ISSN 1330-3651(Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20170415165405

DIFFERENTIALLY PRIVATE REAL-TIME DATA RELEASE BASED ON THE MOVING AVERAGE STRATEGY

Mengang Li, Daiyong Quan, Lu Yu

Original scientific paper
With the development and popularization of mobile-aware service systems, it is easy to collect contextual data such as activity trajectories in daily life. Releasing real-time statistics over context streams produced by crowds of people is expected to be valuable for both academia and business. However, analysing these raw data will entail risks of compromising individual privacy. ε-Differential Privacy has emerged as a standard for private statistics publishing because of its guarantee of being rigorous and mathematically provable. In the mobile-aware service systems, the ultimate goal is not only to protect the user's privacy, but look for a good balance between privacy and utility. To this end, we propose a flexible *m*-context privacy model to ensure user privacy under protection of ε-differential privacy. Experiments using two real-life datasets show that our proposed dynamic allocation of the privacy budget with moving average approximate strategy can work efficiently to release privacy preserved data in real-time.

Keywords: *differential privacy; dynamic allocation; context privacy protection; moving average approximate strategy*

Oslobađanje diferencijalno privatnih podataka u realnom vremenu zasnovano na pokretnoj prosječnoj strategiji

Izvorni znanstveni članak
S razvojem i popularizacijom mobilno-svjesnih (mobile-aware) uslužnih sustava lako je prikupiti kontekstualne podatke kao što su putanje aktivnosti u svakodnevnom životu. Očekuje se da će objavljivanje postojećih statističkih podataka o kontekstualnim strujanjima koje proizvode mase ljudi biti od važnosti i za znanstvenike i za poslovne ljude. Ipak, analiza tih neobrađenih podataka može dovesti do kompromitiranja individualne privatnosti. ε-Differential Privacy pojavila se kao standard za objavljivanje privatnih statističkih podataka zbog toga što garantira preciznost i matematičku dokazivost. Kod mobilno-svjesnih uslužnih sustava krajnji cilj je ne samo zaštita korisnikove privatnosti već i stvaranje balansa između privatnosti i korisnosti. Imajući to u vidu mi predlažemo fleksibilni *m*-kontekst model privatnosti u svrhu osiguranja privatnosti korisnika pod zaštitom ε-diferencijalne privatnosti. Eksperimenti s dva niza podataka iz stvarnog života pokazuju da predložena raspodjela privatnog budžeta primjenom pokretne prosječne aproksimativne strategije može biti efikasna kod objavljivanja privatnih podataka u realnom vremenu.

Ključne riječi: *diferencijalna privatnost; dinamička raspodjela; pokretna prosječna aproksimativna strategija; zaštita konteksta privatnosti*

1 Introduction

Currently, mobile-aware service systems are dramatically increasing the amount of personal data released to service providers as well as to third parties. In order to monitor real-time environmental changes, or to reduce the traffic congestion situation in large cities, many mobile awareness system real-time release aggregate information. Personal data have been increasingly collected, stored, and analysed. These real-time aggregate data (similar to "how many people in the Shanghai Bund?") can be provided to government departments as a basis for preventing public safety events, but also for other mobile-aware users to query or share, as the users decide whether to go out with reference.

However, the information with timestamps is sensitive. It may reveal the location of the commuter, the patient suffering from the type of disease. The availability of locations in real time as well as the historical data about user movements even introduces threats such as assault. In order to protect user privacy, these personal data are simply anonymous when real-time aggregate data is released. However, recent research [1] found that even with anonymous techniques, it is still possible to identify individual identities on a very high probability [2].

The privacy of mobile device users is more fragile, and De Montjoye's research [2] shows that 95 % of users can be identified by randomly selecting four time points from anonymous mobile data sets. In order to improve the user's experience, the users are needed to provide more contextual information as successive spatiotemporal points. More contextual information can help the attacker to guess the user's privacy. A method which is good at trade-off between privacy and utility is needed. And differential privacy is such a privacy protection method. ε-differential privacy (DP) has emerged as a de facto standard for privacy preserving data publishing (PPDP) because of rigorous theoretical guarantees [3, 4]. It ensures that the modification of any single record does not have a significant effect on the outcome of analysis. ε is a positive parameter called privacy budget which is given in advance to control the privacy level. The value of ε is inversely propositional to the privacy level.

To achieve better overall utility, there are several publishing strategies. Statistics on data stream publishing as an approximation strategy has been investigated in earlier research [5, 8]. Instead of directly adding noise to real data, they function by transformation of original data or a query structure to achieve better overall utility. Another major strategy is choosing an appropriate noisy data which was previously published and republishing it if it is "close to" the real statistics which we want to publish. How close the real and noisy data will be measured by MAE. This strategy can be divided into two sub-strategies: the first one is to simply employ the adjacent noisy data (can be shorted as Adj). The second one is to search the most similar noisy data on the timeline (shorted as MMD). These two strategies are dull; we need a flexible strategy, so that data publishers have more strategies to choose.

We propose a moving average strategy (abbreviated as MA(k)) in this paper. By searching for multiple approximate noise data on the timeline, the average of these noise data is the most similar noisy data. The

parameter k is the backtracking interval. On the real dataset, the results reveal that our MA strategy is the generalization of the above strategies of the Adj and MMD.

The contributions of this paper are threefold:

- We propose a way to dynamically allocate privacy budgets. It is sufficient to adaptively adjust privacy budget allocation dependent on underlying data distribution to achieve good performance.
- We propose a moving average strategy that is the generalization of existing strategies. When $k = 1$ our strategy degrades into the Adj strategy, and when $k = t-1$, our strategy degrades into the MMD strategy. This provides a flexible way to improve the utility of published data.
- We evaluate our strategy on the real dataset. Moreover, we compare our strategy with the existing strategies. The results reveal that our strategy is the generalization of existing strategies.

The rest of this paper is organized as follows. In Section 2, we describe our definitions, notations, and assumptions. The proposed privacy model is described in Section 3. Section 4 is the experiment of our proposed strategy on the real dataset. Section V presents the related work. Finally, Section 4 provides our concluding remarks.

2 Problem definition

In this section, we present a way to dynamically allocate privacy budgets. Then we describe our definitions, notations, and theorems.

2.1 Preliminary

Differential privacy was proposed by Dwork et al. [9]. According to the original definition of differential privacy [9], D, D' are two possible neighbour databases that differ in one row that is modified. A randomized function K (that acts as the privacy protection mechanism) provides ε-differential privacy. R represents all possible outputs of K. K satisfies ε-differential privacy, if for any $r \in R$ and any two neighbour databases D, D', we have the following.

$$\Pr\left[K(D) = r\right] \leq \exp(\varepsilon) \cdot \Pr\left[K(D') = r\right] \quad (1)$$

In Inequality (1), ε is a privacy budget given in advance. It is used to control the privacy level. When $\varepsilon = 0$, it means that there is no disclosure. We achieve the perfect privacy protection and the attacker's reasoning attack results are the same as random guesses.

A widely used method to achieve differential privacy is the Laplace mechanism [10], which adds random noise to actual data to prevent the disclosure of sensitive information. The amount of noise added to achieve the differential privacy is closely related to global sensitivity. Sensitivity reflects the effect of input changes on the output. For any function $f : D \to R^d$, the global sensitivity is defined as: $\Delta f = max_{D_1,D_2} \|f(D_1) - f(D_2)\|$.

Where D_1, D_2 are two possible neighbour databases that differ in one row. d represents the query dimension of function. R representing the mapped real space. For any function $f : D \to R^d$, if the output of the algorithm satisfies the following equation: $A(D) = f(D) + \langle Lap_1(\Delta f/\varepsilon), \cdots, Lap_d(\Delta f/\varepsilon)\rangle$, then we can achieve ε-differential privacy where $Lap_i(\Delta f/\varepsilon)(1 \leq i \leq d)$ is the Laplace variable which is independent of each other.

The user context is where the user publishes the data on a specific timestamp. For example, suppose a traffic service that regularly publishes the number of passengers at each location (real-time statistics), and the attendant's presence at a particular location is on a given timestamp. Similarly, in the real-time statistical hottest topic, the participation of the user in the social platform on the topic is the context.

Let t be the current timestamp, $set_of_context$be all the context of the collection, and $|set\_of\_context| = C\#$ be the total number of contexts, set_of_users be the set for the total number of users, $|set\_of\_users| = U\#$ be the total number of users. For any timestamp $i \in [1, t]$, the corresponding contextual data table is C_i and the corresponding statistical release of the real value is r_i. The length of the $U\#$ user's context sequence is $\mathrm{m} = \langle \mathrm{m}1, \cdots, m_{U\#}\rangle$. At the timestamp t, the u user's context stream can be expressed as $c_{u,t} = \left\{\langle u, i, status\rangle, \cdots, \langle u, i, status'\rangle\right\}$, which contains a valid timestamp set of $T_{u,t}$.

2.2 Definitions

In order to define m context privacy in accordance with standard differential privacy, we must first clarify the relationship between the data and the definition of the adjacent context prefix on each timestamp.

Definition 2.1 (Neighbouring dataset at each timestamp) If two datasets C_i, C_i' are collected at timestamp $i \in [1, t]$ and differ in a single status of user u, then we say that C_i, C_i' is a pair of neighbouring datasets with respect to u.

For the infinite stream of the adjacent relationship, we use the stream prefix to represent. A context stream prefix corresponds to all data of the infinite context streams up to the current timestamp t. That is, the context stream prefix $S_t = \{C_1, \cdots, C_t\}$corresponds to the infinite context steam $\{C_1, \cdots, C_i, \cdots\}$ to all data for the current timestamp t.

$$K(z|x) \leq e^{\varepsilon \cdot d(x,x')} \cdot K(z|x') \quad (2)$$

Definition 2.2 (m-context neighbouring stream prefixes): Let$S_t = \{C_1, \cdots, C_t\}$ and $S_t' = \left\{C_1', \cdots, C_t'\right\}$ be two context stream prefixes ending with the current timestamp t. S_t and S_t' are m-context stream prefixes neighbouring each other if one is obtained from another by modifying all status in any one m-context $c_{u,k} = \left\{\langle u, i, status\rangle, \cdots, \langle u, k, status'\rangle\right\}$. We say that S_t and S_t' are neighbouring with respect to $c_{u,k}$.

Definition 2.3 (m-context ε-differential privacy): Let M be an algorithm that takes prefixes of context streams

$S_t = \{C_1, \cdots, C_t\}$ as inputs. Let $O_t = \{o_1, \cdots, o_t\}$ be a possible perturbed output stream of M. If for any m-context neighbouring S_t and $S_t^{'}$, the following holds,

$$\Pr\left[M\left(S_t\right) = O_t\right] \leq e^{\varepsilon} \cdot \Pr\left[M\left(S_t^{'}\right) = O_t\right] \tag{3}$$

Then we say that M satisfies m-context ε-differential privacy.

Differential privacy protection mechanism strikes the balance between the protection level and the data availability (utility). In this paper, we use the mean absolute error MAE as our usability measure.

Definition 2.4 (Utility metrics): The $set\_o\_fcontext$ is the set of the context. Its size is $|set\_o\_fcontext| = C\#$. r_i and o_i are the real statistics and noise value on the time stamp i, respectively, and R and O are the true statistical and noise values on all timestamps, respectively. Then, the average absolute error MAE on each timestamp is :

$$MAE\left(R, O\right) = \frac{1}{T \cdot C\#} \cdot \sum_{i=1}^{T} \sum_{j=1}^{C\#} \left|r_i\left(j\right) - o_i\left(j\right)\right| \tag{4}$$

3 Privacy model

In this section we present a m-context privacy model. Specifically, to satisfy m context privacy, the sum of the privacy budget assigned to any single m context must not be greater than the total privacy budget ε. Fig. 1 illustrates the model that we assume in this paper.

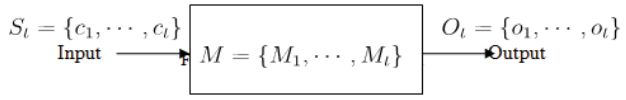

Figure 1 *m*-context privacy model

3.1 *m*-context privacy model

Theorem 1. Let M be an integrated algorithm which takes prefixes of streams $S_t = \{c_1, \cdots, c_t\}$ as inputs, and $O_t = \{o_1, \cdots, o_t\}$ as outputs. M consists of a series of sub mechanisms $\{M_1, \cdots, M_t\}$, each M_i takes c_i as inputs, and outputs noisy data o_i with independent randomness. Presume M_i satisfies ε-differential privacy and ε_i is a privacy budget of M_i, if the following inequality holds,

$$\forall u, \forall k, \sum_{i \in \tau_{u,k}} \varepsilon_i \leq \varepsilon \tag{5}$$

then M satisfies m-context ε-differential privacy.

3.2 Methodology to achieve *m*-privacy

This section details the dynamic allocation of the privacy budget of our proposed m-privacy mechanism (satisfying ε- differential privacy).

The simplest solution is to distribute the privacy budget evenly within each context of the m window. Literature [5, 10] originally proposed the method, referred to as UNIFORM. UNIFORM will serve as a benchmark for our approach of dynamic allocating the privacy budget.

There is no way to optimize evenly distributed. We propose a dynamic allocation scheme similar to that of [5] [10]. The specific algorithm is as follows:

Algorithm 1: Dynamic Budget Allocation: DA

Input: context stream C_1,…, C_t; users set {us_1,…, us_t}; Protected context length m; Total privacy budget ε; Previously allocated budget ε_1,…, ε_{t-1};

Output: Noisy statistics o_t; Allocated budget ε_t on the timestamp t;

1. Allocate fixed budget $\varepsilon_{t,1} \leftarrow \varepsilon/(2*m)$, and initialize the temporary budget $\varepsilon_{t,2} \leftarrow 0$
2. Initialize the spent budget $\varepsilon_{spend} \leftarrow 0$
3. For each user u, calculate the following:
the context stream $C_{u,t}$ at time t, the timestamp set $T_{u,t}$ contained in the context stream, all the privacy budget $\varepsilon_u \leftarrow \sum \varepsilon_i, i \in T_{u,t}$ in the timestamp set $T_{u,t}$, and save the largest ε_u to ε_{spent} ,i.e., $\varepsilon_u = \max_{u \in US_t} \left\{ \sum_{i \in \mathrm{T}_{u,t}} \varepsilon_i \right\}$, $\varepsilon_{spent} \leftarrow \varepsilon_u$
4. Calculate the remaining privacy budget: $\varepsilon_t^r \leftarrow \varepsilon/2 - \varepsilon_{spent}$
5. Allocate in an exponentially decreasing manner: $\varepsilon_{t,2} \leftarrow \varepsilon_t^r/2$
6. $\varepsilon_t \leftarrow \varepsilon_{t,1} + \varepsilon_{t,2}$
7. Moving Average republic Strategy
8. return o_t,ε_t

3.3 Moving average republic strategy

In order to save privacy budget allocation, we propose a moving average strategy. If the real statistics on the current timestamp are similar to the previously released noise data, we assume that the noise data (Or a combination of these noise data) is "fit" to republish. Thereby we can save the privacy budget allocation on that timestamp. Specifically, at the time stamp t, we use the moving average of the statistics that have been published for a period of time to approximate the noisy statistics to be published. The moving average method is suitable for near-term forecasting. The moving average of the published noise data is defined as follows:

$$\overline{o}_{t,k} = \frac{1}{k} \sum_{i=0}^{k-1} o_{t-1} \tag{6}$$

where the parameter k is the length of the backward searching. The key condition for triggering our strategy is the distance between the moving average and the statistical value to be published. If it is below a certain threshold, it is republished with the moving average of the most recent published statistics, avoiding the privacy budget allocation on that timestamp, thus avoiding the addition of noise and improving the availability of the published value. The distance is expressed as:

$$dis = MAE\left(r_t, \overline{o}_{AM}\right) + Lap\left(2/(m * \varepsilon_{t,1}/2)\right) \tag{7}$$

With the distance value, our moving average republic strategy is shown in Algorithm 2:

Algorithm 2: Moving Average Republic Strategy: MA(k)

Input: Real statistics r_t; Parameter k; Privacy budget $\varepsilon_{t,1}$ and $\varepsilon_{t,2}$;

Output: Noisy data o_t or Moving average approximation $\overline{o}_{AM}$;

1. Back search for the most approximate value $\overline{o}_{AM}$ between the moving average of $o_{t-k}, \cdots o_{t-1}$ and the real statistical value of r_t

2. Calculate the distance

$dis = MAE\left(r_t, \overline{o}_{AM}\right) + Lap\left(2/(m * \varepsilon_{t,1}/2)\right)$

3. if

$dis \leq 2/(m * \varepsilon_{t,2})$ then $o_t \leftarrow o_{AM}$

else

$o_t \leftarrow r_t + \left\langle Lap\left(\frac{2m}{\varepsilon_t}\right)\right\rangle^{|C\#|}$

4. return o_t

Algorithm 2 involves a search query that requires access to k real statistics, which requires a private budget $\varepsilon_{t,1}/2$. This is similar to the approximation strategy Adj [5], and it is different from the MMD[10].

4 Experiments

In this section, we design experiments to evaluate our proposed algorithm described in the previous sections and use linear distribution as a baseline to compare our proposed moving average republic strategy with the Adj and MMD strategies. Our ultimate goal is not only to protect the user's mobility patterns privacy, but also to look for a good balance between privacy and utility.

4.1 Dataset

We perform our evaluation on the widely used dataset: GeoLife [11, 12, 13]. This trajectory dataset can be used in many research fields, such as mobility pattern mining, user activity recognition, location-based social networks, location privacy, and location recommendation. The GeoLife GPS Trajectories dataset contains 17621 traces from 182 users, moving mainly in the northwest of Beijing, China, in a period of over three years (from April 2007 to August 2012). Another dataset is the T-Drive [16]. This data set is shared by Prof. Xie Xun and Prof. Zheng Yu [14, 15], researchers at Microsoft Asia Research Institute. The data set contains 10 357 taxis a week of track samples, the total number of points is about 15 million, the total distance of the track to reach 9 million km.

Table 1 Dataset

Type #	Table Column Head		
	Timestamps Amt.	*Data Points Amt.*	*Users Amt.*
Geolife	1440	240 990	170
T-Drive[a]	886	37 225	2698

4.2 Utility evaluation experiment 1

In the case where the overall budget is fixed to 1, we fix the daily publication value of 15, which means that an experiment spans 96 timestamps and m ranges from 10 to 100.

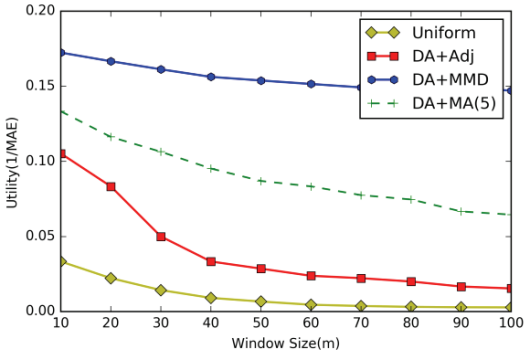

Figure 2 Comparison of the Utility by varying m

Fig. 2 illustrates that larger the m, the more timestamps are allocated, the smaller the privacy budget allocated on each timestamp, and the higher the privacy level, and the greater the amount of noise added, the worse the utility . The yellow curve in the lower left part of Fig. 2 is the linear distribution of the privacy budget using UNIFORM as our utility baseline. Intuitively, the utility curves (red line, green lines, and blue line) for dynamic allocation re-publishing strategies are above the baseline. This means that the data utility with the dynamic allocation privacy budget and approximate strategy is better.

The utility curve (red curve) that used the Adj strategy is lower than the utility curve (blue curve) that used the MMD strategy. The reason is that the adjacent release of the value is not always the most approximate value.

The green curve in the middle part of Fig. 2 is the utility curve obtained by our proposed MA strategy, which illustrates that the utility is better than the Adj strategy, and is inferior to MMD strategy.

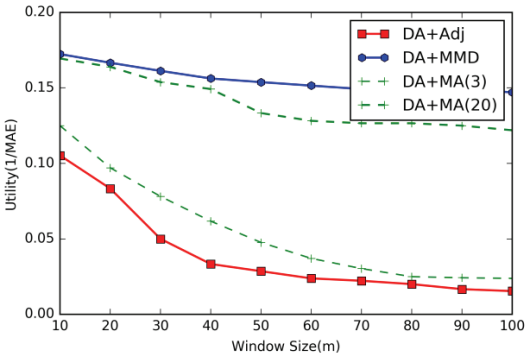

Figure 3 Comparison of the Utility by varying m and k

In Fig. 3, with the greater k value (k = 20), the green dotted line is close to the blue line. With the smaller k value (k = 3), the green dotted line is closer to the red line. In fact, when k = 1, our strategy degrades into the Adj strategy, $k = t - 1$, our strategy degenerates into the MMD strategy. This provides a flexible way to select the approximate strategy.

4.3 Utility evaluation experiment 2

In the case where the m is fixed, we evaluate the utility by varying the overall privacy budget. In our experiment 2, the values of the privacy budget are 0.0001, 0.001, 0.01, 0.1 and 1. Our ultimate goal is not only to protect the user's mobility patterns privacy, but also to look for a good balance between privacy and utility.

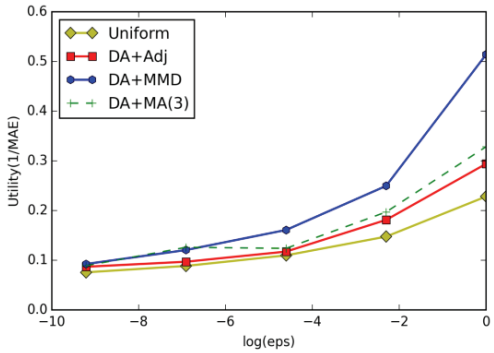

Figure 4 Comparison of the Utility by varying ε

In Fig. 4, the horizontal coordinates are the logarithm of the privacy budgetε, and the vertical coordinates are the logarithmic value that measures the utility of the MAE. As shown in Fig. 3, in the case of fixed m, the smaller the overall privacy budget, the higher the level of privacy protection, accordingly, the greater the amount of noise added, the worse the utility.

The yellow curve in the lower right part of Fig. 3 is the linearity of the distribution of the privacy budget using UNIFORM as our utility baseline. Intuitively, the utility curves are above the baseline, and they are using the dynamic allocation privacy budget, plus the re-publishing strategy.

As the value of the privacy budget is 0.0001, 0.001, 0.01, 0.1, and 1 (the trend is increasing), the utility of the dynamic allocation privacy budget and the approximate strategy under the same privacy guarantee is increasing, especially the MMD Strategy. The green line is the MA strategy that we proposed, which is between the blue and red lines. In addition, the green line is more similar to the red line (the Adj strategy).

4.4 Time complexity analysis

In order to compare our proposed strategy with the existing strategy, we conducted time complexity analysis. The time complexity of UNIFORM is $O(|C\#|)$. For strategy Adj, the time complexity is also $O(C\#)$ because of computing MAE. For strategy MMD, the most time-consuming operation is the comparing r_t with n_i where $i \in [1, t-1]$. Therefore, the overall complexity of MMD is $O(t * |C\#| + t * \log t)$. While our proposed strategy MA(k), the time complexity is $O(k * |C\#| + k * \log k)$. Obviously, when $k = 1$ our strategy time complexity equals $O(|C\#|)$, which is the time consuming for the strategy Adj, and when $k = t - 1$, our strategy time complexity equals $O(t * |C\#| + t * \log t)$, which is the time consuming for the strategy MMD.

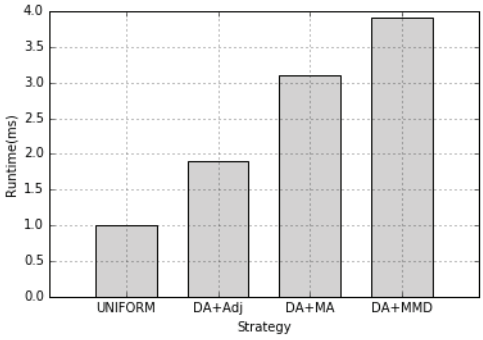

Figure 5 Runtime of each strategy on WorldCup98

In Fig. 5, the runtime of each strategy on real dataset shows that the performance of the proposed strategy is better than the MMD strategy but inferior to Adj strategy.

5 Related work

The literature related to DP provides rich results including application of DP to streaming data [3, 4, 5, 6]. In the setting of streaming data, differential privacy comes with two privacy definitions: user-level and event-level privacy [3], [4]. Roughly speaking, for trajectory data streams, user-level privacy means to protect the whole trajectory history of any user, and event-level privacy only promises to protect any single spatiotemporal data point. A new streaming data privacy model of w-event privacy [5] was proposed recently to strike a nice balance between two former privacy definitions. The model emphasizes protection of data points belonging to every w contiguous timestamps in a sliding window. W-event privacy is not sufficient to protect trajectory streams. [7] proposes a flexible privacy model of ℓ-trajectory privacy to ensure every length of ℓ trajectories under protection of ε-differential privacy. L-model is a flexible model that adopts a dynamic budget allocation based on approximation strategies (Adj and MMD as two different approximation strategies).

Approximation strategies have been investigated in earlier research, such as histogram publishing [15, 16, 17], and statistics on data stream publishing [18, 19]. Instead of directly adding noise to real data, they function by transformation of original data or a query structure to achieve better overall utility.

Literature [7] chooses an appropriate noisy data which was previously published and republish it if it is close to the real statistics. The Adj strategy is to simply employ the adjacent noisy data, and the MMD strategy is to search the most similar noisy data on the timeline. Both strategies use only a single data point and the overall utility remains to be further optimized. We can consider the combination of past points to improve the utility. Moreover, there is a need for a flexible strategy to accommodate these two strategies.

6 Conclusions

In the paper, we explored the potential of approximate strategy to dynamic allocation of the privacy budget over infinite context streams. We struck the balance between the privacy and utility.

First, we present an m-context privacy model which satisfies ε-differential privacy definition. By dynamic allocation of the privacy budget with moving average approximate strategy, our proposed DA+MA(k) can work efficiently to release privacy preserved data in real-time.

Second, we designed experiments to evaluate our proposed algorithm and compare our proposed moving average republic strategy with the Adj and MMD strategies. The experiments conducted with real dataset show that when $k = 1$ our strategy degrades into the Adj strategy, and when $k = t - 1$, our strategy degrades into the MMD strategy. It provides a flexible way to improve the utility of published data.

Third, we quantitatively evaluated the time complexity of our proposed strategy and the competitor strategies (Adj and MMD strategies). The result showed that our proposed strategy MA(k), the time complexity is $O\left(k * \left|C\#\right| + k * \log k\right)$. Obviously, when $k = 1$ our strategy time complexity equals $O\left(\left|C\#\right|\right)$, which is time consuming for the strategy Adj, and when $k = t - 1$, our strategy time complexity equals $O\left(t * \left|C\#\right| + t * \log t\right)$, which is time consuming for the strategy MMD.

As future work, we will explore more flexible approximate strategy. One interesting branch is how to specify different strategy to different sensitive data.

Acknowledgements

This work was supported by the Special Funding Project of Ministry of Education (B09C1100020). We would like to thank the anonymous reviewers for their very insightful and inspiring comments that helped to improve the paper.

7 References

[1] Sweeney, L. k-anonymity: a model for protecting privacy. World Scientific Publishing Co. Inc, 2002.

[2] Montjoye, Y. A. D.; Hidalgo, C. A.; Verleysen, M.; Blondel, V. D. Unique in the crowd: the privacy bounds of human mobility. // Scientific Reports. 3, 6(2013), p. 1376. https://doi.org/10.1038/srep01376

[3] Dwork, C. Differential privacy: a survey of results. // International Conference on Theory and Applications of MODELS of Computation, Springer-Verlag. Vol. 4978, (2008), pp. 1-19. https://doi.org/10.1007/978-3-540-79228-4_1

[4] Dwork, C. Differential privacy in new settings. // in SODA'10.

[5] Kellaris, G.; Papadopoulos, S.; Xiao, X.; Papadias, D. Differentially private event sequences over infinite streams. // Proceedings of the Vldb Endowment. 7, 12(2014), pp. 1155-1166. https://doi.org/10.14778/2732977.2732989

[6] Mir, D.; Muthukrishnan, S.; Nikolov, A.; Wright, R. N. Pan-private algorithms via statistics on sketches. // Thirtieth ACM Sigmod-Sigact-Sigart Symposium on Principles of Database Systems. (2011), pp. 37-48. https://doi.org/10.1145/1989284.1989290

[7] Barthe, G.; Danezis, G.; Gregoire, B.; Kunz, C.; Zanella-Beguelin, S. Verified Computational Differential Privacy with Applications to Smart Metering. // IEEE Computer Security Foundations Symposium. (2013), pp. 287-301. https://doi.org/10.1109/csf.2013.26

[8] Fan, L.; Xiong, L.; Sunderam, V. FAST: differentially private real-time aggregate monitor with filtering and adaptive sampling. // ACM SIGMOD International Conference on Management of Data. (2013), pp. 1065-1068. https://doi.org/10.1145/2463676.2465253

[9] Dwork, C. Differential privacy. // in Automata, languages and programming. Springer, (2006), pp. 1-12.

[10] Cao, Y.; Yoshikawa, M. Differentially Private Real-Time Data Release over Infinite Trajectory Streams. // IEEE International Conference on Mobile Data Management, IEEE Computer Society. 99, 1(2015), pp. 68-73. https://doi.org/10.1109/mdm.2015.15

[11] Zheng, Y.; Li, Q.; Chen, Y.; Xie, X.; Ma, W. Y. Understanding mobility based on GPS data. // ACM International Conference on Ubiquitous Computing. (2008), pp. 312-321. https://doi.org/10.1145/1409635.1409677

[12] Zheng, Y.; Zhang, L.; Xie, X.; Ma, W. Y. Mining interesting locations and travel sequences from GPS trajectories. // ACM International Conference on World Wide Web. (2009), pp. 791-800. https://doi.org/10.1145/1526709.1526816

[13] Zheng, Y.; Xie, X.; Ma, W. GeoLife: A Collaborative Social Networking Service among User, Location and Trajectory. // IEEE Data Eng. Bull. 33, 49(2010), pp. 32-40.

[14] Yuan, J.; Zheng, Y.; Xie, X.; Sun, G. Driving with knowledge from the physical world. // ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. (2011), pp. 316-324. http://research.microsoft.com/en-us/projects/tdrive/ https://doi.org/10.1145/2020408.2020462

[15] Xu, J.; Zhang, Z.; Xiao, X.; Yang, Y.; Yu, G.; Winslett, M. Differentially private histogram publication. // The VLDB Journal. 22, 6(2013), pp. 797-822. https://doi.org/10.1007/s00778-013-0309-y

[16] Zhang, X. Towards Accurate Histogram Publication under Differential Privacy. // Siam International Conference on Data Mining, 2014. https://doi.org/10.1137/1.9781611973440.68

[17] Zhang, X.; Meng, X.; Chen, R. Differentially Private Set-Valued Data Release against Incremental Updates. // DASFAA 2013**:** Database Systems for Advanced Applications. (2013), pp. 392-406.

[18] Chan, T. H.; Shi, E.; Song, D. Private and Continual Release of Statistics. // International Colloquium Conference on Automata, Languages and Programming, Springer-Verlag. (2011), pp. 405-417. https://doi.org/10.1145/2043621.2043626

[19] Fan, L.; Xiong, L.; Sunderam, V. FAST: differentially private real-time aggregate monitor with filtering and adaptive sampling. // ACM SIGMOD International Conference on Management of Data. (2013), pp. 1065-1068. https://doi.org/10.1145/2463676.2465253

Authors' addresses

Menggang Li
Transportation Engineering and Theoretical Economics,
Beijing Jiaotong University, Siyuan East Building,
Haidian District, Beijing 100044, China
E-mail: mgli@bjtu.edu.cn

Daiyong Quan
Data Privacy in Intelligent Traffic, China Centre for Industrial Security Research, Beijing Jiaotong University, 7th Teaching Building, Beijing 100044, China,

Lu Yu
School of Economic and Management of Beijing Jiaotong University, Siyuan East Building,
Haidian District, Beijing 100044, China