Privacy preservation with unequal data exchange strategy in participatory sensing*

Tianqi Zhang¹, Rong Zhang² and Jin Wang*¹

¹School of Information Science and Technology, Nantong University, Nantong, Jiangsu, 226001, China
²School of Information Science and Technology, Nantong University, Nantong, Jiangsu, 226001, China
*Corresponding author’s e-mail: wj@ntu.edu.cn

Abstract. In participatory sensing, participants contribute sensing data to task center. However, the task center is not always trustworthy. It may try to profile the participants by data mining, which is a great threat to the privacy of participants. To solve this problem, three collaborative data exchange strategies are studied. Participants exchange sensing data before upload to the task center. The mixed data protect the participants’ privacy from data mining by the task center. The simulations show that the unequal data exchange strategy is more efficient than the full exchange strategy because it considers both the tradeoff between privacy preservation and the cost of data exchange.

1. Introduction
With the rapid development of smartphone technology in recent years, the mobile phones have been equipped with a large number of embedded sensors (such as microphone, camera and accelerometer) and provided with advanced processing and storage capacities. According to the report published by Global System for Mobile Communications Association (GSMA) in February 2017, the number of global mobile phone users will break through 5 billion in 2017. Such a great number of mobile phones have provided an unprecedented space coverage and thorough deployment without any additional cost of communication facilities. The activities that users utilize mobile devices like mobile phones as the sensing devices to perceive the environment are referred to as participatory sensing. The participatory sensing system has the excellent characteristics of being convenient, efficient and low in cost, which has stimulated a mass of new applications, including collection and sharing of personal diet [1], perception and real-time display of urban noise pollution [2], collection and sharing of bicycling experience[3], etc. In the applications of the participatory sensing system, the sensing data are marked with information such as the collection time and position coordinates[4], and such data are uploaded to the task center under the control of the applications. Without the protection mechanism, these data might disclose private information of the users, such as the haunts, residential place and work place. Design of the user location privacy preservation mechanism is the key to reducing and eliminating these privacy risks.

The participatory sensing system provides sensing data by means of numerous participants. Early researches on the participatory sensing system mainly focused on how to stimulate participants to contribute data [5] and how to improve data quality. Xiao et al.[6] designed a new incentive mechanism which encourages the participants to provide high-quality sensing data and applied this
mechanism into the transport trip quality measurement system. Chang et al.[7] proposed a scheme, named PURE, which can estimate the global regression model for low quality data with participatory sensing.

However, some scholars have recently begun to realize the importance of privacy protection for participatory sensing. Niu et al.[8] introduced E-cent, an exchangeable unit bearer currency, to the participatory sensing. And participants use the E-cent to take part in tasks anonymously. Wang et al.[9] proposed a framework, in which the cluster head collaborates with cluster members to change pseudonyms and reputation values. This framework preserves privacy using pseudonyms and maintains the incentive mechanism in the meanwhile. Liu et al.[10] proposed a mechanism named PriRe, which first measures privacy risks based on user preferences towards data sharing in participatory sensing systems and then make the recommendation according to the measurement. They develop a privacy-preserving rewarding scheme which allows campaign administrators to reward users for the data they contribute. Dimitriou et al.[11] developed a rewarding scheme which allows campaign administrators to reward users for the data they contribute. Anonymous tokens are used to preserve privacy. Messaoud et al.[12] proposed a trust-worthy entity to distort the collected data before its release.

This paper investigates the data collection process in participatory sensing. To prevent the untrustworthy task center from infringing the privacy of the participants through data mining technique, this paper proposes a data collaborative exchange method. Information collected by the participants includes their path information. This paper put forwards that when two participants are encountered, they will exchange the collected sensing data through cooperation to confuse the actual path information of the participants and further preserve their privacies.

2. System Model

Figure 1. Roles in the participatory sensing system

The model of the participatory sensing system application is as shown in Figure 1. It mainly comprises three characters, namely, (1) task provider; (2) task center; and (3) participant. It is assumed that the participatory sensing system application does not restrain the real-time uploading of the sensing data, and the participants can customize the strategy for delayed uploading. It is assumed that the mobile device can automatically collect the sensing data (such as sound, picture and speed) and each sensing report is composed of a triple \((t, l, s)\), wherein \(t\) refers to the time of perception, \(l\) refers to the location of perception and \(s\) refers to the data of perception. The participants will upload the mixed data but not directly upload their own sensing data into the task center. Purpose of this idea is to eliminate the relation between the spatio-temporal environment (namely, time and location) and ID of the participants.
3. Privacy Preservation Against Task Center

3.1 Problem Description
In the participatory sensing system, the task center can directly access the data uploaded by the participants and saved in the task center. These data comprise the driving position, speed, location, time and other information of the participants. When the task center is attacked, the attacker can directly access these data and speculate the daily paths of the participants according to the sensing data of the participants and further speculate some personal private information of the participants.

This paper mainly adopts the collaborative data exchange strategy to confuse the actual paths of the participants so as to preserve the privacy of the participants. This strategy will firstly assess the credibility of the opposite party and then decide whether to exchange the data or not. If the data need to be exchanged, corresponding exchange strategy will be adopted as the exchange protocol. This paper studies three kinds of exchange strategies.

3.2 Collaborative Data Exchange Strategies
The three data exchange strategies are specifically as follows:

Collaborative data exchange strategy A: full exchange strategy. When the participants encounter with each other, they will exchange all the data respectively collected by them. If the quantity of the data packages collected by the participants is different, the exchanged quantity can be disymmetric.

Collaborative data exchange strategy B: unequal exchange strategy. Each participant will randomly and independently decide the exchanged quantity and the data exchanged. Compared with the full exchange strategy, this strategy exchanges party of the data, and which data are exchanged shall be decided by the participants themselves.

Collaborative data exchange strategy C: equal exchange strategy, namely, the quantity of exchanged sensing data is equal. The participants will negotiate to decide the exchanged quantity $n$ before the exchange, $n$ will be randomly selected according to the exchange quantity required by each participant, and then data exchange is completed after agreed by both parties. Compared with the unequal exchange strategy, this method can ensure symmetry of the exchange. That is, the confusion degree of the data is equal, and expenses will be equally distributed among the participants.

![Figure 2. The real path](image1)

![Figure 3. The mixed path after data exchange](image2)

The exchange strategy is specifically described based on the cases in Figure 2 and Figure 3. Assume that the actual driving paths of participant A and B are as shown in Figure 2 and they encounter with each other at location 1. When participant A encounters with participant B at location 1, they will exchange the sensing data they collected, and the quantity shall be decided by the exchange strategy. Besides, assume that all the sensing data collected by participant A and B are exchanged. When they are encountered, participant A and participant B continuously drive following their paths and collect the sensing data until the participants get to the final destination. By using the collaborative exchange strategy, the participants exchange the sensing data collected by them with other participants.
and then upload the data into the server of the task center. As shown in Figure 3, based on these mixed data, the server of the task center considers that the driving path of participant A is from the company of A to the home of B. The driving path of participant B is from the bank to the home of A. In this example, most of the actual paths followed by the participants are concealed in the server of the task center, and therefore the privacy of the participants is preserved.

3.3 Evaluation

In the simulations, the total number of participants is 20. Assume that the number of the initial data packages of each vehicle is 10, and the speed of each vehicle in collecting the data packages is equal, which is 10 km/h. The experiment lasts for 6 hours. It is assumed that the vehicles will be encountered at a certain probability \( P \) and they collaboratively exchange the data after the encounter. Each experiment is repeated for 10 times, and the average value of the 10 results is taken as the final result.

(1) Comparison on Privacy Preservation of Three Strategies

This simulation is mainly to evaluate the privacy preservation capacities of the three exchange strategies. In this paper, the mixedness is used to represent the degree of data mixing after executing an exchange strategy, which is the percentage between the exchanged data quantity and the total quantity of collected data. The higher the mixedness is, the higher of privacy preservation degree will be. The average mixedness \( JD \) is defined as

\[
JD = \frac{1}{n} \sum_{i=1}^{n} \frac{S_i}{M_i}
\]

(1)

Where \( S \) refers to the quantity of exchanged data, \( M \) refers to the total quantity of collected data and \( n \) refers to the total number of participants.

![Figure 4. The mixedness comparison of three data exchange strategies](image)

It can be known from Figure 4 that the average mixedness gradually increases along with the increase of the encounter probability. When the encounter probability reaches a certain value, the mixedness will gradually tend to be stable. Meanwhile, it can be seen that among the three exchange strategies, the average mixedness of the full exchange strategy is the highest, which is followed by the unequal exchange strategy, while the effect of the equal exchange strategy is the worst.

(2) Comparison on Cost of Three Strategies

This simulation is mainly to evaluate the cost generated by the three data exchange strategies. Before the data exchange, the participants will conduct internal exchange first. The so-called internal exchange is that the participant will exchange the collected data packages and then exchange a small amount of data packages with other neighboring participants, so that the privacy preservation can be better. In this paper, it is assumed that the number of data packages for each internal exchange is 10; the vehicles will be encountered once every 1 hour; the expense of internal exchange for a unit data
package is 1, the expense generated in the exchange with other participants is 2. If the costs of internal mixture are the same, the total costs shall be expressed as $C$.

\[ C = \sum_{i=0}^{n} I_i + \sum_{j=0}^{m} J_j \]  

(2)

Where in $S$ refers to the quantity of exchanged data, refers to the total quantity of collected data and $n$ refers to the total number of participants.

Where in $I$ refers to the cost of internal exchange, $n$ refers to the quantity of internal exchange, $J$ refers to the cost generated in the exchange with companions, $m$ refers to the quantity of internal exchange, and $N$ refers to the number of exchange times.

![Figure 5. The cost comparison of three data exchange strategies](image)

The experimental results are as shown in Figure 5. With the increase of the number of exchange times, the costs of the three strategies are gradually increased. Among the three exchange strategies, the cost increasing of the full exchange strategy is the largest along with the increase of exchanges. The cost of the other two strategies rise slowly, and that of the unequal exchange strategy is relatively higher.

The above two experimental results indicate that the mixedness of the full exchange strategy is the best. The premise of participant’s upload of the data package is that the mixedness shall reach a certain value. This strategy requires fully exchanging the collected data packages, therefore a large amount of costs required from other participants, and this is unreasonable. Under the extreme circumstance that participants carrying thousands of data packages encounter and exchange with the participants carrying only a few data packages, a great amount of cost will be generated to the latter after the exchange. By contrast, the cost of the unequal data exchange strategy are less, but the mixedness of this strategy will slightly decline within an acceptable range. The cost of the equal exchange strategy are the least, but the exchange conditions proposed by this strategy are strict, which have restricted the exchange quantity and reduced the mixedness. Under the scene assumed in this paper, the participants pay particular emphasis on privacy preservation and their own benefits. In conclusion, the mixedness and cost of the unequal data exchange strategy are both within the acceptable range of the participants, thus it is considered in this paper that this strategy can better conform to the demands of the participants.

4. Conclusion

In the participatory sensing system, uploading the data with spatio-temporal information may reveal the privacy of the participants. This paper proposes a collaborative data exchange method which eliminates the relation between the sensing data and the ID of the collectors and preserved the location privacy of the participants. This paper studies on three exchange strategies to evaluate their effects of privacy preservation. The simulations indicate that the exchange strategy can effectively solve the
incredibility threaten of the task center. Besides, the mixedness and cost of the unequal exchange strategy are both within the acceptable range of the participants. This strategy can be easily integrated into other privacy protection schemes to further improve their privacy protection effect.

References

[1] Reddy S., Parker A., Hyman J., Burke J., EstrinD., Hansen M. (2007) Image browsing, processing, and clustering for participatory sensing: lessons from a DietSense prototype. In: Proceedings of the 4th workshop on Embedded networked sensors. Cork. pp. 13–17.

[2] Rana R. K., Chou C. T., KanhereS. S., BulusuN., Hu W. (2010) Ear-phone: an end-to-end participatory urban noise mapping system. In Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks. Stockholm. pp. 105–116.

[3] Eisenman S. B., MiluzzoE., Lane N. D., Peterson R. A., Ahn-G.-S., Campbell A. T. (2009) BikeNet: A mobile sensing system for cyclist experience mapping. ACM Transactions on Sensor Networks (TOSN), 6(1):1–6.

[4] Christin, D., Reinhartd, A., Kanhere, S. S., Hollick, M. (2011). A survey on privacy in mobile participatory sensing applications. Journal of Systems and Software, 84(11):1928-1946.

[5] Restuccia, F., Das, S. K., Payton, J. (2016). Incentive mechanisms for participatory sensing: survey and research challenges. ACM Transactions on Sensor Networks (TOSN), 12(2):1-13.

[6] Xiao, Z., Lim, H. B., &Ponnambalam, L. (2017). Participatory sensing for smart cities: a case study on transport trip quality measurement. IEEE Transactions on Industrial Informatics, 13(2):759-770.

[7] Chang, S., Zhu, H., Zhang, W., Lu, L., Zhu, Y. (2016). PURE: Blind Regression Modeling for Low Quality Data with Participatory Sensing. IEEE Transactions on Parallel and Distributed Systems, 27(4):1199–1211.

[8] NiuX., Wang J., Ye Q., Zhang Y. (2018). A Privacy-Preserving Incentive Mechanism for Participatory Sensing Systems. Security and Communication Networks, 2018:1–15.

[9] Wang J., Wang Y., GuX., Chen L., Wan J.(2018). ClusterRep: A cluster-based reputation framework for balancing privacy and trust in vehicular participatory sensing. International Journal of Distributed Sensor Networks, 14(9): 155014771880329.

[10]Liu R., Liang J., Gao W., Yu R. (2018). Privacy-based recommendation mechanism in mobile participatory sensing systems using crowdsourced users’ preferences. Future Generation Computer Systems, 80: 76–88.

[11]DimitriouT.(2018). Privacy-respecting reward generation and accumulation for participatory sensing applications. Pervasive and Mobile Computing, 49:139–152.

[12]MessaoudR. B., SghaierN., Moussa M. A.,Ghamri-DoudaneY. (2019). Privacy Preserving Utility-Aware Mechanism for Data Uploading Phase in Participatory Sensing.IEEE Transactions on Mobile Computing, 18(9):2160–2173.