B-DP: Dynamic Collection and Publishing of Continuous Check-In Data with Best-Effort Differential Privacy

Youqin Chen 1, Zhengquan Xu 1,*, Jianzhang Chen 2,3 and Shan Jia 1

1 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; chenyouqin@whu.edu.cn (Y.C.); jias@whu.edu.cn (S.J.)
2 College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China; jzchen@fafu.edu.cn
3 School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

* Correspondence: xuzq@whu.edu.cn

Abstract: Differential privacy (DP) has become a de facto standard to achieve data privacy. However, the utility of DP solutions with the premise of privacy priority is often unacceptable in real-world applications. In this paper, we propose the best-effort differential privacy (B-DP) to promise the preference for utility first and design two new metrics including the point belief degree and the regional average belief degree to evaluate its privacy from a new perspective of preference for privacy. Therein, the preference for privacy and utility is referred to as expected privacy protection (EPP) and expected data utility (EDU), respectively. We also investigate how to realize B-DP with an existing DP mechanism (KRR) and a newly constructed mechanism (EXP_Q) in the dynamic check-in data collection and publishing. Extensive experiments on two real-world check-in datasets verify the effectiveness of the concept of B-DP. Our newly constructed EXP_Q can also satisfy a better B-DP than KRR to provide a good trade-off between privacy and utility.

Keywords: B-DP; DP; check-in data; dynamic collection and publishing

1. Introduction

The explosive progress of mobile Internet and location technology, LBS (Location Based Service) applications, including Brightkite, Gowalla, Facebook and other social network platforms, generate a large number of check-in data every day. Check-in data generally include information such as time, locations, PoI (Points of Interest) attributes, mood and comments, and hence the check-in data has become a carrier of a user’s life trajectory and interest tendency [1–4]. However, a data analyst’s mining and analysis of the check-in data may directly or indirectly expose the sensitive information of a data provider [5–9]. There have been many privacy protection methods [10–16]. Some of them [10,11] rely on specific attack assumptions and background knowledge, and some methods [12–15] are based on differential privacy (DP) [17]. DP provides provable privacy protection, which is independent of the background knowledge and computational power of an attacker. The protection level of DP is evaluated by privacy budget [17]. When the privacy budget is relatively small, it has strong privacy protection, but the utility is often poor [17]. With the gradual integration of DP on practical applications, utility has become the bottleneck of its development and popularization.

In general, there is a contradiction between privacy and utility and it is necessary to be a trade-off [18,19]. In [19], the authors discussed a monotone trade-off in the semi-honest model. Therein, when the utility becomes worse, the privacy protection becomes stronger, and on the other hand, when the utility gets better, the privacy protection gets weaker. In many other DP theoretical studies, including strict $\epsilon$-DP [17] and relaxed $(\epsilon, \delta)$-DP [20], they often provide privacy priority and then make data more available or the best available,
which is a kind of trade-off with satisfying utility as much as possible under the privacy guarantee. Unfortunately, the applications of DP in real-world do not seem to follow this principle completely. One of the best examples is the four applications in Apple’s MacOS Sierra (version 10.12), i.e., Emojis, New words, Deeplinks and Lookup Hints. When they collect the data, the privacy budget is set to only 1 or 2 per each datum, but the overall privacy budget for the four applications is as high as 16 per day [21]. Furthermore, Apple renews the available privacy budget every day, which would result in a potential privacy loss of 16 times the number of days that a user participated in DP data collection for the four applications [21]. It is far beyond the reasonable protection scope of DP [22].

Based on the above facts, when there exists a contradiction between privacy and utility, privacy is no longer a priority as suggested in the DP theoretical studies, but the most desirable way is to balance the preference for privacy and utility, where the preference for privacy and utility is referred to as expected privacy protection (EPP) and expected data utility (EDU), respectively. However, few researchers have proposed solutions to reasonably balance EPP and EDU except the authors in [23]. They proposed an adaptive DP and its mechanisms in a rational model, which can achieve a balance between the approximate EDU and the EPP by adding conditional filtering noise [23]. If the privacy protection intensity under the balance of the approximate EDU that is satisfied by the data analyst is not the expectation of the data provider, then it still cannot meet the EPP of the data provider. In addition, the absolute value range of the conditional filtering noise belongs to (0.5,1.5), which makes it easy to be attacked by background knowledge. Therefore, best-effort differential privacy (B-DP) is proposed to make the EDU satisfied first and then the EPP satisfied as much as possible in this paper. We face the following two basic challenges at least.

1. If the EDU is to be satisfied first, then privacy protection may be no longer to be guaranteed by DP, how does it evaluate the guarantee degree of satisfying EPP as much as possible under B-DP?
2. If there is a reasonable metric for the guarantee degree of satisfying EPP as much as possible under B-DP, does it exist an implementation mechanism (or algorithm) to realize B-DP?

With the challenges of B-DP above, this paper explores a typical application with dynamic collection and publishing of continuous check-in data, where the check-in scenario is a semi-honest model with an honest but curious data collector. Each check-in user visiting a POI generates a check-in state and perturbs his check-in state to a POI Center for his privacy protection, where the POI Center is a data collector. The frequency of check-in users are calculated by POI Center according to the received check-in states, which is approximately the check-in data distribution and used for publishing to data analysts. We assume that one check-in state is perturbed to only one check-in state and each publishing is required to satisfy the EDU first and then satisfy EPP as much as possible in the dynamic publishing, and moreover, the privacy to be protected is the check-in state of a user and the utility to be realized is the distribution of the check-in data with relative error as its metrics (see Section 4.1 for more details). In fact, since the relative error is used as a metric of the published distribution, it needs a distribution dependent privacy protection mechanism (or implementation) in order to satisfy the EPP as much as possible under the constraint of EDU. In addition, since each publishing is required to satisfy the EDU first and then satisfy EPP as much as possible in the dynamic publishing, it needs an algorithm to make the privacy protection under the constraint of EDU to be satisfied continuously as much as possible in the process of dynamic publishing. Therefore, this mechanism or algorithm will be proposed from a new perspective, which is different from the existing methods in literature.

1.1. Our Contributions

The main contributions of this paper are concluded as follows.
1. A privacy protection concept of B-DP and two metrics of privacy guarantee degree are put forward. B-DP discussed in this paper is an expansion of the concept of DP, which can satisfy the EDU first and then provide the EPP as much as possible to be useful for real-world applications. It uses two new metrics including the point belief degree (see Definition 4) and the regional average belief degree (see Definition 5) to quantify the degree of privacy protection for any expected privacy budget (see Section 4.2), rather than for DP itself by the privacy budget $\epsilon$ to evaluate only one EPP with the expected privacy budget equal to $\epsilon$. In addition, the regional average belief degree can be used as the average guarantee degree of the EPP in a region including multiple expected privacy budgets. To the best of our knowledge, it is a new discussion and definition of B-DP that is different from the existing literature, and it uses two new metrics to explore and analyze the performance of privacy from a new perspective of the preference for privacy.

2. An EXP$_Q$ mechanism is proposed (see Definition 10). The newly constructed EXP$_Q$ mechanism can be used to the categorical data for privacy protection, which smartly alters the privacy budget based on its probability in the data distribution to make itself to realize a better B-DP compared to the existing KRR mechanism [24,25]. Thereby, it also verifies that B-DP can be better realized to provide a good trade-off between privacy and utility.

3. The dynamic algorithm with the implementation algorithms of two perturbation mechanisms is proposed to realize the dynamic collection and publishing of continuous check-in data and meanwhile to satisfy B-DP. The two perturbation mechanisms include the newly constructed EXP$_Q$ and a classical DP mechanism KRR [25,26] (a simple local differential privacy (LDP) mechanism). We take KRR as an example to show how to realize B-DP based on the existing DP mechanisms for the categorical data. Moreover, the number of domain values of both KRR and EXP$_Q$ is more than 2 and both the randomized algorithms based on them only take one value as input and one value as output. In addition, the dynamic algorithm can also be used to other applications of social behavior except check-in data.

1.2. Outline

The remainder of this paper is organized as follows: Section 2 summarizes the related work on the trade-off methods, utility metrics of relative error and LDP mechanisms. Section 3 presents conceptual background of DP and details of KRR mechanism and utility metrics. Section 4 introduces the system model, the relevant definitions of B-DP, including two metrics of the guarantee degree, etc., and model symbolization of the check-in data. Section 5 introduces the design and implementation of B-DP mechanisms and Section 6 describes the design of B-DP mechanism algorithm in the dynamic collection and publishing. Section 7 provides the experimental evaluation of the dynamic collection and publishing algorithm based on both two B-DP mechanisms. Finally, we provide a discussion and conclusion in Section 8.

2. Related Work

DP has become a research hotspot in the field of privacy protection since Dwork [12] proposed it in 2006. The model of DP starts from the traditional centralization [15,18], gradually grows to be distributed [27], and develops to be localization [24,28] and even to be personalized localization [29] and so on. It is not only the evolution process of DP technique, but also the comprehensive embodiment of the gradual integration of DP technique with real-world applications. However, no matter how it evolves, the two themes running through DP are privacy and utility [18], which is also focused by this paper. Table 1 summarizes the mainly related work from the perspective of privacy and utility priority as well as their metrics, the used privacy mechanism and the focusing problem with EPP and EDU. It will be divided into three categories to show its details.
Table 1. Comparison of existing literature with the method proposed in this paper.

| Selected Papers       | Mechanism          | Utility First | Privacy First | Privacy Metrics | Utility Metrics | EPP & EDU Involved |
|-----------------------|--------------------|---------------|---------------|-----------------|-----------------|-------------------|
| Katrina et al. [30]    | Laplace            | Yes           | No            | Central DP      | Absolute error  | No                |
| Liu et al. [23]       | Gauss with         | Yes           | No            | Central DP      | Relative error  | Yes, but it may not provide EPP as much as possible |
|                       | conditional        |               |               |                 |                 |                   |
|                       | filtering noise    |               |               |                 |                 |                   |
| Maryam et al. [31]    | Laplace            | No            | Yes           | Central DP      | Relative error  | No                |
| Xiao et al. [18]      | Laplace            | No            | Yes           | Central DP      | Relative error  | No                |
| Kairouz et al. [25]   | W-RR               | No            | Yes           | LDP             | KL divergence   | No                |
| Erlingsson et al. [28]| RAPPOR             | No            | Yes           | LDP             | Standard deviation | No |
| Bassily et al. [32]   | S-Hist             | No            | Yes           | LDP             | Absolute error  | No                |
| Chen et al. [29]      | PCEP               | No            | Yes           | PLDP            | KL divergence   | No                |
|                       |                    |               |               |                 | /relative error |                   |
| Kairouz et al. [24,25]| KRR                | No            | Yes           | LDP             | KL divergence   | No                |
| Our paper             | EXP_Q              | Yes           | No            | (Local) B-DP    | Relative error  | Yes, and it provides EPP as much as possible |

1. Trade-off model with utility first. The majority of DP research is based on the trade-off model with privacy first, while there are few relevant ones on the trade-off model with utility first. Therein, Katrina et al. [30] proposed a generalized “noise reduction” framework based on the modified “Above Threshold” algorithm [33] to minimize the empirical risk of privacy (ERM) on the premise of utility priority, but the scheme is only applicable to the framework that minimizes the empirical risk of privacy, where the privacy minimized may not be able to meet the EPP. Liu et al. proposed firstly that DP satisfies the monotonic trade-off between privacy and utility and its associated bounded monotone trade-off under the semi-honest model. They showed that there is no trade-off under the rational model, while unilateral trade-off could lead to utility disaster or privacy disaster [18,23,34]. They also presented an adaptive DP and its mechanisms under the rational model, which can realize the trade-off between approximately EDU and EPP by adding conditional filtering noise [23], but the mechanisms are probably not able to meet the expectation of data provider for privacy protection and are easily attacked by background knowledge because of the adding conditional filtering noise. Most importantly, the above two utility-first research [23,30] do not provide a quantitative metrics of the unmet privacy protection or the unmet degree of EPP, whereas this paper presents two detailed quantitative metrics including the point belief degree and the regional average belief degree to evaluate the privacy from a new perspective of preference for privacy.

2. Utility metrics of relative error. Maryam et al. [31] presented DP in real-world applications, which discussed how to add Laplace [12] noise from a view of utility. They studied the relationship between the cumulative probability of noise and the privacy level in Laplace mechanism and combined with the relative error metrics to discuss how to use a DP mechanism reasonably without losing the established utility. However, the literature does not delve into the details that how the guarantee degree of privacy protection will be changed when utility is satisfied. Xiao et al. [18] presented a DP publishing algorithm on a batch query using resampling technique of correlation noise to reduce noise added and improve data utility. When the algorithm picks the priority items each time, it is based on the intermediate results with noise, and the intermediate results with noise are not enough to reflect the original order of data. In this way, there is a bias in adjusting the privacy budget allocation, which may cause the query items that should be optimized to be not optimized, thus affecting the utility of published data. However, the literature is a classical example of optimizing utility with privacy first, which runs counter to the theme of this paper. In addition, the above two schemes are essentially based on the central DP and use continuous
Laplace mechanism, which are different from the LDP (discrete) data statistics and release required by the check-in application in this paper. Therefore, these schemes cannot be directly applied to the applications this paper considers.

3. LDP mechanisms. In 1965, Warner first proposed the randomized response technique (W-RR) to collect statistical data on sensitive topics and keep the sensitive data of contributing individuals confidential [35]. Although W-RR can strictly satisfy \( \epsilon \)-LDP [25] in one survey statistics, multiple collections on the same survey individuals will weaken the privacy protection intensity [12]. Therefore, Erlingsson et al. [28] used a double perturbation scheme combining permanent randomized response with instantaneous randomized response, namely, RAPPOR, to expand the application of W-RR, and it has been used by Google in Chrome browser to collect users’ behavior data. In addition, RAPPOR also uses Bloom Filter technology [36] as the encoding method, which maps the statistical attributes into a binary vector. Finally, the mapping relation and Lasso regression method [37] are combined to reconstruct the frequency statistics corresponding to the original attribute string. Due to the high communication cost of RAPPOR, Bassily et al. [32] proposed the S-Hist method. In the method, each user first encodes his attributes, then randomly selects one of the bits and uses the randomized response technique to perturb it, and finally sends the result of the perturbation to the data collector, so as to reduce the communication cost. Chen et al. [29] proposed a PCEP mechanism and designed a PLDP (personalized LDP) applied to spatial data with it, aiming to protect the users’ location information and count the number of users in the area. Therein, the privacy budget of the scheme is determined by the users’ personalization, and hence the utility depends on the users’ individual behavior settings. In addition, the mechanism combines the S-Hist [32] method and adopts the random projection technique [38]. Although it can greatly reduce the communication cost, it still has the problem of unstable query precision. Based on the check-in application with multiple check-in spots in this paper, the KRR mechanism [24,25] just easily fits this application with no prior data distribution knowledge, but it is not very good for B-DP. In addition, DP has already been studied in these applications, such as social networks [39,40], recommender systems [41], data publishing [42–44], deep learning [45], reinforcement learning [46] and federated learning [47].

3. Preliminaries

In this section, the key notations used in this paper are given in Table 2.

### 3.1. Differential Privacy (DP)

Differential privacy (DP), broadly speaking, is a privacy protection technique that does not depend on an attacker’s background knowledge and computational power [17,20,48]. It can be generally divided into central DP and LDP depending on whether it is based on a trusted data collector [33]. The formal definitions of these two types of DP are given as follows.

**Definition 1** ((\( \epsilon, \delta \))-Central) DP [17,20]). A randomized algorithm \( M \) and a set \( S \) of all possible outputs of \( M \), for a given dataset \( D \) and any adjacent dataset \( D' \) that differ on at most one record, if \( M \) satisfies the following inequality, then it is said that \( M \) satisfies \((\epsilon, \delta)\)-Central DP.

\[
P[M(D) \in S] \leq e^\epsilon \times P[M(D') \in S] + \delta,
\]

where \( P[\cdot] \) represents the risk of privacy disclosure and is controlled by the randomness of algorithm \( M \), the parameter \( \epsilon \) is called privacy budget that represents the level of privacy protection, and \( \delta \) represents the probability of failure to satisfy \( \epsilon \)-(central)DP. When \( \delta = 0 \), \( M \) satisfies the \( \epsilon \)-(central) DP.
**Table 2. Notations.**

| Symbol     | Description                                                                 |
|------------|-----------------------------------------------------------------------------|
| EDU        | Expected data utility                                                       |
| EPP        | Expected privacy protection                                                 |
| $\epsilon_e$ | Expected privacy budget                                                   |
| Region($e_e$) | Expected privacy protection region around $e_e$                           |
| $\eta$     | Expected data utility                                                       |
| $\epsilon_n$ | The privacy budget of a differential privacy mechanism that just meets the expected data utility $\eta$ |
| $C_{e_e}$   | Point belief degree of $e_e$                                                |
| $C_{Region(e_e)}$ | Regional average belief degree of Region($e_e$)  |
| $p$         | Original data distribution                                                  |
| $\hat{p}$  | Perturbed data distribution                                                 |
| $\hat{\hat{p}}$ | Estimated data distribution                                                |
| $S$         | Check-in state space                                                        |
| $h(S)$      | Original check-in counts vector                                             |
| $\hat{h}(S)$ | Perturbed check-in counts vector                                            |
| $\hat{\hat{h}}(S)$ | Estimated check-in counts vector                                            |
| $Q$         | Perturbation probability matrix                                             |
| $\eta_{ij}$ | The perturbation probability of the original check-in state $S_j$ to the check-in state $S_i$ |
| KRR         | $k$-ary randomized response mechanism                                       |
| EXP$_Q$     | Perturbation mechanism                                                      |
| $\gamma$   | Privacy setting parameter                                                   |
| $\gamma\eta$ | Privacy setting parameter with satisfying the expected data utility $\eta$ |
| $\kappa_n$ | The parameter of privacy protection intensity change point                  |
| $w$         | Modified estimate parameter                                                |
| Re$\theta$  | Update threshold parameter                                                 |
| $\text{err}(p, \hat{p})$ | The maximum relative error between $p$ and $\hat{p}$                |

In the check-in application of this paper, the POI Center is an honest and curious data collector, even if the POI Center or other attackers can obtain the check-in state submitted by a user, they cannot conclusively infer the original check-in state of the user. If $\mathcal{K}$ can satisfy $(\epsilon, \delta)$-LDP to protect the check-in state of the user, then it needs to meet the following definition.

**Definition 3** (Check-in state of $(\epsilon, \delta)$-LDP). A user $u$ generates a check-in in a POI, whose check-in state variable is denoted as $s^u$ with $s^u \in \{S_1, S_2, \cdots, S_n\}$. Assume that the original check-in state of $u$ is $S_j$ or $S_j'$ for $j, j' \in [1, n]$. Moreover, $S_j$ and $S_j'$ generate the same check-in state $S_i$ for $i \in [1, n]$ after being perturbed by a randomized algorithm $\mathcal{K}$, respectively, and the perturbed check-in state variable is $\hat{s}^u$ with $\hat{s}^u \in \{S_1, S_2, \cdots, S_n\}$. If there exists an $\epsilon \in \mathbb{R}^+$ such that $\mathcal{K}$ satisfies the following constraints for $i, j, j' \in [1, n]$,

$$ P(s^u = S_i | s^u = S_j) \leq e^\epsilon P(s^u = S_i | s^u = S_j') + \delta, \tag{3} $$

where, $P(s^u = S_i | s^u = S_j)$ and $P(s^u = S_i | s^u = S_j')$ are the perturbation probabilities of the original check-in states $S_j$ and $S_j'$ to the check-in state $S_i$, respectively, then $\mathcal{K}$ will enable the check-in state to satisfy $(\epsilon, \delta)$-LDP. When $\delta = 0$, $\mathcal{K}$ satisfies $\epsilon$-LDP.
Property 1 (Parallel composition [49]). Assume that randomized algorithms are $K_1, K_2, \ldots, K_n$ and their privacy budgets are $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$, respectively, then for the disjoint datasets $D_1, D_2, \ldots, D_n$, an algorithm $K(K_1(D_1), K_2(D_2), \ldots, K_n(D_n))$ provides $\max(\epsilon_i)$-(local) DP, and the level of privacy protection it provides depends on the largest privacy budget.

3.2. KRR Mechanism

KRR is a LDP mechanism [24,25], which satisfies the following probability distribution,

$$P(y|x) = \frac{1}{e^{\epsilon} + k - 1} \begin{cases} e^{\epsilon}, y = x \\ 1, y \neq x \end{cases}$$

where $x, y \in \chi$ and $|\chi| = k$.

KRR is a more general form of the randomized response mechanism of W-RR, that is, when $k = 2$, KRR degenerates into W-RR.

3.3. Utility Metrics

In this paper, the worst relative error of POIs in check-in statistics will be used to measure the overall utility of the check-in application, where the calculation formula of the relative error of POI $i$ is as follows,

$$\text{err}(r_i, r_i^*) = \frac{|r_i - r_i^*|}{\max\{r_i, \phi\}}$$

where $r_i^*$ is the estimated result of POI $i$ check-in statistics after LDP protection, $r_i$ is the real check-in result of the POI $i$, and the parameter $\phi$ is a constant to avoid the situations that $r_i = 0$ causes the denominator to be 0 or $r_i$ is too small [18,50,51]. For the convenience of analysis, this paper will use the relative root mean square error for the utility metrics approximately, the specific formula is as follows,

$$\text{err}(r_i, r_i^*) = \sqrt{\frac{\zeta(r_i, r_i^*)}{\max\{r_i, \phi\}}}$$

where $\zeta(r_i, r_i^*)$ is the expectation of the mean square error between the real statistical result $r_i$ and the statistical estimate result $r_i^*$ after LDP protection, and the parameter $\phi$ is defined as above. Then, the formula for calculating the maximum relative error of $n$ POIs is as follows,

$$\text{err}(r, r^*) = \max(\text{err}(r_i, r_i^*)) = \max\left(\frac{\sqrt{\zeta(r_i, r_i^*)}}{\max\{r_i, \phi\}}\right).$$

As above, $r_i, r_i^*$ not only can represent data distribution, but also can represent frequency or counts.

4. Problem Formulations

4.1. System Model

As shown in Figure 1, there are three types of participants, namely, check-in users (data providers), POI Center (data collector), and data analysts (for example, POI managers) in the check-in model. Each check-in user visiting a POI generates a check-in state and sends it to POI Center through a terminal with the check-in APP, where each check-in state corresponds to a count and the check-in state belongs to category data. POI Center calculates the counts and frequency of check-in users’ visiting POIs according to the received check-in states, where frequency is approximately the check-in data distribution and used for publishing to data analysts. In addition, it is assumed that each check-in user is independent of each other and only one check-in state is submitted by a check-in user in one publishing. It is also assumed that the check-in scenario is a semi-honest model, in which the POI Center is an honest but curious data collector, and the check-in state of a
user is sensitive. Hence, a user will adopt a perturbation mechanism (for example, LDP mechanism) to perturb his check-in state for his privacy protection, and then sends it to the POI Center. Therein, it is assumed that one check-in state is perturbed to only one check-in state.

![POI check-in model](image)

**Figure 1.** POI check-in model. Therein, $S_i$, $S_j$ and $S_k$ represent check-in states. $\hat{h}(S)$ and $\hat{p}$ represent the check-in counts and the check-in frequency (data distribution) in perturbation phase, respectively, while $\tilde{h}(S)$ and $\tilde{p}$ represent the check-in counts and the check-in frequency (data distribution) in construction phase, respectively. $K$ represents a perturbation mechanism. The more details can also be seen in Section 4.3.

In this paper, we focus on the dynamic collection and publishing of continuous check-in data with both privacy and utility requirements, where the privacy to be protected is the check-in state of a user and the utility to be realized is the distribution of the check-in data with relative error as its metrics. Therein, the privacy refers to EPP, which is the preference for privacy of a user, and the utility refers to EDU, which is the preference for utility of a data analyst. Moreover, each publishing is required to satisfy the EDU first and then satisfy EPP as much as possible in the dynamic publishing. Thereby, we adopt B-DP based on the LDP model including perturbation, aggregation, reconstruction and publishing, and we also need to have the process of initializing or updating the perturbation mechanism $K$ at least to make every publishing to satisfy the EPP as much as possible under the EDU satisfied first in the dynamic publishing, as shown in Figure 1.

### 4.2. The Related Concepts of B-DP

In the concept of best-effort differential privacy (B-DP), there is an expected privacy protection (EPP) and an expected data utility (EDU), respectively. When the two cannot be satisfied simultaneously, the EDU should be satisfied first and the EPP should be satisfied as much as possible. Since the protection level of DP is evaluated by privacy budget [17], the preference for privacy also refers to the preference for the privacy budget in the B-DP. Hence, the EPP refers to a data provider’s preference for the privacy budget and we define this privacy budget as the expected privacy budget symbolized as $\epsilon$. We use $\text{Region}(\epsilon)$ to symbolize the expected privacy protection region, which refers to a data provider’s preference for a region including multiple expected privacy budgets.

We use $\eta$ to symbolize the EDU. In this paper, the expectation of the maximum relative error of Formula (7) is used to measure data utility. When the expectation of the maximum relative error of Formula (7) is less than or equal to $\eta$, it means that the EDU is satisfied;
when equal, it means that the EDU is just satisfied. The privacy budget of a DP mechanism that just satisfies the EDU $\eta$ is symbolized as $\epsilon_\eta$.

**Definition 4** ($C_{\epsilon^e}$-Point belief degree). It defines the guarantee degree of EPP under the expected privacy budget $\epsilon^e$, which can be provided by the $\epsilon_\eta$-DP mechanism, as the point belief degree, and the symbol is denoted as $C_{\epsilon^e}$. Moreover, $C_{\epsilon^e} = \sum_{j=1}^{n} p_i \chi(\epsilon_i, \epsilon^e)$, where $n$ represents the number of POIs in check-in application, $p_i$ represents the probability of POI, perturbed by $\epsilon_\eta$-DP mechanism, $\epsilon_i$ represents the actual privacy budget of POI, and $\chi(\epsilon_i, \epsilon^e)$ represents an indicator function for whether the EPP is satisfied, which is defined as follows,

$$\chi(\epsilon_i, \epsilon^e) = \begin{cases} 0, & \epsilon_i > \epsilon^e \\ 1, & \epsilon_i \leq \epsilon^e \end{cases}$$

(8)

**Definition 5** ($C_{\text{Region}(\epsilon^e)}$-Regional average belief degree). The average guarantee degree of the EPP under the expected privacy protection region $\text{Region}(\epsilon^e)$, which can be provided by the $\epsilon_\eta$-DP mechanism, is defined as the regional average belief degree, and the symbol is denoted as $C_{\text{Region}(\epsilon^e)}$. When $\text{Region}(\epsilon^e) = \{\epsilon_1, \epsilon_2, \ldots, \epsilon_K\}$ and $\epsilon_1 < \epsilon_2 < \ldots < \epsilon_K$ for $K \geq 2$, it defines

$$C_{\text{Region}(\epsilon^e)} = \frac{1}{\epsilon_K - \epsilon_1} \sum_{k=1}^{K-1} (\epsilon_k - \epsilon_{k+1}) C_{\epsilon_k},$$

(9)

where $C_{\epsilon_k}$ can refer the definition of point belief degree.

**Definition 6** ($(\eta, C_{\epsilon^e})$-B-DP). The DP mechanism that just satisfies the EDU $\eta$ with the point belief degree $C_{\epsilon^e}$ of the expected privacy budget $\epsilon^e$ is defined as $(\eta, C_{\epsilon^e})$-B-DP. Therein, the $(\eta, C_{\epsilon^e})$-B-DP, where the point belief degree $C_{\epsilon^e}$ is maximum, is defined as $(\eta, C_{\epsilon^e})$-Best-B-DP.

**Definition 7** ($(\eta, C_{\text{Region}(\epsilon^e)})$-B-DP). The DP mechanism that just satisfies the EDU $\eta$ with the regional average belief degree $C_{\text{Region}(\epsilon^e)}$ of the expected privacy protection region $\text{Region}(\epsilon^e)$ is defined as $(\eta, C_{\text{Region}(\epsilon^e)})$-B-DP. Therein, the $(\eta, C_{\text{Region}(\epsilon^e)})$-B-DP, where the regional average belief degree $C_{\text{Region}(\epsilon^e)}$ is maximum, is defined as $(\eta, C_{\text{Region}(\epsilon^e)})$-Best-B-DP.

Note that, generally, B-DP includes both central B-DP and local B-DP, which depends on whether it is based on a trusted data collector the same as the DP. This paper focuses on local B-DP.

### 4.3. Model Symbolization

Let $POI_i$ with $i \in [1, n]$ represent $n$ POIs in check-in scenario, and the check-in state space is $S = \{S_1, S_2, \ldots, S_n\}$ where $S_i$ is the check-in state of $POI_i$. Let $s^u, s^v, s^w \in S$ be variables of the original check-in state, the perturbed check-in state and the estimated check-in state of the user $u$, respectively. Let $p$, $\tilde{p}$ and $\hat{p}$ be the probability distributions of the original check-ins, the perturbed check-ins and the estimated check-ins, respectively, where $p = [p_1, p_2, \ldots, p_n]^T$, $\tilde{p} = [\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_n]^T$ and $\hat{p} = [\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n]^T$. Assume that it is the same probability distribution law for all the users, that is, $p_i = P(s^u = S_i)$, $\tilde{p}_i = P(s^v = S_i)$ and $\hat{p}_i = P(s^w = S_i)$ for any $i \in [1, n]$ and $u$. $h(S) = [h(S_1), h(S_2), \ldots, h(S_n)]^T$, $\tilde{h}(S) = [\tilde{h}(S_1), \tilde{h}(S_2), \ldots, \tilde{h}(S_n)]^T$ and $\hat{h}(S) = [\hat{h}(S_1), \hat{h}(S_2), \ldots, \hat{h}(S_n)]^T$ represent the original check-in counts vector, the perturbed check-in counts vector and the estimated check-in counts vector with $m \in N^+$ users, respectively.

**Definition 8** (Random perturbation and perturbation probability matrix $Q$). The process for any user $u$ to change check-in state from $S_i$ to $S_j$ with a certain perturbation probability is called random perturbation, and the perturbation probability is denoted as $q_{ij} = P(s^w = S_j | s^v = S_i)$ with $S_i, S_j \in S$. The matrix composed of $q_{ij}$ for any $i, j \in [1, n]$ is called the perturbation probability matrix $Q$, where $Q = (q_{ij})_{n \times n}$ and $\sum_{i=1}^{n} q_{ij} = 1$ for any $j \in [1, n]$. 


Therefore, the perturbed probability distribution \( \tilde{p} \), the original probability distribution \( p \) and the perturbation probability matrix \( Q \) have the following relationship

\[
\tilde{p} = Qp.
\]

From Equation (10), it can be seen that \( \tilde{p}_i = \sum_{j=1}^{n} q_{ij}p_j \) for any \( i \in [1,n] \). Obviously, \( \tilde{p}_i \) and \( p_i \) are not always equal, and hence the result of the perturbation is biased. Assume that \( Q \) is always reversible and its inverse matrix is defined as \( R = Q^{-1} = (r_{ij})_{n \times n} \). Therefore, it can get the following theorem.

**Theorem 1.** The check-in counts vector \( h(S) \) is perturbed by the perturbation probability matrix \( Q \) to obtain the perturbed check-in counts vector \( \tilde{h}(S) \) and then it is corrected by the inverse matrix \( R \). The estimated check-in counts vector \( \hat{h}(S) = R\tilde{h}(S) \) satisfies \( E[\hat{h}(S)] = h(S) \).

**Proof.** Since \( E[\hat{h}(S)] = E[R\tilde{h}(S)] = E[RQh(S)] \), and \( RQ = I \), it has \( E[\hat{h}(S)] = E[h(S)] = h(S) \). Therefore, the result follows. \( \square \)

Theorem 1 states that the estimated check-in counts vector \( \hat{h}(S) \) obtained after the correction of the inverse matrix \( R \) is an unbiased estimate of the original check-in counts vector \( h(S) \).

Here, the relative root mean square error \( err(h(S_i), \hat{h}(S_i)) \) of the original check-in counts \( h(S_i) \) and the estimated check-in counts \( \hat{h}(S_i) \) for \( n \) POIs and \( m \) users on \( POI_i \) can be calculated as follows,

\[
err(h(S_i), \hat{h}(S_i)) = \frac{\sqrt{Var[\hat{h}(S_i)]}}{E[\hat{h}(S_i)]}, \tag{11}
\]

where \( \phi = 1 \) and \( Var[\hat{h}(S_i)] \) can be calculated as follows,

\[
Var[\hat{h}(S_i)] = \sum_{j=1}^{n} r_{ij}^2(\sum_{k=1}^{n} q_{jk}h(S_k) - h(S_i)). \tag{12}
\]

**Theorem 2.** The relative root mean square error between the original probability \( p_i \) of \( POI_i \) check-ins and the estimated probability \( \tilde{p}_i \) of \( POI_i \) check-ins is \( err(p_i, \tilde{p}_i) = \frac{\sqrt{Var[h(S_i)]}}{E[h(S_i)]} \), where \( i \in [1,n] \).

**Proof.** The relative root mean square error between the original probability \( p_i \) of \( POI_i \) check-ins and estimated probability \( \tilde{p}_i \) of \( POI_i \) check-ins can be represented as follows,

\[
err(p_i, \tilde{p}_i) = \frac{\sqrt{E[(p_i - \tilde{p}_i)^2]}}{p_i} = err(h(S_i), \hat{h}(S_i)). \tag{13}
\]

Therefore, \( err(p_i, \tilde{p}_i) = \frac{\sqrt{Var[h(S_i)]}}{E[h(S_i)]} \), where \( i \in [1,n] \). \( \square \)

According to Formulas (11)–(13), it can be known \( err(p_i, \tilde{p}_i) \), \( Q \) and \( p \) are related. If \( max(err(p_i, \tilde{p}_i)) = \eta \) means that the EDU is just satisfied, then \( Q \) should satisfy the following constraints according to B-DP:

\[
\begin{align*}
max(err(p_i, \tilde{p}_i)) & \leq \eta, i \in [1,n], \\
q_{ij} & \leq e^{\epsilon_i}q_{ij'}, \forall i, j, j' \in [1,n], \epsilon_i \geq 0, \\
\sum_{i=1}^{n} q_{ij} & = 1, \forall j \in [1,n].
\end{align*} \tag{14}
\]

Assume that there is a randomized algorithm \( \mathcal{K} \) with a perturbation probability matrix \( Q \), which can provide the expected privacy budget \( \epsilon_e \) with the point belief degree


\[ C_{e_i} = \sum_{i=1}^{n} \hat{p}_i \hat{K}(e_i, e_f) \]. Then, if \( K \) wants to satisfy \( e_\eta = \max(e_i, C_{e_i}) \)-Best-B-DP, it should still need to maximize \( C_{e_i} \). Therefore, \( K \) should satisfy the following optimization problem.

\[
\begin{align*}
\text{maximize} & \quad C_{e_i} \\
\text{subject to} & \quad \max\{\text{err}(p_i, \hat{p}_i)\} \leq \eta, i \in [1, n], \\
& \quad q_{ij} \leq e^{\epsilon_i} q_{ij'}, \forall i, j, j' \in [1, n], \epsilon_i \geq 0, \\
& \quad \sum_{i=1}^{n} q_{ij} = 1, \forall j \in [1, n].
\end{align*}
\] (15)

Similarly, it is assumed that \( K \) can provide the expected privacy protection region \( \text{Region}(e_\alpha) = \{e_{\alpha_1}, e_{\alpha_2}, \cdots, e_{\alpha_k}\} \) with the regional average belief degree \( C_{\text{Region}(e_\alpha)} = \frac{1}{e_{\alpha_k} - e_{\alpha_1}} \sum_{k=1}^{k-1} (e_{\alpha_k} - e_{\alpha_{k-1}}) e_{\alpha_k} \), where \( e_{\alpha_k} \) is the point belief degree of the expected privacy budget \( e_\alpha \). Then, if \( K \) wants to satisfy \( (e_\eta = \max(e_i, C_{\text{Region}(e_\alpha)})) \)-Best-B-DP, it should still need to maximize \( C_{\text{Region}(e_\alpha)} \). Therefore, \( K \) should satisfy the following optimization problem.

\[
\begin{align*}
\text{maximize} & \quad C_{\text{Region}(e_\alpha)} \\
\text{subject to} & \quad \max\{\text{err}(p_i, \hat{p}_i)\} \leq \eta, i \in [1, n], \\
& \quad q_{ij} \leq e^{\epsilon_i} q_{ij'}, \forall i, j, j' \in [1, n], \epsilon_i \geq 0, \\
& \quad \sum_{i=1}^{n} q_{ij} = 1, \forall j \in [1, n].
\end{align*}
\] (16)

From the above optimization Equations (15) and (16), in each optimization problem, it can be concluded that the perturbation probability matrix \( Q \) contains \( n^2 \) unknown variables, \( n^3 \) inequality constraints, \( n \) equality constraints and one EDU \( \eta \) constraint. Therefore, directly solving the optimization problem is a huge challenge, especially in the case of a large domain size \( n \). Therefore, two simplified models are considered in this paper and we will present the details one by one in the following section.

5. Design and Implementation of B-DP Mechanism

This section includes the design of two B-DP mechanisms and their implementation algorithms. One is based on a classical LDP mechanism KRR, and the other is based on the newly constructed mechanism EXPQ in this paper. The number of domain values of both two mechanisms is more than 2. Moreover, we combine three data distributions with the typical non-uniformity and two B-DP mechanisms to directly show and analyze the two metrics proposed in this paper, including the point belief degree and the regional average belief degree.

5.1. B-DP Mechanism Based on KRR

Without prior knowledge of the data distribution, we assume that it is uniform. Setting \( \epsilon_i = \epsilon_j = \epsilon_\eta, q_{ii} = p, q_{ij} = q \) for \( i, j \in [1, n], j \neq i \) and \( q_{ii} \geq q_{ij} \), then it has \( p = \frac{e^{\epsilon_\eta}}{e^{\epsilon_\eta} + n - 1}, q = \frac{1}{e^{\epsilon_\eta} + n - 1} \), and \( \text{err}(p, \hat{p}) = \eta \). Therefore, the privacy budget of KRR here is not arbitrary, which is constrained by the EDU of \( \text{err}(p, \hat{p}) = \eta \).

**Definition 9 (\( e_\eta \)-KRR).** *KRR that just meets the EDU \( \eta \) is called \( e_\eta \)-KRR.*

Here, it can also derive the following theorem.

**Theorem 3.** *If there exists \( e_\eta \)-KRR, then it satisfies \( e = e_\eta \)-LDP. Moreover, the point belief degree of \( e_\eta \)-KRR is*

\[
C_{e_\eta} = \begin{cases} 
0, & \epsilon_e < e_\eta \\
1, & \epsilon_e \geq e_\eta
\end{cases}
\] (17)
Proof. If there exists $\epsilon_\eta$-KRR, then it has $p = \frac{e^{\epsilon_\eta}}{e^{\epsilon_\eta} + n - 1}, q = \frac{1}{e^{\epsilon_\eta} + n - 1}$. Since $e = ln\left(\frac{p}{q}\right) = \epsilon_\eta$, it satisfies $e = \epsilon_\eta$-LDP. According to $C_\epsilon = \sum_{i=1}^{n} T_i \chi(\epsilon_i, \epsilon_\eta)$, when $\epsilon_\eta < \epsilon$, $\chi(\epsilon_i, \epsilon_\eta) = \chi(\epsilon_i, \epsilon_\eta) = 0$, it has $C_\epsilon = 0$; when $\epsilon_\eta \geq \epsilon$, $\chi(\epsilon_i, \epsilon_\eta) = \chi(\epsilon_i, \epsilon_\eta) = 1$, it has $C_\epsilon = 1$. Hence, the result follows.

Thus, it is impossible for $\epsilon_\eta$-KRR to provide the EPP $\epsilon_\eta$, or provide the EPP $\epsilon_\eta$ with a 100% satisfaction. Therefore, what KRR can achieve is two distinct jumps of EPP with or without guarantee, that is, it is not a good B-DP mechanism.

5.2. B-DP Mechanism Based on EXP$_Q$

Since relative error is used as utility metrics about the check-in data distribution in this paper, and a privacy budget of DP usually determines absolute error which is the numerator of relative error, thus the privacy budget of every POI should vary with its probability in the distribution of check-ins, that is, the value of the privacy budget should be reduced when the corresponding probability in the data distribution becomes larger, and increased when the corresponding probability in the data distribution becomes smaller. In this way, the small amounts of check-ins can also satisfy the EDU, while the large amounts of check-ins can also satisfy the EPP in priority, so as to better realize B-DP. It defines the following perturbation mechanism EXP$_Q$.

**Definition 10** (Perturbation mechanism EXP$_Q$). Given the data distribution $p = [p_1, p_2, \ldots, p_n]^T$, where $p_1 \geq p_2 \geq \cdots \geq p_n$. Call the randomized algorithm with Q as perturbation mechanism EXP$_Q$ if Q satisfies $q_{ij} = e^{-\gamma u(i,j)}$, where $q_{ij}$ is the probability that the check-in state is perturbed from $S_j$ to $S_i$ for $i, j \in [1, n]$ and $\gamma \geq 0$ is the privacy setting parameter, where $u(j, i)$ satisfies Formulas (18)–(20), and $\kappa_n \in [0, n]$ is the parameter of privacy protection intensity change point.

(i) When $\kappa_n = 0$,

$$u(j, i) = \begin{cases} 0, & i = j \\ 1 + p_{n-i+1}, & i \neq j \end{cases}$$

(18)

(ii) When $\kappa_n \in [1, n-1]$,

$$u(j, i) = \begin{cases} 0, & i = j \\ 1 - p_i, & i \neq j \text{ and } i \leq \kappa_n \\ 1 + p_{n-i+\kappa_n+1}, & i \neq j \text{ and } i > \kappa_n \end{cases}$$

(19)

(iii) When $\kappa_n = n$,

$$u(j, i) = \begin{cases} 0, & i = j \\ 1 - p_i, & i \neq j \end{cases}$$

(20)

**Theorem 4.** In perturbation mechanism EXP$_Q$, its Q satisfies the following properties.

(i) When $\kappa_n = 0$ and the normalized factor perturbing from POI to a POI is $\Omega_j = 1 + \sum_{k=1,k \neq j}^{n} e^{-\gamma(1+p_{n-k+1})}$ for $i, j \in [1, n]$,

$$q_{ij} = \begin{cases} \frac{1}{\Omega_j}, & i = j \\ e^{-\gamma(1+p_{n-i+1})}/\Omega_j, & i \neq j \end{cases}$$

(21)
Based on the definition of B-DP, where the indicator function provides the protection of relaxed (18), it cannot well realize B-DP. EXP is one of the classical mechanisms of LDP, but B-DP mechanisms are given in Algorithms 1 and 2, respectively.

### 5.3. Implementation of B-DP Mechanism

For the check-ins scenario, two B-DP mechanisms based on KRR and EXP are proposed and realized in this paper. KRR is one of the classical mechanisms of LDP, but it cannot well realize B-DP. EXP is newly proposed in this paper, which can not only provide the protection of approximately optimal (18, Cε)-Best-B-DP, but also provides the protection of relaxed (ε, 1 − Cε)-LDP to satisfy the EDU. The pseudo codes of the two B-DP mechanisms are given in Algorithms 1 and 2, respectively.
Algorithm 1 B-DP mechanism based on KRR.

Input: Probability distribution \( \mathbf{p} = [p_1, p_2, \cdots, p_n]^T \), sample size \( m \) and expected data utility (EDU) \( \eta \)

Output: Privacy budget \( \epsilon_\eta \) and perturbation probability matrix \( \mathbf{Q} \)

1: Initialize \( \epsilon_\eta > 0 \), iteration step size \( \Delta \epsilon_\eta > 0 \) and the worst utility \( \text{MaxRE} = 1 \);
2: while \( \text{MaxRE} > \eta \) do
3: \( \mathbf{Q} \) is constructed by KRR with \( \epsilon_\eta \);
4: According to the relative error formula, the current worst relative error \( \text{CurrentMaxRE} = \max_{i \in [1,n]} \left( \sqrt{\sum_{j=1}^{n} r_{ij}^2 (\sum_{k=1}^{n} \eta_k p_k) - m p_i} \right) \) is obtained, where \( \mathbf{R} = (r_{ij})_{n \times n} = \mathbf{Q}^{-1} \), \( \mathbf{Q} = (q_{ij})_{n \times n} \), referred to Formulas (11)–(12) and Theorems 1 and 2 for details.
5: if \( \text{CurrentMaxRE} < \text{MaxRE} \) then
6: \( \text{MaxRE} = \text{CurrentMaxRE} \);
7: end if
8: if \( \text{MaxRE} > \eta \) then
9: \( \epsilon_\eta = \epsilon_\eta + \Delta \epsilon_\eta \);
10: end if
11: end while
12: return \( \epsilon_\eta \) and \( \mathbf{Q} \)

5.4. Case Analysis of Point Belief Degree and Regional Average Belief Degree

The above description theoretically analyzes two metrics, including the point belief degree and the regional average belief degree, on the two B-DP mechanisms based on KRR and EXP\(_Q\). In order to show the two metrics more clearly, the following of this section will use three data distributions with typical non-uniformity for analysis. For simplicity, in the following of this section, the KRR-based B-DP mechanism is represented by KRR and the EXP\(_Q\)-based B-DP mechanism is represented by EXP\(_Q\), including the diagram descriptions.

(1) Three data distributions with typical non-uniformity.

The data distribution in this section is set as Pareto distribution, where the discrete case of Pareto distribution satisfies \( p_j \propto \frac{1}{x_j^{\theta+1}} \) for \( j \in [1,n] \), \( \theta > 0 \) and \( x_j > 0 \). Three data distributions of \( n = 20 \), \( \theta = 1.55, 1.17, 0.52 \) are shown in Figure 2 and are, respectively, denoted as P1, P2 and P3, where \( x_j = x_1 + (j - 1) \Delta x \), \( x_1 = 2 \), \( \Delta x = 0.2 \). Figure 2 shows both ordered and disordered cases of Pareto distribution, where the disordered case illustrates that the identification of scenic spots is independent of the order of probability. It also shows the corresponding Gini coefficient of P1, P2 and P3, which is calculated according to the method of Gini mean difference [52]. Gini coefficient is used to indicate the degree of unevenness of data distribution. There exists a quantitative relationship between Pareto distribution parameter and Gini coefficient in Table 3. As shown in Figure 2, the data distribution of \( \theta = 1.55 \) is pretty uneven, and the data distribution of \( \theta = 1.17 \) is relatively reasonably uneven, while the data distribution of \( \theta = 0.52 \) is relatively even.

Table 3. Gini coefficient VS. Pareto parameter of \( \theta \).

| Pareto Distribution | \( \theta \) | Gini Coefficient |
|---------------------|-------------|-----------------|
| P1                  | 1.55        | 0.4471          |
| P2                  | 1.17        | 0.3884          |
| P3                  | 0.52        | 0.2784          |
Algorithm 2 R-DP mechanism based on EXP_Q.

**Input:** Probability distribution $p = [p_1, p_2, \ldots, p_n]^T$, sample size $m$, expected data utility (EDU) $\eta$, expected privacy budget $\epsilon_i$ (or expected privacy protection region $Region(\epsilon_i) = \{\epsilon_{e_1}, \epsilon_{e_2}, \ldots, \epsilon_{e_\kappa}\}$ with $\epsilon_{e_1} < \epsilon_{e_2} < \cdots < \epsilon_{e_\kappa}$)

**Output:** Privacy setting parameter $\gamma_\eta$, the parameter of privacy protection intensity change point $\kappa_\eta$, perturbation probability matrix $Q$ and actual privacy budget $\epsilon_i$ of POI $i$ for $i \in [1, \eta]$

1. Initialize privacy setting parameter $\gamma_0 > 0$ and the iteration step size $\Delta \gamma_\eta > 0$;
2. Initialize $\kappa_\eta = n$ and $tag = 0$, where $tag$ is used to identify whether it exists a comparatively reasonable result or not;
3. while $\kappa_\eta \geq 0$ do
   4. Initialize $\gamma_\eta = \gamma_0, \epsilon_i = 0$ for $i \in [1, \eta]$, the worst utility $MaxRE = 1$ and $C_{e_i} = 0$ (here, the initialization of regional average belief degree $C_{Region(\epsilon_i)}$ is also uniformly recorded as $C_{e_i} = 0$);
   5. while $MaxRE > \eta$ do
      6. $Q$ is constructed by EXP_Q with $p, \gamma = \gamma_\eta$ and $\kappa_\eta$, where the row represents the perturbed check-in state and the column represents the original check-in state;
      7. According to $Q$, use $\ln{\frac{\max(q_{ij})}{\min(q_{ij})}}$ to update the value $\epsilon_i$, where $Q = (q_{ij})_{n \times n}$ and $i, j, j' \in [1, n]$;
   8. According to the relative error formula, the current worst relative error is obtained $CurrentMaxRE = \max_{i \in [1, \eta]} (\sum_{i=1}^{\eta} (\sum_{j=1}^{n} (\sum_{j=1}^{n} q_{ij}p_{ij} - mp_{ij})^2))$, here $R = (r_{if})_{n \times n} = Q^{-1}$, referred to Formulas (11)–(12) and Theorems 1 and 2 for details;
      9. if $CurrentMaxRE < MaxRE$ then
         10. $MaxRE = CurrentMaxRE$;
      11. end if
      12. if $MaxRE > \eta$ then
         13. $\gamma_\eta = \gamma_\eta + \Delta \gamma_\eta$;
      14. end if
   15. end while
   16. if $\epsilon_i$ is not all zero for $i \in [1, \eta]$ then
      17. Calculate the current point belief degree value according to $C^*_e = \sum_{i=1}^{n} \tilde{p}_i(\epsilon_i, \epsilon_e) = \sum_{i=1}^{n} \sum_{j=1}^{n} q_{ij}p_{ij}(\epsilon_i, \epsilon_e)$, and set $C^*_e = C^*_e$ (or, calculate the current regional average belief degree value according to $C_{Region(\epsilon_i)} = \frac{\sum_{i=1}^{\kappa} \sum_{j=1}^{\kappa} (\epsilon_{e_{k+1}} - \epsilon_{e_k})C_{e_i}}{\epsilon_{e_k} - \epsilon_{e_1}}$, where $C_{e_k} = \sum_{i=1}^{n} \tilde{p}_i(\epsilon_i, \epsilon_{e_k}) = \sum_{i=1}^{n} \sum_{j=1}^{n} q_{ij}p_{ij}(\epsilon_i, \epsilon_{e_k})$, and set $C^*_e = C_{Region(\epsilon_i)}$);
         18. if $C^*_e < C^*_e$ then
            19. Update $tag = 1, C_e = C^*_e, \gamma_{opt} = \gamma_\eta, \kappa_{opt} = \kappa_\eta, Q_{opt} = Q$ and $\epsilon_{i_{opt}} = \epsilon_i (i \in [1, \eta])$;
         20. end if
      21. end if
      22. Update $\kappa_\eta = \kappa_\eta - 1$;
   23. end while
   24. if $tag == 0$ then
      25. Update $\kappa_\eta = 0$, and update the value $\epsilon_i$ according to Step 7;
   26. else
      27. Record $\gamma_\eta = \gamma_{opt}, \kappa_\eta = \kappa_{opt}, Q = Q_{opt}$ and $\epsilon_i = \epsilon_{i_{opt}} (i \in [1, \eta])$;
   28. end if
29. return $\gamma_\eta, \kappa_\eta, Q$ and $\epsilon_i$ for $i \in [1, \eta]$. 


(2) Point belief degree

In the point belief degree $C_{e_e}$ in KRR, let $e_e = e_\eta$ ($e_\eta$ is used as the expected privacy budget or used as a basis for division of the expected privacy protection region, just for better comparison between KRR and EXP$_Q$) determined by $e_\eta$-KRR (see definition of $e_\eta$-KRR and Algorithm 1 for details), which equals the $e_e$-coordinate of the jump point shown by the dotted line in Figure 3. It is also combined with the same $e_e$ and $\eta$ to determine the perturbation probability and related parameters with EXP$_Q$ (see Algorithm 2 for details). For example, when the EDU $\eta = 0.1$, the point belief degree $C_{e_e}$ of KRR and EXP$_Q$ is shown in Figure 3.

From Figure 3, it can be seen that under the same EDU, if the expected privacy budget of any data provider is $e_e \geq e_\eta$, KRR can provide $e_e$-DP with belief degree of 1. On the other hand, if the expected privacy budget of any data provider is $e_e < e_\eta$, its belief degree is 0. However, in EXP$_Q$, if the expected privacy budget of any data provider is $e_e \geq e_\eta$, it indicates that it cannot satisfy the EPP when $e_e$ is closer to $e_\eta$, and when $e_e$ is large enough, it can also provide $e_e$-DP with belief degree of 1. Conversely, if the expected privacy budget of all data providers is $e_e < e_\eta$, it indicates that it can satisfy the EPP when $e_e$ is closer to $e_\eta$, and the degree of providing $e_e$-DP is greater when $e_e$ is closer to $e_\eta$. Therefore, in the case of EDU first, EXP$_Q$ can provide a privacy guarantee degree between 0 and 1 for the EPP, while KRR can only provide either 0 or 1. Moreover, EXP$_Q$ can provide more privacy protection than KRR, especially when the EDU and the EPP are contradictory, and when the EPP of all data providers is not fully (partially) satisfied.

Figure 2. Pareto distribution.

![Figure 2. Pareto distribution.](image.png)

Figure 3. Point belief degree ($C_{e_e}$) with $\eta = 0.1$.

![Figure 3. Point belief degree ($C_{e_e}$) with $\eta = 0.1$.](image.png)
(3) Regional average belief degree

In the regional average belief degree $C_{Region(e)}$ in KRR, maximizing $C_{Region(e)}$ equals to maximize $C_P$, and hence it is the same as Algorithm 1. According to the approximately optimal expected privacy budget $\epsilon_\eta$ under satisfying the EDU $\eta$, the data provider’s expected privacy protection region $Region(e)$ can be roughly divided into three categories: $\{\forall e \in Region(e) > \epsilon_\eta\}$, $\{\forall e \in Region(e) < \epsilon_\eta\}$, and $\{\{\epsilon_\eta\} \subset Region(e)\}$.

Similarly, EXPQ can provide different levels of optimal privacy protection for the three categories of expected privacy protection region (see Algorithm 2 for details). Generally speaking, the regional average privacy protection degree of EXPQ in region $\{\forall e \in Region(e) > \epsilon_\eta\}$ is less than or equal to that of KRR. However, the regional average privacy protection degree of EXPQ in region $\{\forall e \in Region(e) < \epsilon_\eta\}$ is greater than or equal to that of KRR. For region $\{\{\epsilon_\eta\} \subset Region(e)\}$, it may exist the situation where there is a contradiction between the EPP and the EDU. As shown in Figure 4, there is the regional average belief degree $C_{Region(e)}$ of both mechanisms with $Region(e) = [1, 1.001, 1.002, \cdots, 4]$ and the data distributions $P_1$, $P_2$ and $P_3$, respectively, where $\epsilon_\eta$ is determined by $\epsilon_\eta$-KRR with $\eta = 0.1$ (see the dotted line in Figure 4 where the value of $\epsilon_e$ whose $C_{Region(e)}$ is the first non-zero value is equal to $\epsilon_\eta$).

As can be seen from Figure 4, under the same EDU and the same expected privacy protection region, EXPQ is more capable of offering data providers with a certain degree of privacy protection than KRR.

6. B-DP Dynamic Collection and Publishing Algorithm Design

Algorithms 1 and 2 are implemented with KRR and EXPQ under the known data distribution, and moreover, the point belief degree and the regional average belief degree under B-DP are analyzed. In real-world, there is often no prior data distribution at the beginning or accurate prior data distribution cannot be obtained. This means the implementation of two B-DP mechanisms of Algorithms 1 and 2 cannot be directly applied to the collection and publishing of continuous check-in data with relative error as utility metrics. Therefore, this paper designs an iterative update algorithm to adaptively update the data distribution in order to realize the two B-DP mechanisms, so as to adaptively realize B-DP dynamic collection and publishing of continuous check-in data. See pseudo codes of Algorithm 3 for more details. Therein, Algorithms 1 or 2 is a main part of Algorithm 3.
Algorithm 3 B-DP dynamic collection and publishing of check-in data algorithm—(KRR/EXP_Q).

Initialization process: The data collector initializes the perturbation probability matrix \( Q \), and the estimated data distribution \( \hat{p}(0) = [\hat{p}_1(0), \hat{p}_2(0), \ldots, \hat{p}_n(0)]^T \), and the perturbation probability matrix \( Q \) is broadcasted to the data provider.

1: For KRR, initialize \( p_i = \frac{1}{n} \) for \( i \in [1, n] \) and \( \epsilon_\eta = \ln(\frac{1+\sqrt{n^2+n-1}}{n}) \); For EXP_Q,

initialization \( p_i = \frac{1}{n} \) for \( i \in [1, n] \), \( \kappa_n = 0 \) and \( \gamma = \frac{n}{n+1} \ln(\frac{1+\sqrt{n^2+n-1}}{1}) \);

2: \( Q \) is constructed according to KRR/EXP_Q;

3: Initialize \( \hat{p}(0)_i = 0 \) for \( i \in [1, n] \);

4: \( Q \) is broadcasted to the data provider.

Perturbation process
1: The data provider uses \( Q \) to perturb the check-in data;
2: The perturbed check-in data are sent to the data collector.

Statistics and update processes
(1) Statistical process: including aggregation and reconstruction procedures
1: After collecting the check-in data of time slice \( T \), the data collector carries out frequency statistics to get the perturbed data distribution \( \tilde{p} \). Assuming that the current time slice is the \( t \)th, it is recorded as \( \tilde{p}_t \);
2: From the inverse estimation formula \( \hat{p}_t = Q^{-1}\tilde{p}_t \), it obtains the estimated distribution \( \hat{p}_t \);
3: Correcting the estimated data distribution \( \hat{p}_t \), it gets the corrected estimate data distribution \( \hat{p}^{(t)} \) of the \( t \)th time slice (to be released);
if \( \hat{p}^{(t-1)} \neq 0 \) then
\( \hat{p}^{(t)} = (1 - w)\hat{p}^{(t-1)} + w\hat{p}\), where \( w \in (0, 1) \) is the corrected estimate parameter of a positive real number;
else
\( \hat{p}^{(t)} = \hat{p}_t \);
end if

(2) Update process
1: The data collector calculates maximum relative error \( Re = \max(\frac{\hat{p}^{(t)} - \hat{p}^{(t-1)}}{\hat{p}^{(t-1)}}) \) based on \( \hat{p}^{(t)} \) and \( \hat{p}^{(t-1)} \);
2: Initialize \( Re_{threshold} \in (0, 1) \), which is a update threshold parameter of a positive real number;
if \( Re > Re_{threshold} \) then
2-1: Start a process of updating \( Q \) to get a new \( Q \), as shown in Algorithm 1/Algorithm 2, where the input data distribution of Algorithm 1/Algorithm 2 is \( \hat{p}^{(t)} \);
2-2: The data collector broadcasts the new \( Q \) to the data provider.
end if

Since the original data distribution is assumed to be uniform during initialization, it is possible to calculate the privacy setting parameter \( \epsilon_\eta \) or \( \gamma \) with a closed-form expression that satisfies the EDU \( \eta \), as shown in the example with EXP_Q below. According to Corollary A1 of Appendix A, in the case of uniform data distribution, EXP_Q degenerates into KRR.
Let $\kappa_n = 0$, the probability $q_{ij}$ of $Q$ be calculated as follows, where $\gamma_\eta$ is $\gamma$ that makes $\max(\text{err}(p_i, \hat{p}_i)) = \eta$ true.

$$q_{ij} = \begin{cases} e^{\gamma_\eta \left(1 + \frac{1}{n}\right)} & i = j \\ e^{\gamma_\eta \left(1 + \frac{1}{n}\right) + n - 1} & i \neq j \end{cases}$$

(25)

Let $p = \frac{e^{\gamma_\eta \left(1 + \frac{1}{n}\right)}}{e^{\gamma_\eta \left(1 + \frac{1}{n}\right) + n - 1}}, q = \frac{1}{e^{\gamma_\eta \left(1 + \frac{1}{n}\right) + n - 1}},$ and the inverse matrix $R$ of $Q$ can be expressed as

$$r_{ij} = \begin{cases} 1 & i = j \\ \frac{1 - p - q}{p - q} & i \neq j \end{cases}$$

(26)

Therefore, $p$, $q$ and $\gamma_\eta$ that satisfy the EDU $\eta$ can be calculated. Since $p \geq q$, it has $q = -\frac{1}{n} \sqrt{\frac{n - 1}{mp^2 + n - 1}} + \frac{1}{n}$ and $p = \frac{n - 1}{n} \sqrt{\frac{n - 1}{mp^2 + n - 1}} + \frac{1}{n}$. Hence, $\gamma_\eta = \frac{n}{n + 1} \ln \left(\frac{p}{q}\right)$, and $\gamma = \gamma_\eta = \frac{n}{n + 1} \ln \left(1 + \left(n - 1\right) \sqrt{\frac{n - 1}{mp^2 + n - 1}}\right)$.

### 7. Experimental Evaluation of B-DP Dynamic Collection and Publishing Algorithm

In this paper, the check-in data uses relative error as its utility metrics and the implementation of the two B-DP mechanisms based on KRR and $\text{EXP}_Q$ needs to rely on the data distribution. Therein, the number of domain values of both KRR and $\text{EXP}_Q$ is more than 2, and moreover, both the randomized algorithms based on them only take one value as input and one value as output. Thereby, KRR and $\text{EXP}_Q$ are fit for the check-in perturbation model we consider in this paper. In this section, we evaluate the performance of the dynamic algorithm based on the two B-DP mechanisms in terms of validity and robustness as well as privacy and utility. For simplicity, in the following of this section, we use KRR and $\text{EXP}_Q$ to represent B-DP mechanism based on KRR and B-DP mechanism based on $\text{EXP}_Q$ in the dynamic algorithm, respectively, including the diagram descriptions.

#### 7.1. Experimental Settings

(1) Datasets

Two datasets with real-world data from location-based social networking platforms are used to verify the algorithms.

**Brightkite [6]:** It contains 4,491,143 check-ins over the period of April 2008–October 2010. In this paper, we used the check-ins of June 2008–September 2008, September 2009–December 2009 and January 2010—April 2010 from Brightkite to construct three types of check-in distributions with different uniformity degree according to the unified longitude and latitude division method, which are, respectively, abbreviated as B1, B2 and B3, and the number of regions is 12.

**Gowalla [6]:** It contains 6,442,890 check-ins over the period of February 2009–October 2010. We used the check-ins of January 2010–April 2010 of Gowalla to construct three types of check-in distributions with different uniformity degree according to different partitioning methods of longitude and latitude, which are, respectively, abbreviated as G1, G2 and G3, and the number of regions is 25, respectively.

The average data distribution and the corresponding Gini coefficient of the data are shown in Figure 5 and Table 4, respectively. Therein, Gini coefficient is used to indicate the degree of unevenness of data distribution, which is calculated according to the method of Gini mean difference [52].
Figure 5. The average data distribution of Brightkite VS. Gowalla.

Table 4. Gini coefficient of data in Brightkite and Gowalla.

| Datasets  | Data Distributions | Gini Coefficient |
|-----------|--------------------|------------------|
| Gowalla   | G1 0.2357          |                  |
|           | G2 0.3488          |                  |
|           | G3 0.4465          |                  |
| Brightkite| B1 0.2849          |                  |
|           | B2 0.3488          |                  |
|           | B3 0.4628          |                  |

Figure 5 and Table 4 both show that the daily check-in data in two datasets fluctuates greatly, meaning a high diversity. We verify the effectiveness of our algorithms on these real-world datasets in our experiment.

(2) Utility/Privacy Metrics

Utility Metrics: The utility uses the maximum relative error as its metrics (see Section 3.3 for details). In this paper, it uses the mean and deviation of the maximum relative error to evaluate the same EDU between KRR and EXPQ in the dynamic algorithm.

Privacy Metrics: The privacy uses two new metrics including the point belief degree and the regional average belief degree (see Definitions 4 and 5 for details). In this paper, it needs to compare the privacy guarantee degree about the expected privacy protection (EPP) under the same expected data utility (EDU) using these two privacy metrics between KRR and EXPQ in the dynamic algorithm.

(3) Parameter Settings

We evaluate our solutions through experiments using two real-world datasets. The experiments are performed on an Intel Core CPU 2.50-GHz Windows 10 machine equipped with 8 GB of main memory by matlab. In the experiments, the total check-in amount of statistical validity is $m = 100,000$. Three kinds of EDU are $\eta = 0.1, 0.08$ and $0.05$. The expected privacy protection region is $\text{Region}(c_\epsilon) = [1, 1.001, 1.002, \ldots, 4]$ or $\text{Region}(c_\epsilon) = [1, 1.001, 1.002, \ldots, 10]$. The modified estimate parameter $\epsilon_0$ is set as Table 5. The update threshold parameter is $\epsilon_{\text{threshold}} = 0.02$, and the remaining relevant parameters $\epsilon_0$ and $\Delta \epsilon_\eta$ are set to 0.5 and 0.005, respectively.
Table 5. All kinds of modified estimate parameter $w$ used in the dynamic algorithm.

| Mechanisms | Brightkite | Gowalla |
|------------|------------|---------|
|            | B1         | B2      | B3      | G1 | G2 | G3 |
| EXP_Q      | 0.045      | 0.06    | 0.15    | 0.18 | 0.25 | 0.4 |
| KRR        | 0.035      | 0.055   | 0.15    | 0.18 | 0.25 | 0.4 |

7.2. Validity and Robustness Evaluation

The performance of validity and robustness of the corresponding dynamic algorithm with KRR and EXP$_Q$ is examined through the dynamic statistics process with two real-world datasets.

Figures 6 and 7 show the mean values and deviations of the maximum relative error $\text{err}(p, \hat{p})$ under the three kinds of EDU ($\eta = 0.1, 0.08$ and $0.05$), which are shown by the statistics of the corresponding data subsets under B1, B2, B3, G1, G2 and G3 according to the frequency of once a day. Moreover, the frequency of each day is different and each result is repeated 10 times. In both Figures 6 and 7, the horizontal axis of each graph represents the number of time slices in continuously, and the vertical axis represents the maximum relative error $\text{err}(p, \hat{p}) = \max\text{err}(p_i, \hat{p}_i)$ between the original data distribution and the estimated data distribution for any $i \in [1, n]$. As can be seen from the left small graphs of Figures 6 and 7, the corresponding dynamic algorithm with KRR and EXP$_Q$ can converge quickly and maintain the corresponding unified convergence stable state under different data distributions of B1, B2, B3, G1, G2 and G3. This verifies that the dynamic algorithm has a good validity and robustness.

7.3. Utility and Privacy Evaluation

The performance of utility and privacy of the corresponding dynamic algorithm with two B-DP mechanisms based on KRR and EXP$_Q$ is also examined through the dynamic statistics process with two real-world datasets. As can also be seen from the right small graphs of Figures 6 and 7, a part of the left small graphs of Figures 6 and 7, it shows clearly that the dynamic algorithm can satisfy the utility even during the dynamic process.

In addition, Figures 8 and 9 show the point belief degree and the regional average belief degree of each subset of two datasets under the three kinds of EDU ($\eta = 0.1, 0.08$ and $0.05$). In Figures 8 and 9, the horizontal axis of each graph represents the EPP with different expected privacy budget $\epsilon_e$ and the vertical axis represents the guarantee degree of the EPP satisfied. It shows that the guarantee degree of the EPP satisfied varies with the data distribution and EDU. For example, from the point belief degree of all small graphs in the left of Figures 8 and 9, the guarantee degree of the EPP satisfied becomes better until its value up to 1 when the expected privacy budget becomes bigger, and the more evener distribution can support the EPP with the smaller $\epsilon_e$ to provide a better privacy protection in the same EDU (as Figure 10 shown). The smaller value of EDU, i.e., the lower utility, can generally support the EPP with the smaller $\epsilon_e$ to provide a better privacy protection in the same data distribution (as Figure 11 shown).

In Figure 10, for EXP$_Q$ on G1, G2 and G3 with the given EDU (such as $\eta = 0.1$), it shows clearly that the minimum $\epsilon_e$ with $C_{\epsilon_e} > 0$ is the smallest on G1 and the largest on G3. According to Table 4, G3 is pretty uneven, while G1 is relatively even. It is the same for EXP$_Q$ on B1, B2 and B3, KRR on B1, B2 and B3 as well as on G1, G2 and G3. In Figure 11, for EXP$_Q$ on G1, G2 and G3 with the given data distribution (such as G1), it also shows clearly that the minimum $\epsilon_e$ with $C_{\epsilon_e} > 0$ is the smallest on the EDU with $\eta = 0.1$ and the largest on the EDU with $\eta = 0.05$. Similar trends can be observed for EXP$_Q$ on B1, B2 and B3, KRR on B1, B2 and B3 as well as on G1, G2 and G3.
Figure 6. The mean and deviation of $err(p, \hat{p})$ on B1, B2 and B3 subsets of Brightkite. Three $\eta$ settings of EDU, including 0.1, 0.08 and 0.05, are compared. (a) B1; (b) B2; (c) B3.
Figure 7. The mean and deviation of $err(p, \hat{p})$ on G1, G2 and G3 subsets of Gowalla. Three $\eta$ settings of EDU, including 0.1, 0.08 and 0.05, are compared. (a) G1; (b) G2; (c) G3.
Figure 8. The belief degree on B1, B2 and B3 subsets of Brightkite, where the belief degree includes the point belief degree and the regional average belief degree. (a) B1; (b) B2; (c) B3.
Figure 9. The belief degree on G1, G2 and G3 subsets of Gowalla, where the belief degree includes the point belief degree and the regional average belief degree. (a) G1; (b) G2; (c) G3.
Figure 10. The minimum $\epsilon_e$ with $C_{\epsilon_e} > 0$ based on the same EDU ($\eta$) and different data distributions, where $C_{\epsilon_e}$ is the point belief degree on the EPP of $\epsilon_e$, and moreover, $\eta = 0.1, 0.08$ and 0.05 represent three kinds of EDU.

Figure 11. The minimum $\epsilon_e$ with $C_{\epsilon_e} > 0$ based on the same data distribution and different EDU ($\eta$), where $C_{\epsilon_e}$ is the point belief degree on the EPP of $\epsilon_e$, and moreover, $\eta = 0.1, 0.08$ and 0.05 represent three kinds of EDU.

For the regional average belief degree, the similar results can be concluded from all small graphs in the right of Figures 8 and 9. Moreover, in Figure 12, it shows the maximum difference of $C_{\text{Region}(\epsilon_e)}$ in EXP$_Q$ minus $C_{\text{Region}(\epsilon_e)}$ in KRR with different $\eta$ on each subset. It shows that the more unevener the data distribution is, the more bigger the maximum difference is. It means that EXP$_Q$ is more adapt to the unevener data distribution than KRR.

Furthermore, in order to be more objective evaluation of the privacy performance of KRR and EXP$_Q$, it extends to use the privacy metrics of DP to compare the $\epsilon_\eta$ on each subset shown in Table 6, where $\epsilon_\eta$ refers to the privacy budget of a DP mechanism that just satisfies the EDU $\eta$. As can be seen from Table 6, except for $\eta = 0.08$ and $\eta = 0.1$ on B3, all the $\epsilon_\eta$ of EXP$_Q$ is a little greater than those of KRR, which means that EXP$_Q$ provides a little worse DP than KRR. However, EXP$_Q$ could provide better B-DP from a new perspective of preference for privacy and utility than KRR to provide a good trade-off between them.
Figure 12. The maximum difference of $C_{\text{Region}(\epsilon_e)}$ in EXP$_Q$ minus $C_{\text{Region}(\epsilon_e)}$ in KRR with different $\eta$, where $C_{\text{Region}(\epsilon_e)}$ is the regional average belief degree on the region of $\text{Region}(\epsilon_e) = [1, 1.001, 1.002, \cdots, 4]$ or $\text{Region}(\epsilon_e) = [1, 1.001, 1.002, \cdots, 10]$, and $\eta = 0.1, 0.08$ and $0.05$ represents three kinds of EDU.

Table 6. $\epsilon_e$ on each subset, where $\eta = 0.1, 0.08$ and $0.05$ represent three kinds of EDU.

| $\eta$ | Mechanisms | Data Distributions |
|-------|------------|--------------------|
|       |            | B1 | B2 | B3 | G1 | G2 | G3 |
| 0.1   | EXP$_Q$    | 1.738 | 1.874 | 6.627 | 2.928 | 4.429 | 5.774 |
|       | KRR        | 1.55 | 1.83 | 6.71 | 2.65 | 4.09 | 5.695 |
| 0.08  | EXP$_Q$    | 2.017 | 2.128 | 6.967 | 3.204 | 4.688 | 6.072 |
|       | KRR        | 1.91 | 2.01 | 7.065 | 2.945 | 4.505 | 5.955 |
| 0.05  | EXP$_Q$    | 2.568 | 2.751 | 7.995 | 3.858 | 5.451 | 7.041 |
|       | KRR        | 2.435 | 2.57 | 7.95 | 3.66 | 5.375 | 6.995 |

8. Discussions and Conclusions

This paper proposes a concept of best-effort differential privacy (B-DP) with the expected data utility (EDU) satisfied first and then with the expected privacy protection (EPP) satisfied as much as possible, and designs two new metrics including point belief degree and regional average belief degree to measure the guarantee degree of satisfying the EPP. Moreover, we also provide implementation algorithms, including the corresponding dynamic algorithm of two B-DP mechanisms based on KRR and a newly constructed mechanism EXP$_Q$. Extensive experiments on two real-world check-in datasets verify the effectiveness of the concept of B-DP. It also verifies that the dynamic algorithm has a good validity and robustness, and can satisfy the utility even during the dynamic process. Besides, EXP$_Q$ is more adapt to the unevener data distribution and satisfies a better B-DP than KRR to provide a good trade-off between privacy and utility.

Specifically, the point belief degree measures the guarantee degree of privacy protection for any one expected privacy budget, and the regional average belief degree measures the average guarantee degree of the EPP in a region including multiple expected privacy budgets. Compared with the $(\epsilon, \delta)$-DP, the latter can measure only one EPP with the expected privacy budget equal to $\epsilon$ and cannot directly measure the average guarantee degree of the EPP, that is, the $(\epsilon, \delta)$-DP can only measure the guarantee degree of the EPP when $\epsilon_e = \epsilon$, i.e., $1 - \delta$. In addition, many real-world applications can only provide an approximate value of $\epsilon_e$ as their EPP, and hence a neighborhood interval with $\epsilon_e$ can be
regarded as their EPP. Therefore, the regional average belief degree introduced in this paper is very necessary.

Moreover, two B-DP mechanisms based on KRR and newly constructed EXPQ in this paper are applied to the dynamic collection and publishing of check-in data with relative error as its utility metrics. Therein, KRR itself does not depend on the data distribution, but the dynamic collection and publishing algorithm with B-DP mechanism based on KRR needs to, where the privacy setting parameter has to be adjusted with the influence of data distribution to realize the utility guaranteed firstly in real time. In addition, EXPQ itself is dependent on the data distribution to realize some of its outputs having strong privacy protection and some having weak privacy protection, which is different from KRR to provide consistent privacy protection intensity. Thus, the dynamic collection and publishing algorithm based on these two B-DP mechanisms needs to depend on the data distribution, and then it has to face the challenges of algorithm validity and robustness with unknown data distribution. Fortunately, the experimental results have already verified that the algorithm can solve both challenges and is promising for the typical application of check-in data.

Besides, if the scenic spots use EXPQ for privacy protection, the data provider may be more inclined to visit these scenic spots with a large number of visitors, because the regions where these scenic spots are located may have a stronger privacy protection. Compared with the algorithms based on the existing DP mechanisms with consistent privacy protection intensity to realize B-DP, such as KRR in this paper, they maybe do not achieve the EPP at all, but the algorithm based on EXPQ newly proposed in this paper can achieve the EPP partly at least, that is, EXPQ can satisfy a better B-DP to provide a good trade-off between privacy and utility.

In a word, although the B-DP dynamic collection and publishing algorithm based on KRR or EXPQ is not necessarily perfect, it fully proves the feasibility of the concept of B-DP in this paper. It is not only a great step forward for the basic theory of DP, but also provides two feasible solutions for the implementation of DP in practical applications. The two solutions take check-in data as an example, but are not limited to it. They can also be used to other category data for privacy protection where the perturbation model is one input and one output. In the future work, we will make a further discussion on other mechanisms with binary inputs in LDP, where the perturbation model can support one input is perturbed to multiple outputs, such as RAPPOR, and design them to achieve better B-DP. Moreover, it is an interesting problem about correlated B-DP.

Author Contributions: The problem was conceived by Y.C. and Z.X. The theoretical analysis and experimental verification were performed by Y.C., Z.X. and J.C. Y.C. and J.C. wrote the paper. S.J. reviewed the writing on grammar and structure of the paper. All authors have read and agreed to the published version of the manuscript.

Funding: Natural Science Foundation of China (No.41971407), China Postdoctoral Science Foundation (No.2018M633354), Natural Science Foundation of Fujian Province, China (Nos.2020J01571, 2016J01281) and Science and Technology Innovation Special Fund of Fujian Agriculture and Forestry University (No.CXZX2019119S).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Two publicly available datasets were analyzed in this study. Both datasets can be found here: http://snap.stanford.edu/data/loc-gowalla.html and http://snap.stanford.edu/data/loc-brightkite.html (accessed on 5 March 2022).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Theorem 5

Proof. (1) From Definition 11, \(\epsilon_\eta = \max_{1 \leq i \leq n}(\epsilon_i)\). According to the definition of LDP, it can be seen that \(\epsilon_\eta^{-\text{EXPQ}}\) satisfies \(\epsilon = \epsilon_\eta^{-\text{LDP}}\).
If it wants to proof (2) and (3) of Theorem 5, it needs the following theorems and corollary first.

**Theorem A1.** In perturbation mechanism EXP,Q, where i, j ∈ [1, n] and κ_i ∈ [0, n], it has
(a) q_{ij} ≥ q_{i,j}; if j_1, j_2 ∈ [1, n], i_1 ≤ i_2 and j_1, j_2 ≠ i, then q_{i,j_1} ≥ q_{i,j_2};
(b) q_{ii} ≥ q_{ij}; if j_1, j_2 ∈ [1, n], j_1 ≤ j_2 and j_1, j_2 ≠ i, then q_{i,j_1} ≥ q_{i,j_2};
(c) for j_1 ≤ j_2, q_{i,j_1} ≥ q_{i,j_2};
(d) for i ≤ j_2, q_{i,j} ≥ q_{i,j_2}, where q_{i,j}, \hat{q}_{i,j} are the distribution probabilities of POI_{i_1} and POI_{i_2} after perturbation, respectively.

**Proof.** See Appendix B. □

**Theorem A2.** In perturbation mechanism EXP,Q, it satisfies the following inequalities, where i ∈ [1, n]:
(1) For κ_n ∈ [1, n − 1],
(i) if 1 ≤ i ≤ κ_n and j, j’ ∈ [1, n], then \( e^{-\gamma(1-p_{i+1})} ≤ \frac{q_{ij}}{q_{ij'}} ≤ e^{\gamma(1-p_{i+1})} \);
(ii) if κ_n < i ≤ n − 1, κ_n ≤ n − 2 and j, j’ ∈ [1, n], then \( e^{-\gamma(1+p_{n−i+κ_n})} ≤ \frac{q_{ij}}{q_{ij'}} ≤ e^{\gamma(1+p_{n−i+κ_n})} \);
(iii) if κ_n < i = n and j, j’ ∈ [1, n], then \( e^{-\gamma(1+p_{n+1})} ≤ \frac{q_{ij}}{q_{ij'}} ≤ e^{\gamma(1+p_{n+1})} \);
(2) For κ_n = 0,
(i) if 1 ≤ i ≤ n − 1 and j, j’ ∈ [1, n], then \( e^{-\gamma(1+p_{n−1})} ≤ \frac{q_{ij}}{q_{ij'}} ≤ e^{\gamma(1+p_{n−1})} \);
(ii) if i = n and j, j’ ∈ [1, n], then \( e^{-\gamma(1+p_{n})} ≤ \frac{q_{ij}}{q_{ij'}} ≤ e^{\gamma(1+p_{n})} \);
(3) For κ_n = n,
(i) if 1 ≤ i ≤ n − 1 and j, j’ ∈ [1, n], then \( e^{-\gamma(1−p_{i+1})} ≤ \frac{q_{ij}}{q_{ij'}} ≤ e^{\gamma(1−p_{i+1})} \);
(ii) if i = n and j, j’ ∈ [1, n], then \( e^{-\gamma(1−p_{n})} ≤ \frac{q_{ij}}{q_{ij'}} ≤ e^{\gamma(1−p_{n})} \).

**Proof.** See Appendix C. □

From Theorems A1 and A2, it has following Corollary A1.

**Corollary A1.** For any i ∈ [1, n], let \( \epsilon_i \) be the actual privacy budget provided by EXP,Q for POI_i, it has the following properties.
(1) For κ_n ∈ [1, n − 1],
(i) if 1 ≤ i ≤ κ_n, then \( \epsilon_i = \gamma(1−p_{i+1}) \);
(ii) if κ_n < i ≤ n − 1 and κ_n ≤ n − 2, then \( \epsilon_i = \gamma(1+p_{n−i+κ_n}) \);
(iii) if κ_n < i = n, then \( \epsilon_n = \gamma(1+p_{n+1}) \).
(2) For κ_n = 0,
(i) if 1 ≤ i ≤ n − 1, then \( \epsilon_i = \gamma(1+p_{n−1}) \);
(ii) if i = n, then \( \epsilon_n = \gamma(1+p_{1}) \).
(3) For κ_n = n,
(i) if 1 ≤ i ≤ n − 1, then \( \epsilon_i = \gamma(1−p_{i+1}) \);
(ii) if i = n, then \( \epsilon_n = \gamma(1−p_{n}) \).

Let \( \gamma_\eta = \gamma_\eta \), where \( \gamma_\eta \) is the privacy setting parameter \( \gamma \) of EXP,Q that satisfies the EDU err(p, \hat{p}) = \eta. According to Theorem A2, the actual privacy budget for each POI_i can be set to \( \epsilon_i = \phi(\gamma_\eta, p) \), which is a function of \( \gamma_\eta \) and p. It is easy to obtain that \( \epsilon_i \) is a monotonic non-decreasing function according to Theorem A2 and Corollary A1 for a fixed \( \kappa_n \).

Therefore, the \( \epsilon_\eta = \max_{1 \leq i \leq n}(\epsilon_i) \)-EXP,Q satisfies (2) and (3) of Theorem 5 that can be proved as follows.
(2) When $\kappa_n$ is fixed and the point belief degree of $\epsilon_p$-EXP$_Q$ is $C_{e_\epsilon} = \sum_{i=1}^n \bar{p}_i \chi(\epsilon_i, e_{\epsilon}),$ if $C_{e_\epsilon}$ is maximized, then $\epsilon_p$-EXP$_Q$ satisfies $(\epsilon_p, C_{e_\epsilon})$-Best-B-DP. According to Corollary A1, it can be known that the larger $p_i$ is, the smaller $\epsilon_i$ is, indicating that in the same $\epsilon_\epsilon$, $\epsilon_i$ can satisfy $\epsilon_\epsilon$ at first, and the proportion that it satisfies is $\bar{p}_i$. According to Theorem A2, when $p_i$ is larger, $\bar{p}_i$ will always be larger too. Hence, $C_{e_\epsilon}$ is also maximized under a fixed $\kappa_n$.

(3) According to Definition 10 and Corollary A1, in EXP$_Q$, the data distribution $p$ satisfies $p_1 \geq \cdots \geq p_i$. When for any user $C \in \{1, 2, \cdots, i\}$, $R_2 = \{i+1, i+2, \cdots, n\}$, then $P(k \in R_1) = C_{e_\epsilon}, P(k \in R_2) = 1 - C_{e_\epsilon}$. Hence, for any user $u$ and $k, j', j \in [1, n]$, it has

\[
P(\ln(\frac{q_{kj'}}{q_{kj}})) \leq \epsilon_{\epsilon}
\]

\[
P(\ln(\frac{q_{kj'}}{q_{kj}})) \leq \epsilon_{\epsilon}\big|k \in R1\big) + P(k \in R1)
\]

\[
+ P(\ln(\frac{q_{kj'}}{q_{kj}})) \leq \epsilon_{\epsilon}\big|k \in R2\big)P(k \in R2)
\]

\[
= 1 \cdot C_{e_\epsilon} + 0 \cdot (1 - C_{e_\epsilon})
\]

\[
= C_{e_\epsilon}.
\]

Therefore, it is easy to obtain

\[
q_{kj'} \leq e^{\epsilon_\epsilon}q_{kj} + (1 - C_{e_\epsilon})
\]

for any user $u$ and $k, j', j \in [1, n]$.

Therefore, according to Definitions 2, 3 and Theorem A2, $\epsilon_p$-EXP$_Q$ with the point belief degree $C_{e_\epsilon}$ satisfies $(\epsilon_\epsilon, 1 - C_{e_\epsilon})$-LDP.

**Appendix B. Proof of A1**

**Proof.** Because of the perturbation mechanism EXP$_Q$, the check-in data distribution $p = [p_1, p_2, \cdots, p_n]^T$ satisfies $p_1 \geq p_2 \geq \cdots \geq p_n$. In accordance with the $\kappa_n \in [1, n-1], \kappa_n = n$ and $\kappa_n = 0$, it can be discussed for three cases. For $\kappa_n \in [1, n-1]$, if the perturbation mechanism EXP$_Q$ satisfies Theorem A1, then the other two cases, $\kappa_n = n$ and $\kappa_n = 0$, are also proved by using the same method.

For $\kappa_n \in [1, n-1]$, it can be discussed as follows based on Theorem 4.

(a) For any $i \in [1, n]$, the probability $q_{ij}$ that the check-in state of $POI_j$ is perturbed to that of $POI_i$ satisfies the following cases.

(i) For $i \leq \kappa_n$ and $i \neq j$, it has $q_{ij} = \frac{1}{\Omega_i} \geq q_{ij} = \frac{e^{-\gamma(1-p_i)}}{\Omega_i}$. Moreover, for $i_1 \leq i_2, i_1, i_2 \neq j$ and $i_1, i_2 \leq \kappa_n$, it has $q_{i_1j} = \frac{e^{-\gamma(1-p_{i_1})}}{\Omega_i} \geq q_{i_2j} = \frac{e^{-\gamma(1-p_{i_2})}}{\Omega_i}$.

(ii) For $i > \kappa_n$ and $i \neq j$, it has $q_{ij} = \frac{1}{\Omega_i} \geq q_{ij} = \frac{e^{-\gamma(1-p_{i_2})}}{\Omega_i}$. Moreover, for $i_1, i_2 > \kappa_n, i_1 \leq i_2, i_1, i_2 \neq j$, it has $q_{i_1j} = \frac{e^{-\gamma(1+p_{i_1}-q_{i_2n+1})}}{\Omega_i} \geq q_{i_2j} = \frac{e^{-\gamma(1+p_{i_2}-q_{i_2n+1})}}{\Omega_i}$.

(iii) For $i_1 \neq j, i_1 \leq \kappa_n, i_2 \neq j$ and $i_2 > \kappa_n$, it has $q_{i_1j} = \frac{e^{-\gamma(1-p_{i_1})}}{\Omega_i} \geq q_{i_2j} = \frac{e^{-\gamma(1+p_{i_2}-q_{i_2n+1})}}{\Omega_i}$.

From the above discussions (i – iii), the property (a) in this theorem holds for $i_1, i_2 \neq j$ and $\kappa_n \in [1, n-1].$

(b.1) For $j \in [1, n]$, the probability $q_{ij}$ that the check-in state of $POI_j$ is perturbed to that of $POI_i$ satisfies the following cases.
(i) For \( i, j \leq \kappa_n \) and \( j \neq i \), it has

\[
q_{ij} = \frac{1}{\Omega_j} = \frac{e^{-\gamma(1-p_i)} - e^{-\gamma(1-p_j)}}{e^{-\gamma(1-p_i)} - e^{-\gamma(1-p_j)}}
\]

\[
e^{-\gamma(1-p_i)}(1 + \sum_{k=\kappa_n}^{n} e^{-\gamma(1-p_k)} + \sum_{k=\kappa_n+1}^{n} e^{-\gamma(1+p_{n-k}+\kappa_n+1)})
\]

\[
e^{-\gamma(1-p_i)}(1 + \sum_{k=\kappa_n}^{n} e^{-\gamma(1-p_k)} + \sum_{k=\kappa_n+1}^{n} e^{-\gamma(1+p_{n-k}+\kappa_n+1)})
\]

\[
= \frac{1 + \sum_{k=\kappa_n}^{n} e^{-\gamma(1-p_k)} + \sum_{k=\kappa_n+1}^{n} e^{-\gamma(1+p_{n-k}+\kappa_n+1)})}{\Omega_j} = q_{ij}.
\]

(ii) For \( j \neq i, i \leq \kappa_n \) and \( j > \kappa_n \), it has

\[
q_{ij} = \frac{1}{\Omega_j} = \frac{e^{-\gamma(1-p_j)}}{e^{-\gamma(1-p_j)}}
\]

\[
e^{-\gamma(1-p_j)}(1 + \sum_{k=1, k \neq j}^{\kappa_n} e^{-\gamma(1-p_k)} + \sum_{k=\kappa_n+1}^{n} e^{-\gamma(1+p_{n-k}+\kappa_n+1)})
\]

\[
e^{-\gamma(1-p_j)}(1 + \sum_{k=1, k \neq j}^{\kappa_n} e^{-\gamma(1-p_k)} + \sum_{k=\kappa_n+1}^{n} e^{-\gamma(1+p_{n-k}+\kappa_n+1)})
\]

\[
= \frac{1 + \sum_{k=1, k \neq j}^{\kappa_n} e^{-\gamma(1-p_k)} + \sum_{k=\kappa_n+1}^{n} e^{-\gamma(1+p_{n-k}+\kappa_n+1)})}{\Omega_j} = q_{ij}.
\]

(iii) For \( j \neq i, i > \kappa_n \) and \( j \leq \kappa_n \), it has

\[
q_{ij} = \frac{1}{\Omega_j} = \frac{e^{-\gamma(1+p_{n-j}+\kappa_n+1)}}{e^{-\gamma(1+p_{n-j}+\kappa_n+1)}}
\]

\[
e^{-\gamma(1+p_{n-j}+\kappa_n+1)}(1 + \sum_{k=1}^{\kappa_n} e^{-\gamma(1-p_k)} + \sum_{k=\kappa_n+1, k \neq j}^{n} e^{-\gamma(1+p_{n-k}+\kappa_n+1)})
\]

\[
e^{-\gamma(1+p_{n-j}+\kappa_n+1)}(1 + \sum_{k=1}^{\kappa_n} e^{-\gamma(1-p_k)} + \sum_{k=\kappa_n+1, k \neq j}^{n} e^{-\gamma(1+p_{n-k}+\kappa_n+1)})
\]

\[
= \frac{1 + \sum_{k=1}^{\kappa_n} e^{-\gamma(1-p_k)} + \sum_{k=\kappa_n+1}^{n} e^{-\gamma(1+p_{n-k}+\kappa_n+1)})}{\Omega_j} = q_{ij}.
\]

(iv) For \( j \neq i, i > \kappa_n \) and \( j > \kappa_n \), it has
\[ q_{ii} = \frac{1}{\Omega_i} = \frac{e^{-\gamma(1+p_{i,-i}+\kappa_i)}}{e^{-\gamma(1+p_{i,-i}+\kappa_i+1)\Omega_i}} \]

\[ = \frac{e^{-\gamma(1+p_{i,-i}+\kappa_i+1)}}{e^{-\gamma(1+p_{i,-i}+\kappa_i+1)+\sum_{k=1}^{\kappa_i+1} e^{-\gamma(1-p_k)}+\sum_{k=1}^{\kappa_i+1} e^{-\gamma(1+p_{i,-i}+\kappa_i+1)1)}} \]

\[ \geq \frac{e^{-\gamma(1+p_{i,-i}+\kappa_i+1)}}{e^{-\gamma(2+p_{i,-i}+\kappa_i+1)+\sum_{k=1}^{\kappa_i} e^{-\gamma(1-p_k)}+\sum_{k=1}^{\kappa_i} e^{-\gamma(1+p_{i,-i}+\kappa_i+1)1}}} \]

\[ \geq \frac{1}{\Omega_i} = \frac{e^{-\gamma(1+p_{i,-i}+\kappa_i+1)}}{e^{-\gamma(1+p_{i,-i}+\kappa_i+1)+\sum_{k=1}^{\kappa_i} e^{-\gamma(1-p_k)}+\sum_{k=1}^{\kappa_i} e^{-\gamma(1+p_{i,-i}+\kappa_i+1)1}}} \]

\[ = q_{ij}. \]

From the above discussions (i – iv), it can be seen that \( q_{ii} \geq q_{ij} \) for \( j \in [1, n] \).

(ii) For \( j_1 \leq j_2, j_1, j_2 \neq i \) and \( j_1, j_2 \in [1, n] \), the probability \( q_{j_1 j_2} \) that the check-in state of \( POI_{j_1} \) is perturbed to that of \( POI_{j_2} \) and the probability \( q_{j_2 j_1} \) that the check-in state of \( POI_{j_2} \) is perturbed to that of \( POI_{j_1} \) satisfy the following cases.

(i) For \( i \leq \kappa_i \) and \( j_1, j_2 \leq \kappa_{j_2}, \) it has

\[ q_{j_1 j_2} = \frac{e^{-\gamma(1-p_i)}}{\Omega_{j_2}} \]

\[ = \frac{e^{-\gamma(1-p_i)}}{1+\sum_{k=1}^{\kappa_i} e^{-\gamma(1-p_k)}+\sum_{k=\kappa_i+1}^{n} e^{-\gamma(1+p_{i,-i}+\kappa_i+1)}} \]

\[ \geq \frac{1}{\Omega_{j_2}} = \frac{e^{-\gamma(1-p_i)}}{1+\sum_{k=1}^{\kappa_i} e^{-\gamma(1-p_k)}+\sum_{k=\kappa_i+1}^{n} e^{-\gamma(1+p_{i,-i}+\kappa_i+1)}} \]

\[ = q_{j_2 j_1}. \]

(ii) For \( j_1, j_2 > \kappa_i \) and \( i \leq \kappa_i, \) it has

\[ q_{j_1 j_2} = \frac{e^{-\gamma(1-p_i)}}{\Omega_{j_2}} \]

\[ = \frac{e^{-\gamma(1-p_i)}}{1+\sum_{k=1}^{\kappa_i} e^{-\gamma(1-p_k)}+\sum_{k=\kappa_i+1, k \neq j_1}^{n} e^{-\gamma(1+p_{i,-i}+\kappa_i+1)}} \]

\[ \geq \frac{1}{\Omega_{j_2}} = \frac{e^{-\gamma(1-p_i)}}{1+\sum_{k=1}^{\kappa_i} e^{-\gamma(1-p_k)}+\sum_{k=\kappa_i+1, k \neq j_1}^{n} e^{-\gamma(1+p_{i,-i}+\kappa_i+1)}} \]

\[ = q_{j_2 j_1}. \]
(iii) For \(i \leq \kappa_n, j_1 \leq \kappa_n\) and \(j_2 > \kappa_n\), it has
\[
q_{ij_1} = \frac{e^{-\gamma(1-p_i)}}{\Omega_{j_1}}
\]
\[
= \frac{e^{-\gamma(1-p_i)}}{1 + \sum_{k=k=1,k\neq j_1}^n e^{-\gamma(1-p_k)} + \sum_{k=\kappa_n+1}^{\kappa_n} e^{-\gamma(1+p_{\kappa_n-k+\kappa_n+1})}}
\]
\[
= \frac{1 + \sum_{k=k=1,k\neq j_1}^n e^{-\gamma(1-p_k)} + \sum_{k=\kappa_n+1}^{\kappa_n} e^{-\gamma(1+p_{\kappa_n-k+\kappa_n+1})}} {e^{-\gamma(1-p_i)}}
\]
\[
\geq \frac{1 + \sum_{k=k=1,k\neq j_1}^n e^{-\gamma(1-p_k)} + \sum_{k=\kappa_n+1}^{\kappa_n} e^{-\gamma(1+p_{\kappa_n-k+\kappa_n+1})}} {\Omega_{j_2}}
\]  
\[\eqno{(A9)}\]

From the above discussions (i - iii) of (b.2), when \(i \leq \kappa_n, j_1, j_2 \neq i, j_1 \leq j_2\) and \(j_1, j_2 \in [1, n]\), \(q_{ij_1} \geq q_{ij_2}\). When \(i > \kappa_n\) for \(j_1, j_2 \neq i, j_1 \leq j_2\) and \(j_1, j_2 \in [1, n]\), we can use the similar method to get \(q_{ij_1} \geq q_{ij_2}\). Then it has \(q_{ij_1} \geq q_{ij_2}\) for \(i, j_1, j_2 \in [1, n]\), \(j_1 \leq j_2\) and \(j_1, j_2 \neq i\).

Hence, from (b.1 ~ b.2), when \(\kappa_n \in [1, n-1]\), the part (b) of Theorem A1 holds.

(c) According to the property (b) and its proof process, it is easy to get \(\tilde{\pi}_j \geq 1,\) i.e., \(q_{ij_1} \geq q_{ij_2}\) holds for \(j_1 \leq j_2\) and \(j_1, j_2 \neq i\).

(d) Since \(\tilde{\pi}_j = \sum_{i=1}^n q_{ij}p_j\) for \(i \in [1, n]\), it has \(i_1 \leq i_2\) that implies \(p_{i_1} \geq p_{i_2}\), and it has Formula (A10). According the property (a) \(q_{ij_1} \geq q_{ij_2}\) and Formula (A10), it can be seen that \(\sum_{j=1,i\neq j \neq j_2}^n (q_{ij_1} - q_{ij_2})p_j \geq 0\).

\[
\tilde{\pi}_1 - \tilde{\pi}_2 = \sum_{j=1}^n q_{ij_1}p_j - \sum_{j=1}^n q_{ij_2}p_j
\]
\[
= \sum_{j=1, j \neq i_1, j \neq i_2}^n (q_{ij_1} - q_{ij_2})p_j + q_{i_1 i_1}p_{i_1} + q_{i_2 i_2}p_{i_2} - q_{i_1 i_2}p_{i_1} - q_{i_2 i_2}p_{i_2}
\]  
\[\eqno{(A10)}\]

If it wants to prove \(\tilde{\pi}_1 \geq \tilde{\pi}_2\) always stands up for \(i_1 \leq i_2\), then it just has to prove \((q_{i_1 i_1} - q_{i_2 i_2})p_{i_1} - (q_{i_2 i_2} - q_{i_1 i_1})p_{i_2} \geq 0\). Similarly, it only wants to discuss the following case \(\kappa_n \in [1, n-1]\), and the other two cases, that is, \(\kappa_n = 0\) and \(\kappa_n = n\) can also be proved by using the same method.

(i) For \(1 \leq i_1 \leq i_2 \leq \kappa_n\), it has
\[
(q_{i_1 i_1} - q_{i_2 i_2})p_{i_1} - (q_{i_2 i_2} - q_{i_1 i_1})p_{i_2}
\]
\[
= \frac{1 - e^{-\gamma(1-p_{i_1})}}{\Omega_{i_1}} p_{i_1} - \frac{1 - e^{-\gamma(1-p_{i_2})}}{\Omega_{i_2}} p_{i_2}
\]
\[
\geq \frac{1 - e^{-\gamma(1-p_{i_1})}}{\Omega_{i_2}} p_{i_1} - \frac{1 - e^{-\gamma(1-p_{i_1})}}{\Omega_{i_2}} p_{i_2}
\]
\[
= \frac{1 - e^{-\gamma(1-p_{i_1})}}{\Omega_{i_2}} (p_{i_1} - p_{i_2}) \geq 0.
\]  
\[\eqno{(A11)}\]
(ii) For $1 \leq i_1 \leq \kappa_n$ and $n \geq i_2 > \kappa_n$, it can also obtain

\[
\begin{align*}
(q_{i_1 i_1} - q_{i_2 i_1}) p_{i_1} & - (q_{i_1 i_2} - q_{i_1 i_2}) p_{i_2} \\
& = \frac{1 - e^{-\gamma(1+p_{n-i_2+k_n+1})}}{\Omega_{i_1}} p_{i_1} - \frac{1 - e^{-\gamma(1-p_{i_2})}}{\Omega_{i_2}} p_{i_2} \\
& \geq \frac{1 - e^{-\gamma(1-p_{i_2})}}{\Omega_{i_1}} p_{i_1} - \frac{1 - e^{-\gamma(1-p_{i_1})}}{\Omega_{i_2}} p_{i_2} \\
& \geq \frac{1 - e^{-\gamma(1-p_{i_1})}}{\Omega_{i_2}} p_{i_1} - \frac{1 - e^{-\gamma(1-p_{i_1})}}{\Omega_{i_2}} p_{i_2} \\
& = \frac{1 - e^{-\gamma(1-p_{i_1})}}{\Omega_{i_2}} (p_{i_1} - p_{i_2}) \geq 0.
\end{align*}
\]

(A12)

(iii) For $\kappa_n < i_1 \leq i_2 \leq n$, it can also get

\[
\begin{align*}
(q_{i_1 i_1} - q_{i_2 i_1}) p_{i_1} & - (q_{i_1 i_2} - q_{i_2 i_2}) p_{i_2} \\
& = \frac{1 - e^{-\gamma(1+p_{n-i_2+k_n+1})}}{\Omega_{i_1}} p_{i_1} - \frac{1 - e^{-\gamma(1+p_{n-i_2+k_n+1})}}{\Omega_{i_2}} p_{i_2} \\
& \geq \frac{1 - e^{-\gamma(1+p_{n-i_2+k_n+1})}}{\Omega_{i_1}} p_{i_1} - \frac{1 - e^{-\gamma(1+p_{n-i_2+k_n+1})}}{\Omega_{i_2}} p_{i_2} \\
& \geq \frac{1 - e^{-\gamma(1+p_{n-i_2+k_n+1})}}{\Omega_{i_2}} p_{i_1} - \frac{1 - e^{-\gamma(1+p_{n-i_2+k_n+1})}}{\Omega_{i_2}} p_{i_2} \\
& = \frac{1 - e^{-\gamma(1+p_{n-i_2+k_n+1})}}{\Omega_{i_2}} (p_{i_1} - p_{i_2}) \geq 0.
\end{align*}
\]

(A13)

From the above discussions ($i \sim iii$), the part (d) in Theorem A1 is true. Therefore, from (a $\sim$ d), the result follows.

Appendix C. Proof of Theorem A2

Proof. It is only to deal with case (1) here, and a similar statement can be made of cases (2) or (3). From Theorem A1, it can be seen that $q_{ij} \geq q_{ij}$ for $i, j \in [1, n]$ and it has $q_{ij} \geq q_{ij}$ for $j \leq j_i, j, k \in [1, n]$ and $j, k \neq i$. Hence, it has the following cases to discuss for $\kappa_n \in [1, n - 1]$.

(i) If $i \leq \kappa_n$ and $j, j' \in [1, n]$, then it has Formula (A14).

\[
\begin{align*}
q_{ij} & \leq q_{ij} = \frac{1}{\Omega_n} \\
& = e^{\gamma(1-p_{i_1})} \frac{1}{\Omega_n} \\
& = e^{\gamma(1-p_{i_1})} \frac{1 + e^{-\gamma(1-p_{i_1})} + \sum_{k=1, k \neq i} e^{-\gamma(1-p_{i_1})} + \sum_{k=x_1+1, k \neq i} e^{-\gamma(1+p_{n-x_1+k_1})}}{1 + e^{-\gamma(1+p_{n-1})} + \sum_{k=x_1+1, k \neq i} e^{-\gamma(1+p_{n-x_1+k_1})}} \\
\end{align*}
\]

(A14)
Therefore, according Formula (A14), if it wants to show $\frac{q_{ij}}{q_{ij'}} \leq e^{\gamma(1-p_{i+1})}$, then it needs to show

$$1 + e^{-\gamma(1-p_i)} - e^{\gamma((p_i-p_{i+1})} \leq e^{\gamma(1-p_{i+1})} e^{-\gamma(1+p_{k+1})},$$

i.e.,

$$1 + e^{-\gamma(1-p_i)} - e^{\gamma(p_i-p_{i+1})} - e^{\gamma(p_i-p_{i+1})} e^{-\gamma(1+p_{k+1})} \leq 0 \quad (A15)$$

always stands up. Let $f(x) = 1 + e^{-\gamma(1-x)} - e^{\gamma(x-p_{i+1})} - e^{\gamma((x-p_{i+1})} e^{-\gamma(1+p_{k+1})}$, where $x \in [0, p_{k+1}]$, $\gamma > 0$ and $0 \leq p_{i+1} \leq x$. If it takes the derivative of $f(x)$ with respect to $x$, it gets $f(x)' = \gamma e^{-\gamma(1-x)} - \gamma e^{-\gamma(1+p_{k+1})} e^{\gamma(x-p_{i+1})} \leq 0$, i.e., $f(x)$ decreases monotonically with $x$. It implies that $f(x) \leq f(0) = 0$ always stands up. Therefore, the inequation $1 + e^{-\gamma(1-p_i)} - e^{\gamma(p_i-p_{i+1})} - e^{\gamma(p_i-p_{i+1})} e^{-\gamma(1+p_{k+1})} \leq 0$ always stands up for $p_i \in [0, p_{k+1}]$. From the above discussion, it can be seen that

$$\frac{q_{ij}}{q_{ij'}} \leq \frac{q_{ii}}{q_{in}} \leq e^{\gamma(1-p_{i+1})}, \quad (A16)$$

Similarly, it has

$$\frac{q_{ij}}{q_{ij'}} \geq \frac{q_{ii}}{q_{ii}} \geq \frac{q_{in}}{q_{ii}} \geq e^{-\gamma(1-p_{i+1})}, \quad (A17)$$

(ii) For $n > i > k_n$, $k_n \leq n - 2$ and $j, j' \in [1, n]$, it has

$$\frac{q_{ij}}{q_{ij'}} \leq \frac{q_{ii}}{q_{in}} = \frac{1}{\gamma(1+p_{i+1})} \Omega_n \quad (A18)$$

Hence, if we want to show $\frac{q_{ij}}{q_{ij'}} \leq e^{\gamma(1+p_{n-i+k_n})}$ according the Formula (A18), then it needs to show that $1 + e^{-\gamma(1+p_{n-i+k_n})} \leq e^{\gamma(p_{n-i+k_n}-p_{n-i+k_n+1})(1 + e^{-\gamma(1+p_{k+1})})},$ i.e.,

$$1 + e^{-\gamma(1+p_{n-i+k_n})} - e^{\gamma(p_{n-i+k_n}-p_{n-i+k_n+1})(1 + e^{-\gamma(1+p_{k+1})})} \leq 0 \quad (A19)$$

always stand up. Let $f(x) = 1 + e^{-\gamma(1+x)} - e^{\gamma(p_{n-i+k_n}-x)(1 + e^{-\gamma(1+p_{k+1})})}$, where $x \in [p_{n-i+k_n+1}(f \in [k_n + 1, n])$, $\gamma > 0$ and $p_{n-i+k_n} \geq x$. If we take the derivative of $f(x)$ with respect to $x$, we get $f(x)' = -\gamma e^{-\gamma(1+x)} + \gamma e^{\gamma(p_{n-i+k_n}-x)(1 + e^{-\gamma(1+p_{k+1})})} \geq 0$, that is, $f(x)$ increases monotonically with $x$. Then $f(x) \leq f(p_{k_n+1}) = 1 + e^{-\gamma(1+p_{k_n+1})} - e^{\gamma(p_{n-i+k_n}-p_{n-i+k_n+1})(1 + e^{-\gamma(1+p_{k+1})})} \leq 0$ always stand up. Therefore, the inequation $1 + e^{-\gamma(1+p_{n-i+k_n})} - e^{\gamma(p_{n-i+k_n}-p_{n-i+k_n+1})(1 + e^{-\gamma(1+p_{k+1})})} \leq 0$ always stand up for $p_{n-i+k_n+1} \in [p_{n-i+k_n+1}(f \in [k_n + 1, n]).$

Similarly, it has

$$\frac{q_{ij}}{q_{ij'}} \geq \frac{q_{ii}}{q_{ii}} \geq \frac{q_{in}}{q_{ii}} \geq e^{-\gamma(1+p_{n-i+k_n})}, \quad (A20)$$

\text{Entro}y 2022, 24, 404
(iii) For $i = n$ and $j, j' \in [1, n]$, it has

$$\frac{q_{nj}}{q_{nj'}} \leq \frac{q_{nn}}{q_{n(n-1)}} = \frac{1}{e^{\gamma(1+p_{n+1})}} \frac{1}{\Omega_{n-1}}$$

$$\leq e^{\gamma(1+p_{n+1})} \frac{\Omega_{n-1}}{\Omega_n} \frac{1 + e^{-\gamma(1+p_{n+1})} + \sum_{k=1}^{n} e^{-\gamma(1-p_{k})} + \sum_{k=1}^{n} e^{-\gamma(1-p_{k})} + \sum_{k=1}^{n} e^{-\gamma(1-p_{k})} + \sum_{k=1}^{n} e^{-\gamma(1-p_{k})}}{1 + e^{-\gamma(1+p_{n+2})} + \sum_{k=1}^{n} e^{-\gamma(1-p_{k})} + \sum_{k=1}^{n} e^{-\gamma(1-p_{k})} + \sum_{k=1}^{n} e^{-\gamma(1-p_{k})}} \leq e^{\gamma(1+p_{n+1})}.$$  

(A21)

Similarly, it has

$$\frac{q_{nj}}{q_{nj'}} \geq \frac{q_{nn}}{q_{n(n-1)}} \geq e^{-\gamma(1+p_{n+1})}.$$  

(A22)

From the above discussion of $(i) \sim (iii)$, it can be seen that the part (1) of Theorem A2 holds.

References

1. Patil, S.; Norcie, G.; Kapadia, A.; Lee, A.J. Reasons, rewards, regrets: Privacy considerations in location sharing as an interactive practice. In Proceedings of the 8th Symposium on Usable Privacy and Security (SOUPS), Washington, DC, USA, 11–13 July 2012; pp. 1–15.

2. Patil, S.; Norcie, G.; Kapadia, A.; Lee, A. “Check out where I am!”: Location-sharing motivations, preferences, and practices. In Proceedings of the Extended Abstracts on Human Factors in Computing Systems (CHI), Austin, TX, USA, 5–10 May 2012; pp. 1997–2002.

3. Lindqvist, J.; Cranshaw, J.; Wiese, J.; Hong, J.; Zimmerman, J. I’m the mayor of my house: Examining why people use foursquare-a social-driven location sharing application. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI), Vancouver, BC, Canada, 7–12 May 2011; pp. 2409–2418.

4. Guha, S.; Birmelho, J. Can you see me now?: Location, visibility and the management of impressions on foursquare. In Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI), Munich, Germany, 27–30 August 2013; pp. 183–192.

5. Gruteser, M.; Grunwald, D. Anonymous usage of location-based services through spatial and temporal cloaking. In Proceedings of the 1st International Conference on Mobile Systems, Applications and Services (MObSYS), San Francisco, CA, USA, 5–8 May 2003; pp. 31–42.

6. Cho, E.; Myers, S.A.; Leskovec, J. Friendship mobility: User movement in location-based social networks. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD), San Diego, CA, USA, 21–24 August 2011; pp. 1082–1090.

7. Huo, Z.; Meng, X.; Zhang, R. Feel free to check-in: Privacy alert against hidden location inference attacks in GeoSNs. In Proceedings of the International Conference on Database Systems for Advanced Applications (DASFAA), Wuhan, China, 22–25 April 2013; pp. 377–391.

8. Naghizadeh, E.; Bailey, J.; Kulik, L.; Tanin, E. How private can I be among public users? In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UBICOMP), Osaka, Japan, 7–11 September 2015; pp. 1137–1141.

9. Rossi, L.; Williams, M.J.; Stich, C.; Musolesi, M. Privacy and the city: User identification and location semantics in location-based social networks. In Proceedings of the 9th International AAAI Conference on Web and Social Media (ICWSM), Oxford, UK, 26–29 May 2015; pp. 387–396.

10. Sweeney, L. k-anonymity: A model for protecting privacy. Int. J. Uncertain. Fuzz. 2002, 10, 557–570. [CrossRef]

11. Machanavajjhala, A.; Kifer, D.; Gehrke, J.; Venkitasubramaniam, M. l-diversity: Privacy beyond k-anonymity. ACM Trans. Knowl. Discov. Data. 2007, 1, 3–54. [CrossRef]

12. Dwork, C.; McSherry, F.; Nissim, K.; Smith, A. Calibrating noise to sensitivity in private data analysis. In Proceedings of the 3rd Theory of Cryptography Conference (TCC), New York, NY, USA, 4–7 March 2006; pp. 265–284.

13. Hay, M.; Rastogi, V.; Miklau, G.; Dan, S. Boosting the accuracy of differentially private histograms through consistency. arXiv 2010, arXiv:0904.0942v5.

14. Xiao, X.; Wang, G.; Gehrke, J. Differential privacy via wavelet transforms. IEEE Trans. Knowl. Data Eng. 2011, 23, 1200–1214. [CrossRef]
44. Ren, X.; Yu, C.M.; Yu, W.; Yang, S.; Yang, X.; McCann, J.A.; Yu, P.S. LoPub: High-dimensional crowdsourced data publication with local differential privacy. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2151–2166. [CrossRef]

45. Chamikara, M.A.P.; Bertok, P.; Khalil, I.; Liu, D.; Camtepe, S.; Atiquzzaman, M. Local differential privacy for deep learning. *IEEE Internet Things* **2020**, *7*, 5827–5842.

46. Ye, D.; Zhu, T.; Cheng, Z.; Zhou, W.; Yu, P.S. Differential advising in multiagent reinforcement learning. *IEEE Trans. Cybern.* **2020**, *in press*.

47. Ying, C.; Jin, H.; Wang, X.; Luo, Y. Double Insurance: Incentivized federated learning with differential privacy in mobile crowdsensing. In Proceedings of the 2020 International Symposium on Reliable Distributed Systems (SRDS), Shanghai, China, 21–24 September 2020; pp. 81–90.

48. McSherry, F.; Talwar, K. Mechanism design via differential privacy. In Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS), Providence, RI, USA, 21–23 October 2007; pp. 94–103.

49. 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 (SIGMOD), New York, NY, USA, 29 June–2 July 2009; pp. 19–30.

50. Garofalakis, M.; Kumar, A. Wavelet synopses for general error metrics. *ACM Trans. Database Syst.* **2005**, *30*, 888–928. [CrossRef]

51. Vitter, J.S.; Wang, M. Approximate computation of multidimensional aggregates of sparse data using wavelets. *ACM Sigmod Rec.* **1999**, *28*, 193–204. [CrossRef]

52. Gini, C. Measurement of inequality of incomes. *Econ. J.* **1921**, *31*, 124–126. [CrossRef]