electronics

Article

Differentially Private Timestamps Publishing in Trajectory

Liang Yan [1,2], Hao Wang [1,3], Zhaokun Wang [1], Tingting Wu [1], Wandi Fu [1,3] and Xu Zhang [1,3,*]

[1] Department of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2] Chongqing Planning and Natural Resources Information Center, Chongqing 401147, China
[3] Key Laboratory of Tourism Multisource Data Perception and Decision, Ministry of Culture and Tourism, Chongqing 400065, China
* Correspondence: zhangx@cqupt.edu.cn

Abstract: In recent years, location-based social media has become popular, and a large number of spatiotemporal trajectory data have been generated. Although these data have significant mining value, they also pose a great threat to the privacy of users. At present, many studies have realized the privacy-preserving mechanism of location data in social media in terms of data utility and privacy preservation, but rarely have any of them considered the correlation between timestamps and geographical location. To solve this problem, in this paper, we first propose a *k*-anonymity-based mechanism to hide the user's specific time segment during a single day, and then propose an optimized truncated Laplacian mechanism to add noise to each data grid (the frequency of time data) of the anonymized time distribution. The time data after secondary processing are fuzzy and uncertain, which not only protects the privacy of the user's geographical location from the time dimension but also retains a certain value of data mining. Experiments on real datasets show that the *TDP* privacy-preserving model has good utility.

Keywords: data publishing; trajectory data; privacy preserving; differential privacy

check for updates

Citation: Yan, L.; Wang, H.; Wang, Z.; Wu, T.; Fu, W.; Zhang, X. Differentially Private Timestamps Publishing in Trajectory. *Electronics* **2023**, *12*, 361. https://doi.org/10.3390/electronics12020361

Academic Editor: Cheng-Chi Lee

Received: 20 December 2022
Revised: 5 January 2023
Accepted: 7 January 2023
Published: 10 January 2023

1. Introduction

With the rapid development of mobile communication technology and intelligent terminal equipment, as well as the increased popularity of the Internet, various Internet applications have been fully integrated into people's daily lives [1]. These include Weibo, WeChat, Zhihu, Instagram, Facebook, Twitter, and other social media platforms. Users from different regions provide channels for sharing experiences and opinions, accelerate and expand the dissemination of information, soften public opinion, etc. However, in social networks, it takes a long time to meet the personalized needs of users and deliver accurate advertisements, during which personal information and behavioral records are collected on a regular basis, resulting in issues such as data access and usage with an increased risk of leaking sensitive user information. In recent years, privacy on social media has experienced an increased risk of data breaches. In 2018, Facebook held information on more than 50 million users. Instagram has millions of users, many of whose personal data such as email addresses and phone numbers were leaked, and the Zynga social gaming platform had 218 million players in 2019. Therefore, regulating the collection of information on personal data and user behavior in social networks is crucial to better protect personal information, which involves the industry. The issue of user privacy security in social media has garnered extensive attention around the world, among the general public and academic circles alike. On social media platforms, one can respond to users' requests to help solve some of their problems. Although some measures have been taken to protect the confidentiality of user information, the problem of user privacy disclosure may also occur. Sanctions and restrictions lack relevance and flexibility, leading to a lack of information privacy security and exacerbating user concerns. The cumulative effect of concerns over information privacy will further influence users' perceptions of social media, thus affecting the future use of social media and user intent and behavior in social networks [2]. The number of physical users is increasing day by day, and user information on social media is hidden. Confidentiality issues are also a concern of the academic community. As a fundamental concept in information privacy research, data privacy is considered a hot topic, and it also has an essential impact on companies that are encouraged to share personal information as their original intention. Social network owners should be aware of user information privacy issues that may arise. It is especially important to address the root causes of user privacy concerns. Therefore, to review the current status of the privacy of social media user information, protect the status quo, and protect the privacy of social media user information, the key factors are researched and analyzed to establish a social media information security system. Implementing such a system will also help increase user information privacy and security awareness.

From a user perspective, our research shows that, while most Facebook users understand the nature of privacy settings [3], they do not use the available privacy settings. In social networking, what factors influence user behavior and readiness to adopt privacy settings? Relevant studies have shown that demographic characteristics represent the main factor influencing the privacy of social media users. A set of factors, mainly gender and cultural factors, influence the perceptions of social network users. Privacy intent is important, and women are more likely than men to log in and perform basic tasks such as deleting photo tags, uploading photos, and requesting friends, while more frequently using privacy settings. Currently, compared with social media users who promote a culture of collectivism, in individualistic cultures, users are more willing to establish confidentiality. In addition, the external environment is also a factor that influences the privacy settings of social media users. The ability to obtain and read privacy-related news, information, and anecdotal stories about privacy breaches can boost the awareness of social media users regarding user-defined privacy settings. In addition, the user privacy settings also involve other areas such as self-efficacy, perceived privacy literacy, privacy management, and perception control.

The differential privacy problem involves statistical problems in databases. Currently, differential privacy has been applied in the data analysis of social networks, and a large number of network analysis techniques have been developed [4–8]. However, social networks face many challenges in privacy protection. Social network data are very large and can be modeled as a graph. Social relationships in social networks are highly connected and do not represent a single individual. The dependence among users in statistical databases may seriously weaken the privacy guarantee brought about by the differential privacy mechanism.

Although various prioritization solutions toward mitigating differentially private release for trajectory data in social media have been proposed, current schemes are still afflicted by the following challenges:

- Weak privacy preservation: Perturbing trajectory data can protect individual's accurate location, but trajectory has a time property. The adversary can still know the habit of people just by observing the time of an individual's trajectory, e.g., we can know the home location of people by observing their location in the morning or evening, even if the trajectory is perturbed.
- Low-level utility: State-of-the-art methods try to add noise into the latitude and longitude of the trajectory to perturb the real value. However, a noisy trajectory means the loss of accuracy, which has a negative effect on trajectory mining tasks. In real-world applications, we do not want this scene to happen.
- These challenges imply that a novel mechanism for differentially private release of trajectory data in social media is in high demand. With respect to the first challenge, to lift the weak privacy preservation, we attempt to perturb the time pattern based on the k-anonymity algorithm. We use *k*-anonymity to generalize the time range of trajectory. Then, the attacker cannot infer the habit of people just by observing the pattern of the

trajectory's time. For the second challenge, we add noise into the time of trajectory instead of the latitude and longitude. In this way, we can guarantee high-level utility because we do not change the value of the trajectory.

On the basis of these considerations, we propose an effective differentially private release solution for trajectory data, including a novel concept of "time differential privacy" (*TDP*) and a truncated Laplacian mechanism to conduct *TDP*. To the best of our knowledge, *TDP* is the first differential privacy technique for trajectory data release that renders the protection of the time pattern of trajectory. The original contributions of this paper are as follows:

- To defend against inference attacks launched by adversaries who can observe the time pattern of trajectory, we first formalize the notion of *TDP*, and then we show the possibility of achieving *TDP* guarantees by augmenting a *k*-anonymity- and truncated Laplace-based mechanism. Our *TDP* solution can protect an individual's habits and does not need to perturb the real value of the trajectory.
- A *k*-anonymity-based algorithm is proposed to hide the real-time pattern of individuals, which is different from the original *k*-anonymity algorithm. Then, the data in the original single time period are anonymously hidden for a whole day. In this way, we can hide individual's accurate time at a specific position even if the adversary knows their position.
- A truncated Laplacian-based mechanism is proposed to add noise to the frequency of temporal data, and the added noise is deleted during the recovery process. By this way, we can guarantee that the noise added in the time domain has no effect on the trajectory mining utility while preserving the time privacy. It is theoretically proven that this mechanism satisfies differential privacy.

In this paper, Section 2 introduces the related work of differential privacy in trajectory protection, Section 3 introduces the preliminary knowledge, Section 4 introduces our proposed solution, Section 5 gives the utility and privacy analysis, Section 6 describes the results of the experimental evaluation, and a summary of this study and potential future research work is provided in Section 7.

2. Related Work

In 2006, Wang et al. [9] first revealed that social networks are somewhat strange; teenagers know that government agencies collect their data but still share ideas and personal information on social networks. This phenomenon is known as the "privacy paradox", which refers to the relationship between the actual protection of people's privacy and users' perceived privacy concerns [10]. The contradiction in privacy risk perceptions is an important manifestation of the extent of privacy in social networks. Huo [11] believed that people often need to give up social media privacy for impression management, while social media owners agree and even encourage users to share; this paradox essentially stems from the general behavior of users in transferring privacy, which poses challenges to user privacy and the natural notion that privacy should be protected. Currently, social media privacy research mainly focuses on privacy [12] and privacy technologies. Personal behavior occurs in two dimensions. Privacy technology refers to the technologies used to protect privacy in social networks and sensitive data, such as personal interests, spatial location, and body hotspots. The development of encryption algorithms and mathematical models emphasizes the anonymity of individuals, for example, in access control fields [13]. Anonymity means the desensitization of private information. Anonymous algorithms such as *k*-anonymity, *l*-diversity, and *m*-invariance are used to hide user identity and protect privacy [14]. Access control refers to adding features such as access subjects and permissions to ensure privacy protection, mainly including independent interview access control, mandatory access control, risk-based adaptive access control systems, and other model approaches [15].

Considering the characteristics of the trajectory, Xu et al. [16] proposed a trajectory data protection method that meets the differential privacy to protect the user's trajectory data privacy. First, the weighted multipoint judgment method is introduced to find the infection point in the track by setting the threshold. Second, the density of each trajectory point is calculated to determine the initial clustering center point. In addition, the improved differential privacy k-means method is used to deal with the privacy protection of trajectory data.

Privacy protection of trajectory data is an important issue. In recent years, some mechanisms have been proposed for different privacy protection issues at the same time. Mahdi Abadi [17] proposed PDP-SAG to solve the problem of the sensitive nature of spatiotemporal trajectory, which is a differential private mechanism combining the generalization and personalization of sensitive attributes. On the basis of the privacy descriptor of the track data record, the sensitive attribute values in the track data record are summarized and a new personalized difference dedicated tree structure is defined. Each track is determined by the generalized sensitive attribute values of the track data record.

Research on privacy technologies is generally biased toward "passive" security [18], lacking information on privacy behaviors and privacy preferences in the context of social networking. Analytics can easily lead to privacy bias. Master user privacy behavior from a perceptual perspective can be used to analyze the user's views on privacy risks and privacy values, as well as awareness and cognitive attributes such as privacy threats [19]. Wu et al. [20] studied leveraging attribution, organizational justice, and rebuilding trust from a privacy perspective. A complex theory was used to build a privacy-aware computational model. Bi et al. [21] analyzed the gap between users' perceptions of privacy value and privacy behavior. The authors provided personalized services by formulating differentiated privacy protection mechanisms. Ma et al. [22] discussed the "privacy paradox" from a theoretical level with an in-depth analysis of phenomena and their causes. Hua et al. [23] used mollusks, an Asian game that analyzes the benefits, costs, and losses of different types of privacy behaviors.

From the user's point of view, scholars have studied the willingness of users to read privacy policies [24], taking into account the factors and characteristics of privacy policies in terms of the users themselves. Geng et al. [25] used a structural equation model to analyze survey results and found that users' motivation to read (motivation) and their reading ability positively influenced the reading readiness of the users. Using cognitive load theory, Hay et al. [26] studied the effect of how information on users' reading abilities is presented and found that graphical privacy policies are more readable and understandable than text-based policies. Day et al. [27] adopted the eye-tracking method. By examining how users read the privacy policy, it was found that a privacy policy automatically presented by default is easier to read. In addition, researchers have also studied user demographics [28], educational level [29], information privacy, and the text length of privacy policies. The influences of different factors, such as the scope, typing style, and location, on users' readiness to read privacy policies were analyzed. To increase user intent to read privacy policies, it is hoped that the "informed consent" of a privacy policy will be achieved; therefore, relevant research on compliance with privacy policies will become more meaningful and effective.

Although anonymity-based and DP-based schemes are proposed to protect an individual's trajectory privacy, current methods just protect locations of the trajectory, which face the problem that an attacker can still know the home location of people by observing their location in the morning or evening, even if the trajectory is perturbed. Moreover, a noisy trajectory leads to a loss of accuracy, which has a negative effect on trajectory mining tasks. In this paper, we show the possibility of achieving *TDP* guarantees by augmenting a *k*-anonymity- and truncated Laplace-based mechanism. Our *TDP* solution can protect an individual's habit and does not need to perturb the real value of trajectory.

3. Preliminaries

3.1. k-Anonymity

The k-anonymity privacy protection is based on generalization and suppression. By more generally and abstractly describing data or not publishing certain data, each data record cannot be distinguished, at least from other $k - 1$ data records; thus, user privacy is preserved.

The k-anonymous privacy-preserving technology divides the attributes of the user data tuples into four types:

- Identifier attribute: the identity attribute that can identify an individual;
- Quasi-identifier attribute: an attribute that can identify an individual's identity when linked with other data tables;
- Sensitive attributes: attributes that need to be kept confidential when data are published;
- Non-sensitive attributes: attributes that can be disclosed and have no effect on the privacy of the user, also known as ordinary attributes.

The definition of k-anonymity is provided below.

Definition 1. *(k-Anonymous): The original data table is T (A1, A2, . . . , An), the data table after anonymization is RT (A1, A2, . . . , An), and QID is the corresponding quasi-identifier. If each sequence value in RT[QID] appears at least k times (k > 1) in RT[QID], the data table RT satisfies k-anonymity.*

3.2. Differential Privacy

Differential privacy is a privacy-preserving technology based on data distortion. By adding random noise to each data item in the dataset, the data are distorted to achieve privacy protection. At the same time, the processed data are still required to maintain some statistical properties of data mining and other operations.

Definition 2. *(ε-Differential privacy* [28]*): Define the sibling datasets D_1 and D_2, with a difference in at most one data record. There is a random algorithm M, and range (M) is the value range of the algorithm. If $\forall S \subseteq$ range (M), there are*

$$\Pr[M(D_1) \in S] \le \exp(\varepsilon) \times \Pr[M(D_2) \in S]. \tag{1}$$

Then, the algorithm satisfies the ε-differential privacy-preserving equation, where ε represents the privacy budget, and Pr $[\cdot]$ represents the disclosure risk of the event. A larger ε value represents a smaller degree of privacy protection.

Definition 3. *(Sensitivity* [29]*): Define a query function f: $D \to R^d$ with a sensitivity of*

$$\Delta f = \max||f(D_1) - f(D_2)||_1, \tag{2}$$

where D is the dataset, R^d is the d-dimensional real number space mapped by the query function f, and $\|f(D_1) - f(D_2)\|$ is the first-order norm distance of the dataset and query result.

Definition 4. *(Truncated Laplacian Distribution* [30–35]*).*

Given the privacy parameters $0 < \delta < \frac{1}{2}$, where δ indicates the success rate of attack, privacy budget $\varepsilon > 0$, and the query sensitivity $\Delta f > 0$, the probability density function of the truncated Laplacian distribution f_{TLap} is defined as

$$f_{TLap}(x) = \begin{cases} Be^{-\frac{|x|}{\lambda}}, x \in [-A, A] \\ 0,\ otherwise \end{cases}, \tag{3}$$

where $\lambda = \frac{\Delta f}{\varepsilon}$, and λ is the noise scale. $A = \frac{\Delta f}{\varepsilon}\log\left(1 + \frac{e^{\varepsilon}-1}{2\delta}\right)$ is the range of the noise value. B is a multiple factor to make f_{TLap} a probability density function, which can be obtained by

$$B = \frac{1}{2\lambda\left(1 - e^{-\frac{A}{\lambda}}\right)} = \frac{1}{2\frac{\Delta f}{\varepsilon}\left(1 - \frac{1}{1+\frac{e^{\varepsilon}-1}{2\delta}}\right)}. \tag{4}$$

Random algorithm M satisfies

$$M(D) = f(D) + Lap(\lambda). \tag{5}$$

Then, the algorithm M satisfies the ε-differential privacy-preserving model.

3.3. Attack Model

Before defining the attack model, we first introduce two concepts: the DBSCAN algorithm and histogram.

Concept 1 (DBSCAN algorithm). The DBSCAN algorithm is a typical density clustering algorithm that regards clusters as high-density areas separated by low-density areas and can find clusters of arbitrary shapes and sizes in noisy spatial data.

Concept 2 (histogram). A histogram is an accurate graphical representation of digital data distribution, which is widely used in data publishing, data mining, and analysis. In response to this problem, in this paper, we define a histogram query attack mode.

Histogram query attack: The DBSCAN algorithm is used to cluster a preprocessed dataset, and the locations users often visit, such as for working and living, likely belong to the cluster. As shown in Figure 1, four clusters are identified in the GPS track of user001 of the Geolife GPS trajectory dataset of Microsoft Research Asia. Figure 2 shows the time distribution of the GPS sample points in each cluster. From 9 a.m. to 6 p.m., it is clear that users stay in cluster 1, while, from 12 a.m. to 8 a.m., users tend to stay in cluster 2 and cluster 4. Therefore, it can be inferred that the working point of user001 is cluster 1, while the living point is cluster 2, and cluster 4 may be considered a secondary living point. The real working point and living point of user001 can be obtained by displaying the working and living points on the map. Using more complex heuristics can more accurately infer the user's working and living points, such as the user's location on weekends and workdays.

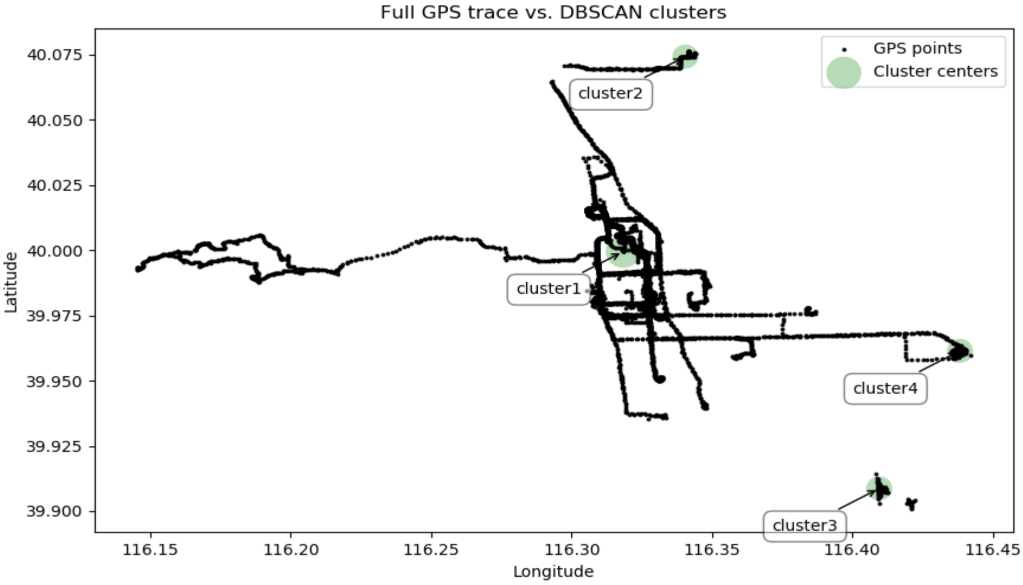


Figure 1. The illustration of user trajectory clustering.

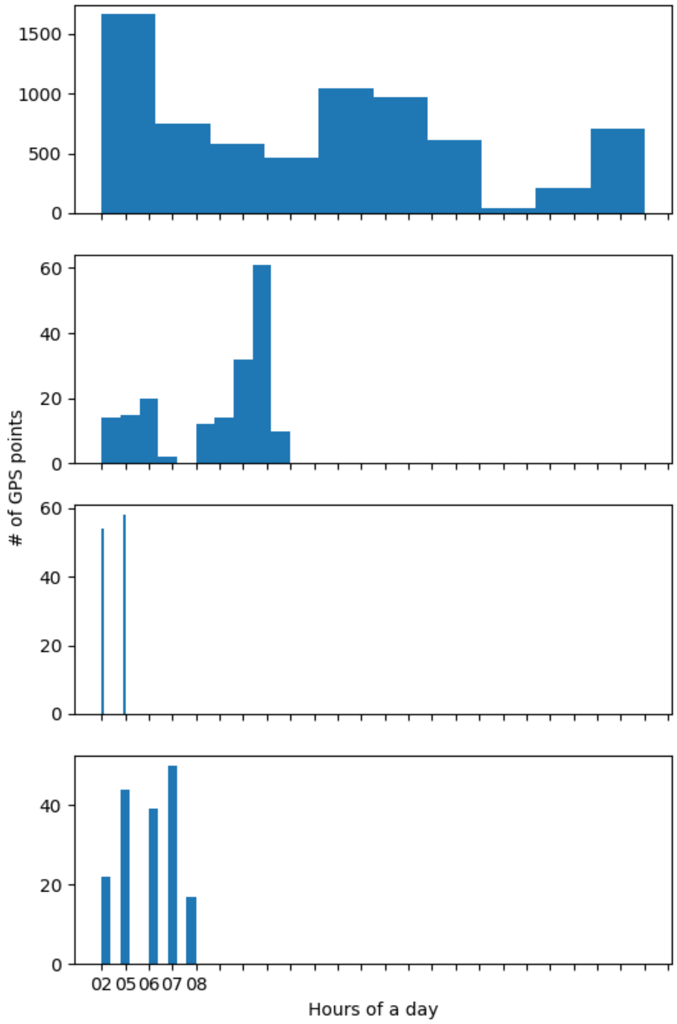


Figure 2. The illustration of user trajectory clustering.

From the above example, we can know that, by observing the clusters of time, the attacker can infer individual's habits, e.g., the working place and hours. Current methods attempt to add noise into the location to protect individual's privacy, but they ignore the disclosure of time pattern. The motive of the attack model is to obtain important information about users, provide favorable research data for attackers, and achieve a better attack effect.

4. Methodology

4.1. Problem Definition

In this paper, we want to hide the time pattern of timestamps of locations generated by individuals. If we use $T = \{t_1, t_2, \cdots, t_n\}$ to denote the set of timestamps, T_i and $T_{i'}$ are arbitrary two subsets of T. They correspond to the high frequency part of T if we transform T to the frequency domain using Fourier transformation. According to the definition of differential privacy, the leakage risk of time pattern is

$$TDP(A) = \sup_{t_i, T_i, T_{i'}} \ln \frac{\Pr[A(t_i \in T_i)]}{\Pr[A(t_i \in T_{i'})]}, \tag{6}$$

where A is the random algorithm, TDP is the abbreviation of timestamps differential privacy.

If the leakage is bounded by ε,

$$TDP(A) \leq \varepsilon,$$

i.e.,

$$\sup_{t_i, T_i, T_{i'}} \ln \frac{\Pr[A(t_i \in T_i)]}{\Pr[A(t_i \in T_{i'})]} \leq \varepsilon. \quad (7)$$

If the above condition is satisfied, *TDP* is considered to have achieved the privacy level of ε, i.e., ε-differential privacy. The challenge in differentially private publication for timestamps is to design mechanisms to satisfy the above definition while guaranteeing high-level utility.

4.2. Sketch of TDP

Before fleshing out the details of *TDP*, which is our solution for differentially private timestamps publishing, a sketch is presented. *TDP* consists of two stages for differential privacy preservation. The first stage is time anonymization to create several same high-frequency parts of timestamps, which can hide the pattern of timestamps. Following this, a truncated Laplace time perturbation mechanism is proposed to perturb the timestamps for each copy of the high-frequency part. In this way, we can hide the repeated high-frequency parts.

4.2.1. Time Anonymization

If we transform the timestamps of original data to the frequency domain, there are a few clusters of the timestamps. These clusters are the accurate clustering timestamps, which disclose the time pattern. From the perspective of original data, these timestamps mean that the person always stays at these positions. To protect these timestamps, we propose a *k*-anonymity-based mechanism to perturb the timestamps. In this way, the centralized data segments in the first half of the day are anonymously distributed throughout the day, and the frequency in the histogram covers almost all time periods.

4.2.2. Truncated Laplace Time Perturbation

Since the frequency of the timestamps is regular after the using of our time anonymization mechanism, the attacker can use this characteristic to sanitize the redundant timestamps which are generated by our *k*-anonymity-based mechanism. To overcome this flaw, we utilize DP to randomize the frequency. By this means, the attacker cannot know the real timestamps.

We discuss time anonymization and truncated Laplace time perturbation in detail in the next two subsections.

4.3. Time Anonymization

In this section, we demonstrate the time anonymous mechanism. Time anonymization uses the *k*-anonymity-based algorithm to copy several timestamps, which are clustered to hide the timestamp pattern. If we add noise into the timestamps directly, the clustering results do not change much. Next, we first give some definitions related to the time anonymization.

Definition 4. *(k-Anonymous timestamps): Suppose the original trajectory dataset is Tra* $= \{(t_1, p_1, s_1), (t_2, p_2, s_2), \cdots, (t_n, p_n, s_n)\}$*, where* t_i*,*p_i*, and* s_i *are the timestamp, position, and speed generated by the GPS device respectively. Among them, we can know that* t_i *is the identifier, while* p_i *and* s_i *are quasi-identifiers (QIs). If there are k subsets of Tra, whose elements only include the timestamps, e.g.,*$\{T_1, T_2, \ldots, T_k\}$*, and if* t_i *is an element of* $\{T_1, T_2, \ldots, T_k\}$*, these are k-anonymous timestamps.*

From Definition 4, we can know that, if we design a mechanism to satisfy Definition 4, the adversary cannot be sure where the specific timestamp t_i is. That is to say, we hide the timestamp t_i in the *k*-timestamp cluster. Next, we present the mechanism to realize *k*-anonymous timestamps.

Consider the time dataset $t_i = \{t_i | t_i \in [t_{\text{min}}, t_{\text{max}}], t_{\text{min}} < t_{\text{max}}\}$ and a subset $T_0 \in T$, $T_0 = \{t_i | t_i \in [t_1, t_2], t_{\text{min}} \le t_1 \le t_2 \le t_{\text{max}}\}, \frac{t_{\text{max}} - t_{\text{min}}}{t_2 - t_1} \ge 2$. There may be a case where the data interval length of the subset T_0 is much shorter than T. This means that the data are concentrated in the interval $[t_1, t_2]$. Under this circumstance, k-anonymity can be adopted. The idea is to anonymize the data and hide the anonymization of a specific time of the day to achieve the effect of privacy protection.

The specific processing method is as follows:

Take $k \in \left[1, \frac{t_{\text{max}} - t_{\text{min}}}{t_2 - t_1}\right]$ and k is an integer and assume there are n time data in T_0.

If k = 1, the time data are more evenly distributed at this time, so no anonymous processing is performed.

If $k \in (1, \frac{t_{\text{max}} - t_{\text{min}}}{t_2 - t_1}]$, then the following data $t_i = t_0$ are calculated for each group in T_0 as follows:

$$\begin{cases} t_i = t_0 + S \times \frac{t_{\text{max}} - t_{\text{min}}}{k-1} \\ \quad S = (i-1)\%k \\ \quad i = 1, 2, 3 \cdots n \end{cases}, \tag{8}$$

where t_0 represents the starting time, S is the remainder of $(i-1)$ and k, and n represents a period when nodes are divided into n segments.

Indeed, k-anonymity in our solution indicates that we want to make "k" copies of the user's time intervals to hide the real-time pattern of the individual, which is different from the original k-anonymity algorithm. Thus, we do not apply the k-anonymity directly; instead, we first judge whether the value of "k" set by the user is suitable for the timestamp protect. Because the timestamps of a day are fixed, we should guarantee that the perturbed timestamps fall into the timestamp range. For example, if $k \notin (1, \frac{t_{\text{max}} - t_{\text{min}}}{t_2 - t_1}]$, then we cannot use k-anonymity. If the value of "k" is determined by the user and it is suitable, then we perform the steps of the algorithm in Equation (8). Next, we summarize the detailed process of time anonymization; Algorithm 1 is its pseudo-code in practice.

Time Anonymization Process

Input: GPS data $T\{t_i \,|\, t_0 \le t_i \le t_{\text{max}}\}$, t_0 and t_{max} as the initialized and biggest timestamps, respectively, k, and ε.

Output: confidence, perturbed GPS data.

Determine whether the expected k value of the initial input is reasonable on the basis of the input time dataset; if it is not reasonable, change it to a reasonable k;

After the value of k is determined, perform anonymous processing on the dataset T according to the k-anonymity implementation method in Section 4.2.

Determine the corresponding Laplacian noise distribution and acceptability according to the input privacy protection budget ε and the individual's acceptable error interval.

Add Laplacian noise to the histogram frequency of the time data.

Apply the truncated Laplacian mechanism to optimize the out-of-bound frequency.

Observe the corresponding data changes.

The pseudo-code of the algorithm is shown in Algorithm 1.

4.4. *Truncated Laplace Time Perturbation*

In Section 4.2, we proposed the time anonymization mechanism to hide the timestamps. However, from the perspective of an adversary, if the time anonymization makes k copies of the timestamps clusters, then it is easy to distinguish whether k copies are the same. Thus, in this section, we propose another truncated Laplace time perturbation mechanism to perturb each of the k timestamps clusters to further hide an individual's information.

In this section, we first propose the definition of truncated Laplace time perturbation, which satisfies DP with a constrained bound. Then, we demonstrate our implement mechanism based on truncated Laplace distribution, which can guarantee DP while limiting the perturbed error into a fixed bound.

Algorithm 1: Time anonymization

```
Input: GPS data $T\{t_i\,|t_0 \leq t_i \leq t_{\max}\}$, $k$, $\varepsilon$.
Output: confidence, perturbed GPS data.
while $k > \frac{t_{max}-t_{min}}{t_2-t_1}$ or $k < 1$ then
    input $k$
end while
for $t_0 \in T\{t_i\,|t_0 \leq t_i \leq t_{\max}\}$do
    $k-\mathrm{ano}(t_0)$
end for
for $t_1 \in T\{t_i\,|t_0 \leq t_i \leq t_{\max}\}$ do
    $\mathrm{Laplace}(t_1)$
end for
for $t_2 \in T\{t_i\,|t_0 \leq t_i \leq t_{\max}\}$ do
    $\mathrm{Truncated}-\mathrm{Laplace}(t_2)$
end for
```

4.4.1. Definition of Truncated Laplace Time Perturbation

As discussed in Section 1, the SP may always feature a noisy error. In this case, the definition of DP is not appropriate. In this section, we propose the notion of truncated Laplace time perturbation with constraints. First of all, we give the definition of a constrained bound.

Definition 5. *(Constrained bound): Denote z as a random noise generated by the privacy preservation method; z is limited by the bound α, and α is the length of the day, for example, the seconds or minutes, i.e., $|z| \leq \alpha$, where z is generated with ε-DP. Then, we can say that the absolute error of z is α.*

Unlike the definition of DP, truncated Laplace time perturbation considers the constrained noise. Truncated Laplace time perturbation must limit the noise to a fixed bound meanwhile satisfying ε-DP. We give its definition below.

Consider two arbitrary timestamps datasets, T_i and $T_{i'}$, which have the same admeasurement, but differ in terms of the record to be protected. Then, the random perturbation mechanism A satisfies ε-DP if A makes all results T_i on two arbitrary timestamps datasets T_i and $T_{i'}$ satisfy

$$\sup_{t_i, T_i, T_{i'}} \ln \frac{\Pr[A(t_i \in T_i)||z| \leq \alpha]}{\Pr[A(t_i \in T_{i'})|z| \leq \alpha]} \leq \varepsilon, \tag{9}$$

where $T_i \subseteq Range(A)$, $Range(A)$ is the value range of random algorithm A. $Pr[\cdot]$ indicates the probability density function (PDF), and ε represents the privacy budget parameter.

4.4.2. Truncated Laplace Time Perturbation Mechanism

Definition 5 gives the formal definition of a constrained bound. Then, we propose a truncated Laplace mechanism to realize constrained bound timestamps DP in practice. The truncated Laplace mechanism is shown in Definition 6.

Definition 6. *(Truncated Laplace perturbation mechanism): A Laplace noise z that conforms to the following distribution satisfies constrained bound timestamps DP:*

$$f(z) = \frac{1}{2\lambda(1-e^{-\alpha/\lambda})} e^{-\frac{|z|}{\lambda}}, z \in [-\alpha, \alpha], \tag{10}$$

where $\lambda = \frac{\Delta f}{\epsilon}$, α is the length of the day (it is also the noise bound), and z is a random noise generated by the truncated Laplace perturbation mechanism.

Definition 6 gives the form of noise that can provide a bounded noisy constraint. Compared to the standard Laplace distribution, the PDF of the truncated Laplace distribution has an extra factor $(1 - e^{-\alpha/\lambda})^{-1}$. The function of this factor is to make the cumulative distribution function of truncated Laplace be 1. Furthermore, the sample of noise is limited by the bound. We can implement this noise form in a post hoc way by discarding the out-of-range cleaned results of the conventional Laplace mechanism until the inbound value is obtained.

Algorithm 2 shows the working flow of the truncated Laplace time perturbation mechanism. In Algorithm 2, we take advantage of our proposed truncated Laplace time perturbation mechanism to perturb the timestamps and limit the noise to the bound (the length of the day). Finally, we publish the perturbed timestamps to protect the real ones.

Algorithm 2: Truncated Laplace time perturbation mechanism

```
Input: Privacy budget $\varepsilon$, timestamps $T = \{t_1, t_2, \cdots, t_n\}$, query function $\Delta f$.
Output: Noise variables $Z = \{z_1, z_2, \cdots, z_n\}$, perturbed query results $T'$.
for each round $k \leftarrow 1, \cdots, n$
    1. Select a timestamp $t_k$;
    2. Compute noise scale parameter $\lambda = \Delta f / \varepsilon$;
    3: Generate truncated Laplace noise $z_k$ according to the PDF in Definition 6 with bound $\alpha$;
    4: Compute the noisy response $t'_k = t_k + z_k$;
end for
return $T'$.
```

5. Security and Utility Analysis

In Section 4, we proposed the anonymity- and truncation-based mechanisms to limit the noisy to a fixed bound. In this section, we analyze the security and utility of our mechanisms. Specifically, in terms of security, we prove that *TDP* also meets the privacy definition of baseline DP. For utility analysis, we deduce the change in noisy variance, which is a base index to measure the performance of utility.

5.1. Security Analysis

In this section, we prove that *TDP* satisfies the requirement of DP. Indeed, the definition of *TDP* is a hard version of DP, i.e., it can also satisfy the definition of DP. We first prove that *TDP* also meets DP, as shown in Theorem 1.

The use of the k-anonymity algorithm and Laplacian mechanism adds Laplacian noise to the data, improves their security, and greatly affects their validity. We define the privacy protection indicators according to the characteristics of the algorithm.

The privacy protection budget ε is used to control the probability ratio of the algorithm to obtain the same output in adjacent datasets, reflecting the level of the privacy protection of the algorithm and the security of the data. A smaller ε denotes a higher level of privacy protection and more secure data. Then, data acceptance is defined by the Laplacian noise probability distribution function.

According to the Laplacian mechanism, the noise we added obeys the Laplacian distribution; if the acceptable error range for the data is $[-\alpha,\alpha]$, we can calculate data acceptance (Accept) as follows:

$$Accept = \int_{-a}^{a} \frac{1}{2b} e^{-\frac{|x|}{b}}. \tag{11}$$

Confidence: First, after using the k-anonymity algorithm, the data in the original single time period are anonymously hidden for a whole day, which means that the original period is divided into k periods. Evidently, the accuracy of the data is only $1/k$ of the original. The Laplacian mechanism is used to achieve differential privacy to further improve data security. Here, data acceptance can be used to express this mechanism. We analyzed the k-anonymity and the correlation of the Laplacian mechanism. The two parameters Δf and ε

in the Laplacian distribution in the *TDP* algorithm have no relationship with k; thus, the value of k does not affect the noise generation. Similarly, k is not affected by the parameters Δf and ε, and we can obtain that $\text{cov}(k, Accept) = 0$. The final data credibility can be calculated as follows:

$$Confidence = \frac{Accept}{k}. \tag{12}$$

Theorem 1. *The truncated Laplace time perturbation mechanism can also preserve ε/k-DP.*

Proof. The proof is similar to that of DP.

$$\begin{aligned}
&\frac{p_T(z)}{p_{T'}(z)} = \prod_{i=1}^{n}\left(\frac{\exp(-\frac{\epsilon|q_i(T)-z_i|}{\triangle f})}{\exp(-\frac{\epsilon|q_i(T')-z_i|}{\triangle f})}\right)/k \\
&= \prod_{i=1}^{n}\exp\Big(\frac{\epsilon(|q_i(T')-z_i|-|q_i(T)-z_i|)}{\triangle f}\Big)/k \\
&\leq \prod_{i=1}^{n}\exp\Big(\frac{\epsilon|q_i(T')-q_i(T)|}{\triangle f}\Big)/k \\
&= \exp\Big(\frac{\epsilon||Q(T)-Q(T')||_1}{\triangle f}\Big)/k \\
&\leq \exp(\varepsilon/k).
\end{aligned}$$

$\frac{p_T(z)}{p_{T'}(z)} \geq \exp(-\varepsilon/k)$ follows by symmetry. □

5.2. Utility Analysis

In this section, we analyze the foundational statistical properties mean and variance to measure the utility loss of our mechanisms.

The standard Laplacian mechanism is a symmetrical distribution with mean 0, and the methods we used are also symmetrical. Thus, our mechanism does not change the mean of noise, as shown in Theorem 2.

Theorem 2. *The mean of the random variables generated by our truncated Laplacian mechanism is 0.*

Proof. Assume z is a random variable that conforms to the truncated Laplacian distribution. Then, the mean of z is

$$E(z) = \int_{-\alpha}^{\alpha} zf(z)dz = \int_{-\alpha}^{\alpha} \frac{z}{2\lambda(1-e^{-\alpha/\lambda})}e^{-\frac{|z|}{\lambda}}dz.$$

Let $\frac{z}{\lambda} = y$; then, we have

$$E(z) = \int_{-\alpha}^{\alpha} \frac{\lambda}{2(1-e^{-\alpha/\lambda})}ye^{-|y|}dy = 0.$$

□

Theorem 2 demonstrates that the mean of our proposed truncated Laplacian mechanism is the same as that of the standard Laplacian mechanism. In addition, we know that the variance of the standard Laplacian distribution is $2\lambda^2$. Theorem 3 demonstrates the variety of variance using our proposed mechanism.

Theorem 3. *Given a random variable z that conforms to the truncated Laplacian distribution, the variance of z is*

$$\sigma_z^2 = \frac{\lambda^2}{1-e^{-\alpha/\lambda}}[2 - e^{-\alpha/\lambda}((\frac{\alpha}{\lambda})^2 + 2\frac{\alpha}{\lambda} + 2)]. \tag{13}$$

Proof. The calculation equation of variance of random variable z is

$$\sigma_z^2 = \int_{-\alpha}^{\alpha} z^2 f(z)dz = \int_{-\alpha}^{\alpha} \frac{z^2}{2\lambda(1-e^{-\alpha/\lambda})} e^{-\frac{|z|}{\lambda}} dz.$$

Let $\frac{z}{\lambda} = y$; then, we have

$$\begin{aligned}
&\sigma_z^2 = \int_{-\alpha/\lambda}^{\alpha/\lambda} \tfrac{\lambda^2}{2(1-e^{-\alpha/\lambda})} y^2 e^{-|y|} dy \\
&= \int_0^{\alpha/\lambda} \tfrac{\lambda^2}{1-e^{-\alpha/\lambda}} y^2 e^{-y} dy \\
&= \tfrac{\lambda^2}{1-e^{-\alpha/\lambda}} e^{-y}(-y^2-2y-2)|_0^{\alpha/\lambda} \\
&= \tfrac{\lambda^2}{1-e^{-\alpha/\lambda}} [2-e^{-\alpha/\lambda}((\tfrac{\alpha}{\lambda})^2+2\tfrac{\alpha}{\lambda}+2)] \\
&= \tfrac{2\lambda^2}{1-e^{-\alpha/\lambda}} \left[1-e^{-\alpha/\lambda}\left(\tfrac{1}{2}(\tfrac{\alpha}{\lambda})^2+\tfrac{\alpha}{\lambda}+1\right)\right] \\
&< 2\lambda^2.
\end{aligned}$$

□

Theorem 3 indicates that the variance of the truncated Laplace random variable is smaller than that of the traditional Laplace mechanism, which may affect the quantity of applications.

6. Experimental Evaluation

6.1. Experiment Datasets and Setup

Geolife: The trajectory dataset of the Geolife project collected trajectory data from 182 volunteers for 5 years (from April 2007 to August 2012), provided by Microsoft Research Asia. Each GPS track is composed of a sequence of time stamps, including latitude and longitude, altitude, time, and other information. This dataset contains 17,621 trajectory data, with a total length of 1,292,951 km and a total duration of 50,176 h.

T-Drive Taxi: This dataset describes the GPS trajectory data of 8602 taxis in Beijing, China in May 2009. The track area covers a rectangular area between latitude and longitude (39.788 N, 116.148 W) and (40.093 N, 116.612 W), with an area close to 34 km × 40 km. The sampling frequency of the trajectories in the dataset ranges from 30 s to 5 min, containing about 4.3 million passenger records, each of which is composed of interpolated sequences with intervals of about 30 s.

The *TDP* algorithm was implemented in Python programming language and run on a Windows 10 platform with a 3.15 GHz CPU and 8.00 GB RAM. The datasets used in this paper were the Geolife Trajectories 1.3 dataset and the T-drive Taxi Trajectories dataset.

Experimental Results and Analysis

To test the effectiveness of the algorithm, we evaluated the algorithm using real spatiotemporal data, compared the *TDP* algorithm with traditional *k*-anonymity, DP-AVG [17], and CIM [21], and used the mean square error and data credibility analyses to evaluate the effectiveness and security of the algorithm from two perspectives.

(1) Intuitive Diagram after *TDP* Algorithm Encryption

Figure 3 presents an analysis diagram of the Geolife Trajectories 1.3 dataset, and Figure 4 provides an analysis diagram of the T-drive Taxi Trajectories dataset. From the visual effects of the *TDP* algorithm in Figures 3 and 4, it can be seen that, after running the *k*-anonymity algorithm of the original data, the centralized data segments in the first half of the day were anonymously distributed throughout the day, and the frequency in the histogram covered almost all time periods. After further using the *TDP* algorithm, the frequency distribution of the histogram was more randomized, and it became difficult to obtain regular data.

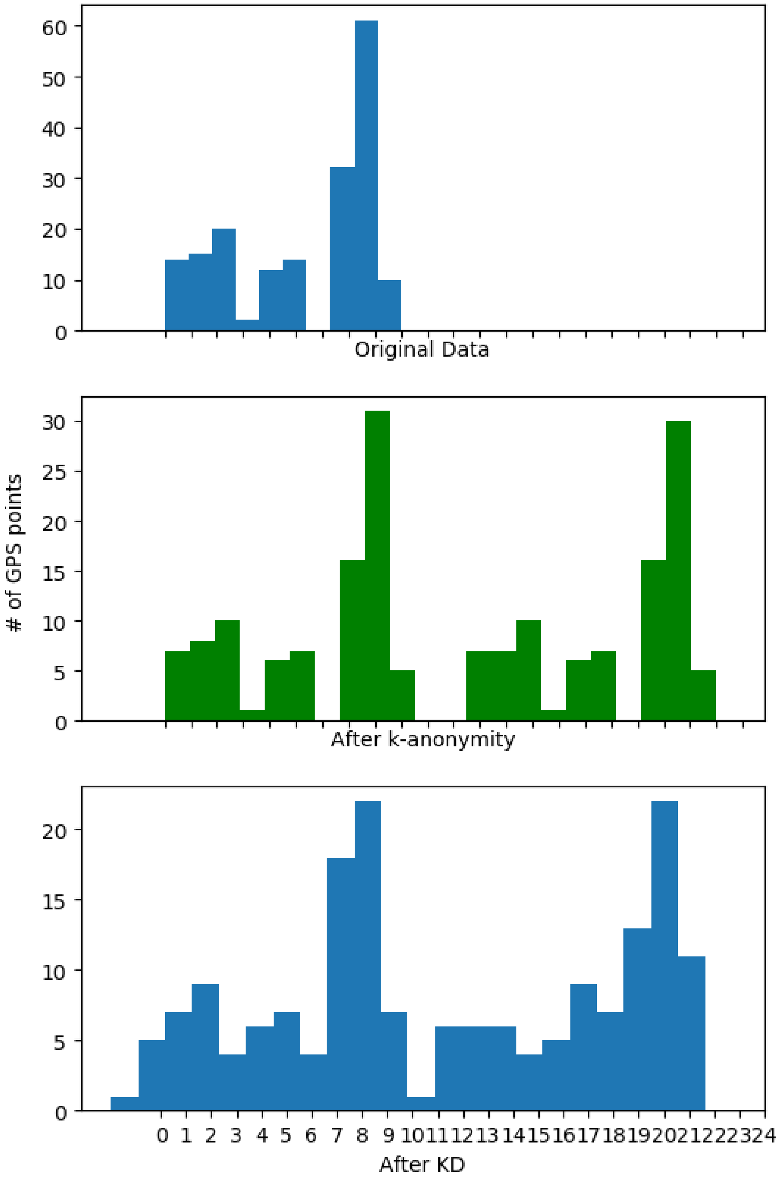


Figure 3. *TDP* for Geolife Trajectories dataset.

From Figures 3 and 4, we can know that, if we transform the timestamps of original data to the frequency domain, there are a few clusters of timestamps. These clusters are the accurate clustering timestamps, which disclose the time pattern. From the perspective of original data, these timestamps mean that the person always stays at these positions. To protect these timestamps, we propose the *k*-anonymity-based mechanism to perturb the timestamps, as shown in the second panel "after *k*-anonymity". In this way, the centralized data segments in the first half of the day are anonymously distributed throughout the day, and the frequency in the histogram covers almost all time periods. Furthermore, since the frequency of the timestamps is regular, the attacker can use this characteristic to sanitize the redundant timestamps which are generated by our *k*-anonymity-based mechanism. To overcome this flaw, we utilize DP to randomize the frequency, as shown in the third panel "after KD". By this means, the attacker cannot know the real timestamps.

(2) Data Security Results and Analysis

We adopted multiple *k*-values to calculate the confidence level. To compare the confidence level of the two datasets, we used $[-0.5,\ 0.5]$ for the confidence interval and k values 2, 3, 4, and 5, as shown below.

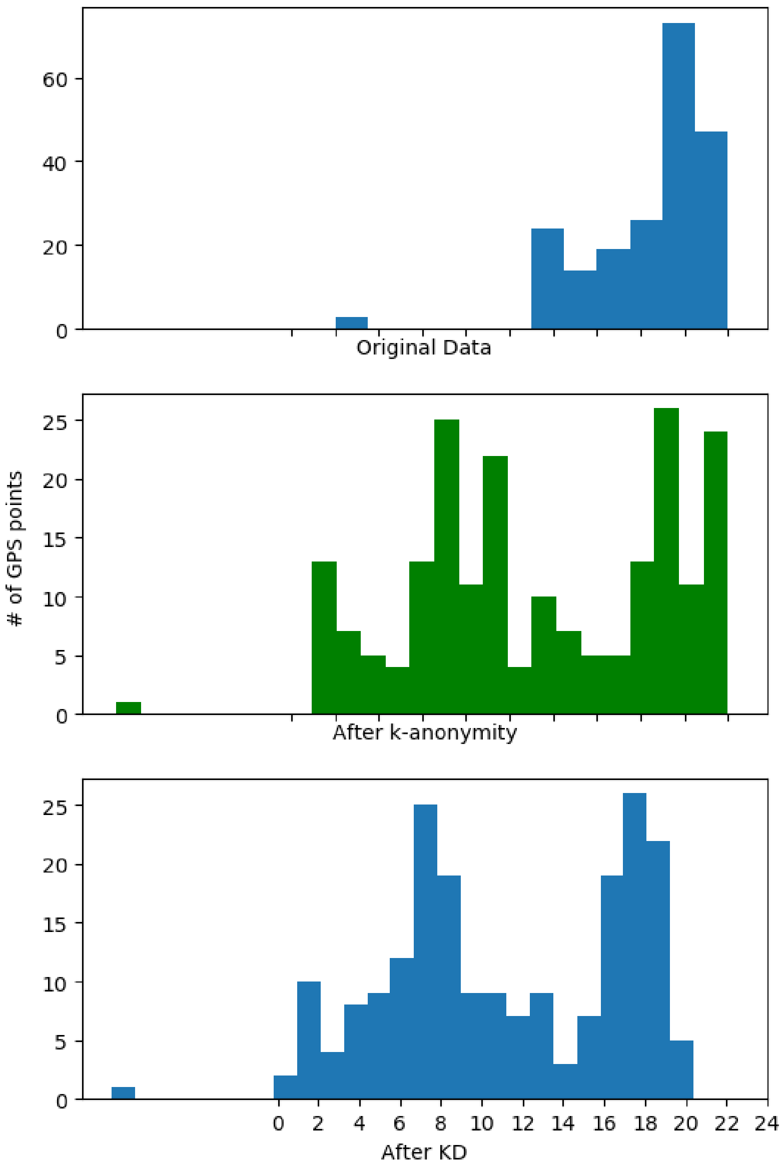
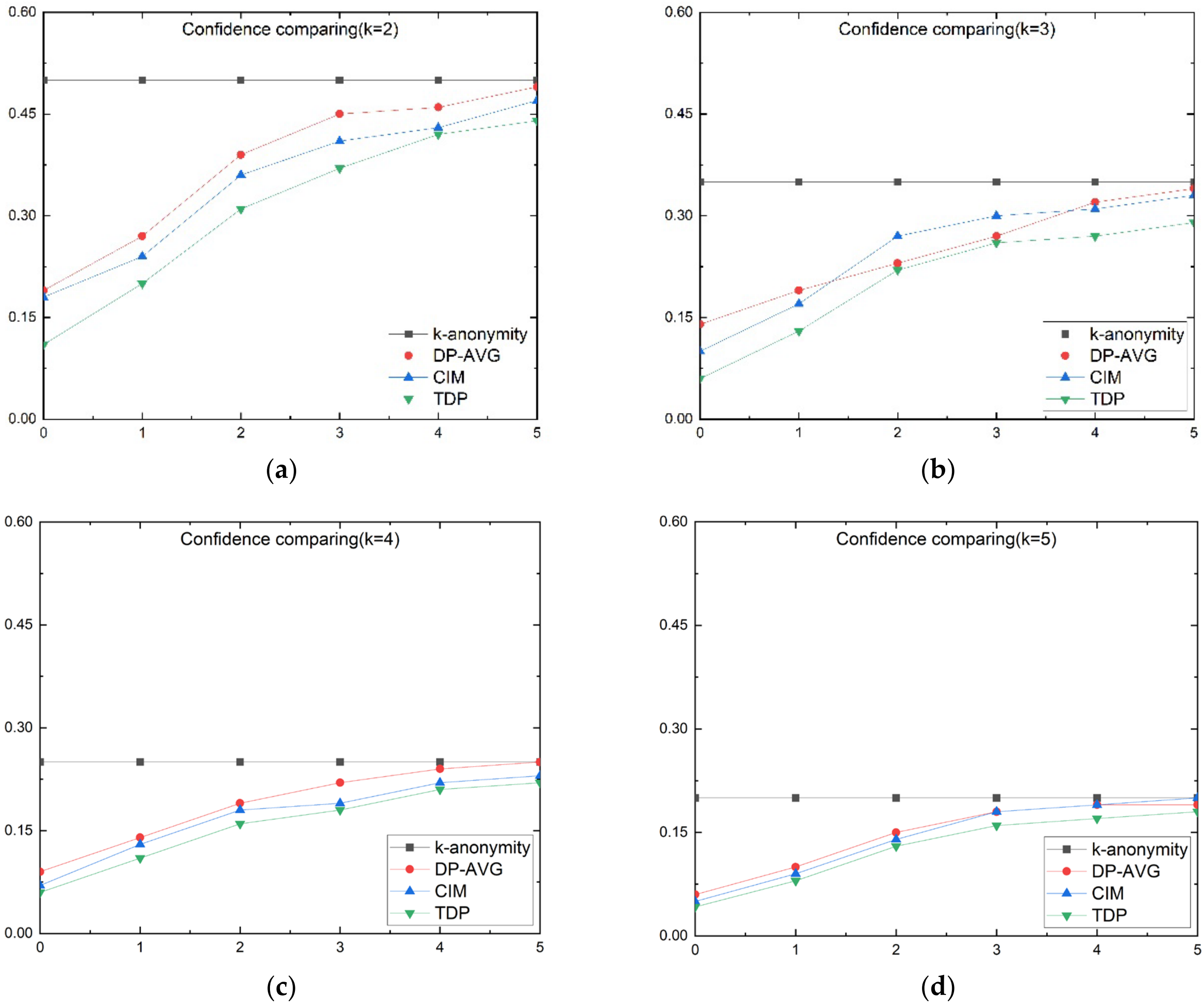


Figure 4. *TDP* for T-drive Taxi Trajectories dataset.

By observing Figure 5a–d, we can see that the data credibility of *TDP* was close to the *k*-anonymous data credibility as ε increased, but the reliability of the data was always higher than that obtained by the *TDP* algorithm. Fox example, when *k* = 2, the confidences of DP-AVG, CIM, and *TDP* were 0.19, 0.18, and 0.11; *TDP* improved by 38.9% compared with the current optimal method CIM. When *k* = 5, we can also observe this trend; in this case, *TDP*'s confidence was 0.042, compared with CIM's 0.05, with *TDP* improving by 16%. It can be seen that the *TDP* algorithm had a good effect on the security of the data and could effectively protect the security of the data.

From the above experimental results, it can be seen that the reliability of the data was higher than that obtained using the *k*-anonymity algorithm, which further indicates that the *k*-anonymity algorithm has a good protection effect on the data, and it has also high data security.



(**a**) (**b**) (**c**) (**d**)

Figure 5. Confidence comparison diagram. (**a**) *k* = 2. (**b**) *k* = 3. (**c**) *k* = 4. (**d**) *k* = 5.

(3) Data Validity Results and Analysis

For effectiveness analysis, since the *TDP* algorithm itself is a random algorithm, we performed multiple experiments and calculated the variance of each experiment to avoid accidents, thus negatively affecting the accuracy of the experimental results.

Figure 6 represents the comparison of the mean squared error of time–frequency after using the *k*-anonymity, DP-AVG, CIM, and *TDP* algorithms for Geolife Trajectories 1.3 data, while Figure 7 represents the time–frequency after using the *k*-anonymity, DP-AVG, CIM, and *TDP* algorithm for T-drive Taxi Trajectories data, which can be used for a comparison of the magnitude of the mean square error. In Geolife, we can observe that the MSEs of *k*-anonymity, DP-AVG, CIM, and *TDP* were 112, 160, 150, and 140, respectively. Among them, *TDP* had the smallest MSE except for *k*-anonymity. The same trend can also be observed in the T-drive Taxi dataset. In the T-drive Taxi dataset, the MSEs of *k*-anonymity, DP-AVG, CIM, and *TDP* were 130, 170, 160, and 150 respectively, as shown in Figure 7. In both datasets, it can be seen that, after repeated use of the *TDP* algorithm, the variance of time–frequency was slightly larger than that in the *k*-anonymity algorithm, and the validity of the data was affected to some extent. To improve its effectiveness, we optimized the algorithm using a truncated Laplacian distribution.

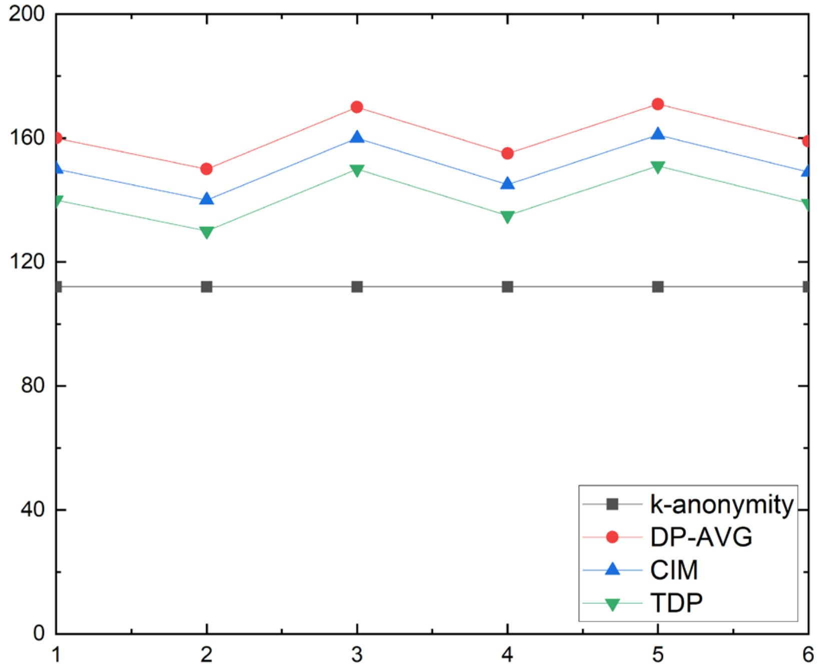


Figure 6. MSE comparison diagram for Geolife Trajectories dataset.

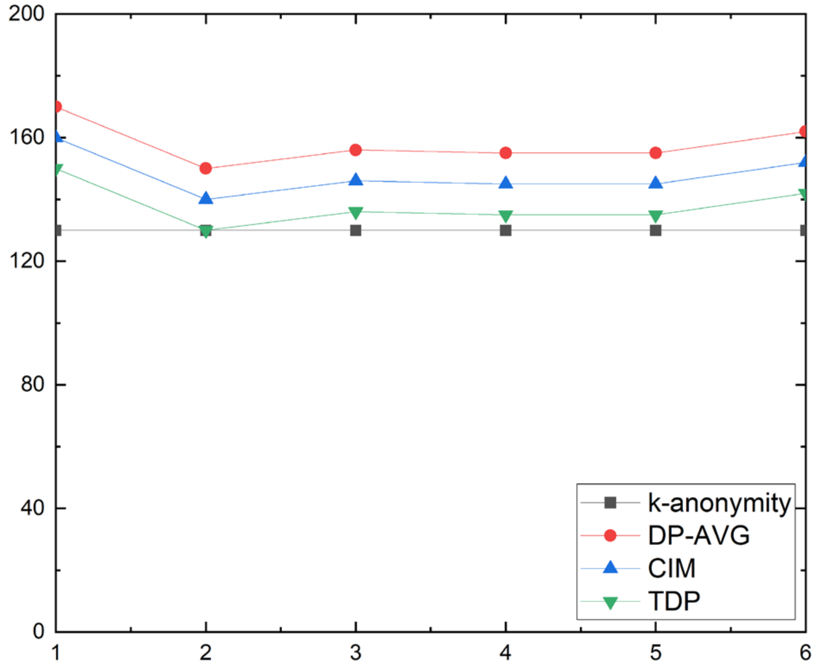


Figure 7. MSE comparison diagram for T-drive Taxi Trajectories dataset.

(4) Optimization Effect of Truncated Laplace

By comparing Figures 8 and 9 with Figures 3 and 4, it can be observed that special data were processed, which definitely had a positive impact on our data mining. Then, we compared the mean square error of the two datasets, the results of which are shown in Figures 10 and 11.

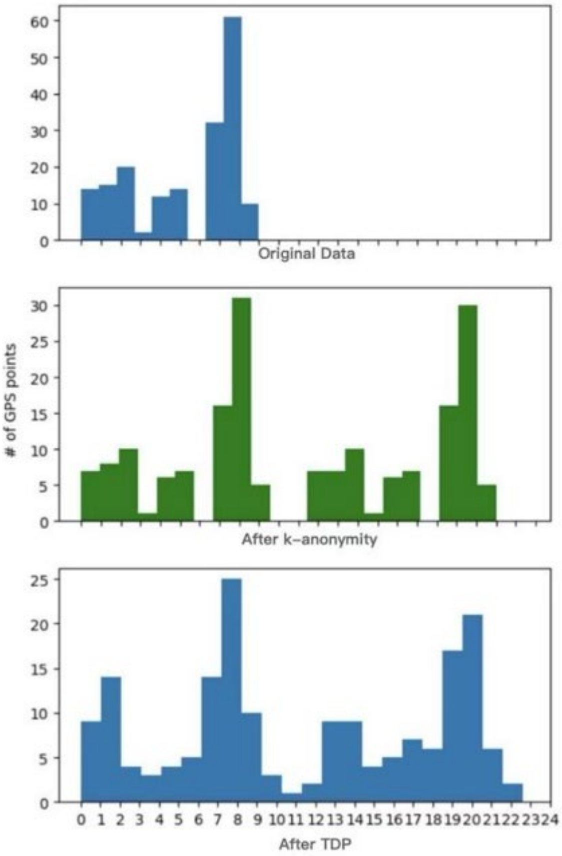


Figure 8. Diagram of time distribution after encrypting data with various algorithms for the Geolife Trajectories dataset.

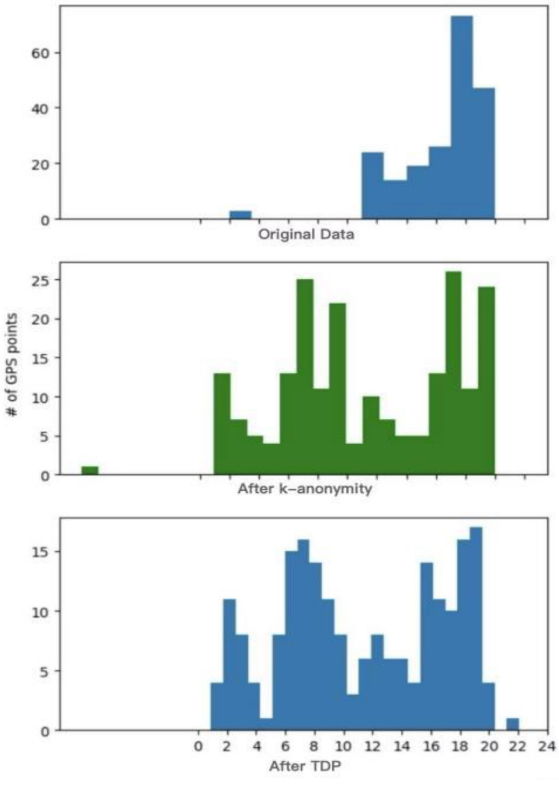


Figure 9. Diagram of time distribution after encrypting data with various algorithms for T-drive Taxi Trajectories dataset.

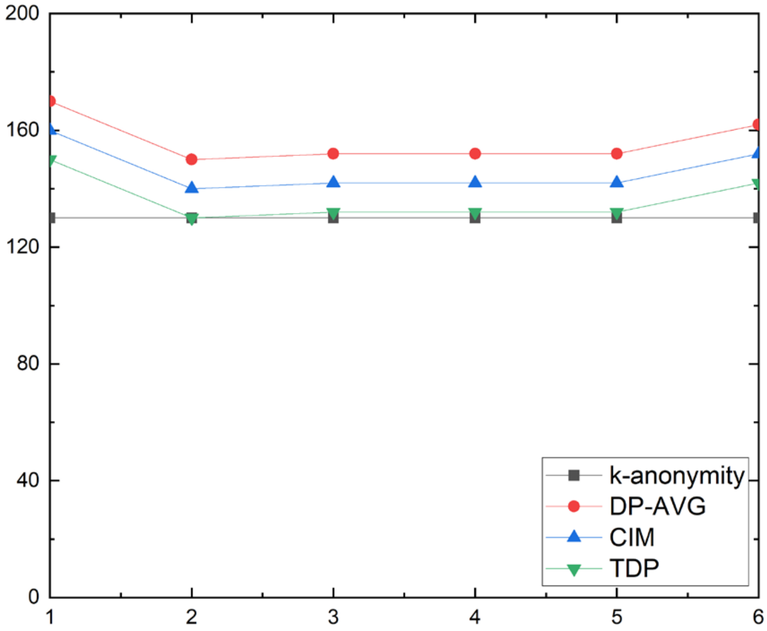


Figure 10. MSE comparison diagram for Geolife Trajectories dataset (using *k*-anonymity and *TDP*).

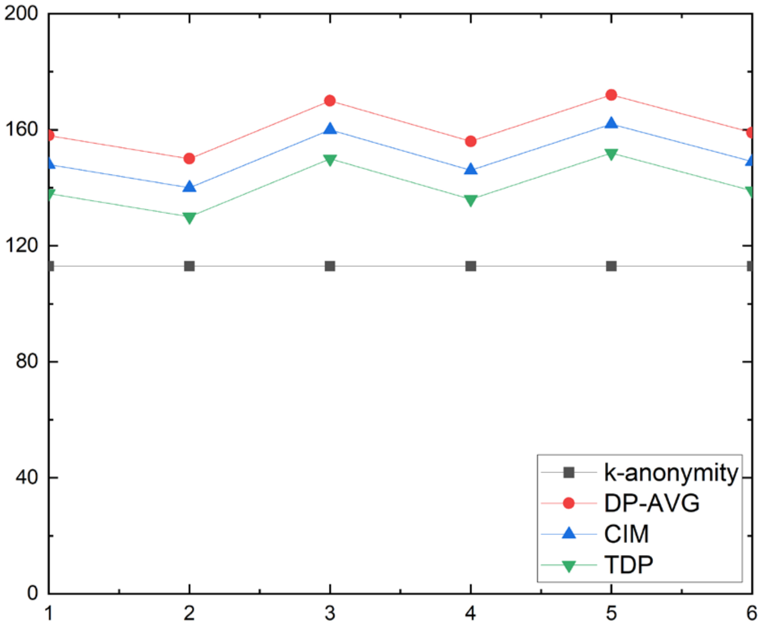


Figure 11. MSE comparison diagram for T-drive Taxi Trajectories dataset (using *k*-anonymity and *TDP*).

It can be seen from Figures 10 and 11 that, after several experiments, for the data in both the Geolife Trajectories 1.3 and the T-drive Taxi Trajectories datasets, the mean square error after using the truncated Laplacian mechanism was smaller. Thus, the effectiveness was further improved. In Geolife, we can observe that the MSEs of *k*-anonymity, DP-AVG, CIM, and *TDP* were 130, 170, 160, and 150, respectively. Among them, *TDP* had the smallest MSE except for *k*-anonymity. The same trend can also be observed in the T-drive Taxi dataset. In the T-drive Taxi dataset, the MSEs of *k*-anonymity, DP-AVG, CIM, and *TDP* were 113, 158, 148, and 138, respectively, as shown in Figure 11.

6.2. Experimental Results Summary

In this experiment, confidence was used for a security comparison, and the mean square error was used for effectiveness analysis. Compared with the state-of-the-art algorithms, our solution *TDP* had a higher value of confidence and a larger mean square error. Specifically, *TDP* improved confidence by 38.9% compared with the current optimal method CIM when $k = 2$. Even when $k = 5$, *TDP* improved by 16.0% compared with the current optimal method CIM. Furthermore, in terms of utility evaluation, the MSE of *TDP* as 140 in the Geolife dataset, improving by 6.7% compared with CIM. In the other dataset T-drive Taxi, the MSE of *TDP* was 138, improving by 6.8% compared with current optimal CIM method. Evidently, it had a better level of security; however, it also lost a certain degree of effectiveness and suffered from defects of special data. The use of the truncated Laplacian mechanism can effectively solve the problem of special data and achieve differential privacy, compared with the ordinary Laplacian mechanism, and the truncated Laplacian mechanism that achieves differential privacy has a higher effectiveness.

7. Conclusions

In this paper, we proposed a new *TDP* model to protect the time pattern data in spatiotemporal trajectory data. This algorithm combines the idea of k-anonymity and the protection of time data frequency through differential privacy, making the processed time data ambiguous. Additionally, it generates uncertainty, meaning that, when the attacker analyzes the data, they cannot obtain accurate information from the user but can only access the information obtained from data mining with a certain degree of credibility. Furthermore, we used the truncated Laplacian mechanism to process the out-of-bounds data to improve the validity of the data. Since the current spatiotemporal data protection strategy only provides thorough location data protection with the use of differential privacy, the protection of time data should be explored further.

In the future, we will continue to study two aspects:

1. The widespread use of k-anonymity should be further considered, since, when the time data are widely distributed, the parameter k can only take the value of 1, which imposes limitations on our privacy protection. Therefore, we intend to optimize the mechanism to further improve it when the data are extensively applied.
2. When adding noise to the time–frequency histogram, although we adopted a truncation method in this study to improve data availability, it is still worth considering how to further improve this. In the future, we will consider combining data distribution characteristics to find more reasonable partition noising schemes.

Author Contributions: Validation, W.F.; Formal analysis, H.W.; Investigation, Z.W.; Data curation, T.W.; Writing—original draft, L.Y.; Project administration, X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Cooperation Project of Chongqing Municipal Education Commission (HZ2021008), the China Postdoctoral Science Foundation (2021M693929), the Special Fund of Chongqing Postdoctoral Research Projects (2021XM3009), the Open Fund of State Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University (20S02), and the PhD Starts Fund Project of Chongqing University of Posts and Telecommunications (A2019-302).

Data Availability Statement: Data is unavailable due to privacy or ethical restrictions.

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

References

1. Wang, H.; Gao, Q.; Li, H.; Wang, H.; Yan, L.; Liu, G. A Structural Evolution-Based Anomaly Detection Method for Generalized Evolving Social Networks. *Comput. J.* **2022**, *65*, 1189–1199. [CrossRef]
2. Zhao, J.; Zhang, Y.; Li, X.; Ma, J. Trajectory privacy preserving method based on trajectory frequency suppression. *J. Comput. Sci.* **2014**, *37*, 2096–2106.

3. Wang, H.; Wang, H. Correlated Tuple Data Release via Differential Privacy. *Inf. Sci.* **2021**, *560*, 347–369. [CrossRef]
4. Yue, S. Research on Privacy Preserving of Web Application System Data Publishing Based on *k*-Anonymity. Master's Thesis, Beijing University of Posts and Telecommunications, Beijing, China, 2018.
5. Zheng, Z.; Lu, H.; Duan, Y. Privacy preserving generalization algorithm for data sets with the same sensitive value. *J. Zhengzhou Univ.* **2018**, *50*, 35–40.
6. Chen, X. Research and Implementation of Data Anonymity Privacy Preserving Method. Master's Thesis, Jiangsu University of Science and Technology, Zhenjiang, China, 2018.
7. Yang, Z.; Ning, B.; Li, Y. Personal privacy preserving method of spatiotemporal data based on κ-generalization technology. *J. East China Norm. Univ.* **2017**, *5*, 174–185.
8. Yu, Q.; Wang, Y.; Ye, Z.; Zhang, S.; Chen, C. Privacy preserving algorithm for trajectory data release based on optimized local suppression. *Comput. Eng.* **2019**, *46*, 112–118.
9. Wang, Z. Research on Path Privacy Preserving Method Based on Location Service. Ph.D. Thesis, Hainan University, Haikou, China, 2015.
10. Wu, X. Research on Privacy Preserving Methods of Fake Trajectory. Master's Thesis, China University of Science and Technology, Taipei City, Taiwan, 2016.
11. Huo, Z.; Meng, X. A path data publishing method satisfying differential privacy. *J. Comput. Sci.* **2018**, *41*, 400–412.
12. Xiong, P.; Zhu, T.; Wang, T. Differential privacy and its application. *J. Comput. Sci.* **2014**, *37*, 101–122.
13. Li, W.; Zhang, X.; Cao, G.; Li, S.; Zhang, Q. Data hierarchical fusion release mechanism based on differential privacy. *Minicomput. Syst.* **2019**, *40*, 2252–2256.
14. Dwork, C. Differential Privacy: A Survey of Results. In *Theory and Applications of Models of Computation*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2006; Volume 26, pp. 1–12.
15. Chen, R.; Fung, B.C.M.; Desai, B.C. Differentially private trajectory data publication. *arXiv* **2020**, arXiv:1112.2020.
16. Xu, Q.; Chen, Z. Trajectory Data Protection based on Differential Privacy k-means. In Proceedings of the 39th Chinese Control Conference, Shenyang, China, 27–29 July 20202; Volume 7, pp. 7649–7654.
17. Abadi, M. PDP-SAG: Personalized Privacy Protection in Moving Objects Databases by Combining Differential Privacy andSensitive Attribute Generalization. *IEEE Access* **2019**, *7*, 85887–85902.
18. Jiang, K.; Shao, D.; Bressan, S.; Kister, T.; Tan, K.L. Publishing trajectories with differential privacy guarantees. In Proceedings of the 25th International Conference on Scientific and Statistical Database Management, Baltimore, MD, USA, 29–31 July 2013; pp. 1–12.
19. Zhang, L.; Liu, Y.; Wang, R. Data publishing technology based on differential privacy in location big data service. *J. Commun.* **2016**, *37*, 46–54.
20. Wu, Y.; Chen, H.; Zhao, S.; Liang, W.; Wu, Y.; Li, C.; Zhang, X. A differential privacy trajectory preserving mechanism based on spatiotemporal correlation. *J. Comput. Sci.* **2018**, *41*, 309–322.
21. Bi, X.; Liang, Y.; Shi, H.; Tian, H. A secondary privacy preserving method for anonymous location based on privacy preference. *J. Shandong Univ.* **2017**, *52*, 75–84.
22. Ma, Y.; Zhang, L. LBS group nearest neighbor query based on differential privacy. *Comput. Sci.* **2017**, *44*, 336–341.
23. Hua, J.; Gao, Y.; Zhong, S.B. Differentially private publication of general time-serial trajectory data. In Proceedings of the 2015 IEEE Conference on Computer Communications (INFOCOM), Hong Kong, China, 26 April–1 May 2015; pp. 549–557.
24. Hu, D.; Zhan, H. Predictable privacy preserving method of differential disturbance user trajectory. *Minicomput. Syst.* **2019**, *40*, 1286–1290.
25. Geng, Q.; Ding, W.; Guo, R.; Kumar, S. Privacy and Utility Tradeoff in Approximate Differential Privacy. *arXiv* **2018**, arXiv:1810.00877.
26. Hay, M.; Li, C.; Miklau, G.; Jensen, D. Accurate estimation of the degree distribution of private networks. In Proceedings of the 2009 Ninth IEEE International Conference on Data Mining, Miami Beach, FL, USA, 6–9 December 2009; pp. 169–178.
27. Day, W.-Y.; Li, N.; Lyu, M. Publishing graph degree distribution with node differential privacy. In Proceedings of the 2016 International Conference on Management of Data, San Francisco, CA, USA, 26 June–1 July 2016; pp. 123–138.
28. Kasiviswanathan, S.P.; Nissim, K.; Raskhodnikova, S.; Smith, A. Analyzing graphs with node differential privacy. In *Theory of Cryptography Conference*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 457–476.
29. Nissim, K.; Raskhodnikova, S.; Smith, A. Smooth sensitivity and sampling in private data analysis. In Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing, San Diego, CA, USA, 11–13 June 2007; pp. 75–84.
30. Sala, A.; Zhao, X.; Wilson, C.; Zheng, H.; Zhao, B.Y. Sharing graphs using differentially private graph models. In Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, Berlin, Germany, 2–4 November 2011; pp. 81–98.
31. Jiang, L.J.; Guo, L. Mining spatial association rules based on taxi trajectory data. *Geomat. Spat. Inf. Technol.* **2019**, *42*, 56–59.
32. Jian, Y.; Wang, D.; Gao, X.; Wang, R.; Lin, S. Privacy preserving model of anonymous group LBS trajectory based on differential privacy. *Minicomput. Syst.* **2019**, *40*, 341–347.
33. Xia, C.; Hua, J.; Tong, W.; Zhong, S. Distributed K-Means clustering guaranteeing local differential privacy. *Comput. Secur.* **2020**, *90*, 101699. [CrossRef]

34. Hu, Z.; Yang, J. Differential privacy protection method based on published trajectory cross-correlation constraint. *PLoS ONE* **2020**, *15*, e0237158. [CrossRef] [PubMed]
35. Chen, Z.; Wang, Y.; Zhang, S.; Zhong, H.; Chen, L. Differentially private user-based collaborative filtering recommendation based on k-means clustering. *Expert Syst. Appl.* **2021**, *168*, 114366. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.