Privacy-Preserving Personal Sensitive Data in Crowdsourcing

Ke Xu, Kai Han(✉), Hang Ye, Feng Gao, and Chaoting Xu

School of Computer Science and Technology/Suzhou Institute for Advanced Study,
University of Science and Technology of China, Hefei, People's Republic of China
{kexu,yehang,gf940312,xct1996}@mail.ustc.edu.cn, hankai@ustc.edu.cn

Abstract. Spatial crowdsourcing system refers to sending various location-based tasks to workers according to their positions, and workers need to physically move to specified locations to accomplish tasks. The workers are restricted to report their real-time sensitive position to the server so as to keep in coordination with the crowdsourcing server. Therefore, implementing crowdsourcing system while preserving the privacy of workers sensitive information is a key issue that needs to be tackled. We discard the assumption of a trustworthy third party cellular service provider (CSP), and further propose a local method to achieve acceptable results. A differential privacy model ensures rigorous privacy guarantee, and Laplace mechanism noise is introduced to preserve workers sensitive information. Finally, we verify the effectiveness and efficiency of the proposed methods through extensive experiments on real-world datasets.

Keywords: Crowdsourcing · Differential privacy
Sensitive information

1 Introduction

Nowadays, with the rapid proliferation of all kinds of smartphones and the convenience of mobile Internet, crowdsourcing has emerged as a significant computing technology which utilizes human intelligences. In particular, numerous crowdsourcing-based platforms, such as CrowdFlower [1], Gigwalk [15], Gmission [5] and etc., which leverages the wisdom of crowd to perform the specialized assignment appropriately and accurately. This new framework encourages active workers to participate in to perform specified tasks that are vicinity to the required locations. The crowd of workers have shift their conventional idea of data consumers to the role of gathering data to gain some deserved rewards (e.g., money, reputation). In crowdsourcing system, smartphone users are engaged to provide pervasive and inexpensive tasks of data collecting and computing eventually. The application of crowdsourcing has developed incredibly. It has been widely used in ride sharing, traffic or environment monitoring.

Specifically, the roles in the whole crowdsourcing system are categorized into three types: crowding platform (i.e., server), crowdworkers (i.e., workers), crowdsourcer (i.e., requester) [3]. The platform is responsible for distributing atomic

S. Chellappan et al. (Eds.): WASA 2018, LNCS 10874, pp. 509–520, 2018.
https://doi.org/10.1007/978-3-319-94268-1_42

tasks or viral tasks to workers and in charge of the data collecting job. The workers are the ones who are concentrating on finishing the small units of work in return for monetary payment. The responsibility of crowdsourcer is aiming to carrying out computationally hard tasks and divide them into several subtasks. In [16], the tasks on crowdsourcing platform can be published in two distinct modes: Worker Selected Tasks (WST) and Server Assigned Tasks (SAT). On the one hand, in WST mode [15], online workers are allowed to select arbitrary tasks delivered by the crowdsourcer in vicinity without the permission of crowdsoucing platform. On the other hand, in SAT mode [6], the workers are restricted to report their real-time position to the server so as to keep in coordination with the crowdsourcing server, then the server will decide how to allocate tasks reasonably in terms of some optimal functions. Meanwhile, in WST mode, workers are unnecessary to track the locations of workers while the crowdsourcing platform are never interrupted to follow the tracks of workers in SAT mode, so the issues on the privacy of these workers needs to be protected are rather hot problems.

Differential Privacy (DP) [11,25] is a relatively new notion of privacy, and also is one of the most popular privacy notations. It is actually implemented by noise mechanism which adds a random noise to the output data. With the privacy definition, To et al. [19] developed a new framework for protecting workers' locations by introducing the cellular service provider (CSP) [20] as a trustworthy third party. private spatial decomposition, which partition the geospatial data into smaller regions and obtain statistics on the points with each region, and designed to enhance the accuracy of the entire crowdsourcing system [7]. Fan et al. [12,23] divide the space into four equal subspaces using quadtree. Recursively partitioning on quadtree is highly efficient compared with partitioning of kd-tree [21,22]. [4,13] consider the privacy concerns that are hard to solve and propose a flexible optimization framework that can be adjusted to trade-offs with the joint efforts of platform and workers. As far as we know, there already have been some related works protecting workers location information on the crowdsourcing system, however, there are few works currently pay scalable attention to privacy-preserving of workers sensitive information, which discards the entirely reliable third party. It is a novel notion that proposed appropriately in accordance with the characteristics of reality, so achieving a desired privacy-preserving result is not a non-trivial problem.

In this paper, we formulate the privacy protection strategies without compromising the third party from a particular worker to a crowdsourcing platform as a non-trivial problem that follows two criteria: (1) efficiency of our proposed method, (2)utility of our method. Note that the aforementioned standards remains a secret for us, which motivates us to consider the fundamental factors in crowdsourcing system. We show that these two criteria can hardly be optimized simultaneously. We immediately divide the data publishing into three part: data preprocessing, information filtering and noise addition. In the first period, we generate a most suitable data structure to storage workers' sensitive information. Next, we filter out the insensitive information by exponent mechanism of differential privacy. Finally, we further add appropriate noises after information filtering to ensure the privacy leakage problem. In summary, the main contributions of our work are listed as follows:

- We identify the specific challenges of privacy-preserving in crowdsourcing system, and we further develop a model that illustrates this issue.
- We abandon the assumption that the third party CSP is rather convincing and adopt a local method to publish sensitive data sets.
- We conduct both extensive numerical evaluations and performance analysis to show the effectiveness and efficiency of our designed method using real-world datasets, and analyze the key factors associated with hierarchical method.

The paper is structured as follows. We present related preliminaries in Sect. 2. Next we develop our model to solve the problem in details in Sect. 3. We discuss the experimental results and analyze crucial factors in Sect. 4, respectively. Section 5 summarizes related work. Finally, we conclude the work in Sect. 6.

2 Preliminaries

Intuitively, Differential Privacy (DP)has grown as the standard in privacy protection, thanks to its strong mathematical guarantees rooted in related statistical analysis. DP ensures the attacker fail to deduce whether a particular individual in or not in the original data, thereby protecting the workers' privacy.

Definition 1 ((ε, δ)-differential privacy) [11]**:** Let D and D' be two neighboring datasets which differ on at most one record, denoted as $|D\varDelta D'| = 1$, a randomized mechanism M: D→ R, $\varOmega$(M) be the set of all possible outputs of M in D and D', algorithm M gives (ε,δ)-differential privacy if:

$$\Pr[M(D) \in \varOmega] \leq exp(\varepsilon) \times Pr[M(D') \in \varOmega] + \delta \tag{1}$$

The parameter ε is called privacy budget, which controls the level of privacy guarantee. The smaller ε is, the higher security becomes. If $\delta = 0$, the randomized mechanism M gives ε -differential privacy by its strictest definition. Thus, (ε,δ)-differential privacy in some degree provide freedom to violate strict differential privacy for some low probability events.

Theorem 1 ***(Sequential Composition)*** [17]**:** Say we get a set of privacy algorithms M = $\{M_1, M_2, ..., M_m\}$. For each M_i satisfies a ε_i-differential privacy guarantee for the same dataset, M will provide $\sum_{i=1}^{m}\varepsilon_i$-differential privacy.

Sequential composition undertakes the privacy guarantee for a combination of the entire differential privacy process. When a set of randomized mechanisms have been conducted on the same dataset, the total privacy budget is the sum of all privacy budgets.

Theorem 2 ***(Parallel Composition)*** [18]**:** Say we get a set of privacy algorithms M = $\{M_1, M_2, \ldots, M_m\}$. For each M_i satisfies a ε_i-differential privacy guarantee on a disjoint subset of the whole dataset, M will provide max (ε_1,ε_2,. . . ,ε_m)-differential privacy.

Parallel composition corresponds to situation where a quantity of private mechanisms are applied to a disjoint dataset. Consequently, the privacy guarantee only depends on the largest privacy budget.

Definition 2 (Sensitivity) [10]**:** Given neighboring datasets D and D', for a query function f: D→ R, the sensitivity of f is defined as

$$\varDelta f = \max_{(D,D')} |f(D) - f(D')|_1 \tag{2}$$

Sensitivity $\varDelta$f is closely related to the query f. It is regarded as the maximal differential between the query results on neighboring datasets. Currently, two basic mechanisms are widely used to guarantee differential privacy: the Laplace mechanism and the Exponential mechanism.

Definition 3 (Laplace mechanism) [10]**:** Given dataset D and a function f: D→ R, $\varDelta$ f is the sensitivity of f, representing the maximal value on the output of f when deleting any tuple in D. The randomized algorithm

$$M(D) = f(D) + Laplace(\frac{\varDelta f}{\varepsilon}) \tag{3}$$

satisfies ε-differential privacy. We use Laplace(x) to represent the noise sampled from a Laplace distribution with a scaling of x.

Definition 4 (Exponential mechanism) [17]**:** Let (q, r) be a function of dataset D that measures the quality of output r ∈ Range, $\varDelta$ q represents the sensitivity of r. The exponential mechanism M satisfies ε-differential privacy if:

$$M(D) = (r : Pr[r \in Range] \propto exp(\frac{\varepsilon q(D,r)}{2\varDelta q})) \tag{4}$$

For non-numeric queries, differential privacy utilizes the exponential mechanism to randomized the results.

3 Designed Model

We consider the problem of privacy-preserving spatial crowdsourcing task assignment in the SAT mode. The crux of our method is to how to choose data structure to storage workers' sensitive information and apply a differential privacy mechanism to each worker's location information. As mentioned in Sect. 1, data structure selection is a non-trivial step. Previous literature assumes that the third party Cell Service Provider (CSP) is completely convinced, but this may not be the case in real-world scenarios. Our proposed method adopts a novel model to protect workers' locations. In this way, the approach recommends tasks to each worker with a better success ration and a stronger privacy guarantee. Figure 1 shows the basic model of the proposed framework consisting of each worker's three components:

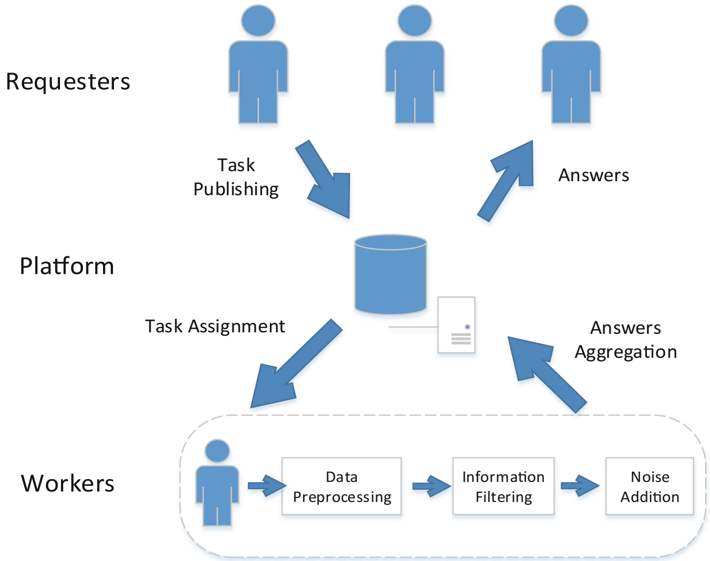


Fig. 1. Our system model for task recommendation in crowdsourcing systems

Data Preprocessing: In this component, each worker collects various statistics periodically in the background. After that, they will preprocess the detailed data for a short time. The sensitive information of each worker are not delivered to the third party, so this step can protect the private context information of participating workers well.

Information Filtering: In this component, based on the statistics preprocessing component, each worker then select the most sensitive location information. Note that workers are allowed to decide how much private information they are willing to share with others. Therefore, we set a constant threshold θ and eliminate those location information that are below this threshold. We may achieve an ideal result according the Exponent mechanism to complete sensitive information selection.

Noise Addition: In this component, based on the information filtering and Laplace mechanism, each worker then adds suitable noises to these sensitive information. It goes without saying that the noises are subjected to laplace distribution. After adding some noise in the original data, we achieve a series of brand new datasets.

3.1 Detailed Explanation

We just described the basic system model for task recommendation in Spatial Crowdsourcing system. Next we will represent a description of the details in these components and explain why we take these steps to tackle the barriers in the entire crowdsourcing systems.

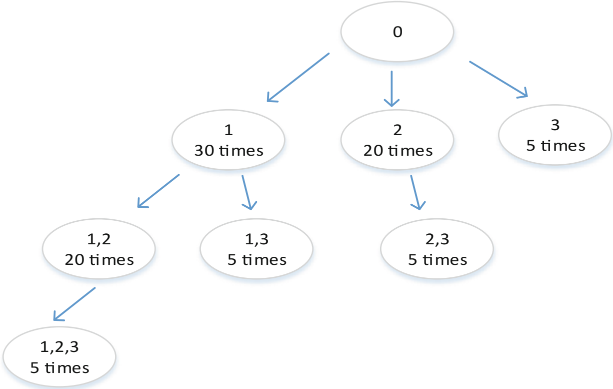


Fig. 2. Creating a Trie-Tree

Data Preprocessing: We aim to provide good efficiency, privacy and utility in our proposed framework. Each worker firstly achieve a transaction database D including the work's identity, the frequency of locations they visited during a long period and when the worker arrive at these locations. Apparently, there is no problem that we ignore some detailed location information such as the longitude and latitude of locations. Then we randomly select several items (e.g., I items) to represent the whole database D. After finishing that, it's vital for us to determine which data structure to storage our location information. We make a thorough decision to choose Trie-Tree to represent workers' context information. It's nontrivial for us to draw a solid conclusion that we select the data structure Trie-Tree. We summarize the reasons as follows: (1) It is most suitable to maintain the link between the location data and overwrite the original data set as well; (2) It reduces the number of noise addition in that we add noise into each node rather than each original data. It disturbs true visit frequency in each node, we are required to add noises only once into each node. In other words, all the original data sets in node are covered with some noise, it is unnecessary for us to add extra noises into each original data. Therefore, workers can protect his/her location privacy at a relatively low cost of utility. Figure 2 shows an example describing how we store workers' sensitive location information. We assume that the root node in level 0, it is clearly see that there are $\sum_{i=1}^{n} C_n^i = 2^i - 1$ nodes in the Trie-Tree and all the 1st-items in the level 1. Here, there are four different nodes in level 1 and total $2^3 - 1 = 7$ nodes in the tree. More specifically, we take node 1, 2 as an example, it denotes the combination of node 1 and node 2 in level 1, the constant value 20 means the least number of visiting times in the node. As we all know, node 1 is visited 30 times and node 2 is 20, so we achieve the minimum value 20 in node set 1, 2. Likewise, it is rather easy to get all the other nodes in our Trie-Tree.

Information Filtering: After representing detailed location data sets in Trie-Tree, we continue to finish sensitive location information selection based on Exponent mechanism. We firstly traverse the entire tree by level and eliminate those nodes in which the constant value is smaller than the specified threshold θ. By the way, we determine how much θ would be according the subsequent experiments. Furthermore, we naturally derive the nodes set S in which the visiting times of each node are above θ. Finally, we filter out n nodes in set S according to the exponent mechanism of differential privacy. The specific operations are as follows:

Step 1: Input the sensitive location information set S, then we successively take out each node in it and mark them in turn:

$$tag(S, s_i) = q(s_i) \tag{5}$$

where q(s_i) denotes the actual visiting frequency of node s_i.

Step 2: calculate the weight of each node:

$$s_i.w = exp(\frac{\varepsilon_1 \times tag(S, s_i)}{2\Delta tag}) \tag{6}$$

Where ε_1 means the privacy budget of exponent mechanism, the function Δtag means the sensitivity of node s_i.

Step 3: randomized figure out the top n nodes according the following equation, then we make up a new set C:

$$Pr(s_i) = \frac{s_i.w}{\sum_{i=1}^{n} s_i.w} \tag{7}$$

The goal we design the equation is to follow the exponent mechanism. s_i.w is derived from step 2.

In this section, we adopt exponent mechanism to finish information selection because of the merits of exponent mechanism. Not only we can evaluate the privacy protection by privacy budget, but it improves the efficiency of our proposed algorithm according to filtering out insensitive location information.

Noise Addition: We carry on adding suitable noises into elements in set C followed by information selection. That is to say, noises which are subjected to laplace distribution are appended into the top n nodes in C. Similarity, we use privacy budget ε_2 to evaluate privacy leakage in that the approach improves the efficiency of our algorithm and strengths the utility of data sets. The details are as follows:

$$q(c_i) = q(c_i) + laplace(\frac{\Delta q}{\varepsilon_2}) \tag{8}$$

where ε_2 means the privacy budget of laplace mechanism, Δq is the global sensitivity of function q. In this paper, both function tag and q mean the sensitive location visiting frequencies of each worker. In order to understand the laplace mechanism, we list the probability density function of laplace mechanism:

$$Pr(x, \lambda) = \frac{e^{-\frac{|x|}{\lambda}}}{\lambda} \tag{9}$$

where λ denotes that there is no correlation between noises and database, it merely concerned with sensitivity of function and privacy parameters. Here, x means the actual visiting frequencies.

We eventually formulate a completely new set E and then publish it to our crowdsourcing platform. At this point, We successfully complete the process of privacy preserving and deliver it to the unreliable third party.

4 Experimental Study

4.1 Experiments Setup

We use a real-world dataset: Gowalla. It contains the check-in history of users in allocation-based social network. It includes some detailed data such as the type of each user, the longitude and latitude of the location and the time when users visit the locations. For our experiments, we assume that these users are workers of the spatial crowdsourcing system, and their locations are those of the most recent check-in points. We transfer the original data into a database D in which it records the visiting frequencies of users in a month. The algorithms are implemented in java 8, and the experiments were performed on Intel(R) Core(TM) i7 2.40 GHz CPU and 8 GB main memory.

4.2 Experimental Results

We evaluate our proposed method from the following two aspects: *efficiency* and *utility*. When each worker receives a series of recommended tasks, he/she decides to choose the appropriate task to conduct. Workers may spend a lot of time finishing information filtering when the size of task set is substantial. Thus, the efficiency of information filtering is directly related to the Trie-Tree. The recommendation system should create Trie-Tree for a short time to ensure thee efficiency of information filtering. Furthermore, utility represents the accuracy and data protection level. From the perspective of the crowdsourcing platform, the utility is expected accuracy of data protection. Meanwhile, the utility is the degree of privacy preserving from the perspective of the workers. The utility for both stakeholders is closely related to the privacy protection level.

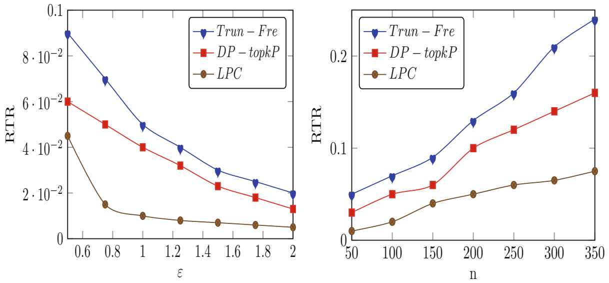
Efficiency. We analyze the efficiency of our method from timeliness of creating a Trie-Tree and extracting data from the tree. We vary the number of nodes from 8 to 512 to observe the situation of creating an entire Trie-Tree. t_{create} denotes how long we need to build a tree. As is shown in the following Table 1, We apparently see that building a tree is not a time-consuming process during the whole period of privacy protection in spite of the size of nodes set.

Table 1. Time to create Trie-Tree

Number of nodes	8	32	64	128	256	512
$T_{create}/\times 10^{-6}s$	9	21	33	59	114	146

Table 2. Time to extract sensitive information.

ε_1	0.01	0.02	0.10	0.40	0.90	1.30
n/s	3120	4382	41297	48633	52614	75621



(a) RTR of varying n (b) RTR of varying ε

Fig. 3. RTR of varying n and ε.

From the Table 2, we also see that the efficiency of extracting data from tree is so high, because it depends on the essential structure and characteristics of Trie-Tree. In some degree, the time of extracting data grows longer as the privacy budget becomes bigger.

Utility. First of all, we use Local Protection in Crowdsourcing (LPC) to represent our method. Then we show the performance on ratio of rejecting true nodes (RTR), which significantly represents the utility of algorithms. Figure 3(a) plots the RTR and number of nodes with n ranging from 50 to 350. At the same time, we fix privacy level to 0.5. For each n, the x-axis represents the number of nodes in set, and the y-axis represents the ratio. The ratio increases as the size of nodes set grows bigger. LPC runs better than Trun-Fre [2] and DP-topkP [24]. This is reasonable because we add noises into sets rather than nodes merely.

Secondly, we analyze the ratio of RTR and the privacy protection level. Since noise needs to be added to provide (ε,δ)-differential privacy, the RTR is achieve at the cost of accuracy. Before the experiment is conducted, we strictly control the number of nodes to be 150. We illustrate the trade-off between the privacy parameter ε and RTR in Fig. 3(b). It shows the RTR and number of nodes with ε ranging from 0.5 to 2.0. For each ε, the x-axis represents the privacy budget in our algorithm, and the y-axis represents the ratio. Our method outperforms than Trun-Fre [2] and DP-topkP [24] as well.

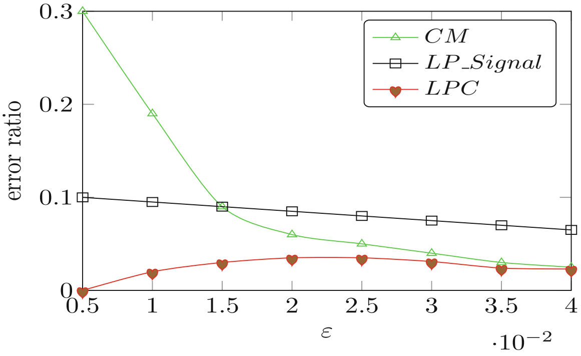


Fig. 4. Error of varying ε.

Lastly, we compare our novel method LPC with previous existing approaches CM [9] and LP_Signal [14]. Define F_{before} denotes the frequencies of locations before adding noises and F_{after} means frequencies after doing that. Thus error signifies the ratio of F_{before} and the differences between F_{after} and F_{before}:

$$Error = \frac{\|F_{after} - F_{before}\|_2}{\|F_{before}\|_2} \tag{10}$$

The experiment results are shown in the Fig. 4. It is clear for us to see that the impact of varying privacy levels on the errors. We can observe that the minimum error in our method compared with other existing algorithms. Moreover, whether the privacy budget ε is high or not, our method also achieve the steady and desired results. Since the error is inevitable with a small range, we conclude the expected errors of our privacy-preserving approach is almost close the optimal one.

5 Related Work

In this section, we review some previous work related to our problem in this literature. Differential privacy [8,11] is a strict privacy definition that is independent of prior knowledge. With the definition of privacy, To et al. [19] proposed a framework for protecting workers' locations by illustrating the cellular service provider (CSP) as a third party. The partition algorithm generates a private spatial decomposition (PSD) is widely conducted by the following works such as dividing grids [23], creating kd-trees [21,22] and quadtree [12]. However, the main shortcoming of these methods is that the privacy-preserving depends on the third party in some degree.

6 Conclusion

Privacy issues are increasingly becoming concerning with the popularity of spatial crowdsourcing. In this paper, we introduced a novel differentially private approach for spatial crowdsourcing, which enables the participation of various workers without compromising their sensitive information privacy. The magnitude of noise is minimized by fully utilizing the given privacy budget, which is crucial for efficiency and utility of our method. To ensure effective privacy protection, we select the Trie-Tree to storage workers' sensitive information rather than sending them to the completely trust third party. Because a trustworthy data collector is merely an assumption, which contradicts the reality and common sense. Comparisons between our method and existing approaches that privacy-preserving effects is dramatically enhanced by a series of experiments.

Acknowledgments. This work is partially supported by National Natural Science Foundation of China (NSFC) under Grant No. 61772491, No. 61472460, and Natural Science Foundation of Jiangsu Province under Grant No. BK20161256. Kai Han is the corresponding author.

References

1. CrowdFlower. http://crowdower.com
2. Bhaskar, R., Laxman, S., Smith, A., Thakurta, A.: Discovering frequent patterns in sensitive data. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 503–512. ACM (2010)
3. Cao, C.C., She, J., Tong, Y., Chen, L.: Whom to ask?: jury selection for decision making tasks on micro-blog services. Proc. VLDB Endow. **5**(11), 1495–1506 (2012)
4. Cao, C.C., Tong, Y., Chen, L., Jagadish, H.: WiseMarket: a new paradigm for managing wisdom of online social users. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 455–463. ACM (2013)
5. Chen, Z., Fu, R., Zhao, Z., Liu, Z., Xia, L., Chen, L., Cheng, P., Cao, C.C., Tong, Y., Zhang, C.J.: gMission: a general spatial crowdsourcing platform. Proc. VLDB Endow. **7**(13), 1629–1632 (2014)
6. Cheng, P., Lian, X., Chen, L., Han, J., Zhao, J.: Task assignment on multi-skill oriented spatial crowdsourcing. IEEE Trans. Knowl. Data Eng. **28**(8), 2201–2215 (2016)
7. Cormode, G., Procopiuc, C., Srivastava, D., Shen, E., Yu, T.: Differentially private spatial decompositions. In: 2012 IEEE 28th International Conference on Data Engineering (ICDE), pp. 20–31. IEEE (2012)
8. Dwork, C.: Differential privacy: a survey of results. In: Agrawal, M., Du, D., Duan, Z., Li, A. (eds.) TAMC 2008. LNCS, vol. 4978, pp. 1–19. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-79228-4_1
9. Dwork, C.: Differential privacy in new settings. In: Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 174–183. SIAM (2010)
10. Dwork, C.: A firm foundation for private data analysis. Commun. ACM **54**(1), 86–95 (2011)
11. Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., Naor, M.: Our data, ourselves: privacy via distributed noise generation. In: Vaudenay, S. (ed.) EUROCRYPT 2006. LNCS, vol. 4004, pp. 486–503. Springer, Heidelberg (2006). https://doi.org/10.1007/11761679_29

12. Fan, L., Xiong, L., Sunderam, V.: Differentially private multi-dimensional time series release for traffic monitoring. In: Wang, L., Shafiq, B. (eds.) DBSec 2013. LNCS, vol. 7964, pp. 33–48. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39256-6_3
13. Gong, Y., Wei, L., Guo, Y., Zhang, C., Fang, Y.: Optimal task recommendation for mobile crowdsourcing with privacy control. IEEE Internet Things J. **3**(5), 745–756 (2016)
14. Jia, O., Jian, Y., Shaopeng, L., Yubao, L.: An effective differential privacy transaction data publication strategy. J. Comput. Res. Dev. **10**, 007 (2014)
15. Kazemi, L., Shahabi, C.: GeoCrowd: enabling query answering with spatial crowdsourcing. In: Proceedings of the 20th International Conference on Advances in Geographic Information Systems, pp. 189–198. ACM (2012)
16. Kazemi, L., Shahabi, C., Chen, L.: GeoTruCrowd: trustworthy query answering with spatial crowdsourcing. In: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 314–323. ACM (2013)
17. McSherry, F., Talwar, K.: Mechanism design via differential privacy. In: 48th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2007, pp. 94–103. IEEE (2007)
18. McSherry, F.D.: Privacy integrated queries: an extensible platform for privacy-preserving data analysis. In: Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data, pp. 19–30. ACM (2009)
19. To, H., Ghinita, G., Shahabi, C.: A framework for protecting worker location privacy in spatial crowdsourcing. Proc. VLDB Endow. **7**(10), 919–930 (2014)
20. Wang, J., Liu, S., Li, Y., Cao, H., Liu, M.: Differentially private spatial decompositions for geospatial point data. China Commun. **13**(4), 97–107 (2016)
21. Xiao, Y., Gardner, J., Xiong, L.: DPCube: releasing differentially private data cubes for health information. In: 2012 IEEE 28th International Conference on Data Engineering (ICDE), pp. 1305–1308. IEEE (2012)
22. Xiao, Y., Xiong, L., Yuan, C.: Differentially private data release through multidimensional partitioning. In: Jonker, W., Petković, M. (eds.) SDM 2010. LNCS, vol. 6358, pp. 150–168. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15546-8_11
23. Xiong, P., Zhang, L., Zhu, T.: Reward-based spatial crowdsourcing with differential privacy preservation. Enterp. Inf. Syst. **11**(10), 1500–1517 (2017)
24. Zhang, X., Wang, M., Meng, X.: An accurate method for mining top-k frequent pattern under differential privacy. J. Comput. Res. Dev. **51**(1), 104–114 (2014)
25. Zhu, T., Li, G., Zhou, W., Philip, S.Y.: Differentially private data publishing and analysis: a survey. IEEE Trans. Knowl. Data Eng. **29**(8), 1619–1638 (2017)