Utility-optimized location privacy scheme with geo-indistinguishability

Ke Zhu[a], Pengfei Yu[b], Xuehong Chen[c*]
[a] State Grid Corporation of China, Beijing 100031, China; [b] State Grid Smart Grid Research Institute Co. LTD., Beijing 102209, China; [c] China Industrial Control Systems Cyber Emergency Response Team, Beijing 100000, China

ABSTRACT

Geographic location privacy protection is an important research content of privacy protection in recent years. Local differential privacy mechanism is one of the mainstream geographic location privacy protection models. among ε-Geo-indistinguishability has become the basic research standard of local differential privacy geographic location protection scheme using distance measurement. Combining the ideas of ε-Geo-indistinguishability and Utility-Optimized Local Differential Privacy, we propose a differential privacy concept Utility-Optimized- ε-Geo-indistinguishability which meets the different degrees of privacy protection in different geographical locations. This is the first time in the field of local differential privacy geographic location protection based on distance measurement. At the same time, we propose a localized differential privacy mechanism called Utility-Optimized Planar Laplace Mechanism that can meet Utility-Optimized-ε-Geo-indistinguishability. Theoretical analysis and experiments based on real data sets show that the experimental effect of our proposed mechanism is better than the existing local differential privacy geographic protection mechanism.

Keywords: Geo-indistinguishability, location privacy, differential privacy, utility

1. INTRODUCTION

In the environment of the interconnection of all things, the rapid development of mobile communication technology and mobile computing have made earth shaking changes in people's life. Mobile devices capable of networked communication can freely send and receive data in a wireless environment, which brings users a convenient and fast service experience. In recent years, mobile smart devices have become a standard configuration for urban residents. Mobile phones, tablets and smart mobile wearable devices such as smart watches have become an indispensable part of people's modern life. Many mobile phone giants have launched their own mobile smart wearable devices and iterated quickly, such as apple, Huawei, Xiaomi and so on. According to the statistics of well-known research institute strategy analytics[1], in the second quarter of 2021 alone, the global shipment of smart watches has reached 18 million units, and the global shipment of smart phones is 314.2 million. It can be seen that mobile smart devices have penetrated into thousands of households. At the same time, according to the latest research report of Emarketer[2], in 2020, the average time spent on mobile phones by Chinese people will reach 3 hours and 16 minutes a day, and that of Americans will reach 3 hours and 10 minutes a day, which is enough to say that mobile smart devices have established a solid and inseparable relationship with human daily life. The user carries the smart device with him. According to the data obtained by the sensor of the smart device, the user can determine his location in real time, and then use it to interact with the location-based service application[3-4], This kind of location-based service (LBS) has brought people a convenient and practical user experience and is deeply favoured by users. Additionally, location information is useful for some commercial applications. For example, the number of electric vehicles at charging piles could be used to support business decision. However, electric vehicle location may contain personal private information, which cannot be collected directly.

Differential privacy mechanism based on distance measurement is one of the main methods of differential privacy in geographic location privacy protection. ε-Geo-indistinguishability as the de facto standard of local differential privacy geographic location protection scheme using distance measurement, the Plane Laplace mechanism to realize its requirements has become the basis of the existing local differential privacy protection mechanism based on distance measurement. Researchers at home and abroad have made many improvements to its original definition and put forward

* 504531816@qq.com

Third International Conference on Computer Science and Communication Technology (ICCSCT 2022)
edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12506, 125061Q
 · 0277-786X · doi: 10.1117/12.2662560

many meaningful schemes For example, in terms of utility, a new randomization algorithm based on linear programming technology is proposed in[5] In terms of the correlation between individual user locations in the set of repeated locations, Lan *et al.*[6] introduced an R-tree, which caches previously published confusion locations for reuse in future versions[7] Another method combined with relevance is introduced by Zhang *et al.*[8], which can reduce the consumption of privacy budget and improve the degree of privacy protection by replacing the actual location with the previously published predicted location.

The Utility-Optimized Local Differential Privacy (ULDP)[9] proposed by Murakami provides us with an inspiration, that is, the existing researchers have not noticed a common fact, that is, different geographical locations have different sensitivities and different privacy protection needs. Combining the ideas of ε-Geo-indistinguishability and Utility-Optimized Local Differential Privacy, we propose a differential privacy concept called Utility-Optimized-ε-Geo-indistinguishability which meets the different degrees of privacy protection in different geographical locations. And we propose a localized differential privacy mechanism called Utility-Optimized Planar Laplace Mechanism (UPL).

The content of this paper is arranged as follows. Section I introduces relevant work at home and abroad; Section II introduces the preparatory knowledge required; Section III proposes Utility-Optimized-ε-Geo-indistinguishability and its implementation mechanism UPL; Section IV compares UPL with existing mechanisms through experiments; Section V summarizes and prospects.

2. PRELIMINARIES

Please follow these instructions as carefully as possible so all articles within a conference have the same style to the title page. This paragraph follows a section title so it should not be indented.

2.1 Differential privacy

Differential privacy[10] is defined by Dwork as a formal and provable privacy guarantee. The idea of this mechanism is that the aggregation results calculated for the dataset should be almost the same whether there is a single element in the dataset or not. In other words, the addition or deletion of a single element should not significantly change the probability of any result of the aggregation function. The differential privacy model assumes that the attacker has the maximum background knowledge and the maximum attack ability, and gives a rigorous and quantitative representation and proof of the risk of privacy disclosure, which is defined as follows.

A randomization mechanism $\mathcal{M}: D \to O$ satisfies $\varepsilon -$Differential Privacy, if and only if, for any two adjacent data sets D and D' with only one record difference and any possible output $o \in O$, the following inequality is satisfied:

$$\frac{\Pr\{\mathcal{M}(D) = o\}}{\Pr\{\mathcal{M}(D') = o\}} \leq e^{\varepsilon} \tag{1}$$

2.2 Local differential privacy

The differential privacy technology using centralized architecture is called Centralized Differential Privacy (CDP). These methods need a trusted third party to collect users' data and add noise to them for disturbance processing to ensure users' data privacy, but this assumption is unrealistic in practical application. In this context, Local Differential Privacy (LDP)[11] is proposed and widely used. In LDP, users add noise to their original data locally, and then send the disturbed data to the data collector for further processing. The model has good practicability without the participation of a trusted third party. At the same time, the user's real data is not local, which can ensure the user's data privacy. The formal definition of LDP is as follows.

Given the input set X, the output set Y, any $x, x' \in X, y \in Y$, the randomization mechanism $\mathcal{M}$ is $\varepsilon -$ LDP, if it satisfies

$$\Pr[\mathcal{M}(x) = y] \leq e^{\varepsilon} \cdot \Pr[\mathcal{M}(x') = y] \tag{2}$$

2.3 Utility-optimized local differential privacy

Utility-Optimized Local Differential Privacy[9] is a differential privacy mechanism proposed by Murakami et al. Aiming at the common fact that the degree of privacy protection between data is different, the mechanism is improved on the basis of local differential privacy, so as to optimize the effect of privacy protection. The traditional local differential privacy mechanism regards the sensitivity of all user data as the same, which will lead to excessive interference to user

data. In fact, not all user data is sensitive. Let's give an example of this view. Suppose people need to answer a question now, which involves privacy, such as "do you have depression" or "have you ever violated the law". Obviously, "yes" is a sensitive answer. We need to protect the person who answers, while "no" is insensitive. We don't need to protect their privacy. In this context, Utility-Optimized Local Differential Privacy came into being. The mechanism divides user data into sensitive data and non-sensitive data. The sensitive data mechanism is protected to meet the general definition of local differential privacy, but the non-sensitive data is not protected. Its formal definition is as follows.

Given the input set X and the output set Y, where the sensitive input is Xs, the non sensitive input is Xn, the sensitive output is Yp and the non sensitive output is Yi, the randomized scrambling mechanism $\mathcal{M}$:

(1) For any output $y \in Yi$, there is an input $x \in Xn$ such that $\Pr[\mathcal{M}(x) = y] > 0$ and any $x' \neq x$ has $\Pr[\mathcal{M}(x') = y] = 0$;

(2) For any output $y \in Yp$ and any input x, $x' \in X$, there is $\Pr[\mathcal{M}(x) = y] \leq e^{\varepsilon} \cdot \Pr[\mathcal{M}(x') = y]$.

Then it is said that the perturbation mechanism $\mathcal{M}$ meets ε- ULDP.

2.4 Geo-indistinguishability

ε-Geo-indistinguishability[12] is a privacy concept for geographic location data privacy protection proposed by Andres et al. It believes that the privacy protection level of user data should be related to distance. This concept is an extension of the traditional differential privacy. The core idea is that within the specified range, the user's real geographical location and approximate location are indistinguishable. Its formal definition is as follows.

Given the geographic location data set X, the data set output after being disturbed by the randomization scrambling mechanism $\mathcal{M}$ is Z, if the randomization mechanism $\mathcal{M}$ is satisfied ε-Geo-indistinguishability, then for any$x, x' \in X$,$z \in Z$, and the Euclidean distance $d(x, x') \leq d_{max}$, d_{max} is the protection scope of the mechanism:

$$\mathcal{M}(x)(z) \leq e^{\varepsilon d(x,x')}\mathcal{M}(x')(z) \tag{3}$$

where $\mathcal{M}(x)(z)$ represents the probability of inputting position point x to mechanism $\mathcal{M}$ to obtain position point z, ε is the budget for privacy protection.

3. MECHANISM DESIGN

The existing scheme does not take into account the fact that different geographical locations have different sensitivities and different privacy protection needs. To solve this problem, we propose Utility-Optimized-ε-Geo-indistinguishability. Further, we propose its implementation mechanism called Utility-Optimized Planar Laplace Mechanism (UPL). The mechanism divides the geographical location into sensitive areas and non-sensitive areas to provide better utility.

3.1 Utility-optimized-ε-geo-indistinguishability

Given the geographic location point input set X and the geographic location point output set Y, where the sensitive geographic location point input is Xs, the non-sensitive geographic location point input is Xn, the sensitive geographic location point output is Yp and the non-sensitive geographic location point output is Yi, the randomized scrambling mechanism $\mathcal{M}$:

(1) For any output $y \in Yi$, there is an input $x \in Xn$ such that $\Pr[\mathcal{M}(x) = y] > 0$ and any $x' \in Xn, x' \neq x$ has $\Pr[\mathcal{M}(x') = y] = 0$;

(2) For any output $y \in Yp$ and any input x, $x' \in X$, $d(x, x') \leq d_{max}$, there is $\Pr[\mathcal{M}(x) = y] \leq e^{\varepsilon d(x,x')} \cdot \Pr[\mathcal{M}(x') = y]$.

Then it is said that the perturbation mechanism $\mathcal{M}$ meets Utility-Optimized-ε-Geo-indistinguishability.

3.2 Utility-optimized planar Laplace mechanism

We modified the planar Laplace mechanism and designed Utility-Optimized Planar Laplace Mechanism. The processing flow is as follows.

(1) Pretreatment.

The two-dimensional plane map is divided into regions, and each location point corresponds to an area of the divided map. Set the sensitive area set corresponding to the sensitive location point set Xs as A and the non-sensitive area set corresponding to the non-sensitive location point set Xn as B.

(2) Calculating the privacy budget.

According to the input position point set X, it is determined that the upper limit d_{max} of the diameter range is the maximum value of the distance between any two points in the position point set X, and the privacy budget ε' satisfied by the mechanism is determined according to the privacy budget ε, the radius length precision δ_r in the grid $\mathcal{G}$, the machine precision of the angle δ_θ and the step unit u of the grid $\mathcal{G}$.

$$\varepsilon' + \frac{1}{u} ln \frac{q + 2e^{\varepsilon' u}}{q - 2e^{\varepsilon' u}} \leq \varepsilon \tag{4}$$

(3) Calculating angle.

To determine the angle of the disturbed position point z relative to the initial position point x, we calculate the angle $\theta = unif[0,2\pi)$.

(4) Calculating length.

Determine the radius r of the disturbed position point z relative to the initial position point x. The formula is as follows, W_{-1} is -1 branch of Lambert W function.

$$t = unif[0,1), \qquad r = \frac{-1}{\varepsilon}\left(W_{-1}\left(\frac{t-1}{e}\right) + 1\right)$$

(5) Determining location point.

Determine the position point z after disturbance that $z = x + \langle rcos(\theta), rsin(\theta)\rangle$.

(6) Output.

When the input position point $x \in Xs$ directly output the position point corresponding to the area where the position point z is located; When the input position point $x \in Xn$, if the disturbed position point $z \in A$, the position point corresponding to the area where the position point z is located is output; otherwise, the initial position point x is output.

3.3 Proof

Here, we prove that UPL satisfies Utility-Optimized-ε-Geo-indistinguishability.

In step 6, When the input position point $x \in Xn$, if the disturbed position point $z \in B$, the initial position point x is output. Then first point of Utility-Optimized-ε-Geo-indistinguishability is satisfied.

When the input position point $x \in Xs$ directly outputs the position point corresponding to the area where the position point z is located; When the input position point $x \in Xn$, if the disturbed position point $z \in A$, the position point corresponding to the area where the position point z is located is output. This is exactly the same as Planar Laplace Mechanism which meet ε-Geo-indistinguishability, and ε-Geo-indistinguishability is the second point of Utility-Optimized-ε-Geo-indistinguishability.

4. EXPERIMENT

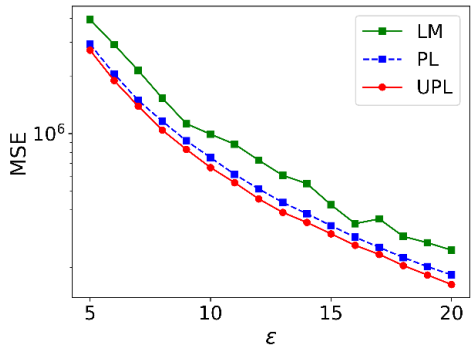
The data sets used in the experiment are yelp Las Vegas data set and the generated uniform data set.

We compare the utility optimized planar Laplacian mechanism UPL with Laplacian mechanism LM and planar Laplacian mechanism PL in the above data sets. The experimental index is utility, i.e., MSE:

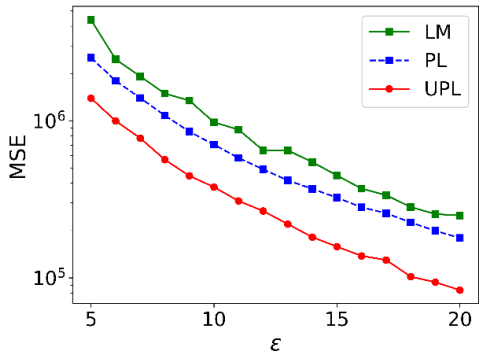
$$MSE(p, p') = \frac{1}{2}\|p - p'\|_2^2 = \frac{1}{2}\sum_{i=1}^{2}(p[i] - p'[i])^2 \tag{5}$$

We set that there is only one sensitive area and it is located at the center of the two-dimensional plane. The side length of the sensitive area is 1/2 of the side length of the whole area. We take the privacy budget ε range of 5 to 20 as the abscissa

of the experiment results. The results are shown in Figure 1. Affected by the real geographical distribution, the curves of the three experimental results are slightly different, but it can be found that the effect of UPL is always better than LM and PL.



(a) Experimental comparison of Yelp map dataset



(b) Experimental comparison of the generated uniform dataset

Figure 1. Experimental comparison of mechanism in each data set.

There are usually multiple sensitive areas in the real geographical location plane. Therefore, next, we conduct experiments on the situation that there are multiple sensitive regions in two-dimensional plane region. We set the number of sensitive areas n as 6, and the side length of each sensitive area is 1/8 of the side length of the whole area. The location of the sensitive area is randomly distributed. Repeat the calculation for 10 times and take the average value. The experimental results on the Yelp dataset are shown in Figure 2. It can be seen that in the case of multi sensitive regions, the utility of UPL is still better than other mechanisms.

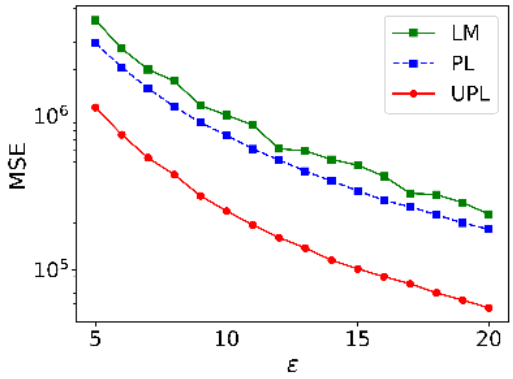


Figure 2. Experimental comparison of various mechanisms in the case of multiple sensitive regions.

5. CONCLUSION

This paper extends the previous research work, first puts forward Utility-Optimized-ε-Geo-indistinguishability, and then designs Utility-Optimized Planar Laplace Mechanism. Theoretical analysis and experiments based on real geographic location data sets show that UPL meets the geographic location indistinguishability of utility optimization, and it is better than other geographic location protection mechanisms that meet the geographic location indistinguishability.

ACKNOWLEDGMENT

This work is supported the science and technology project of State Grid Corporation of China: “Research and Application of Scenario-Driven Data Dynamic Authorization and Compliance Control Key Technology” (Grand No. 5700-202058481A-0-0-00).

REFERENCES

[1] Sui, L., "Global Smartphone Shipments Grew +11% YoY in Q2 2021," Strategy Analytics, 1-2(2021).
[2] Cheung, M.-C., "China Time Spent with Media 2020", eMarketer, 1-3(2020).
[3] Wang, L., Yang D. and Han, X., "Location privacy preserving task allocation for mobile crowd sensing with differential geo-obfuscation," Proc. the 26th International Conference on World Wide Web, 627-636(2017).
[4] To, H. Shihabi, C. and Xiong, L., "Privacy-preserving online task assignment in spatial crowdsourcing with untrusted server," Proc. 34th International Conference on Data Engineering, 833-844(2018).
[5] Smith, S., "Location-based (LBS) service market size", Bfortune Business Insights, 2-5(2019).
[6] Lan, L., Bao, Z. C. and Shu, X. S., "Trajectory position protection algorithm of exchange query under identity Authentication," Journal of Chinese Computer Systems, 42(6), 1340-1344(2021).
[7] Ran, C. X., Lan, T. X. and Long, C. W., "Roadside unit deployment mechanism for urban vehicular networks," Journal of Chinese Computer Systems, 42(3), 140-144(2021).
[8] Zhang, D., Wang, L. and Xiong, H., "4W1H in mobile crowd sensing," IEEE Communications Magazine, 52(8), 42-48(2014).
[9] Takao M., and Yusuke, K., "Utility-optimized local differential privacy mechanisms for distribution estimation," Proc. the 28th USENIX Security Symposium, 1877-1894(2018).
[10] Dwork, C., "Differential privacy," Proc. the 33rd International Colloquium on Automata, Languages and Programming, 16-33(2006).
[11] Duchi, J. C., Jordan, M. I. and Wainwright, M. J., "Local Privacy and Statistical Minimax Rates," Proc. FOCS, 128-140(2013).
[12] Andres, M. E., Bordenabe, N. E. and Chatzikokolakis, K., "Geo-indistinguishability: Differential privacy for location-based systems," Proc. the 2013 ACM SIGSAC conference on Computer & Communications Security, 901-914(2013).