that the privacy mechanism introduces to the real locations of the participants. Therefore, a system needing a high quality of information should avoid the use of privacy mechanisms that introduce large displacements, such as Tessellation, and rather utilize a privacy-preserving mechanism with low displacements, such as Points of Interest with a high number of points or Random Perturbation. In the case where the incentive mechanism is used, the information provided to the final user depends on the available budget, with the quality increasing with the budget. However, there is a point in which the effect of the budget decreases the quality of the estimation. Finally, when privacy and incentive mechanisms work together, the budget available to the incentive mechanism and the average displacements introduced by the privacy mechanism are the factors that affect the quality of the information to the user the most: for systems needing a high quality of estimation, a high budget should be used as well as a privacy mechanism with a low average displacement.

References

1. Agrawal, R., Srikant, R.: Privacy-preserving data mining. In: Proceedings of the 2000 ACM International Conference on Management of Data - SIGMOD, Dallas, Texas, USA, vol. 29, pp. 439–450 (May 2000)
2. Cornelius, C., Kapadia, A., Kotz, D., Peebles, D., Shin, M., Triandopoulos, N.: Anonysense: Privacy-aware people-centric sensing. In: Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services - MobiSys, Breckenridge, Colorado, USA, pp. 211–224 (June 2008)
3. Goovaerts, P.: Geostatistics for Natural Resources Evaluation, volume 4th printing. Oxford University Press, USA (1997)
4. Hoh, B., Gruteser, M., Herring, R., Ban, J., Work, D., Herrera, J.C., Bayen, A.M., Annavaram, M., Jacobson, Q.: Virtual trip lines for distributed privacy-preserving traffic monitoring. In: Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services - MobiSys, Breckenridge, Colorado, USA, pp. 15–28 (June 2008)
5. Jaimes, L., Vergara-Laurens, I., Labrador, M.: A location-based incentive mechanism for participatory sensing systems with budget constraints. In: Proceeding of the 2012 IEEE International Conference on Pervasive Computing and Communications - PERCOM, Lugano, Switzerland (March 2012)
6. Juong-Sik, L., Baik, H.: Sell your experiences: a market mechanism based incentive for participatory sensing. In: IEEE International Conference on Pervasive Computing and Communications - PerCom, Mannheim, Germany, pp. 60–68 (April 2010)
7. Mendez, D., Labrador, M., Ramachandran, K.: Data interpolation for participatory sensing systems. Submitted to the Pervasive and Mobile Computing Journal (Summer 2012)
8. Mendez, D., Perez, A., Labrador, M., Marron, J.: P-sense: A participatory sensing system for air pollution monitoring and control. In: Proceedings of IEEE International Conference on Pervasive Computing and Communications, pp. 344–347 (2011)
9. Moore, R.: Geostatistics in hydrology: Kriging interpolation. Technical report, Mathematics Department, Macquarie University, Sydney (1999)
10. Vergara-Laurens, I., Labrador, M.: Preserving privacy while reducing power consumption and information loss in lbs and participatory sensing applications. In: Proceedings of the 2011 IEEE Global Communication Conference Workshops - GLOBECOM Workshops, Houston, Texas, USA (December 2011)