Differential Privacy for Time Series: A Survey

Yulian Mao[1,2], Qingqing Ye[2 *], Qi Wang[1,3], Haibo Hu[2]
[1]Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China
[2]Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China
[3]Shenzhen Key Laboratory of Safety and Security for Next Generation of Industrial Internet, Shenzhen, China
yulian.mao@connect.polyu.hk, qqing.ye@polyu.edu.hk, wangqi@sustech.edu.cn, haibo.hu@polyu.edu.hk

Abstract

Time series are extensively used in finance, healthcare, IoT, and smart cities. However, in many applications, time series often contain personal information, so releasing them publicly can pose privacy risks. Differential privacy has recently emerged as the state-of-the-art approach for safeguarding data privacy. Unfortunately, adapting differential privacy to time series presents unique challenges compared to other data types due to their large volume, temporal correlations, and dynamic nature. To address users' demands for time series analysis while simultaneously protecting privacy, a significant body of research works have been proposed. The aim of this survey is to summarize these works and provide a holistic view of the DP mechanisms under differential privacy. Furthermore, we will discuss the challenges associated with time series release, especially in one of its most prevalent applications — location based services. Finally, we will explore open challenges and shed light on directions for future research.

1 Introduction

Time series are being generated on a large scale across a wide range of application domains, such as IoT, finance, healthcare monitoring, operational event logs, and smart home sensors. For example, smart home devices such as thermostats and humidity sensors generate continuous time series on environmental conditions, tracking temperature fluctuations and moisture levels to optimize home climate control based on user activities. Additionally, trajectories represent a unique type of time series that contain both spatial and temporal information, such as GPS data tracking the movement of vehicles or individuals. To facilitate the analysis of time series and support various downstream tasks, numerous methods have been proposed, ranging from traditional statistical techniques, such as ARIMA [1] and exponential smoothing [2], to advanced machine learning models, such as long short-term memory (LSTM) networks [3]. However, a key issue in these time series applications is privacy. Since many data sources such as smart home sensors and location trajectories contain individuals' private information, the direct release or analysis of such time series can lead to significant privacy violations. Consequently, developing privacy-preserving mechanisms for time series analysis is essential.

Differential privacy (DP) [4] is a paradigm of privacy-preserving mechanisms that provides a theoretical privacy guarantee and has been further extended to the local setting to accommodate more general scenarios [5, 6].

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

[*] Dr. Qingqing Ye is the corresponding author.

Numerous DP mechanisms have been developed to support various queries [7, 8, 9], including estimating statistics such as frequency [7, 10] and mean [11], ensuring that the privacy of individuals in the dataset is protected even when aggregate information is released. Recently, research has shifted towards more complex applications, such as graph data mining [12] and machine learning problems [13]. Techniques such as differentially private stochastic gradient descent (DP-SGD) [14, 15] have been developed to train machine learning models with differential privacy guarantees, enabling the use of private datasets for tasks like classification and prediction without compromising individuals' privacy. It is worth noting that differential privacy is also being explored in the context of time series [16, 17, 19]. This involves developing new mechanisms capable of handling the features of time series, ensuring that privacy is maintained.

Nevertheless, time series data present more challenges compared to other types of data due to their large volume, temporal correlation, and dynamic nature. The large volume, in particular, poses significant issues for privacy models, as protecting every element of a time series would degrade utility. To address this challenge, three privacy levels have been proposed [19]: event-level privacy, which protects a single element in the time series; w-event level privacy, which provides a privacy guarantee for w consecutive elements; and user-level privacy, which protects all elements associated with an individual. To enhance utility, sampling and filtering-based mechanisms [18], as well as privacy budget allocation strategies [19], have been suggested. Additionally, the correlations between elements can lead to privacy breaches [20], necessitating countermeasures to confine these correlations [21, 22]. Given numerous research efforts in this field, a taxonomy is necessary to summarize the existing works and identify areas for future research.

However, to the best of our knowledge, there is no up-to-date and comprehensive survey specifically for time series under differential privacy. Dwork and Roth [23] coauthored a comprehensive survey on differential privacy, which seems outdated now in terms of state-of-the-art techniques. More recently, there are two surveys from Zhao et al. [24, 25] focusing on the concepts and applications of differential privacy, but they do not extensively cover time series. Miranda-Pascual et al. [26] conducted a survey on trajectory data publication, which mainly talks about downstream tasks with little emphasis on privacy preservation. The most relevant survey on time series under differential privacy is by Katsomallos et al. [27]. However, since it was published in 2019, the paper does not reflect current technical trends, such as LDP, which now accounts for a crucial portion of privacy-preserving time series research.

In this survey, we aim to provide a comprehensive review of research works on time series under differential privacy. The main contents and paper organization are summarized as follows.

- **Section 2: Fundamental concepts of time series and differential privacy.** We first introduce the basic concepts of time series and differential privacy, including the definitions of differential privacy and the composition theorems. Additionally, we elucidate the three privacy levels specifically defined within the context of time series.

- **Section 3: Count queries and corresponding advanced queries.** We begin with an introduction to count queries, and then present two core techniques for their realization: the binary tree-based mechanism [17] and the matrix mechanism [28]. Following this, we discuss advanced queries, such as frequency and histogram estimations, which are based on count queries. Finally, we explore downstream applications derived from count queries.

- **Section 4: Sum/mean queries and downstream applications.** We list sum and mean queries together due to their inherent correlation. Following the introduction of sum and mean queries, we present the developed mechanisms for these queries. Subsequently, we review the literature on downstream applications.

- **Section 5: Time series release.** We classify the literature into two categories: methods based on value perturbation and methods based on synthesis. For value perturbation-based methods, we first review privacy budget allocation strategies and then present the optimization strategies to improve utility. We then introduce the synthesis-based methods, including those based on statistics and generative models. Additionally, we discuss the privacy models of the time series mechanisms and review a line of work that perturbs the temporal order rather than the values to accommodate value-critical scenarios.

- **Section 6: Location based services and trajectory release.** Given that location based services are common applications under differential privacy, we dedicate a section to discussing the relevant literature. To improve the utility, geo-indistinguishability [29] was proposed to constrain the perturbation domain. Moreover, due to the apparent temporal correlation, the relationships between locations need to be considered. Finally, we present mechanisms designed for trajectory release based on perturbation and synthesis.
- **Section 7: Open challenges.** We present a few future research directions for DP-based time series in terms of privacy model, potential correlation-based attacks, complex data type, and learning based problems.

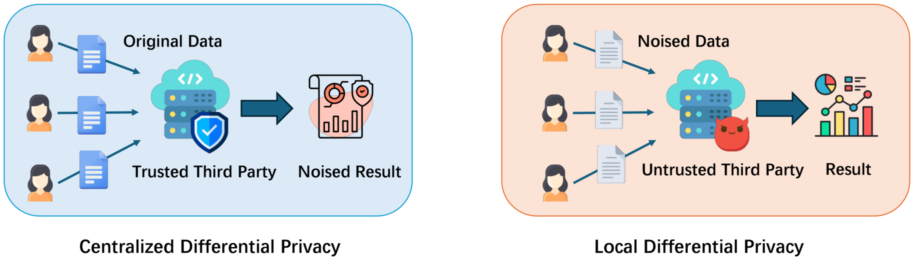
2 Preliminaries

In this section, we first introduce the basic concepts of differential privacy and differential privacy for time series. As for the latter, we mainly focus on different privacy levels of DP mechanisms used to analyze time series.

2.1 Differential Privacy

Differential privacy is a rigorous and practical formalization that provides a quantitative measure of privacy leakage for an individual when participating in a database [4]. In the nearly 20 years since its inception, differential privacy has become widely adopted as a privacy-preserving framework. Many companies, such as Microsoft [30], Google [31], and Apple [32], utilize differential privacy to collect users' data while providing privacy guarantees. Additionally, the US Census Bureau adopted differential privacy for the 2020 decennial census [33].

Based on utilization scenarios, differential privacy can be broadly categorized into centralized differential privacy (CDP) and local differential privacy (LDP) [6]. Centralized differential privacy requires a trusted third party to act as the data curator, collecting data from users and releasing the processed results to the public. The trusted third party is assumed to safeguard private information and not disclose it. However, in many situations, such a trusted third party may not exist. Consequently, local differential privacy has been proposed, allowing data to be sanitized locally before being uploaded. These two different scenarios are depicted in Fig 1, and their formal definitions are provided below.



Centralized Differential Privacy

Local Differential Privacy

Figure 1: An illustration for centralized differential privacy and local differential privacy. In the context of centralized differential privacy, a trusted third party collects original data from users and adds noise to the processed result. In contrast, under local differential privacy, users add noise to their data locally before uploading the noised data to the untrusted third party.

2.1.1 Centralized Differential Privacy (CDP)

Before delving into the formal definition of centralized differential privacy, it is important to first elucidate some underlying concepts. We start with the definition of neighboring datasets.

Definition 2.1: *[Neighboring Datasets [23, 24, 34]] Two datasets D and D' are neighboring if they only differ by only one record. In Unbounded CDP, D can be obtained from D' by adding or removing one record, whereas in Bounded DP, D can be obtained from D' by replacing one record.*

Definition 2.2: *[(ϵ, δ)-Centralized Differential Privacy ((ϵ, δ)-CDP) [23, 24, 34]] A randomized mechanism $\mathcal{M}$ satisfies (ϵ, δ)-centralized differential privacy if and only if for any two neighboring datasets D, D', and any possible output $R \subseteq \mathrm{Range}(\mathcal{M})$, there is*

$$\Pr[\mathcal{M}(D) = R] \leq e^{\epsilon} \cdot \Pr[\mathcal{M}(D') = R] + \delta.$$

When $\delta = 0$, $\mathcal{M}$ satisfies ϵ-centralized differential privacy.

2.1.2 Local Differential Privacy (LDP)

As we aforementioned, local differential privacy is adopted in a local mode. Compared with CDP, LDP requires more added noise to ensure privacy but does not need a trusted third party. Therefore, neighboring datasets in LDP can be any input from users.

Definition 2.3: *[(ϵ, δ)-local differential privacy ((ϵ, δ)-LDP) [24, 34]] A randomized mechanism $\mathcal{A}$ satisfies (ϵ, δ)-local differential privacy if and only if for any inputs v, v', and any possible output $r \subseteq \mathrm{Range}(\mathcal{A})$, there is*

$$\Pr[\mathcal{A}(v) = r] \leq e^{\epsilon} \cdot \Pr[\mathcal{A}(v') = r] + \delta.$$

When $\delta = 0$, $\mathcal{A}$ satisfies ϵ-local differential privacy which is also called pure-LDP [7].

2.1.3 Composition Theorems

Under differential privacy (both CDP and LDP), there are two useful composition theorems [34]: sequential composition and parallel composition.

Definition 2.4: *[Sequential Composition [34]] Given a dataset x, and two mechanisms M_1, M_2 satisfy (ϵ_1, δ_1)-DP and (ϵ_2, δ_2)-DP, respectively, the mechanism $M = (M_1(x), M_2(x))$ satisfies $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$-DP.*

Definition 2.5: *[Parallel Composition [34]] Given a mechanism M satisfy (ϵ, δ)-DP, and the k disjoint separations of the dataset x (i.e., $x_1 \cup x_2 \cup \cdots \cup x_k = x$), the release $M(x_1), M(x_2), \cdots, M(x_k)$ satisfies (ϵ, δ)-DP.*

2.1.4 Pufferfish Privacy

When providing privacy guarantees for correlated time series, differential privacy faces the challenge of excessive noise addition. Specifically, group differential privacy [23] necessitates adding $O(T)$ noise for a correlated time series with length T, leading to significant utility degradation. To address correlated data, pufferfish privacy, a generalized version of differential privacy, was proposed [35]. In addition to the privacy budget ϵ, pufferfish privacy requires three additional parameters [36]: a set of secrets $\mathcal{S}$ representing users' private data, a set of secret pairs $\mathcal{Q} \subseteq \mathcal{S} \times \mathcal{S}$ that must remain indistinguishable, and a class of data distribution Θ indicating the correlation.

Definition 2.6: *[ϵ-Pufferfish Privacy [36]] Give the parameters $\mathcal{S}$, $\mathcal{Q}$, and Θ, a randomized mechanism M satisfies ϵ-pufferfish privacy if $\forall \theta \in \Theta$ with $X \sim \theta$, $\forall (s_i, s_j) \in \mathcal{Q}$, $\forall w \in Range(M)$, there is*

$$\Pr(M(X) = w | s_i, \theta) \leq e^{\epsilon} \cdot \Pr(M(X) = w | s_j, \theta),$$

when $\Pr[\mathrm{s_i}|\theta] \neq 0$ *and* $\Pr[\mathrm{s_j}|\theta] \neq 0$.

2.1.5 Differential Privacy Mechanisms

Given that noise addition from a distribution is a fundamental implementation of differential privacy, this technique is extensively used for statistical estimation. Numerous applications rely on count queries, including frequency estimation, histogram estimation, and top-k item mining. Korolova et al. [37] introduced a mechanism for releasing search log histograms by adding Laplace noise. Xiao et al. [38] proposed a wavelet-based mechanism to release range count queries. In the context of LDP, Wang et al. [7] reviewed existing mechanisms for frequency estimation and introduced two optimized approaches. Li et al. [11] developed a mechanism for estimating numerical data distributions. Wang et al. [9] proposed a prefix extension mechanism to identify the top-k frequent items within a large domain. Beyond count queries, many applications focus on sum or mean queries. Wang et al. [39] devised a mechanism with optimized perturbation probability for numerical data under LDP. Xue et al. [40] introduced a mean estimation mechanism that supports personalized privacy budgets for each user. Zhou et al. [41] developed a mechanism to estimate the mean of sparse vectors. To incorporate data knowledge, Wei et al. [42] proposed an optimized mean estimation mechanism based on data distribution estimated in the initial phase under LDP. Beyond statistical estimations, data release for downstream tasks is another critical application. Ye et al. [43] proposed a mechanism to release edge information for clustering coefficient estimation. Ma et al. [44] developed a mechanism to construct decision trees under LDP, enhancing utility through the adoption of public data. Additionally, since the introduction of Differentially Private Stochastic Gradient Descent (DP-SGD) [14], numerous privacy-preserving learning-based mechanisms have been proposed under differential privacy.

Across the various scenarios, time series represent a unique research field due to their sequential nature and temporal dependencies. This adds complexity to ensuring differential privacy while maintaining data utility. In the following sections, we will introduce the concepts of time series under differential privacy.

2.2 Differential Privacy for Time Series

In this subsection, we will introduce the concept of time series and the specific definitions of differential privacy for time series, including various privacy levels. Additionally, we will provide a concise roadmap of this survey.

2.2.1 Time Series

In general, a time series is regarded an ordered sequence of values with finite length [45], while data streams are continuously generated sequences with infinite length [46]. For ease of presentation, both are referred to as "time series" in this paper, encompassing both finite and infinite settings.

Definition 2.7: *[Time Series [45, 46, 47]] A time series S is an ordered sequences of values, i.e., $S = \{S_{t_1}, S_{t_2}, S_{t_3}, \cdots\}$. For simplicity, the timestamp is usually omitted, and a time series is denoted as $S = \{S_1, S_2, S_3, \cdots\}$. If any element $S_i \in \mathbb{R}$, the time series is called a univariate infinite time series. Otherwise, the time series is a multivariate infinite time series if $S_i \in \mathbb{R}^d$, namely, each element is with d dimensions.*

Note that if a time series has a finite length, it is called a finite time series. Otherwise, it is referred to as an infinite time series.

2.2.2 Privacy Levels

In the context of time series, three major privacy levels have been proposed based on the privacy guarantees. Event-level privacy only protects a single element within a time series, w-event level privacy provides a privacy guarantee for a sequence of w consecutive elements, and user-level privacy protects the entire time series. The corresponding definitions are provided below, with illustrations depicted in Fig. 2.

Definition 2.8: *[Event-Level Adjacent Time Series [19]] For two time series S and S', they are event-level adjacent if*

1) *There exists a timestamp i, $S_i \neq S'_i$;*

2) *For any other timestamp j, $S_j = S'_j$.*

Definition 2.9: *[Event-Level Privacy [19]] A randomized mechanism $\mathcal{M}$ satisfies (ϵ, δ)-event-level differential privacy if and only if for any two event-level adjacent time series S, S', and any possible output $R \subseteq \mathrm{Range}(\mathcal{M})$, there is*

$$\Pr[\mathcal{M}(S) = R] \leq e^{\epsilon} \cdot \Pr[\mathcal{M}(S') = R] + \delta.$$

When $\delta = 0$, $\mathcal{M}$ satisfies ϵ-event-level differential privacy.

Definition 2.10: *[w-Event Level Adjacent Time Series [19]] For two time series S and S', they are w-event level adjacent if*

1) *There exists w consecutive timestamp $\{k, k+1, \cdots, k+w-1\}$ and for any $i \in \{k, k+1, \cdots, k+w-1\}$, $S_i \neq S'_i$;*

2) *For any other timestamp j, $S_j = S'_j$.*

Definition 2.11: *[w-Event Level Privacy] A randomized mechanism $\mathcal{M}$ satisfies (ϵ, δ)-w-event level differential privacy if and only if for any two w-event level adjacent time series S, S', and any possible output $R \subseteq \mathrm{Range}(\mathcal{M})$, there is*

$$\Pr[\mathcal{M}(S) = R] \leq e^{\epsilon} \cdot \Pr[\mathcal{M}(S') = R] + \delta.$$

When $\delta = 0$, $\mathcal{M}$ satisfies ϵ-w-event level differential privacy.

Definition 2.12: *[User-Level Adjacent Time Series] For two time series S and S', they are user-level adjacent if for all timestamps $t_{u_i} = \{t_1, t_2, \cdots, t_k\}$ from any user u_i, there is $S_j \neq S'_j, \forall j \in t_{u_i}$. Note that t_{u_i} can be infinite for infinite time series.*

Definition 2.13: *[User-Level Privacy] A randomized mechanism $\mathcal{M}$ satisfies (ϵ, δ)-user-level differential privacy if and only if for any two user-level adjacent time series S, S', and any possible output $R \subseteq \mathrm{Range}(\mathcal{M})$, there is*

$$\Pr[\mathcal{M}(S) = R] \leq e^{\epsilon} \cdot \Pr[\mathcal{M}(S') = R] + \delta.$$

When $\delta = 0$, $\mathcal{M}$ satisfies ϵ-user-level differential privacy.

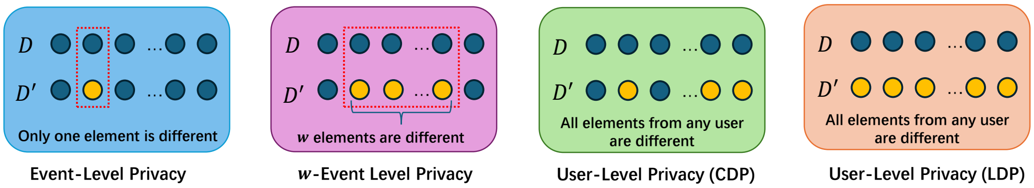


Figure 2: An illustration for privacy levels. In terms of event-level privacy, there is only one element different in the neighboring datasets. While for the w-event level privacy, there are w consecutive elements that differ in the neighboring datasets. For user-level privacy, all the elements from any user can be different in the neighboring datasets. Note that under LDP, the two user-level adjacent time series are from different users, which means the elements could be entirely different.

Obviously, event-level privacy provides the lowest level of privacy but requires the least amount of noise. Conversely, user-level privacy guarantees the strongest privacy but necessitates the largest amount of noise, which can significantly degrade utility.

2.3 Roadmap of This Survey

Since count queries and sum/mean queries are fundamental statistical operations, many advanced queries and downstream applications are derived from them. This survey begins with a review on count queries, followed by a discussion of sum/mean queries. Each subsection on these queries starts with an introduction to the concepts, followed by a review of their downstream applications. Subsequently, we introduce data release mechanisms designed to publicize data for downstream tasks. Given the popularity of location based services in time series applications, we dedicate a separate section to trajectories, reviewing the literature on location perturbation, temporal correlation issues, and trajectory release. To suggest future directions, we propose open challenges related to privacy models, temporal correlation-based attacks, complex data types, and learning-based problems. The roadmap for this survey is illustrated in Fig. 3.

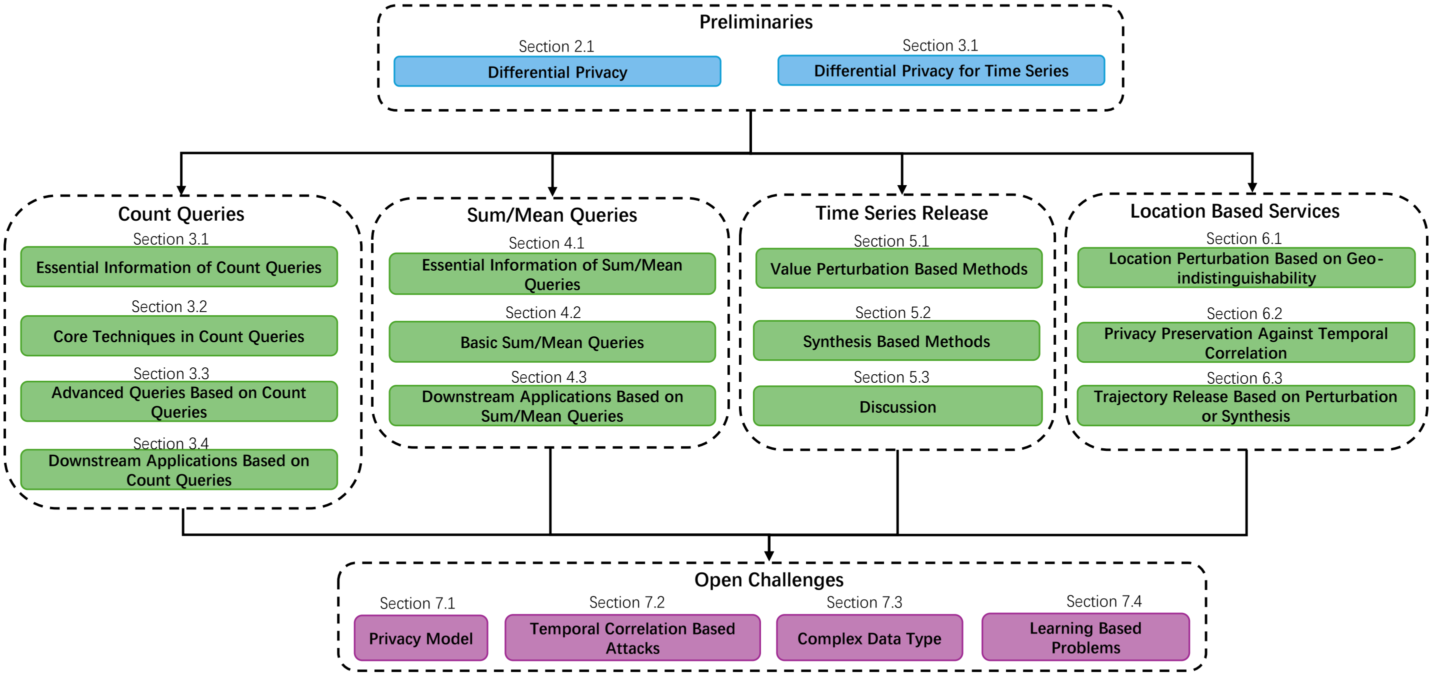

Preliminaries
Section 2.1
Differential Privacy
Section 3.1
Differential Privacy for Time Series
Count Queries
Section 3.1
Essential Information of Count Queries
Section 3.2
Core Techniques in Count Queries
Section 3.3
Advanced Queries Based on Count Queries
Section 3.4
Downstream Applications Based on Count Queries
Sum/Mean Queries
Section 4.1
Essential Information of Sum/Mean Queries
Section 4.2
Basic Sum/Mean Queries
Section 4.3
Downstream Applications Based on Sum/Mean Queries
Time Series Release
Section 5.1
Value Perturbation Based Methods
Section 5.2
Synthesis Based Methods
Section 5.3
Discussion
Location Based Services
Section 6.1
Location Perturbation Based on Geo-indistinguishability
Section 6.2
Privacy Preservation Against Temporal Correlation
Section 6.3
Trajectory Release Based on Perturbation or Synthesis
Open Challenges
Section 7.1
Privacy Model
Section 7.2
Temporal Correlation Based Attacks
Section 7.3
Complex Data Type
Section 7.4
Learning Based Problems

Figure 3: Roadmap of this survey.

3 Differential Privacy for Count Queries

In this section, we introduce the concept of count queries within the framework of differential privacy, a topic that has garnered substantial research attention.

3.1 Essential Information of Count Queries

For a time series S comprising categorical values with length t (t can be ∞), a count query can be formally expressed as follow:

$$\mathrm{F}_{\mathrm{cnt}}(v, S) = \sum_{i=1}^{t} \mathbf{1}_v(x_i),$$

where $\mathbf{1}_v(x_i)$ denotes the indicator function, defined below:

$$\mathbf{1}_v(x_i) := \begin{cases} 1 & \text{if } x_i = v, \\ 0 & \text{if } x_i \neq v. \end{cases}$$

Count queries are fundamental in the context of differential privacy, as they form the basis for other tasks, such as frequency and histogram estimations. Compared with other scenarios, applying count queries to time series presents unique challenges. Time series arrives as a continuous stream, necessitating the injection of a larger amount of noise to provide the desired privacy guarantees due to continuous release of query results.

3.2 Core Techniques in Count Queries

For continuously releasing a finite time series, a naive approach is to release each element with the entire privacy budget ϵ [17]. However, such a method would introduce substantial additive noise, leading to low utility with an error bound of $O(\frac{\sqrt{T}}{\epsilon})$. Dwork et al. [16] proposed the first work to handle binary time series under event-level CDP with a logarithm error bound. However, this mechanism is limited to time series with finite length T. Subsequently, Chan et al. [17] improved the mechanism to support the release of infinite binary time series. Both of these works employ a tree-based method to enhance utility. The binary tree mechanism [17], illustrated in Fig. 4, ensures that each release influences at most one node at each level for finite time series. Consequently, each node only needs to add noise corresponding to the privacy budget $\frac{\epsilon}{(\log T+1)}$. To guarantee a logarithm error bound for infinite time series [17, 48], more binary trees will be construed. Since the update of one element only influences one tree, each tree will be allocated an entire privacy budget. More details can be referred to Fig. 4.

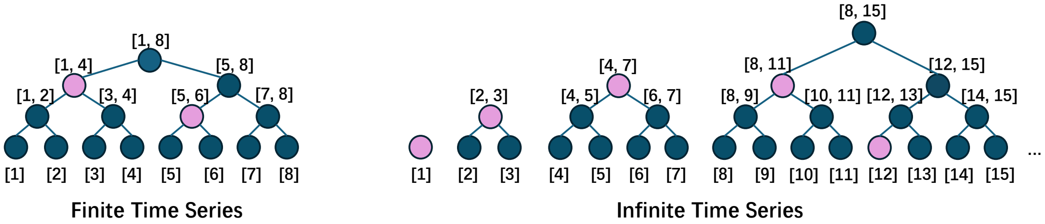


Figure 4: Here is an illustration of binary tree mechanism for time series count query under event-level CDP. For finite time series $[1, 8]$ (i.e., $T = 8$), the count query result for the range $[1, 6]$ is specified as $\hat{\mathrm{F}}_{\mathrm{cnt}}(v, [1, 6]) = \mathrm{F}_{\mathrm{cnt}}(v, [1, 4]) + \mathrm{F}_{\mathrm{cnt}}(v, [5, 6]) + 2Lap(\frac{\log T+1}{\epsilon})$, where $Lap(\cdot)$ is drawn from the Laplace distribution with zero mean. For an infinite time series, multiple binary trees will be constructed. Since a change in any single element only influences one tree, each tree will be allocated an entire privacy budget. The overall privacy budget consumption of the mechanism is ϵ, which can be calculated using parallel composition [34]. For example, $\hat{\mathrm{F}}_{\mathrm{cnt}}(v, [1, 12]) = \hat{\mathrm{F}}_{\mathrm{cnt}}(v, [1, 1]) + \hat{\mathrm{F}}_{\mathrm{cnt}}(v, [2, 3]) + \hat{\mathrm{F}}_{\mathrm{cnt}}(v, [4, 7]) + \hat{\mathrm{F}}_{\mathrm{cnt}}(v, [8, 11]) + \hat{\mathrm{F}}_{\mathrm{cnt}}(v, [12, 12])$.

In addition to the tree-based structure, another approach to handle time series under differential privacy is based on the matrix mechanism [28, 49]. Without privacy concerns, a count query $\mathcal{M}$ for binary time series x with length n can be specified as

$$\mathcal{M}(x) = Mx = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 1 & 1 & 0 & \cdots \\ 1 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} x = \begin{bmatrix} x_1 \\ x_1 + x_2 \\ \vdots \\ \sum_1^n x_i \end{bmatrix}.$$

Based on this formulation, the matrix mechanism is employed to release data streams under (ϵ, δ)-DP for further error reduction. Given a workload matrix M, the strategy matrix R and reconstruction matrix L are first constructed, denoted as $M = LR$. The strategy matrix R is utilized to pre-process the input x. After adding a Gaussian noise vector z to the processed term Rx, a post-processing step L is applied. In summary, for an input time series $x \in \mathbb{R}^n$, the matrix mechanism is denoted as

$$\mathcal{M}_{L,R}(x) = L(Rx + z),$$

where $z \sim N(0, \|R\|^2_{1\to 2} C^2_{\epsilon,\delta}\mathbf{I})$ to ensure (ϵ, δ)-DP, $\|R\|^2_{1\to 2}$ is the maximum of the 2-norm of the columns of the strategy matrix R. The additive mean squared error $\mathrm{err}_{\ell_2^2}$ of the matrix mechanism is

$$\mathrm{err}_{\ell_2^2}(\mathcal{M}_{L,R}, \mathcal{M}, n) = \max_{x\in\mathbb{R}^n} \mathbb{E}\left[\frac{1}{n}\|\mathcal{M}_{L,R}(x) - \mathcal{M}(x)\|_2^2\right] = \frac{1}{n} trace(L^T L)\|R\|^2_{1\to 2} C^2_{\epsilon,\delta}.$$

Hence, the matrix mechanism can be regarded as an optimization problem. For more comprehensive information, we recommend reading the work by Henzinger et al. [49].

3.3 Advanced Queries Based on Count Queries

Query	**Ref.**	**Type**	**Privacy Level**	**Method Primitive**	**Error Bound**
Binary Counting	[16]	finite	event-level CDP	binary tree mechanism	$O(\frac{1}{\epsilon} \cdot (\log^{1.5} t))$
	[17]	infinite	event-level CDP	binary tree mechanism	$O(\frac{1}{\epsilon} \cdot (\log^{1.5} t))$
	[48]	infinite	user-level CDP	binary tree mechanism	$O(\frac{\kappa(D_t)}{\epsilon} \cdot \log^{1.5} t \cdot \log^{1+\theta}(\kappa(D_t)))$
	[49]	infinite	event-level (ϵ, δ)-CDP	matrix mechanism	$O(\frac{4}{\epsilon^2}(\frac{4}{9} + \ln(\frac{1}{\delta}\sqrt{\frac{2}{\pi}}))(1 + \frac{\ln(4n/5))}{\pi})^2)$
Frequency Estimation	[50]	finite	event-level zCDP	multi-branch tree	$O(\tau \log T \sqrt{2(s+1)(t-1)\log(6T/\beta)})$
	[48]	infinite	user-level CDP	binary tree mechanism	$O(\frac{\kappa(D_t)}{\epsilon\theta} \cdot \log^{1+\theta}(\kappa(D_t))) \cdot \log(tR/\beta)$
Distinct Elements Counting	[51]	finite	user-level CDP	bipartite maximum matching	$O(\frac{\ell_*}{\epsilon} \log(\frac{\ell_{\max}}{\beta})$

Table 1: A brief summary of count-based queries, where t indicates the current timestamp, T is the length of the time series, D represents the dataset, $\kappa(D_t)$ denotes the maximum number of elements contributed by any user, τ is the privacy level of zero-concentrated differential privacy (zCDP) [52], θ is any small constant, ℓ is the sensitivity (ℓ_* represents the bounded sensitivity), and β is the confidence parameter. [1]

Based on basic binary count queries, numerous other types of count queries have been proposed, including frequency estimation, frequency moment estimation, and distinct element counting. A brief summary of the literature is presented in Table 1. Cardoso et al. [50] introduced differentially private histograms in the continual observation model with an unknown domain. To facilitate practical implementation, the authors propose a mechanism that continually returns a noisy histogram by aggregating counts at each round and adding noise to them. Dong et al. [48] proposed a mechanism to estimate frequency under user-level CDP. Since a user may contribute multiple elements to the time series, the mechanism first estimates each user's contribution and then applies a truncation process to retain only a limited number of elements per user, marking subsequent items as invalid. Furthermore, the proposed approach reduces the domain of elements to further enhance the utility of frequency estimates. The work by [51] proposes a method to estimate the number of distinct elements in a time series, and obtain the bounds on the true number of unique elements. The paper models the dataset as a bipartite graph and reduces the unique counting process to a max-flow problem, allowing the utilization of standard algorithms for bipartite maximum matching to solve unique counting problem. Furthermore, Kalemaj et al. [53] proposed a mechanism that achieves a logarithmic error bound for releasing distinct elements with insertions and deletions in a finite time series, under item-level differential privacy, which considers neighboring datasets

[1] The error bound of [49] in the table is the L_2 norm error.

differing by more than one element deletion. Epasto et al. [54] presented the first work to release differentially private ℓ_p frequency moments, denoted as $\sum_i f_i^p$. Notably, when $p = 1$, the frequency moment reduces to distinct counting. Practically, Zhang et al. [55] propose DP-SQLP, the first differentially private stream aggregation processing system, which has been implemented for Google Shopping and is planned for future application to Google Trends.

3.4 Downstream Applications Based on Count Queries

The application of differential privacy in count queries primarily involves releasing histograms and monitoring anomalies. Recent research [56] has demonstrated that employing subsampling and filtering techniques can reduce the sensitivity of real-time series, thereby enhancing the utility of differentially private mechanisms applied to such data. To improve utility, these mechanisms [18, 57, 58, 59] typically employ sampling to reduce the number of elements needing protection and filtering to mitigate the impact of added noise. Additionally, some mechanisms leverage pufferfish privacy [60, 61], which introduces less noise while achieving similar privacy guarantees. The details of these mechanisms are as follows. Fan et al. [18] introduced a mechanism for collecting count query results under the FAST framework, which employs filtering and adaptive sampling techniques to satisfy CDP. In a separate work, Fan et al. [57] applied the FAST framework for anomaly detection, specifically for detecting epidemic outbreaks. Li et al. [58] proposed a mechanism that only releases the histogram when it significantly differs from previous values, with the threshold adjusted according to feedback from the control system. Wang et al. [59] proposed a framework called SecWeb, following the idea of FAST under w-event level differential privacy. In addition to adaptive sampling and filtering, SecWeb incorporates dynamic grouping and injects Laplace noise based on the groups rather than individual elements. Wang et al. [8] proposed the first mechanism to achieve almost local differential privacy (LDP) under the w-event privacy model. Their approach involves employing multiple agents to collect data from users and release sanitized data to an untrusted third party. Liang et al. [60] introduced a mechanism for releasing web browsing histograms under the pufferfish privacy framework, which is beneficial for perturbing correlated data. Their proposed mechanism includes a model to quantify privacy leakage arising from temporal correlations and presents three strategies to enhance the model's efficiency: bounding the number of secret pairs, limiting the session length, and avoiding repetitive computations. Ding et al. [61] proposed a mechanism within the framework of pufferfish privacy to make the time and occurrence of an element indistinguishable.

Based on count queries, mechanisms proposed under the local differential privacy (LDP) framework are commonly used to estimate statistics. To conserve the privacy budget under LDP, a memoization technique [62] was proposed, which stores sanitized versions of all values for further release. [63] improved upon memoization by incorporating hashing. However, memoization may leak privacy in the presence of a knowledgeable adversary who can potentially derive changes without prior knowledge. Xue et al. [64] introduced a difference tree-based mechanism that applies fresh perturbation at each timestamp under user-level privacy, enabling the aggregation of statistics without violating changing points. Additionally, [65] proposed a method to reduce the item domain, thereby enhancing utility. Beyond memoization based methods, He et al. [66] proposed a privacy budget allocation strategy to enhance the utility of frequency release under w-event level condensed local differential privacy (CLDP) [67]. In their approach, the allocated privacy budget depends on the predicted elements, determined via a proportional-integral-derivative (PID) controller. In other applications, Feng et al. [68] proposed a mechanism to estimate the distribution of infinite time series while satisfying user-level differential privacy. Their approach achieves reasonable utility by bounding privacy leakage and optimizing the allocation of the privacy budget. Li et al. [69] introduced the first work on collecting the top-k items from a time series while satisfying event-level local differential privacy and adhering to a bounded memory space constraint. Their proposed mechanisms are based on the HeavyGuardian data structure, which maintains the frequently occurring elements while evicting the infrequent ones. Additionally, Gu et al. [70] introduced a mechanism under pattern-level privacy, which is similar to w-event level privacy but does not require successions. To privately release critical patterns (i.e., subsequences of elements in a time series), their mechanism perturbs the existence of each element, thereby providing a privacy guarantee.

4 Differential Privacy for Sum/Mean Queries

This section provides a comprehensive review of the literature concerning sum and mean queries within the differential privacy framework. It explores the concept of sum/mean queries, the basic queries, and the downstream applications.

4.1 Essential Information of Sum/Mean Queries

While preserving the occurrence of an element in a time series is important, maintaining the accuracy of the element's value is equally crucial. To ensure precise results for queries such as sum or mean, small deviations in the perturbation process are necessary. Since the mean is directly correlated to the sum, we discuss sum and mean queries together.

The sum query can be denoted as

$$\mathrm{F}_{sum}(T, S) = \sum_{i=1}^{T} S_i,$$

where T represents the timestamp for sum release and S is the corresponding time Series. The range query on sum is

$$\mathrm{F}_{rsum}((T_1, T_2), S) = \sum_{i=T_1}^{T_2} S_i,$$

where (T_1, T_2) represents the query range, and S is the corresponding time Series.

As for the mean query, it can be denoted as

$$\mathrm{F}_{mean}(T, S) = \frac{1}{T} \sum_{i=1}^{T} S_i,$$

where T represents the timestamp for mean release, and S is the corresponding time Series. Another common mean query is to release the mean at a timestamp from users' time series,

$$\mathrm{F}_{rtm}(t, S) = \frac{1}{n} \sum_{i=1}^{n} S_t^{u_i},$$

where $S_t^{u_i}$ represents the value at timestamp t from the user u_i, and n is the number of users.

4.2 Basic Sum/Mean Queries

There have proposed a line of work to release sum/mean queries under differential privacy, with a brief summary provided in Table 2. The pioneering work by Bolot et al. [71] was the first to study the continual decaying sums problem. They explored three variants: the window sum (range sum query), which releases the sum of W consecutive elements; the exponential decay sum, which releases the sum of elements weighted by an exponential function; and the polynomial sum, which releases the sum of elements weighted by a polynomial function. Henzinger et al. [72] also investigated the continual decaying sum problem. Their work introduced the use of the Gaussian mechanism for adding noise and derived tighter error bounds. In contrast, Dong et al. [48]

[2] The error bound of [72] in the table is the L_2 norm error.

Query	Ref.	Type	Privacy Level	Method Primitive	Error Bound
Sum	[48]	infinite	user-level CDP	binary tree mechanism	$O(\frac{\varphi(D_t)}{\epsilon\theta} \cdot \log^{1.5}(tR) \cdot \log^{1+\theta}(\varphi(D_t)) \cdot \log(1/\beta))$
Window Sum	[71]	finite	event-level CDP	binary tree mechanism	$O(\frac{1}{\epsilon} \log W \frac{1}{\beta})$
	[72]	finite	event-level (ϵ, δ)-CDP	matrix mechanism	$O(2\sigma^2_{\epsilon,\delta}\Delta^2(1 + \frac{\log W}{\pi} + \frac{2}{W})^2)$
Exponential Decay Sum	[71]	finite	event-level CDP	binary tree mechanism	$O(\frac{1}{\epsilon} \log \frac{\alpha}{1-\alpha} \frac{1}{\beta})$
	[72]	finite	event-level (ϵ, δ)-CDP	matrix mechanism	$O(\sigma^2_{\epsilon,\delta}\Delta^2(1 + \frac{1}{\pi} S_{T,2\alpha})^2)$
Polynomial Decay Sum	[71]	finite	event-level CDP	binary tree mechanism	$\Omega(1 - \frac{\epsilon^{c-1}}{\log^{c-1}(1/\beta)})$
	[72]	finite	event-level (ϵ, δ)-CDP	matrix mechanism	$O(\sigma^2_{\epsilon,\delta}\Delta^2(1 + \frac{H_{T,2c-1}}{4})^2)$

Table 2: The table provides a brief summary of sum-based queries, where $\varphi(D_t)$ denotes the maximum contribution from any user at time t, θ is a small constant, W is the window length, β is the confidence parameter, α indicates the exponential decay parameter, c represents the polynomial decay parameter, $H_{T,2c}$ is the generalized Harmonic sum, $S_{T,2\alpha}$ is a defined series sum with $\alpha > 1$.[2]

addressed sum queries by reducing them to count queries, as their approach could only handle the latter. For each timestamp with value x_i, they expand it into R steps, where the first $x_i(w.l.o.g., x_i < R)$ steps are filled with 1, and the remaining steps are filled with a special symbol $\perp$. However, this method requires prior knowledge of the maximum value R that can occur in the time series. [73] proposed a method for answering sum queries with a threshold under a multi-branch tree structure. The threshold is optimized based on the expected squared error between the true result and the estimated one. Their proposed mechanism cannot handle infinite time series, so they claim that most queries focus on a limited range, allowing for truncation of the infinite time series. Additionally, their work introduced the first mechanism to release time series under LDP with a threshold for value truncation. Instead of directly perturbing the values, the mechanisms proposed by Ye et al. [74, 75] perturb the temporal order. This approach makes them naturally adaptable for sum/mean queries while preserving the original values. These methods demonstrate superior performance for calculating moving averages.

4.3 Downstream Applications Based on Sum/Mean Queries

Due to utility considerations, existing mechanisms for downstream applications based on sum and mean queries primarily operate under event-level privacy or w-event level privacy. Perrier et al. [76] introduced a differentially private mechanism for publishing statistics of real-valued time series under event-level privacy, such as moving averages derived from energy data collected through smart meters. Their approach addresses scenarios where the bound on observations is either overly conservative or unknown, which is crucial for real-time monitoring applications. The proposed mechanism optimizes utility by scaling the added noise to the threshold value instead of a potentially larger bound, thereby improving accuracy. To enable real-time computation of the mean at any timestamp from users' time series under w-event LDP, Wang et al. [77] proposed sampling strategy to select important elements and a privacy budget allocation strategy according to the importance of the elements. However, the sampling process may inadvertently reveal some private information due to the intentional selection. Kurt et al. [78] proposed an online anomaly detection method for networks based on the cumulative sum algorithm, satisfying event-level (ϵ, δ)-differential privacy. Their approach adds noise to the statistic at each timestamp from each network node, operating under event-level privacy, and then derives the mean from the data of all nodes. This allows for detecting anomalies by monitoring changes in the released means.

5 Differential Privacy for Time Series Release

Time series release aims to publicly share private time series while preserving privacy. Value perturbation methods add noise to the data values, often using sampling and adaptive budget allocation for utility. While temporal perturbation methods dispatch elements across timestamps to obfuscate event timings, avoiding value distortion but risking empty releases or collisions.

5.1 Value Perturbation Based Methods

Time series release aims to directly publicize the time series for downstream tasks. Since time series release focuses on preserving the accuracy of values, privacy budget allocation is critical for controlling added noise and minimizing distortion. Therefore, a sampling-based method is often employed to select crucial elements according to the tasks, reducing the number of points requiring privacy budget allocation and enhancing utility. Additionally, to further improve the utility of the released data, a post-processing step can be adopted to correct the noisy data using prior knowledge. The outline of time series release is summarized in Fig. 5.

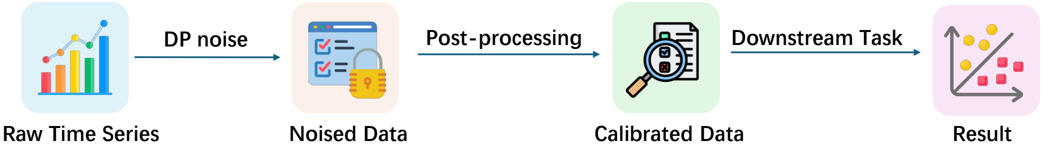


Figure 5: The outline of time series release under differential privacy.

5.1.1 Privacy Budget Allocation Strategies

User-level privacy protects the elements from any user but requires a larger privacy budget for reasonable utility, particularly challenging for time series release. Conversely, event-level privacy safeguards individual elements, yet may not suffice for comprehensive privacy guarantees. To balance the privacy levels, Kellaris et al. [19] first proposed w-event level privacy under CDP. By leveraging w-event level privacy, sanitized time series can offer better privacy guarantees than event-level privacy and higher utility than user-level privacy. To optimize the advantages of w-event level privacy, Kellaris et al. [19] proposed two privacy budget allocation strategies. Building upon the ideas of the budget distribution and absorption strategies in [19], Ren et al. [79] proposed corresponding strategies under LDP framework. To mitigate the utility degradation caused by dividing the privacy budget, the authors divide the users instead, with each user reporting only one timestamp within a w-length window. Several sampling-based methods have been developed to reduce the number of elements requiring a privacy budget, with adaptive budget allocation based on element importance. Wang et al. [80] introduced RescueDP, a scheme for real-time publishing of spatio-temporal crowd-sourced data with w-event level CDP, integrating adaptive sampling, privacy budget allocation, dynamic grouping, perturbation, and filtering techniques. The adaptive sampling component adjusts sampling rates based on data changes, ensuring efficient resource utilization. The privacy budget allocation mechanism dynamically distributes the privacy budget for sampling points across successive timestamps. Zhang et al. [81] proposed Re-DPoctor, a real-time health data releasing scheme ensuring w-event level CDP, enhancing utility with a partition algorithm safeguarding health data patterns and improving privacy through adaptive sampling and budget allocation. He et al. [66] introduced a new privacy concept using condensed local differential privacy (CLDP) [67] for w-event level privacy, aiming to enhance utility. They save privacy budget from empty releases and reallocate it to released elements, and utilize a PID controller-based method for adaptive budget allocation. However, this approach may inadvertently disclose private information through omitted empty points.

5.1.2 Optimization Strategies

Leveraging the correlations present in time series, the pre-processing or post-processing methods can be applied to improve the utility of the perturbed data. Ren et al. [82] addressed privacy challenges in high-dimensional crowdsourced data by proposing LoPub, an LDP data publication algorithm under event-level privacy. They use expectation maximization (EM) and Lasso regression to efficiently estimate multivariate joint distributions, identifying attribute correlations to reduce data dimensionality for distribution learning speed and utility improvements. Wang et al. [83] proposed the methods, LoCop and DR_LoCop, for releasing high-dimensional crowdsourced data under event-level LDP. The methods comprise four integrated components: a transformation component that ensures LDP by hashing and randomizing the data, an estimation component that infers probability distributions from the resulting Bloom filter strings, a computation component that derives marginal distributions and captures dependencies among dimensions, and a sampling component that generates a new synthetic dataset based on the computed distributions and dependencies. Fioretto and Hentenryck [84] introduced OptStream, a method for releasing time series under w-event level, which extends to handling hierarchical streams like energy profiles, making it applicable beyond its original energy domain. OptStream involves four key steps: sampling points for private measurement, perturbing them for privacy, reconstructing non-sampled points, and post-processing with convex optimization to improve accuracy by redistributing added noise. Zhang et al. [85] introduced a method for releasing differentially private sequential data using first-order autoregressive processes under user-level privacy. Their approach estimates unreleased data from previously released data by leveraging learned correlations, without requiring prior knowledge. The estimated data is combined with the observed data and perturbed with calibrated noise at each timestamp, facilitating real-time data release. Li et al. [86] proposed a framework for locally private stream data release that employs shuffling and subsampling techniques. Their approach maintains utility in the context of continual data collection by sampling a subset of users at each timestamp. An optimal sample size is determined to reduce redundant data and enhance utility, and the framework incorporates pre-processing within the shuffler to mitigate bias arising from distributed sampling. Besides pre- or post-processing methods, Bao et al. [87] proposed a mechanism based on the assumption of data fluctuation. Since time series may not change significantly over time, this mechanism formalizes the correlation between elements, allowing a later element to be represented by previous elements. To protect privacy, noise is added to the correlation.

5.2 Synthesis Based Methods

Directly releasing users' data poses significant risks of privacy breaches. An alternative approach is to train a synthesis model under strict privacy conditions and then release the data generated by this model for downstream tasks. To synthesize time series data under DP, one method involves first estimating the relevant statistics and then generating the synthetic data based on these estimations. Additionally, the advent of Generative Adversarial Networks (GANs) under DP [88, 89] has facilitated the use of deep learning algorithms to generate data, thereby enhancing both privacy protection and data accuracy.

5.2.1 Synthesis Based on Statistics

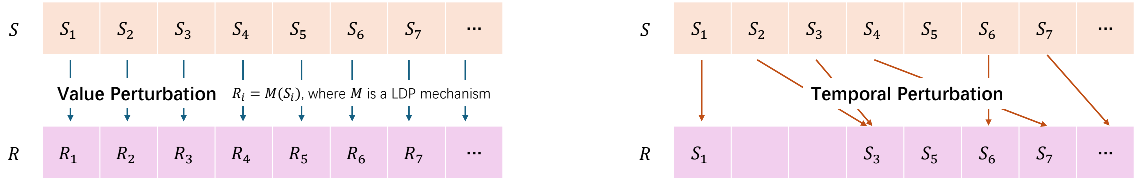
Synthesis mechanisms based on statistical features first capture the statistical characteristics from datasets under DP. Subsequently, new data is generated according to these estimated statistics. He et al. [90] introduced an efficient polynomial-time algorithm for generating online differentially private synthetic data under event-level privacy from a continuous time series within the hypercube $[0, 1]^d$. The algorithm achieves near-optimal accuracy bounds in 1-Wasserstein distance and extends previous work to include Lipschitz queries. By utilizing an online hierarchical partitioning approach and a novel Inhomogeneous Sparse Counting Algorithm, the method maintains strong privacy guarantees while ensuring high utility for infinite time horizons. To achieve a higher privacy level, Bun et al. [91] focused on generating differentially private synthetic data through statistical estimation under user-level CDP. They proposed algorithms that maintain the accuracy of fixed time window and cumulative time queries, ensuring minimal error while preserving privacy. Their approach involves a two-stage process that combines noisy estimates with post-processing techniques to ensure consistency and accuracy in synthetic data generation.

5.2.2 Synthesis Based on Generative Models

In addition to statistics-based mechanisms, another method for synthesizing time series is through Generative Adversarial Networks (GANs). Unlike other data types, time series requires consideration of temporal correlations. Frigerio et al. [92] presented a framework for releasing high-quality open data while ensuring user privacy through DP, addressing both continuous and discrete data. By leveraging deep learning and generative models with long short-term memory networks, the framework maintains data utility and correlations, introducing innovations such as clipping decay to optimize performance. Wang et al. [93] introduced PART-GAN, a privacy-preserving generative model designed for time series augmentation and sharing. PART-GAN combines Conditional and Temporal Generative Adversarial Networks (CT-GAN) with differential privacy mechanisms, enabling the generation of unlimited synthetic data that addresses issues of incomplete and irregularly sampled time series. Torfi et al. [94] proposed a mechanism to generate high-quality synthetic health record data while ensuring privacy using Rényi Differential Privacy (RDP). Their framework combines convolutional autoencoders and convolutional generative adversarial networks (CGAN) to effectively handle both discrete and continuous data, preserving temporal and feature correlations. For specific applications, Lamp et al. [95] introduced GlucoSynth, a novel privacy-preserving GAN framework designed to generate high-quality synthetic glucose traces while maintaining strong differential privacy guarantees. By focusing on preserving the relationships among glucose events (motifs) and temporal dynamics, GlucoSynth addresses the unique challenges of synthesizing glucose data.

5.3 Discussion

The aforementioned release mechanisms modify the values of the original time series, which can degrade utility in value-critical scenarios. For elements in a time series where occurrence indicates more sensitive information, temporal perturbation can be employed to avoid distorting the original values. Ye et al. [74] first proposed a method to achieve temporal perturbation in the local setting, maintaining the privacy guarantee while enhancing utility. As illustrated in Fig. 6, unlike value perturbation that directly modifies the original values by adding LDP noise, temporal perturbation dispatches elements across different timestamps. This obfuscates the precise occurrence times, preventing adversaries from determining the exact event timings. However, temporal perturbation can lead to issues such as delayed releases, empty releases (where no elements are dispatched to certain timestamps), and element substitutions, resulting in missing data. To address these issues, Ye et al. [75] proposed a bi-directional perturbation mechanism that eliminates collisions during the dispatching process, ensuring that elements are only delayed. Furthermore, Mao et al. [96] extended the definition to a metric-based version tailored for anomaly detection, aiming to reduce collisions involving anomalous elements. However, these proposed mechanisms [74, 75, 96] primarily address event-level privacy concerns, leaving room for enhancements to achieve higher levels of privacy.



S
S_1 S_2 S_3 S_4 S_5 S_6 S_7 ⋯
Value Perturbation
$R_i = M(S_i)$, where M is a LDP mechanism
R
R_1 R_2 R_3 R_4 R_5 R_6 R_7 ⋯
S
S_1 S_2 S_3 S_4 S_5 S_6 S_7 ⋯
Temporal Perturbation
R
S_1 S_3 S_5 S_6 S_7 ⋯

Figure 6: Value perturbation-based methods will perturb the original values by adding DP noise. In contrast, temporal perturbation-based methods will dispatch the values to corresponding timestamps for release.

6 Differential Privacy for Location Based Services

Location Based Service (LBS) is a crucial application of mobile computing that provides personalized services based on users' geographical locations. These services, ranging from navigation assistance to location-based recommendations, require continuous collection and analysis of users' location data, often in the form of trajectories.

A trajectory is a specific type of time series that comprises spatial-temporal data. It can be regarded as a sequence of time-ordered points, denoted as $T = \{p_1, p_2, \cdots, p_T\}$, where p_i represents a location and T is the length of the trajectory. Compared with other time series, the correlations within trajectory data are more pronounced due to the constraints imposed by spatial variation. Regarding privacy levels for trajectory data, location privacy corresponds to event-level privacy, providing protection for individual locations, whereas trajectory privacy safeguards the entire trajectory.

In this section, we introduce mechanisms for handling trajectory data under differential privacy, organized according to utility improvement in location perturbation, privacy preservation against temporal correlation, and trajectory release. These mechanisms ensure that the privacy of individual locations and movements is preserved while maintaining the utility of the data for analysis and service provision.

6.1 Location Perturbation Based on Geo-indistinguishability

For meaningful outputs in LBS, the perturbed location should not deviate excessively from the actual one. As illustrated in Fig. 7, constraining the perturbation domain is essential for improving utility; otherwise, a large perturbation domain yields less useful results. For example, perturbing Paris to London is impractical [29]. Therefore, a metric-based privacy notion, ϵ-geo-indistinguishability, is proposed. Specifically, a user's level of privacy is defined as $\ell = \epsilon r$, where r is the radius of the perturbation domain, corresponding to r_i in Fig. 7. Here is the formal definition of geo-indistinguishability.

Definition 6.1: *[Geo-indistinguishability [29]] Given any two locations x and x' ($\mathrm{d}(x, x') \leq r$), a randomized mechanism M satisfies ϵ-geo-indistinguishability iff*

$$\Pr[M(x) \in Z] \leq e^{\epsilon d(x,x')} \Pr[M(x') \in Z],$$

where $Z \subseteq \mathcal{Z}$ is the possible output domain, where $\mathrm{d}(x, x')$ represents a distance measure between x and x'.

Since the inception of geo-indistinguishability, numerous enhancements have been made to improve the privacy notion from various perspectives. To enhance the calculation efficiency, Bordenabe et al. [97] proposed a method to optimize the trade-off between geo-indistinguishability and service quality in location privacy, employing linear programming to minimize service quality loss while ensuring optimal privacy guarantees. By reducing the number of constraints from cubic to quadratic, their approach significantly improves computational efficiency. Building on the concept of geo-indistinguishability, Weggenmann and Kerschbaum [98] introduced the notion of directional privacy, a relaxation of pure differential privacy that performs effectively in the local model. To enhance practical utility, Zhao et al. [99] proposed geo-ellipse-indistinguishability to protect individual location data in directional distribution analysis. This method incorporates the covariance matrix to account for dispersion and orientation of community locations, using elliptical noise instead of circular noise. The proposed mechanisms, based on gamma and multivariate normal distributions, ensure higher probabilities of randomized locations aligning with community location trends while maintaining statistical quality. Liang and Yi [100] further advanced the concept of geo-indistinguishability by introducing concentrated geo-privacy, an update from the CDP version. This approach supports advanced composition mechanisms for high-dimensional data and achieves a lower noise scale, thereby enhancing overall privacy protection while maintaining utility. Zhao and Chen [101] proposed vector-indistinguishability (vector-ind) to enhance location privacy by preserving distance

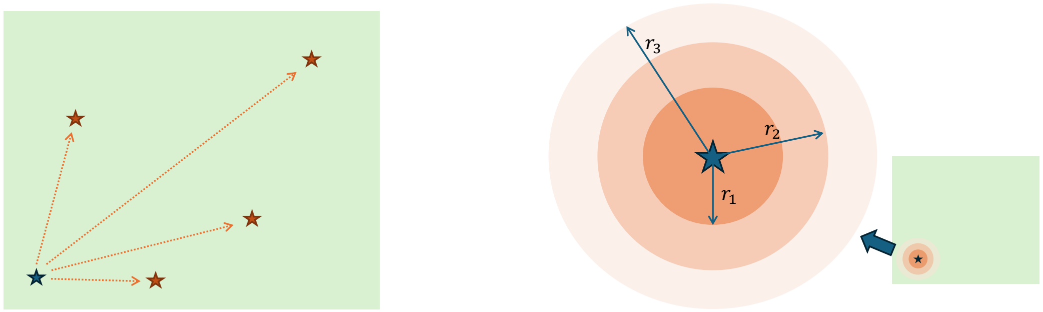


Figure 7: A traditional DP mechanism (illustrated in the left-hand figure) involves a perturbation domain (green area) for a location that is typically large, resulting in low utility for the perturbed location. To improve utility while providing useful services, a metric-based privacy notion, geo-indistinguishability, is introduced to control the perturbation (illustrated in the right-hand figure). A smaller distance r_i leads to higher utility but offers less privacy. Therefore, it is important to balance the trade off between utility and privacy when designing mechanisms under geo-indistinguishability.

and direction dependencies between successive locations. They introduced four mechanisms using Laplace and Uniform distributions to achieve vector-ind, maintaining data utility while ensuring CDP.

Several mechanisms have been proposed to address various privacy issues across numerous scenarios based on geo-indistinguishability. Yu et al. [102] proposed PIVE, a dynamic differential location privacy framework that integrates geo-indistinguishability and expected inference error to protect against inference attacks. PIVE operates in two phases: First, it identifies a protection location set based on user-defined error thresholds and prior knowledge; And then, it generates pseudo-locations within this set, ensuring differential privacy. This approach enables adaptive, personalized privacy settings tailored to individual user needs and location-based service requirements, thereby enhancing both privacy preservation and utility. Cao et al. [103] extended differential privacy to define ϵ-spatiotemporal event privacy and proposed a framework to quantify its protection level in existing location privacy-preserving mechanisms. They demonstrated their framework by adapting the Planar Laplace Mechanism for geo-indistinguishability to ensure spatiotemporal event privacy while maintaining linear computational complexity. Niu et al. [104] introduced Eclipse, a mechanism that combines geo-indistinguishability, k-anonymity, and expected inference error to protect location privacy against long-term observation attacks. Eclipse obfuscates user locations within an anonymity set, minimizing privacy leakage while maintaining service usability and correctness. Qiu et al. [105] tackled the Vehicle-based spatial crowdsourcing Location Privacy (VLP) problem, aiming to minimize travel cost distortion while preserving location privacy over road networks. They redefined geo-indistinguishability based on path distance and approximated the VLP problem as a linear programming formulation through discretization. To improve time efficiency, they proposed a two-layer optimization algorithm and analyzed the trade-off between privacy and quality of service. Haydari et al. [106] proposed a differential privacy-based map-matching algorithm (DPMM) for protecting user privacy in mobility data. DPMM generates privatized link-level location trajectories by incorporating road characteristics such as capacity and functional role. The algorithm adaptively selects the noise level based on link density, effectively balancing privacy preservation and trajectory accuracy.

6.2 Privacy Preservation Against Temporal Correlation

Due to the intrinsic features of location data, temporal correlations pose significant privacy issues when handling locations. Specifically, an adversary can infer information about a location based on its preceding or succeeding elements. For instance, Shao et al. [20] proposed iTracker, a framework designed to recover multiple trajectories from differentially private data using a structured sparsity model. iTracker leverages interdependencies among locations to enhance recovery accuracy, effectively challenging existing Laplace perturbation-based location protection mechanisms. To address privacy risks posed by temporal correlations in location data, numerous mechanisms have been proposed. Xiao and Xiong [107] introduced a solution that preserves location privacy with rigorous differential privacy guarantees by proposing δ-location set, which accounts for temporal correlations in location data. They also introduced the sensitivity hull to capture geometric sensitivity in multidimensional space and presented the planar isotropic mechanism (PIM), an efficient location perturbation method that achieves optimal utility while meeting differential privacy requirements. Cao et al. [21] first investigated the privacy loss of CDP mechanisms under temporal correlations and introduced the concept of Temporal Privacy Leakage (TPL). They proposed an efficient algorithm to calculate TPL and designed methods to convert traditional DP mechanisms to ones that mitigate TPL, ensuring privacy over continuous data releases by bounding the leakage within a defined parameter α. Xiao et al. [22] proposed LocLok, a system that protects user locations with differential privacy by modeling temporal correlations using a hidden Markov model and applying PIM for optimal noise addition. LocLok generates possible locations via the Markov model, perturbs them with PIM, and infers true locations within a set of all possible locations, ensuring robust local privacy even when adversaries have access to historical location data. Liu et al. [108] protected location privacy by analyzing the impact of temporal-spatial correlations and proposing new privacy definitions, introducing Bayesian-based geo-indistinguishability to better evaluate and enhance privacy levels. Their method optimally allocates noise among spatially and temporally correlated locations, effectively protecting sensitive locations within a trajectory while achieving differential privacy. Ma et al. [109] proposed RPTR under w-event level CDP to protect real-time vehicle trajectory data. They employed dynamic sampling and ensemble Kalman filters, utilizing a position transfer probability matrix to infer correlations and ensure accurate predictions while balancing data availability and privacy. Additionally, they introduced a regional privacy weight mechanism to enhance protection in high-density areas, thereby ensuring higher prediction accuracy and adaptability across different scenarios. Cao et al. [110] propose a post-processing framework to enhance the utility of differentially private streaming data releases by leveraging temporal correlations. They modeled the problem as a maximum a posterior estimation, transformed it into a nonlinear constrained programming problem, and used a transition matrix to incorporate both probabilistic and deterministic constraints. Ahuja et al. [111] proposed a method to release histogram information from trajectories under user-level CDP. To enhance utility, they introduced a method based on variational autoencoders to refine the histograms by utilizing the correlations of histograms.

6.3 Trajectory Release Based on Perturbation or Synthesis

As location-based services (LBS) become increasingly integral to everyday applications, ensuring the privacy of users while maintaining the utility of location data remains a critical challenge. Various mechanisms have been proposed to address this issue, each focusing on different aspects of location privacy and data utility. Wang et al. [112] introduced L-SRR, the first LDP framework for location-based services, enhancing utility while ensuring strict privacy. The proposed staircase randomized response mechanism perturbs user locations using optimized probabilities, significantly improving utility for applications such as traffic density estimation and k-nearest neighbor queries. Cunningham [113] introduced a locally differentially private mechanism for trajectory data sharing that integrates public knowledge to enhance utility while ensuring privacy. This mechanism perturbs hierarchically-structured n-grams of trajectory data to capture spatio-temporal relationships, leveraging public data without compromising privacy. Zhang et al. [114] proposed a trajectory perturbation mechanism under user-level LDP that enhances privacy by using adjacent direction information to connect neighboring points. They introduce a two-stage pivot sampling process utilizing bi-directional clues from pivots, and an anchor-based method to restrict the spatial region of trajectories.

Synthetic trajectory generation has emerged as another promising solution, allowing for the publication of useful data without compromising individual privacy. Gursoy et al. [115] presented DP-Star, a framework for publishing trajectory data that ensures differential privacy while maintaining high utility. DP-Star normalizes raw trajectories using representative points, constructs a density-aware grid to preserve spatial densities, and employs a private Markov mobility model to maintain correlations and intra-trajectory mobility. This results in synthetic trajectory datasets that are both privacy-preserving and useful for various data mining tasks. Moreover, Gursoy et al. [116] presented AdaTrace, a scalable location trace synthesizer that achieves statistical privacy, deterministic attack resilience, and strong utility preservation. AdaTrace generates differentially private synthetic traces through a four-phase process: feature extraction, noise injection, and utility-aware synthesis. The synthetic traces preserve utility-critical information and are robust against Bayesian inference, partial sniffing, and outlier leakage attacks, ensuring privacy without significant utility loss. Du et al. [117] introduced LDPTrace, a locally differentially private framework for synthesizing realistic trajectories with minimal computational cost and strong privacy guarantees. LDPTrace captures key movement patterns from users' trajectories, ensuring robust statistical privacy and resilience against attacks. Extensive evaluations demonstrate that LDPTrace generates authentic trajectories without external knowledge, outperforming existing methods in terms of utility and privacy protection. Hu et al. [118] introduced RetraSyn under w-event level LDP, aimed at real-time trajectory synthesis while ensuring data privacy. RetraSyn leverages mobility patterns from trajectory streams and incorporates a global mobility model, dynamic update mechanisms, and Markov-based synthesis to generate realistic trajectories. This framework effectively captures complex spatial-temporal contexts and employs adaptive privacy budget allocation strategies, ensuring authenticity and practicality in diverse real-world scenarios. Sun et al. [119] proposed SPRT, a method for synthesizing private and realistic vehicle trajectories by incorporating geographic structures into differential privacy mechanisms. SPRT constructs a geography-aware grid to capture accurate mobility patterns and defines a moveable constraint based on real-world conditions, enhancing both summary-level statistics and individual-level mobility patterns.

7 Open Challenges

Although many mechanisms have been proposed to handle time series under differential privacy, several issues still need to be addressed. In this section, the challenges will be introduced according to privacy model, potential attacks, data type, and learning based problems.

7.1 Privacy Model

The privacy model is a crucial factor in differential privacy. As aforementioned, there are three privacy levels when handling time series [19]. Event-level privacy guarantees the privacy of a single element in a time series, making it easier to implement since the sensitivity of an individual element in neighboring datasets is simpler to measure. In contrast, user-level privacy provides a higher privacy guarantee and is more practical in real-world applications. However, bounding the sensitivity of a single user's participation is challenging, making the allocation of the privacy budget more complex. Additionally, utility issues become more pronounced when dealing with infinite time series.

Several research works have explored handling infinite time series under user-level privacy with specific conditions. Dong et al. [48] introduced mechanisms for basic queries such as count and sum. These mechanisms are based on event-level privacy approaches and are adapted to user-level privacy by bounding the maximum changes of a single user in a time series. For LDP, Xue et al. [64] proposed a mechanism for count queries, but it requires that the time series does not fluctuate significantly. Feng et al. [68] proposed a strategy that randomly allocates the privacy budget according to a converging sum series.

Therefore, improving the utility of infinite time series under user-level differential privacy is an intriguing future direction. Beyond basic queries, efforts can be made to accommodate specific queries or applications. Key challenges include accurately measuring data sensitivity and effectively allocating the privacy budget. Overcoming these challenges can enhance the practical utility of differential privacy mechanisms for managing infinite time series under user-level privacy.

7.2 Temporal Correlation Based Attacks

Compared to other data types, the correlation in time series is more pronounced. Due to the inherent sequential nature of time series, each element is often directly influenced by its predecessors. Various works have employed the Markov model to capture and represent these correlations under DP [21, 22, 115]. In the context of the Markov model, an element is directly influenced only by its immediate neighboring element. Consequently, the influence of previous elements is implicitly carried forward through the chain of direct dependencies between neighboring elements. Since time series are always modeled explicitly, this simplicity can result in the model overlooking long-term information that extend beyond immediate neighbors. To capture such information, more complex models like the long short-term memory network [3] are needed. Moreover, for specific types of time series such as trajectories, public knowledge can introduce additional privacy issues. For example, certain perturbations may be impossible due to physical world limitations. In summary, compared with single data points, elements in time series are at a greater risk of privacy leakage. This suggests a potential direction for research: attacking existing privacy mechanisms by exploiting these correlations and to design new mechanisms that account for the inherent dependencies in time series.

7.3 Complex Data Type

Most current works can only handle simple time series with high utility, such as one value at each timestamp. However, real-world data are more complex, and mechanisms should be designed to handle this complexity. Here are two examples:

First, sensor data are often multi-dimensional and correlated across each dimension. This complexity requires mechanisms capable of managing and analyzing data with multiple interacting variables. Traditional methods that handle single-dimensional elements at each timestamp are insufficient for capturing the nuances of multi-dimensional sensor data. For example, environmental sensors may collect temperature, humidity, and air pressure simultaneously. Analyzing these factors independently can miss critical interactions and patterns, such as how temperature changes might influence humidity levels during users' activities. Therefore, advanced methods must be developed to process and interpret multi-dimensional sensor data effectively.

Second, the element at each timestamp can be intricate. For instance, social networks change over time, making real-time analysis of such dynamic networks complicated. Managing the evolution of relationships and interactions within the network adds a layer of complexity beyond time series analysis. Additionally, privacy concerns in such contexts extend beyond the temporal dimension to include graph privacy, encompassing node-level privacy and edge-level privacy. Mechanisms must account for these additional privacy requirements to ensure data protection while enabling real-time analysis.

7.4 Learning Based Problems

Current DP mechanisms are primarily designed for basic queries, such as count, mean, and frequency. However, time series without privacy concerns are often used for more complex downstream tasks, such as classification and clustering. These tasks require a deeper understanding and manipulation of the data, going beyond simple statistical queries. To accommodate more practical downstream tasks, it is essential to develop DP mechanisms that can support these sophisticated operations effectively, ensuring both the utility and privacy of the data.

A reasonable solution is to generate synthetic time series from the real dataset. Synthetic approach allows the fundamental features of the time series to be captured while protecting the privacy of the underlying data. Since the features of time series are complex and extend beyond basic statistics, traditional statistical methods are insufficient for capturing these intricate patterns. For example, critical patterns such as seasonal trends, cyclic behaviors, and sudden anomalies are vital in time series analysis but are not adequately addressed by basic statistical methods.

Therefore, learning-based methods are preferable for generating synthetic time series. These methods can model and replicate the complex dependencies and structures inherent in time series. For instance, the method proposed by Lamp et al. [95] for synthesizing glucose traces exemplifies how deep learning can be applied to generate realistic and privacy-preserving synthetic data. Predictably, more and more application-specific mechanisms will be proposed.

8 Conclusion

In this paper, we present a comprehensive survey on handling time series under differential privacy. We begin by introducing the basic concepts of time series and differential privacy, along with relevant definitions. Our survey starts with an exploration of two basic queries: count queries and sum/mean queries. For each query type, we first explain the concept of the basic query, then review the core techniques or developments related to the queries, and finally discuss the advanced queries derived from the basic ones. At the end of each query section, we review the downstream tasks based on these queries. Subsequently, we introduce mechanisms for time series release, categorizing them into value perturbation based methods and synthetic generation based methods. Additionally, we dedicate a separate section to location-based services (LBS), as they are common application scenarios for time series. We review relevant papers for LBS according to two popular privacy issues and the demands of trajectory release. Finally, we illustrate four open challenges and suggest future directions.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant No: 62372122, 62072390, and 92270123), the Research Grants Council, Hong Kong SAR (Grant No: 15203120, 15208923 and 15210023), Shenzhen Key Laboratory of Safety and Security for Next Generation of Industrial Internet
(Grant No. ZDSYS20210623092007023) and Guangdong Provincial Key Laboratory of Brain-inspired Intelligent Computation (Grant No. 2020B121201001)

References

[1] Hyndman, R. & Athanasopoulos, G. Forecasting: principles and practice. (OTexts,2018)

[2] Gardner Jr, E. Exponential smoothing: The state of the art—Part II. International Journal Of Forecasting. **22**, 637-666 (2006)

[3] Shi, X., Chen, Z., Wang, H., Yeung, D., Wong, W. & Woo, W. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances In Neural Information Processing Systems. **28** (2015)

[4] C. Dwork, "Differential privacy," in International colloquium on automata, languages, and programming. Springer, 2006, pp. 1–12.

[5] Ye, Q. & Hu, H. Local differential privacy: Tools, challenges, and opportunities. International Conference On Web Information Systems Engineering. pp. 13-23 (2020)

[6] M. Yang, T. Guo, T. Zhu, I. Tjuawinata, J. Zhao, and K.-Y. Lam, "Local differential privacy and its applications: A comprehensive survey," Computer Standards & Interfaces, p. 103827, 2023.

[7] Wang, T., Blocki, J., Li, N. & Jha, S. Locally differentially private protocols for frequency estimation. 26th USENIX Security Symposium (USENIX Security 17). pp. 729-745 (2017)

[8] Wang, Z., Pang, X., Chen, Y., Shao, H., Wang, Q., Wu, L., Chen, H. & Qi, H. Privacy-preserving crowd-sourced statistical data publishing with an untrusted server. IEEE Transactions On Mobile Computing. **18**, 1356-1367 (2018)

[9] Wang, T., Li, N. & Jha, S. Locally differentially private heavy hitter identification. IEEE Transactions On Dependable And Secure Computing. **18**, 982-993 (2019)

[10] Fu, Y., Ye, Q., Du, R. & Hu, H. Collecting Multi-type and Correlation-Constrained Streaming Sensor Data with Local Differential Privacy. ACM Transactions On Sensor Networks. (2023)

[11] Li, Z., Wang, T., Lopuhaä-Zwakenberg, M., Li, N. & Škoric, B. Estimating numerical distributions under local differential privacy. Proceedings Of The 2020 ACM SIGMOD International Conference On Management Of Data. pp. 621-635 (2020)

[12] Ye, Q., Hu, H., Au, M., Meng, X. & Xiao, X. Towards locally differentially private generic graph metric estimation. 2020 IEEE 36th International Conference On Data Engineering (ICDE). pp. 1922-1925 (2020)

[13] Zhang, J., Zhang, Z., Xiao, X., Yang, Y. & Winslett, M. Functional mechanism: Regression analysis under differential privacy. ArXiv Preprint ArXiv:1208.0219. (2012)

[14] Abadi, M., Chu, A., Goodfellow, I., McMahan, H., Mironov, I., Talwar, K. & Zhang, L. Deep learning with differential privacy. Proceedings Of The 2016 ACM SIGSAC Conference On Computer And Communications Security. pp. 308-318 (2016)

[15] Fu, J., Ye, Q., Hu, H., Chen, Z., Wang, L., Wang, K. & Xun, R. DPSUR: Accelerating Differentially Private Stochastic Gradient Descent Using Selective Update and Release. ArXiv Preprint ArXiv:2311.14056. (2023)

[16] Dwork, C., Naor, M., Pitassi, T. & Rothblum, G. Differential privacy under continual observation. Proceedings Of The Forty-second ACM Symposium On Theory Of Computing. pp. 715-724 (2010)

[17] Chan, T., Shi, E. & Song, D. Private and continual release of statistics. ACM Transactions On Information And System Security (TISSEC). **14**, 1-24 (2011)

[18] Fan, L. & Xiong, L. An adaptive approach to real-time aggregate monitoring with differential privacy. IEEE Transactions On Knowledge And Data Engineering. **26**, 2094-2106 (2013)

[19] Kellaris, G., Papadopoulos, S., Xiao, X. & Papadias, D. Differentially private event sequences over infinite streams. Proc. VLDB Endow.. **7**, 1155-1166 (2014,8), https://doi.org/10.14778/2732977.2732989

[20] Shao, M., Li, J., Yan, Q., Chen, F., Huang, H. & Chen, X. Structured sparsity model based trajectory tracking using private location data release. IEEE Transactions On Dependable And Secure Computing. **18**, 2983-2995 (2020)

[21] Cao, Y., Yoshikawa, M., Xiao, Y. & Xiong, L. Quantifying differential privacy under temporal correlations. 2017 IEEE 33rd International Conference On Data Engineering (ICDE). pp. 821-832 (2017)

[22] Xiao, Y., Xiong, L., Zhang, S. & Cao, Y. Loclok: Location cloaking with differential privacy via hidden markov model. Proceedings Of The VLDB Endowment. **10**, 1901-1904 (2017)

[23] Dwork, C., Roth, A. & Others The algorithmic foundations of differential privacy. Foundations And Trends® In Theoretical Computer Science. **9**, 211-407 (2014)

[24] Zhao, Y. & Chen, J. A survey on differential privacy for unstructured data content. ACM Computing Surveys (CSUR). **54**, 1-28 (2022)

[25] Zhao, Y., Du, J. & Chen, J. Scenario-based Adaptations of Differential Privacy: A Technical Survey. ACM Computing Surveys. **56**, 1-39 (2024)

[26] Miranda-Pascual, À., Guerra-Balboa, P., Parra-Arnau, J., Forné, J. & Strufe, T. SoK: Differentially private publication of trajectory data. Proceedings On Privacy Enhancing Technologies. (2023)

[27] Katsomallos, M., Tzompanaki, K. & Kotzinos, D. Privacy, space and time: A survey on privacy-preserving continuous data publishing. Journal Of Spatial Information Science. **2019**, 57-103 (2019)

[28] Li, C., Miklau, G., Hay, M., McGregor, A. & Rastogi, V. The matrix mechanism: optimizing linear counting queries under differential privacy. The VLDB Journal. **24** pp. 757-781 (2015)

[29] Andrés, M., Bordenabe, N., Chatzikokolakis, K. & Palamidessi, C. Geo-indistinguishability: Differential privacy for location-based systems. Proceedings Of The 2013 ACM SIGSAC Conference On Computer & Communications Security. pp. 901-914 (2013)

[30] B. Ding, J. Kulkarni, and S. Yekhanin, "Collecting telemetry data privately," Advances in Neural Information Processing Systems, vol. 30, 2017.

[31] Ú. Erlingsson, V. Pihur, and A. Korolova, "Rappor: Randomized aggregatable privacy-preserving ordinal response," in Proceedings of the 2014 ACM SIGSAC conference on computer and communications security, 2014, pp. 1054–1067.

[32] A. G. Thakurta, A. H. Vyrros, U. S. Vaishampayan, G. Kapoor, J. Freudiger, V. R. Sridhar, and D. Davidson, "Learning new words," Mar. 14 2017, uS Patent 9,594,741.

[33] C. Dwork, "Differential Privacy and the US Census," in Proceedings of the 38th ACM SIGMOD-SIGACT-SIGAI symposium on principles of database systems, 2019, pp. 1–1.

[34] Li, N., Lyu, M., Su, D. & Yang, W. Differential privacy: From theory to practice. (Springer,2017)

[35] Kifer, D. & Machanavajjhala, A. Pufferfish: A framework for mathematical privacy definitions. ACM Transactions On Database Systems (TODS). **39**, 1-36 (2014)

[36] Song, S., Wang, Y. & Chaudhuri, K. Pufferfish privacy mechanisms for correlated data. Proceedings Of The 2017 ACM International Conference On Management Of Data. pp. 1291-1306 (2017)

[37] Korolova, A., Kenthapadi, K., Mishra, N. & Ntoulas, A. Releasing search queries and clicks privately. Proceedings Of The 18th International Conference On World Wide Web. pp. 171-180 (2009)

[38] Xiao, X., Wang, G. & Gehrke, J. Differential privacy via wavelet transforms. IEEE Transactions On Knowledge And Data Engineering. **23**, 1200-1214 (2010)

[39] Wang, N., Xiao, X., Yang, Y., Zhao, J., Hui, S., Shin, H., Shin, J. & Yu, G. Collecting and analyzing multidimensional data with local differential privacy. 2019 IEEE 35th International Conference On Data Engineering (ICDE). pp. 638-649 (2019)

[40] Xue, Q., Zhu, Y. & Wang, J. Mean estimation over numeric data with personalized local differential privacy. Frontiers Of Computer Science. **16** pp. 1-10 (2022)

[41] Zhou, M., Wang, T., Chan, T., Fanti, G. & Shi, E. Locally differentially private sparse vector aggregation. 2022 IEEE Symposium On Security And Privacy (SP). pp. 422-439 (2022)

[42] Wei, F., Bao, E., Xiao, X., Yang, Y. & Ding, B. AAA: an Adaptive Mechanism for Locally Differential Private Mean Estimation. ArXiv Preprint ArXiv:2404.01625. (2024)

[43] Ye, Q., Hu, H., Au, M., Meng, X. & Xiao, X. LF-GDPR: A framework for estimating graph metrics with local differential privacy. IEEE Transactions On Knowledge And Data Engineering. **34**, 4905-4920 (2020)

[44] Ma, Y., Zhang, H., Cai, Y. & Yang, H. Decision tree for locally private estimation with public data. Advances In Neural Information Processing Systems. **36** (2024)

[45] Esling, P. & Agon, C. Time-series data mining. ACM Computing Surveys (CSUR). **45**, 1-34 (2012)

[46] Silva, J., Faria, E., Barros, R., Hruschka, E., Carvalho, A. & Gama, J. Data stream clustering: A survey. ACM Computing Surveys (CSUR). **46**, 1-31 (2013)

[47] Ruiz, A., Flynn, M., Large, J., Middlehurst, M. & Bagnall, A. The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Mining And Knowledge Discovery. **35**, 401-449 (2021)

[48] Dong, W., Luo, Q. & Yi, K. Continual Observation under User-level Differential Privacy. 2023 IEEE Symposium On Security And Privacy (SP). pp. 2190-2207 (2023)

[49] Henzinger, M., Upadhyay, J. & Upadhyay, S. Almost tight error bounds on differentially private continual counting. Proceedings Of The 2023 Annual ACM-SIAM Symposium On Discrete Algorithms (SODA). pp. 5003-5039 (2023)

[50] Cardoso, A. & Rogers, R. Differentially private histograms under continual observation: Streaming selection into the unknown. International Conference On Artificial Intelligence And Statistics. pp. 2397-2419 (2022)

[51] Knop, A. & Steinke, T. Counting Distinct Elements Under Person-Level Differential Privacy. 37th Conference On Neural Information Processing Systems (NeurIPS 2023). (2023)

[52] Bun, M. & Steinke, T. Concentrated differential privacy: Simplifications, extensions, and lower bounds. Theory Of Cryptography Conference. pp. 635-658 (2016)

[53] Kalemaj, I., Jain, P., Raskhodnikova, S., Sivakumar, S. & Smith, A. Counting Distinct Elements in the Turnstile Model with Differential Privacy under Continual Observation. Advances In Neural Information Processing Systems, NeurIPS 2023. (2023)

[54] Epasto, A., Mao, J., Medina, A., Mirrokni, V., Vassilvitskii, S. & Zhong, P. Differentially private continual releases of streaming frequency moment estimations. ArXiv Preprint ArXiv:2301.05605. (2023)

[55] Zhang, B., Doroshenko, V., Kairouz, P., Steinke, T., Thakurta, A., Ma, Z., Apte, H. & Spacek, J. Differentially Private Stream Processing at Scale. ArXiv Preprint ArXiv:2303.18086. (2023)

[56] Koga, T., Meehan, C. & Chaudhuri, K. Privacy amplification by subsampling in time domain. International Conference On Artificial Intelligence And Statistics. pp. 4055-4069 (2022)

[57] Fan, L. & Xiong, L. Differentially private anomaly detection with a case study on epidemic outbreak detection. 2013 IEEE 13th International Conference On Data Mining Workshops. pp. 833-840 (2013)

[58] Li, H., Xiong, L., Jiang, X. & Liu, J. Differentially private histogram publication for dynamic datasets: an adaptive sampling approach. Proceedings Of The 24th ACM International On Conference On Information And Knowledge Management. pp. 1001-1010 (2015)

[59] Wang, Q., Lu, X., Zhang, Y., Wang, Z., Qin, Z. & Ren, K. Secweb: Privacy-preserving web browsing monitoring

with w-event differential privacy. Security And Privacy In Communication Networks: 12th International Conference, SecureComm 2016, Guangzhou, China, October 10-12, 2016, Proceedings 12. pp. 454-474 (2017)

[60] Liang, W., Chen, H., Liu, R., Wu, Y. & Li, C. A pufferfish privacy mechanism for monitoring web browsing behavior under temporal correlations. Computers & Security. **92** pp. 101754 (2020)

[61] Ding, J., Ghosh, A., Sarkar, R. & Gao, J. Publishing Asynchronous Event Times with Pufferfish Privacy. 2022 18th International Conference On Distributed Computing In Sensor Systems (DCOSS). pp. 53-60 (2022)

[62] Erlingsson, Ú., Pihur, V. & Korolova, A. Rappor: Randomized aggregatable privacy-preserving ordinal response. Proceedings Of The 2014 ACM SIGSAC Conference On Computer And Communications Security. pp. 1054-1067 (2014)

[63] Ding, B., Kulkarni, J. & Yekhanin, S. Collecting telemetry data privately. Advances In Neural Information Processing Systems. **30** (2017)

[64] Xue, Q., Ye, Q., Hu, H., Zhu, Y. & Wang, J. DDRM: A continual frequency estimation mechanism with local differential privacy. IEEE Transactions On Knowledge And Data Engineering. (2022)

[65] Arcolezi, H., Pinzón, C., Palamidessi, C. & Gambs, S. Frequency estimation of evolving data under local differential privacy. ArXiv Preprint ArXiv:2210.00262. (2022)

[66] He, Y., Wang, F., Deng, X., Ni, J., Feng, J. & Liu, S. Ordinal data stream collection with condensed local differential privacy. 2022 IEEE 24th Int Conf On High Performance Computing & Communications; 8th Int Conf On Data Science & Systems; 20th Int Conf On Smart City; 8th Int Conf On Dependability In Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys). pp. 562-569 (2022)

[67] Gursoy, M., Tamersoy, A., Truex, S., Wei, W. & Liu, L. Secure and utility-aware data collection with condensed local differential privacy. IEEE Transactions On Dependable And Secure Computing. **18**, 2365-2378 (2019)

[68] Feng, S., Mohammady, M., Wang, H., Li, X., Qin, Z. & Hong, Y. DPI: Ensuring Strict Differential Privacy for Infinite Data Streaming. ArXiv Preprint ArXiv:2312.04738. (2023)

[69] Li, X., Liu, W., Lou, J., Hong, Y., Zhang, L., Qin, Z. & Ren, K. Local differentially private heavy hitter detection in data streams with bounded memory. Proceedings Of The ACM On Management Of Data. **2**, 1-27 (2024)

[70] Gu, H., Plagemann, T., Benndorf, M., Goebel, V. & Koldehofe, B. Differential Privacy for Protecting Private Patterns in Data Streams. 2023 IEEE 39th International Conference On Data Engineering Workshops (ICDEW). pp. 118-124 (2023)

[71] Bolot, J., Fawaz, N., Muthukrishnan, S., Nikolov, A. & Taft, N. Private decayed predicate sums on streams. Proceedings Of The 16th International Conference On Database Theory. pp. 284-295 (2013)

[72] Henzinger, M., Upadhyay, J. & Upadhyay, S. A unifying framework for differentially private sums under continual observation. Proceedings Of The 2024 Annual ACM-SIAM Symposium On Discrete Algorithms (SODA). pp. 995-1018 (2024)

[73] Wang, T., Chen, J., Zhang, Z., Su, D., Cheng, Y., Li, Z., Li, N. & Jha, S. Continuous release of data streams under both centralized and local differential privacy. Proceedings Of The 2021 ACM SIGSAC Conference On Computer And Communications Security. pp. 1237-1253 (2021)

[74] Ye, Q., Hu, H., Li, N., Meng, X., Zheng, H. & Yan, H. Beyond value perturbation: Local differential privacy in the temporal setting. IEEE INFOCOM 2021-IEEE Conference On Computer Communications. pp. 1-10 (2021)

[75] Ye, Q., Hu, H., Huang, K., Au, M. & Xue, Q. Stateful switch: Optimized time series release with local differential privacy. IEEE INFOCOM 2023-IEEE Conference On Computer Communications. pp. 1-10 (2023)

[76] Perrier, V., Asghar, H. & Kaafar, D. Private continual release of real-valued data streams. ArXiv Preprint ArXiv:1811.03197. (2018)

[77] Wang, Z., Liu, W., Pang, X., Ren, J., Liu, Z. & Chen, Y. Towards pattern-aware privacy-preserving real-time data collection. IEEE INFOCOM 2020-IEEE Conference On Computer Communications. pp. 109-118 (2020)

[78] Kurt, M., Yılmaz, Y., Wang, X. & Mosterman, P. Online privacy-preserving data-driven network anomaly detection. IEEE Journal On Selected Areas In Communications. **40**, 982-998 (2022)

[79] Ren, X., Shi, L., Yu, W., Yang, S., Zhao, C. & Xu, Z. LDP-IDS: Local differential privacy for infinite data streams. Proceedings Of The 2022 International Conference On Management Of Data. pp. 1064-1077 (2022)

[80] Wang, Q., Zhang, Y., Lu, X., Wang, Z., Qin, Z. & Ren, K. RescueDP: Real-time spatio-temporal crowd-sourced data publishing with differential privacy. IEEE INFOCOM 2016-The 35th Annual IEEE International Conference On Computer Communications. pp. 1-9 (2016)

[81] Zhang, J., Liang, X., Zhang, Z., He, S. & Shi, Z. Re-DPoctor: Real-time health data releasing with w-day differential

privacy. GLOBECOM 2017-2017 IEEE Global Communications Conference. pp. 1-6 (2017)

[82] Ren, X., Yu, C., Yu, W., Yang, S., Yang, X., McCann, J. & Philip, S. LoPub : high-dimensional crowdsourced data publication with local differential privacy. IEEE Transactions On Information Forensics And Security. **13**, 2151-2166 (2018)

[83] Wang, T., Yang, X., Ren, X., Yu, W. & Yang, S. Locally private high-dimensional crowdsourced data release based on copula functions. IEEE Transactions On Services Computing. **15**, 778-792 (2019)

[84] Fioretto, F. & Van Hentenryck, P. Optstream: Releasing time series privately. Journal Of Artificial Intelligence Research. **65** pp. 423-456 (2019)

[85] Zhang, X., Khalili, M. & Liu, M. Differentially private real-time release of sequential data. ACM Transactions On Privacy And Security. **26**, 1-29 (2022)

[86] Li, X., Cao, Y. & Yoshikawa, M. Locally Private Streaming Data Release with Shuffling and Subsampling. 2023 IEEE 39th International Conference On Data Engineering Workshops (ICDEW). pp. 125-131 (2023)

[87] Bao, E., Yang, Y., Xiao, X. & Ding, B. CGM: an enhanced mechanism for streaming data collection with local differential privacy. Proceedings Of The VLDB Endowment. **14**, 2258-2270 (2021)

[88] Xie, L., Lin, K., Wang, S., Wang, F. & Zhou, J. Differentially private generative adversarial network. ArXiv Preprint ArXiv:1802.06739. (2018)

[89] Jordon, J., Yoon, J. & Van Der Schaar, M. PATE-GAN: Generating synthetic data with differential privacy guarantees. International Conference On Learning Representations. (2018)

[90] He, Y., Vershynin, R. & Zhu, Y. Online Differentially Private Synthetic Data Generation. ArXiv Preprint ArXiv:2402.08012. (2024)

[91] Bun, M., Gaboardi, M., Neunhoeffer, M. & Zhang, W. Continual Release of Differentially Private Synthetic Data from Longitudinal Data Collections. Proceedings Of The ACM On Management Of Data. **2**, 1-26 (2024)

[92] Frigerio, L., Oliveira, A., Gomez, L. & Duverger, P. Differentially private generative adversarial networks for time series, continuous, and discrete open data. ICT Systems Security And Privacy Protection: 34th IFIP TC 11 International Conference, SEC 2019, Lisbon, Portugal, June 25-27, 2019, Proceedings 34. pp. 151-164 (2019)

[93] Wang, S., Rudolph, C., Nepal, S., Grobler, M. & Chen, S. PART-GAN: Privacy-preserving time-series sharing. Artificial Neural Networks And Machine Learning–ICANN 2020: 29th International Conference On Artificial Neural Networks, Bratislava, Slovakia, September 15–18, 2020, Proceedings, Part I 29. pp. 578-593 (2020)

[94] Torfi, A., Fox, E. & Reddy, C. Differentially private synthetic medical data generation using convolutional GANs. Information Sciences. **586** pp. 485-500 (2022)

[95] Lamp, J., Derdzinski, M., Hannemann, C., Linden, J., Feng, L., Wang, T. & Evans, D. GlucoSynth: Generating Differentially-Private Synthetic Glucose Traces. Advances In Neural Information Processing Systems. **36** (2024)

[96] Mao, Y., Ye, Q., Wang, Q. & Hu, H. Utility-Aware Time Series Data Release with Anomalies under TLDP. IEEE Transactions On Mobile Computing. (2023)

[97] Bordenabe, N., Chatzikokolakis, K. & Palamidessi, C. Optimal geo-indistinguishable mechanisms for location privacy. Proceedings Of The 2014 ACM SIGSAC Conference On Computer And Communications Security. pp. 251-262 (2014)

[98] Weggenmann, B. & Kerschbaum, F. Differential privacy for directional data. Proceedings Of The 2021 ACM SIGSAC Conference On Computer And Communications Security. pp. 1205-1222 (2021)

[99] Zhao, Y., Yuan, D., Du, J. & Chen, J. Geo-ellipse-indistinguishability: community-aware location privacy protection for directional distribution. IEEE Transactions On Knowledge And Data Engineering. (2022)

[100] Liang, Y. & Yi, K. Concentrated geo-privacy. Proceedings Of The 2023 ACM SIGSAC Conference On Computer And Communications Security. pp. 1934-1948 (2023)

[101] Zhao, Y. & Chen, J. Vector-indistinguishability: location dependency based privacy protection for successive location data. IEEE Transactions On Computers. (2023)

[102] Yu, L., Liu, L. & Pu, C. Dynamic Differential Location Privacy with Personalized Error Bounds.. NDSS. (2017)

[103] Cao, Y., Xiao, Y., Xiong, L. & Bai, L. PriSTE: from location privacy to spatiotemporal event privacy. 2019 IEEE 35th International Conference On Data Engineering (ICDE). pp. 1606-1609 (2019)

[104] Niu, B., Chen, Y., Wang, Z., Li, F., Wang, B. & Li, H. Eclipse: Preserving differential location privacy against long-term observation attacks. IEEE Transactions On Mobile Computing. **21**, 125-138 (2020)

[105] Qiu, C., Squicciarini, A., Pang, C., Wang, N. & Wu, B. Location privacy protection in vehicle-based spatial crowdsourcing via geo-indistinguishability. IEEE Transactions On Mobile Computing. **21**, 2436-2450 (2020)

[106] Haydari, A., Chuah, C., Zhang, M., Macfarlane, J. & Peisert, S. Differentially private map matching for mobility trajectories. Proceedings Of The 38th Annual Computer Security Applications Conference. pp. 293-303 (2022)

[107] Xiao, Y. & Xiong, L. Protecting locations with differential privacy under temporal correlations. Proceedings Of The 22nd ACM SIGSAC Conference On Computer And Communications Security. pp. 1298-1309 (2015)

[108] Liu, B., Zhu, T., Zhou, W., Wang, K., Zhou, H. & Ding, M. Protecting privacy-sensitive locations in trajectories with correlated positions. 2019 IEEE Global Communications Conference (GLOBECOM). pp. 1-6 (2019)

[109] Ma, Z., Zhang, T., Liu, X., Li, X. & Ren, K. Real-time privacy-preserving data release over vehicle trajectory. IEEE Transactions On Vehicular Technology. **68**, 8091-8102 (2019)

[110] Cao, X., Cao, Y., Pappachan, P., Nakamura, A. & Yoshikawa, M. Differentially Private Streaming Data Release Under Temporal Correlations via Post-processing. IFIP Annual Conference On Data And Applications Security And Privacy. pp. 184-200 (2023)

[111] Ahuja, R., Zeighami, S., Ghinita, G. & Shahabi, C. A Neural Approach to Spatio-Temporal Data Release with User-Level Differential Privacy. Proceedings Of The ACM On Management Of Data. **1**, 1-25 (2023)

[112] Wang, H., Hong, H., Xiong, L., Qin, Z. & Hong, Y. L-srr: Local differential privacy for location-based services with staircase randomized response. Proceedings Of The 2022 ACM SIGSAC Conference On Computer And Communications Security. pp. 2809-2823 (2022)

[113] Cunningham, T., Cormode, G., Ferhatosmanoglu, H. & Srivastava, D. Real-world trajectory sharing with local differential privacy. ArXiv Preprint ArXiv:2108.02084. (2021)

[114] Zhang, Y., Ye, Q., Chen, R., Hu, H. & Han, Q. Trajectory data collection with local differential privacy. ArXiv Preprint ArXiv:2307.09339. (2023)

[115] Gursoy, M., Liu, L., Truex, S. & Yu, L. Differentially private and utility preserving publication of trajectory data. IEEE Transactions On Mobile Computing. **18**, 2315-2329 (2018)

[116] Gursoy, M., Liu, L., Truex, S., Yu, L. & Wei, W. Utility-aware synthesis of differentially private and attack-resilient location traces. Proceedings Of The 2018 ACM SIGSAC Conference On Computer And Communications Security. pp. 196-211 (2018)

[117] Du, Y., Hu, Y., Zhang, Z., Fang, Z., Chen, L., Zheng, B. & Gao, Y. Ldptrace: Locally differentially private trajectory synthesis. Proceedings Of The VLDB Endowment. **16**, 1897-1909 (2023)

[118] Hu, Y., Du, Y., Zhang, Z., Fang, Z., Chen, L., Zheng, K. & Gao, Y. Real-Time Trajectory Synthesis with Local Differential Privacy. ArXiv Preprint ArXiv:2404.11450. (2024)

[119] Sun, X., Ye, Q., Hu, H., Wang, Y., Huang, K., Wo, T. & Xu, J. Synthesizing realistic trajectory data with differential privacy. IEEE Transactions On Intelligent Transportation Systems. (2023)