An Enhanced Location-Data Differential Privacy Protection Method Based on Filter

Shasha Zhang[1], Haiyan Kang[2](✉), and Dong Yu[2]

[1] School of Computer Science, Beijing Information Science and Technology University, Beijing 100192, China
[2] School of Information Management, Beijing Information Science and Technology University, Beijing 100192, China
kanghaiyan@126.com

Abstract. Location based service (LBS) is the basic function and important application of Internet of things. The disclosure of location data which contains a lot of sensitive information will be a threat for individual. This paper proposed an enhanced location-data differential privacy protection method based on filter. Firstly, noise is added in location-data for differential privacy. Secondly, Kalman is used to predict, correct and optimize the Location-data after the addition of noise, which ensure the optimization to satisfy the differential privacy. Finally, released the processed data and carry out the location query service. Experimental results demonstrate that the proposed algorithm promotes Location-data utility and level of privacy protection.

Keywords: Privacy protection · Kalman filter · Location-data · Location based service · Differential privacy

1 Introduction

The Internet of things (IoT) is a new concept proposed in recent years, which is getting more and more attention. The Internet of things has been widely used, covering such as intelligent transportation, environmental protection, government work, public safety, environmental monitoring and so on. However, the Internet of things faces many security threats at the same time. It is one of the necessary conditions for the wide application to solve the problem of privacy protection in the application of the IoT.

Location based service (LBS) is the basic function and important application of Internet of things. In recent years, the applications and offline products with location-based services as the core are becoming more and more popular. Such as: Hand ring, Runtopia APP, Baidu Map, etc. These sensors and mobile terminals with "inner smart" query the location data and provide it to the server. This provides more extension services to users. However, when users share their location information to the Internet of things for extended service, the location privacy information would be threatened. The query scenario based on the location service is shown in Fig. 1.

Y.-B. Lin et al. (Eds.): SGIoT 2021, LNICST 447, pp. 117–127, 2022.
https://doi.org/10.1007/978-3-031-20398-5_10

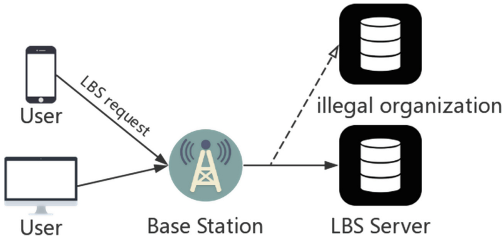


Fig. 1. Query scenarios based on location service

The location data privacy threat means that the location data of the user is stolen by the location transmission device, the location information transmission channel without authorization and so on. An attacker can combine the user location information in the Internet of things with extension services to infer the information of a user, such as personal habits, home addresses, private life or identify a person's true identity through existing background knowledge. Forester Research has conducted a more authoritative sample of private placement in the United States. The result shows that 70% of users think it is necessary to pass legislation to guarantee the privacy of their location. The threat of location data privacy security seriously hinders the market development and commercial prospects of location services. Therefore, it is very important to keep confidential the location information of users while providing services to users.

Currently, most of the privacy-preserving approaches in location-based services are based on k-anonymity or l-diversity [1], and such techniques generalize the user's real location into a region to achieve privacy protection of location information. These approaches provide low precision and large amount of location data to the server, which obviously degrades the quality of service. Dwork et al. [2] proposed a differential privacy-preserving model in 2006, which became a mainstream technique due to its good privacy-preserving strength, by adding random noise to the original query result so that adding or removing a piece of data in the dataset has no effect on the query result, thus making it is difficult for an attacker to infer a particular piece of real data back through multiple queries to achieve privacy protection. However, although existing differential privacy solutions enhance the protection of location data to some extent, the setting of differential privacy noise size can lead to degradation of service quality of service providers.

The main contributions of this paper are as follows: (1) An enhanced location data differential privacy protection algorithm based on filtering technology is proposed. (2) The Kalman filtering mechanism is introduced to optimize the data of differential privacy protection. Thus, the location information obtained by the server is more accurate and the quality of the location service is improved. At the same time, privacy is not disclosed.

2 Related Work

The literature [3–5] proposed a location-based privacy protection technology based on k-anonymity. The position scope of the user is given by generalization to prevent the real location of the user from being exposed. Any location in the position scope could be the user's real location, so the user's real location is protected. However, if the generalization is too large, it will reduce the quality of the location service. C. Y. Chow et al. [6] proposed a k-anonymity protection method of introducing an anonymous device. Not only can this method prevent user location information from being leaked, but also prevent user identity information from being leaked. But this approach increases the server's computing resources and the load on the network. Gedik and Liu [7] proposed that user-defined k values and minimize the size of fuzzy areas. However, this method has a large amount of computation, and the system is not very efficient, and it can only be used when the k value is small. Chen Ping, Feng Yunxia [8] proposed and implemented a mobile phone based location anonymization scheme for continuous query LBS users. The method adopts the independent system structure, and all the anonymous protection calculation is carried on the mobile terminal. Privacy protection can be achieved by constantly sending false queries, without a third party server providing anonymization service. But this method of sending fake inquiries does not have the effect of k-anonymity. And the constant sending of bogus queries not only increases the burden on the server, but also increases the load on the mobile network. Zhang et al. [9] proposed a scheme to enhance user privacy through caching and spatial anonymity in continuous LBS, using multi-level caching to reduce the risk of exposing user's information to untrusted location service providers, however, an attacker may obtain the user's true identity based on an inference attack to obtain a specific user's location. Zhao et al. [10] developed a trustworthiness-based k-anonymity scheme, where The trustworthiness is set based on the location similarity of the mobile user, and the computational overhead of the change scheme is large.

The literature [11] proposes a privacy-preserving algorithm for trajectory suppression based on information entropy, which calculates the minimum cost of suppressing sensitive location points through a function to select a reasonable suppression method for the sequences containing sensitive points in the original dataset. The scrambling technique is to generate false locations from real locations by certain transformations to achieve the purpose of protecting real locations. The literature [12] proposes a maximum-minimum false location selection scheme based on the false location privacy method, which makes it difficult for the attacker to combine the edge information to filter some false locations for privacy protection of location information. All the above privacy protection techniques have some limitations and drawbacks, and attackers can obtain users' location privacy information through long-term observation, mining and analysis [13–15], so these techniques cannot resist relevant attacks background knowledge attacks.

Dwork et al. [2] proposed a differential privacy-preserving model in 2006, which became a mainstream technique due to its good privacy-preserving strength, by adding random noise to the original query results so that adding or removing a piece of data in the dataset has no effect on the query results, thus making it difficult for an attacker to backpropagate a piece of real data through multiple queries to achieve privacy protection. Huo Zheng et al. [16] constructed noisy quadtrees and noisy R-trees for free space and road network space, respectively, and protected location data by adding Laplace noise, but did not consider the interaction between 2 consecutive moments of location data. Yin et al. [17] proposed a location privacy preservation method that satisfies differential privacy constraints to protect location data privacy and maximize the utility of data and algorithms in industrial IoT. Yan et al. [18] proposed an unbalanced quadtree partitioning algorithm based on region uniformity, and accordingly, designed a gradient privacy budget allocation scheme and adjustment method to ensure the effectiveness of the differential privacy model. Zhao et al. [19] combined federal learning and localized differential privacy together, and proposed four localized differential privacy mechanisms to scramble the gradients generated by vehicles, and introduced a three-output mechanism to reduce the communication cost. Tian Feng et al. [20] proposed a personalized differential privacy publishing mechanism for trajectory data, using Hilbert curve to extract the distribution characteristics of trajectory data at each moment to generate location clusters, using sampling mechanism and index mechanism to select the representative elements of each location cluster, and then using the location representative elements to generalize the original trajectory data to generate the trajectory data to be published.

3 Differential Privacy

3.1 Differential Privacy Conceptions

Definition 1. Differential Privacy [21]. A randomized algorithm K, Range(k) denotes the value Range of a random algorithm K, $Pr[z]$ denotes the disclosure risk of event z. For any two neighboring datasets D_1 and D_2, at most one record is different, namely, $|D_1 \Delta D_2|$. Algorithm K is ε-differential privacy if for any datasets D_1 and D_2 differing at most one record, and any possible outputs set D,

$$Pr[K(D_1) = D] \leq e^{\varepsilon} \times Pr[K(D_2) = D] \tag{1}$$

where ε is privacy parameter, also called the privacy budget, e is the natural base.

Definition 2. Global Sensitivity [22]. The global sensitivity of a function $f : D_n \rightarrow R^d$, is

$$\Delta f = max\|f(D_1) - f(D_2)K\|_P \tag{2}$$

where, R is the real space of the mapping, d is the query dimension of function f, P is the norm distance measured by Δf.

3.2 Mechanisms to Achieving Differential Privacy

Laplace Mechanism

Typical differential privacy is implemented by adding a Laplace noise to the output of the query. It defined as follows.

Definition 3. Laplace Mechanism [22]. For a function $f : D_n \rightarrow R^d$, the mechanism satisfies ε-differential privacy, if the output results satisfy.

$$Y(D) = f(D) + Lap\left(\frac{\Delta f}{\varepsilon}\right) \tag{3}$$

where Lap( ) is density function of Laplace.

Exponential Mechanism
For functions where adding noise does not make sense or the output space is non-numeric, the exponential mechanism was proposed by McSherry. This mechanism mainly deals with the non-numerical algorithm of the output results.

Definition 4. Exponential Mechanism [23]. For a score function $q : (D \times O) \rightarrow R$, the mechanisms K satisfies ε-differential privacy, if K satisfy.

$$K(D, u) = \{r : | \Pr[r \in O] \propto \exp\left(\frac{\varepsilon q(D, r)}{2\Delta q}\right)\} \tag{4}$$

If the mechanisms K that outputs r with probability proportional to $exp(\,)$.

4 Location Data Query Differential Privacy Protection Algorithm Based on Filter Mechanism

4.1 Location Data

The location data is the representation value of the original data in the dimension of the location. It is to project the spatially based spatial data into the two-dimensional plane of longitude and latitude representation. Location data query privacy protection means that users can freely query any other location data within the individual location dimension and ensure that the location information of the individual is not leaked.

Definition 5. Location data. Location data is a spatial data. We mapped the spatial location data to the two-dimensional space. Using two-dimensional spatial representation represents location data. Namely, we use $T(x, y)$ for the location data. The location data is only considered with the longitude and dimension of the position, where x and y is the longitude and latitude of position data.

4.2 Location Data Query Privacy Protection Mechanism

The process of differential privacy protection based on filter mechanism is shown in Fig. 2.

The process of the location data differential privacy protection algorithm based on filter mechanism is as follows:

(1) Differential privacy protection for location data p is achieved by adding Laplace noise.
(2) Use filtering technology to predict and revise the sampled data after adding noise. Because the Laplace noise value is randomly generated. There may be larger disturbances to the original location data. It makes the data less practical and affects the quality of query results.

(3) Use Kalman filtering technology to optimize the position data after adding noise. That ensures that the optimization is satisfied with the ε-differential privacy. Obtain the optimized query location data $\hat{p}$ and submit it to the location service provider. The location service provider provides location services based on query location data $\hat{p}$.

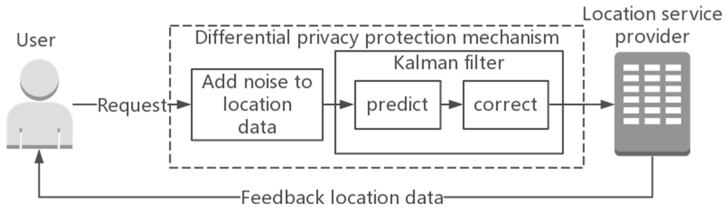


Fig. 2. The process of differential privacy protection based on filter mechanism

Use Kalman filter to predict and correct the location data after adding noise. It can obviously improve the availability of location data and improve the quality of location service. The differential privacy protection mechanism based on filter technology has obvious advantages for the practicality of location data, which will be proved in subsequent experiments.

4.3 Kalman Filtering Mechanism

Kalman filter [24] can be used to predict and modify the location data of noise, which can solve the problem of low service quality caused by adding noise better.

System measurement model:

$$Z = T + v \tag{5}$$

where, Z is the value of the location data T after adding noise. is Gaussian white noise (measurement noise). It's a Gaussian distribution that is expected of 0 and the variance is R:

$$v \sim N(0, R) \tag{6}$$

The system measurement model is very similar to the Laplace noise model, and the difference is that the noise value distribution obeys the Gaussian distribution and the Laplace distribution. Therefore, the Kalman filter can be used in differential privacy protection by using Gaussian noise to simulate the Laplace noise (i.e., the value of R). L. Fan et al. [25, 26] combined with the relevant formula of Kalman filtering, the optimal selection method of variance R is given, $R = 1/\varepsilon^2$, when using Gaussian noise to simulate the Laplace noise. The selection of this parameter will be proved in subsequent experiments.

The use of Kalman filtering in differential privacy is solved in the above text. The main function of Kalman filtering in this paper is to solve the problem of decreasing service quality of service providers due to differential privacy protection. Kalman filtering can reduce the noise value very well, so that the service provider can collect the posterior estimate of the actual location of the user and improve the service quality.

It is assumed that T means kalman prior estimates, $\hat{T}$ means the posterior estimate (the location data value service provider gets). The calculation equation of the posterior estimation in kalman filter is as follows:

$$\overline{T} = T \tag{7}$$

$$\hat{T} = \overline{T} + K_k(Z - \overline{T}) \tag{8}$$

According to the Eq. (8), the posterior estimate can be calculated. The user location data value is obtained by the service provider. It is determined by the sampling point data plus noise value Z, the prior estimator T and K_k. Where, K_k is the Kalman gain, which is constantly changing. Among them, K represents the Kalman gain, which is constantly changing. It minimizes the error and the variance.

Kalman filtering mechanism is the process of predicting and correcting the optimal estimation. Its obvious advantage is that the filtering principle takes into account all available data within the system (the original value and the noise value). The predicted values are the closest to the original data. It can make up for the inaccuracy of the differential privacy protection to the location data and ensure that the individual's true location privacy is not leaked. Subsequent experiments will demonstrate the advantages of using Kalman filtering technology for location data query protection.

Original location data is $T(x, y)$, total privacy budget is ε, distribution the total privacy budget average to each location data. The location data differential privacy protection algorithm based on the filtering mechanism is as follows:

Algorithm: Differential privacy protection algorithm for location data based on filtering mechanism
Input: $T(x, y)$-original location data, ε-privacy budget
Output: T -query location data

Step 1: Add noise to location data, obtain the value $Z_x = x + Lap\left(\frac{0,1}{\varepsilon^2}\right)$, $Z_y = y + Lap\left(\frac{0,1}{\varepsilon^2}\right)$
Step 2: Prediction: $\overline{x} = x$, correction: $\hat{x} = \overline{x} + K_k(Z_x - \overline{x})$
Step 3: Do the Same process for y
Step 4: Get the final query location data $\hat{T} = (\overline{x}, \overline{y})$
Step 5: Repeat Step 1–4. // The program continuously circulation process all the location information in the data set

---

5 Experimental Evaluation

5.1 Experimental Datasets and Environment

The experimental datasets come from a location service social network user accessing the location data from the Datatang [27]. It contains the location data of 270,000 longitude and latitude representations. To simplify the experiment, only ten percent of the data was sampled for comparison. The algorithm uses the Java language implementation, the programming environment is Eclipse8. The experimental environment is Windows7 2. 68 GHz, 4.0 GB, data mining tool is SQL Server 2008R2.

5.2 Validation of Parameter R Selection in Filter Mechanism

In the filter mechanism, the choice of parameter R directly affects the accuracy of the whole mechanism. It mentioned above that variance R of Gaussian noise has a certain relationship with the value of the Laplace noise ε. In this experiment, we compare the different values of R in terms of the privacy budget ε = 0.1, ε = 0.01, and then verify the value formula of optimal R. As shown in Fig. 3, the x axis represents the value of R. The y axis represents the relative error. Obviously, when the privacy budget ε = 0.1, ε = 0.01, the optimal value of R is $R = 10^2$ and $R = 10^4$. The formula of the optimal value of R is verified $R = 1/\varepsilon^2$.

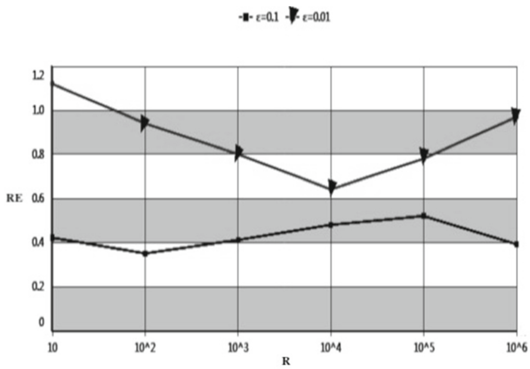
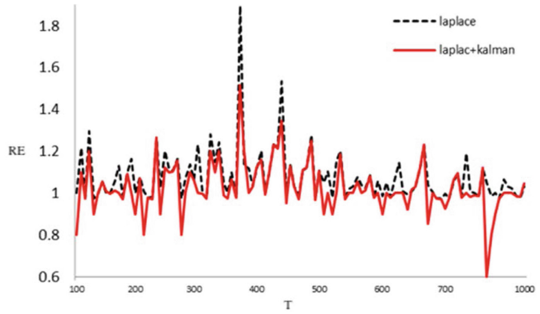


Fig. 3. Kalman filter parameter R selection

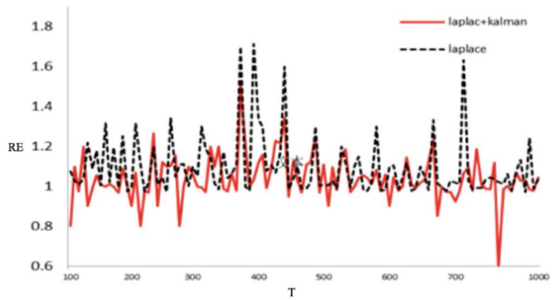
5.3 Experiment and Analysis

Experimental Steps: Firstly, use the location data differential privacy protection algorithm based on filter mechanism. Input: original location data $T(x, y)$ and privacy budget ε = 1,0.1,0.01. Lastly, output the final query location data $\hat{T}$.

Test and Analysis: IN order to verify the advantage of the data practicability, we compare the experiment with traditional differential privacy and noise technology. The privacy budget parameters were selected for comparison with ε = 1, 0.1, 0.01 respectively. The relative error between two protection mechanisms is obtained to judge the merits of the data availability. (When ε = 0.1, ε = 0.01, ε = 1, the data practicality of the three privacy budgets is quite similar and due to space reasons, this paper only gives a comparison of ε = 1), as shown in Fig. 4. Here red solid line represents our method(laplace + kalrman), and black dotted line represents traditional differential privacy (laplace).



(a) Comparison of longitude data query values



(b) Comparison of longitude data query values

Fig. 4. Comparison of the usability of the data after filter

Analysis: (1) In Fig. 4, the practical aspect of the position data treated with filtration technology is significantly higher than that of the noising mechanism. After filtering, the location data is closer to the original value, which can improve the service quality of the location service provider and enable users to query the service more accurately. Figure 4(a) and Fig. 4(b) are approximately overlapped, and the relative error is not obvious because the added noise value is too small and the optimized data optimization is not obvious. (2) In terms of data security, the two methods use differential privacy for protection. Data security is similar. (3) Based on the experiment, the data practicability and security of the location data differential privacy protection algorithm based on the filtering mechanism can achieve the optimal balance of differential privacy protection mechanism.

6 Conclusions

Location-based services can provide users with many conveniences and improve the user experience. But location services must require users to provide their own specific location information, which may pose a serious threat to users' privacy. In this paper, the differential privacy mechanism is applied to the location data query protection, and

Kalman filtering technology is introduced to improve the practicability of location data. It is to focus on practicality and user's location security. The experiments show that the protection mechanism proposed in this paper has a significant improvement in both practicality of data and privacy protection.

Acknowledgment. This work is partially supported by National Social Science Foundation of China (21BTQ079), Humanities and Social Sciences Project of the Ministry of Education (20YJAZH046) and Higher education research projects (2020GJZD02).

References

1. Wang, L., Xiaofeng, M.: Location privacy preservation in big data era: a survey. J. Softw. **25**(04), 693–712 (2014)
2. Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., Naor, M.: Our data, ourselves: privacy via distributed noise generation. In: Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer, Berlin, Heidelberg, pp. 486–503 (2006)
3. Kim, J.S., Li, K.J.: Location K-anonymity in indoor spaces. GeoInformatica **20**(3), 415–451 (2016)
4. Jagwani, P., Kaushik, S.: Privacy in location based services: Protection strategies, attack models and open challenges. In: International Conference on Information Science and Applications, pp. 12–21. Springer, Singapore (2017)
5. Wang, Y.-H., Zhang, H.-L., Yu, X.-Z.: Research on location privacy in mobile internet. J. Commun. **36**(9), 1–14 (2015)
6. Chow, C.Y., Mokbel, M.F.: The new Casper: a privacy aware location-based database server. In: Proceedings of the 23rd IEEE International Conference on Data Engineering, pp. 1499–1500 (2007)
7. Gedik, B., Liu, L.: A customizable k-anonymity model for protecting location privacy. In: ICDCS'05, pp. 620–629 (2005)
8. Chen, L., Feng, Y.-X., Dai, G.-J.: Mobile phone based trajectory anonymization of continuous query LBS users. Appl. Res. Comput. **28**(12), 4653–4656 (2011)
9. Zhang, S., Li, X., Tan, Z., Peng, T., Wang, G.: A caching and spatial K-anonymity driven privacy enhancement scheme in continuous location-based services. Futur. Gener. Comput. Syst. **94**, 40–50 (2019)
10. Zhao, P., et al.: ILLIA: Enabling k-anonymity-based privacy preserving against location injection attacks in continuous LBS queries. IEEE Internet Things J. **5**(2), 1033–1042 (2018)
11. Yifei, W., Luo Yonglong, Y., Qingying, L.Q., Wen, C.: A trajectory privacy protection method based on information entropy suppression. Comput. Appl. **38**(11), 3252–3257 (2018)
12. Jie, W., Chunru, W., Jianfeng, M., Hongtao, L.: False location selection algorithm based on location semantics and query probability. J. Commun. **41**(03), 53–61 (2020)
13. Hu, Z., Yang, J., Zhang, J.: Trajectory privacy protection method based on the time interval divided. Comput. Secur. **77**, 488–499 (2018)
14. Li, J., et al.: Mobile location privacy-preserving algorithm based on PSO optimization. J. Comput. Sci. **41**(5), 1037–1051 (2018)
15. Pan, J., Liu, Y., Zhang, W.: Detection of dummy trajectories using convolutional neural networks. Secur. Commun. Netw. **2019**, 8431074 (2019)
16. Zheng, H., Xiaofeng, M.: A trajectory data publishing method satisfying differential privacy. J. Comput. Sci. **41**(02), 400–412 (2018)

17. Yin, C., Xi, J., Sun, R., Wang, J.: Location privacy protection based on differential privacy strategy for big data in industrial internet of things. IEEE Trans. Industr. Inf. **14**(8), 3628–3636 (2017)
18. Yan, Y., Gao, X., Mahmood, A., Feng, T., Xie, P.: Differential private spatial decomposition and location publishing based on unbalanced quadtree partition algorithm. IEEE Access **8**, 104775–104787 (2020)
19. Zhao, Y., et al.: Local differential privacy-based federated learning for internet of things. IEEE Internet Things J. **8**(11), 8836–8853 (2020)
20. Tian Feng, W., Laifeng, Z.L., Hai, L., Xiaolin, G.: A personalized differential privacy protection mechanism for trajectory data distribution. J. Comput. Sci. **44**(04), 709–723 (2021)
21. Dwork, C.: Differential privacy. In: Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (ICALP), pp. 1–12. Venice, Italy (2006)
22. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Halevi, S., Rabin, T. (eds.) TCC 2006. LNCS, vol. 3876, pp. 265–284. Springer, Heidelberg (2006). https://doi.org/10.1007/11681878_14
23. McSherry, F., Talwar, K.: Mechanism design via differential privacy. In: Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pp. 94–103. Providence, RI, USA (2007)
24. Kalman, R.E.: A new approach to linear filtering and prediction problems. J. Basic Eng. **82**, 35–45 (1960)
25. Liyue, F., Xiong, L.: Real-time aggregate monitoring with differential privacy. In: Proceedings of the 21st ACM International Conference on Information and knowledge management, pp. 2169–2173 (2012)
26. Fan, L., Li, X.: An adaptive approach to real-time aggregate monitoring with differential privacy. IEEE Trans. Knowl. Data Eng. **26**(9), 2094–2106 (2014)
27. Datatang. Social network users access geographic location data. http://www.datatang.com/data/43896. 27 Nov 2015