

---

# Privacy Amplification by Subsampling in Time Domain

---

Tatsuki Koga  
UC San Diego

Casey Meehan  
UC San Diego

Kamalika Chaudhuri  
UC San Diego

## Abstract

Aggregate time-series data like traffic flow and site occupancy repeatedly sample statistics from a population across time. Such data can be profoundly useful for understanding trends within a given population, but also pose a significant privacy risk, potentially revealing *e.g.*, who spends time where. Producing a private version of a time-series satisfying the standard definition of Differential Privacy (DP) is challenging due to the large influence a single participant can have on the sequence: if an individual can contribute to each time step, the amount of additive noise needed to satisfy privacy increases linearly with the number of time steps sampled. As such, if a signal spans a long duration or is oversampled, an excessive amount of noise must be added, drowning out underlying trends. However, in many applications an individual realistically *cannot* participate at every time step. When this is the case, we observe that the influence of a single participant (sensitivity) can be reduced by subsampling and/or filtering in time, while still meeting privacy requirements. Using a novel analysis, we show this significant reduction in sensitivity and propose a corresponding class of privacy mechanisms. We demonstrate the utility benefits of these techniques empirically with real-world and synthetic time-series data.

## 1 INTRODUCTION

Time-series data about people’s behavior can be sensitive, yet may contain information that is highly beneficial for society. As a concrete example, consider

highway traffic data that is released by the California Department of Transportation <sup>1</sup> (Chen et al., 2001). The data contains aggregate statistics (number of cars, average speed, etc.) every five minutes gathered by physical sensors. Just like any other form of sensitive, aggregate data, this traffic data could expose the behavior of a single participant. In an extreme case, if a participant is the *only* driver on a highway (say, late at night), then an adversary could track their movement through the entire road network. However, such data is absolutely vital for society to manage critical road infrastructure and strategize on where to augment it. To this end, we aim to provide a privacy mechanism that releases a sanitized version of a given time-series while preserving its aggregate trends through time.

We adopt differential privacy (DP) (Dwork et al., 2006b), one of the most widely used privacy definitions, as our privacy measure. For data with bounded sensitivity, a first plausible DP mechanism is to add Gaussian (Dwork and Roth, 2014) or Laplace (Dwork et al., 2006b) noise proportional to the sensitivity. Rastogi and Nath (2010) propose a second algorithm that adds noise to the first  $k$  discrete Fourier transform (DFT) coefficients and ignores higher frequencies of the signal. In both cases, the problem is that the sensitivity scales with  $T$ , the number of time steps. If an individual can theoretically participate at each timestep, then the worst-case influence they can have on the time-series grows linearly with  $T$ . Thus, as  $T$  becomes larger, the requisite noise variance increases, and the resulting output loses utility. In this paper, we take a new approach to addressing this longstanding problem using subsampling and filtering techniques from the signal processing domain. We find that random subsampling along with filtering can reduce sensitivity without significant losses of utility.

Our starting point is to observe two key properties of real time-series data, which we exploit to design better mechanisms that add significantly less noise for a given  $T$ . The first is that time-series data is often oversampled. Returning to our example, a transit analyst may be interested in observing how traf-

---

Proceedings of the 25<sup>th</sup> International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151. Copyright 2022 by the author(s).

<sup>1</sup><https://pems.dot.ca.gov/>

fic changes on an hourly basis. This being the case, the five-minute sampling period of the sensor system is excessive and contributes very little to their utility. Indeed, with standard mechanisms, this excessive sampling rate may *damage* their utility since it increases the number of time steps,  $T$ , many times beyond what is needed. The second property is that a single person *almost never* contributes to the aggregate statistics at every time step. It is implausible that an individual passes the same sensor on a highway every five minutes for more than a year.

We formalize these two observations into mathematical assumptions, and exploit them as follows. The first property motivates our use of subsampling, which trades these unnecessary extra time steps in exchange for reduced additive noise. The second property enables us to significantly boost this reduction in additive noise: we know that, feasibly, an individual can only participate in a limited number of time steps to begin with. This allows us to show that w.h.p. they will contribute to an even smaller number of timesteps *after subsampling*, thus allowing us to significantly scale down noise added.

We additionally extend our results to a class of mechanisms that *filter* and subsample the time-series. For the traffic analyst, the process of subsampling may be high variance if traffic changes significantly during an hour. In essence, which sample is chosen during that hour will have a large effect on the shape of the resulting time series. To mitigate this the analyst may wish to average the time series in time (filter) before subsampling. Unfortunately, this significantly complicates the privacy analysis used on subsampling alone. To address this we employ a novel concentration analysis that accounts for the effects of subsampling and filtering simultaneously.

This paper investigates how and in what circumstances subsampling and filtering enable us to achieve a better privacy-utility tradeoff for publishing aggregate time-series data under differential privacy. Ultimately, by exploiting the two reasonable assumptions listed above, we make nontrivial gains in utility while maintaining a strong  $(\epsilon, \delta)$ -DP guarantee. Using a novel concentration analysis, we propose a class of filter/subsample mechanisms for publishing sanitized time-series data. Our experiments indicate that—by using significantly less additive noise than baseline methods—our mechanisms achieve a better privacy-utility tradeoff in general for oversampled time-series.

## 1.1 Related Work

With regards to subsampling and Differential Privacy, prior work has shown that subsampling individuals

(rows) helps amplify privacy (Kasiviswanathan et al., 2011; Beimel et al., 2014, 2013; Bassily et al., 2014; Abadi et al., 2016; Balle et al., 2018); our setting is different from these results in that we consider subsampling time steps which are equivalent to columns or features, which has not been considered in prior work.

The most closely related work in terms of publishing time-series data privately is by Rastogi and Nath (2010). They propose an  $\epsilon$ -DP algorithm that perturbs only the first  $k$  discrete Fourier transform (DFT) coefficients with the Laplace mechanism and zeros the remaining high frequency coefficients, since adding noise to those results in the sanitized time-series being scattered. However, the sensitivity still scales with  $T$  and the algorithm perturbs the first  $k$  coefficient a lot if  $T$  is large. In comparison, we provide an analysis using subsampling that introduces better scaling with  $T$ . Fan and Xiong (2012) utilize the Kalman filter to correct the noisy signal by the Laplace mechanism and adopt adaptive sampling to reduce the published time steps, thereby reducing the number of compositions. Their main focus is real-time data publishing, where we focus on offline data publishing.

Closer to our line of study is differentially private continual observation (Dwork et al., 2010; Chan et al., 2011), which proposes a method to repeatedly publish aggregates (in particular, count) under the continual observation of the data signal. The main difference from our work is that they protect event-level privacy; namely, they try to hide the event that an individual participates in the time-series at a *single* time step by perturbation. We hide participation in the *entire* time-series, assuming that an individual can participate at multiple time steps.

As for DP under filtering, Ny and Pappas (2014) investigate the DP mechanisms for general dynamic systems. The work relates filters to an input with the sensitivity, but it does not address the issue that the sensitivity depends linearly on  $T$ .

More broadly, some work proposes new privacy definitions suitable for traffic (more generally, spatio-temporal) data (Xiao and Xiong, 2015; Cao et al., 2017, 2019; Meehan and Chaudhuri, 2021). Though these definitions are more appropriate for some situations, we believe the classic DP definition is the right choice for our setting, where there is a trusted central authority.

## 2 PRELIMINARIES AND PROBLEM SETTING

Among aggregate time-series data, we focus on a **count** data throughout this paper. More formally, we consider time-series signal  $x \in \mathbb{N}^T$  of length  $T$ , where  $x_t$  corresponds to the count of individuals' participation at time  $t$ . The count signal  $x$  is an output of a function  $\text{count} : \mathcal{X} \rightarrow \mathbb{N}^T$ , where  $\mathcal{X}$  is the input space. We aim to publish the randomized version of  $x$ , denoting as  $z$ , with a privacy guarantee.

The appropriate privacy notion for our setting is differential privacy. We say two datasets  $D, D' \in \mathcal{X}$  are neighboring if they differ on at most an individual's participation.

**Definition 1** ( $(\epsilon, \delta)$ -DP (Dwork et al., 2006a)). *A randomized algorithm  $M$  satisfies  $(\epsilon, \delta)$ -DP if for any neighboring datasets  $D, D'$  and for any  $S \subseteq \text{range}(M)$ ,*

$$\Pr\{M(D) \in S\} \leq \exp(\epsilon) \Pr\{M(D') \in S\} + \delta.$$

One of the most famous DP algorithms is the Gaussian mechanism, which adds the Gaussian noise with a variance proportional to the  $L_2$ -sensitivity (Dwork et al., 2006b). The  $L_2$ -sensitivity of a function  $f$  is the maximum difference between the outputs of  $f$  from any neighboring datasets in terms of  $\ell_2$ -norm.

**Definition 2.** *The  $L_2$ -sensitivity of a function  $f : \mathcal{X} \rightarrow \mathbb{R}^d$  is*

$$\Delta_{2,f} = \max_{D, D'} \|f(D) - f(D')\|_2,$$

where  $D$  and  $D'$  are neighboring datasets and  $\|\cdot\|_2$  is  $\ell_2$ -norm.

**Theorem 1** (Gaussian mechanism (Dwork and Roth, 2014)). *Let  $\epsilon \in (0, 1)$  be arbitrary and  $f : \mathcal{X} \rightarrow \mathbb{R}^d$  be a function. Then, for  $c^2 > 2 \ln(1.25/\delta)$ , the algorithm  $M$ :*

$$M(D) = f(D) + (\xi_1, \dots, \xi_d)^\top$$

satisfies  $(\epsilon, \delta)$ -DP, where  $\xi_i$ 's are drawn i.i.d. from  $\mathcal{N}(0, \sigma^2)$  and  $\sigma \geq \frac{c\Delta_{2,f}}{\epsilon}$ .

As we see from the Gaussian mechanism, the larger  $L_2$ -sensitivity leads to the higher noise variance. Thus, we consider filtering and subsampling the input signal to reduce the sensitivity.

Filtering the signal is a well-used technique in signal processing. In particular, low-pass filters attenuate the effect of high frequencies and make the signal smoother. Let  $h \in \mathbb{R}^T$  be a filtering vector and  $y \in \mathbb{R}^T$  be a filtered signal. We express the filtering operation as  $y_t = \sum_{k=0}^{T-1} x_k h_{(t-k) \bmod T}$  for all  $t \in [T]$ . With a

matrix  $A \in \mathbb{R}^{T \times T}$ , this can be seen as  $y_t = (Ax)_t = \sum_{k=0}^{T-1} a_{tk} x_k$ , where  $(A)_{ij} = a_{ij} = h_{(i-j) \bmod T}$ , for all  $t \in [T]$ . Note that a matrix  $A$  is a circular matrix. From now on, we represent the filter with  $A$  and the filtered signal with  $y = Ax$ . We further assume that the  $\ell_1$ -norm of each row is normalized to 1, i.e.,  $\|A_t\|_1 = 1$  for all  $t \in [T]$ , where  $A_t$  is  $t$ -th row of a matrix  $A$ .

As opposed to *row-wise* subsampling in the past DP literature that can completely discard an individual's participation in a dataset with some probability, we contemplate *column-wise* subsampling. Formally, we define the subsampling operator as  $\text{Subsample} : \mathbb{R}^T \times \mathcal{J} \rightarrow \mathbb{R}^{T'}$ , where  $\mathcal{J} = \{j \subseteq [T]\}$  is the power set of  $[T]$  and  $T'$  is a length of subsampled signal. We define  $J \in \mathcal{J}$  as the random variable of subsampling indices. We consider the Poisson sampling with parameter  $p$  for  $J$ , i.e., for all  $t \in [T]$ ,  $t \in J$  with probability  $p$  independently. Therefore, we express the subsampling to  $x$  with  $j$  as  $\hat{x} = \text{Subsample}(x, j)$ , where  $\hat{x} \in \mathbb{R}^{|j|}$ .

## 3 PRIVACY AMPLIFICATION BY SUBSAMPLING

Recalling the example of the California highway traffic data (Chen et al., 2001), the number of cars passed by the physical sensor is sampled once every five minutes. The sensitivity grows exceedingly large with the number of sampled time steps,  $T$ ; thus, the noise added to satisfy DP becomes too large to convey even daily or weekly trends for baseline mechanisms. However, it is extremely unlikely that a single individual participates at every sensor snapshot in a day. Therefore, by making the mild assumption that an individual participates no more than at  $I < T$  time steps in the signal, we can significantly reduce the sensitivity. Without subsampling, this trivially reduces the ( $L_2$ -) sensitivity by a factor of  $\sqrt{I/T}$ . In this section, we demonstrate how we can reduce the sensitivity even further with high probability through subsampling and propose our  $(\epsilon, \delta)$ -DP algorithm that utilizes subsampling.

### 3.1 Assumption

To make the above assumption precise, we formalize it as follows.

Denote  $i$ -th individual's participation to the dataset  $D$  as  $d_i \in \mathbb{R}^T$ . Thus,  $x_t = \text{count}(D)_t = \sum_i d_{i,t}$ . Without loss of generality, we assume  $d_{i,t} \in \{0, 1\}$  for all  $i$  and  $t$ , i.e., an individual can participate at most once at each time step. It is straightforward to extend our results to the case where participation is at most  $M > 1$  times. The formal statement for limited individual's participation assumption is as follows.

Table 1: Maximum number of check-ins per individual ( $I$ ) for 15 most visited venues in Gowalla (left) and Foursquare (right) datasets.

ID	$I$ ( $T = 725$ )	ID	$I$ ( $T = 2126$ )
1	63	1	208
2	35	2	127
3	39	3	98
4	29	4	225
5	51	5	265
6	39	6	121
7	43	7	191
8	19	8	78
9	27	9	94
10	208	10	182
11	27	11	170
12	34	12	206
13	45	13	94
14	47	14	82
15	75	15	350

**Assumption 1.**  $\sup_i \sum_t d_{i,t} = I$ , i.e., an individual can participate at most  $I$  time steps.

Assumption 1 formalizes the observation that no individuals participate at all time steps, especially when the signal is oversampled. In most cases, the maximum time steps that an individual can participate at are smaller than  $T$ ; thus,  $I < T$ . This assumption is empirically evident in real time-series data. Table 1 shows the maximum number of check-ins per individual for the top 15 most visited venues for Gowalla (Cho et al., 2011) and Foursquare (Yang et al., 2015, 2016) datasets. Similar to our highway example, it is implausible that an individual checks into a certain venue every time step. We see that  $I$  is, at least empirically, much smaller than  $T$ . Since we do not know the exact value of the maximum individual’s participation a priori, we must set  $I$  conservatively in practice. However, even if the assumption does not hold strictly, our guarantee is still valid with gentle privacy degradation as discussed in Section 3.3.1. This assumption yields that  $\Delta_{2,\text{count}} = \sqrt{I}$ .

Although the assumption reduces the  $L_2$ -sensitivity from  $\sqrt{T}$  to  $\sqrt{I}$ , we will show that random subsampling reduces it further in the following sections.

### 3.2 Sensitivity Reduction by Subsampling

With the assumption in mind, the naive approach to achieve DP is to apply the Gaussian mechanism with the sensitivity parameter  $\Delta_{2,\text{count}} = \sqrt{I}$ . The dependency on  $\sqrt{I}$  is undesirable as  $I$ , which tends to be proportional to  $T$ , becomes larger. Therefore, we propose to apply random subsampling to relax this depen-

ency. The intuition behind it is that the probability that subsampled time steps contain all (or most) of an individual’s participation is very low under Assumption 1; thus, we get the further dependency reduction. Formally, we show that Subsample helps reduce the  $L_2$ -sensitivity with high probability, where the probability is over the choice of  $J$ , under Assumption 1.

Throughout the analysis, we assume that  $D = D' \cup \{d_i\}$ , i.e.,  $i$ -th individual is the only one who participates in  $D$  and does not in  $D'$ . Also, we let  $j_i = \{t | d_{i,t} = 1\}$ , time indices that  $i$ -th individual participates.

#### 3.2.1 Sensitivity Reduction by Subsampling Only

We start from examining the sensitivity of  $f_s(\cdot, \cdot) = \text{Subsample}(\text{count}(\cdot), \cdot)$ . Intuitively, under Assumption 1, the probability that  $J$  contains almost all  $I$  time steps is small; thus, we obtain the sensitivity reduction with high probability.

**Theorem 2.** Fix  $I' \leq I$ . Let  $\delta' = \Pr\{|J \cap j_i| > I'\} = \sum_{t=I'+1}^I \binom{I}{t} p^t (1-p)^{I-t}$ . Then, with probability  $1 - \delta'$ ,  $\Delta_{2,f_s} = \sqrt{I'}$ .

The proof follows directly from the tail probability of the binomial distribution with parameters  $I$  and  $p$ . By using the tail bound of the binomial distribution, we can explicitly express the upper bound of  $I'$  for fixed  $\delta'$  with  $I$  and  $p$ .

**Corollary 1.** Let  $\delta' < 1$ . Then, with probability  $1 - \delta'$ ,  $\Delta_{2,f_s} = \sqrt{I'} \leq \sqrt{Ip + \sqrt{\frac{I}{2} \log 1/\delta'}}$ .

The claim is a direct consequence of the Hoeffding’s inequality. We see with high probability that the dependency on  $I$  is replaced by  $Ip$  plus the probably small additional term.

#### 3.2.2 Sensitivity Reduction by Filtering and Subsampling

Next, we consider the case where we filter the signal before subsampling. Some filters have an averaging effect across time steps; thus, filtered signals become smoother and less noisy in general.

Again, we go back to our running example of the CA highway traffic data. The signal is very oversampled for an analyst who is only interested in how traffic trends change every two hours. Changes on the order of five minutes do not contribute to their utility, but magnifies the sensitivity significantly. By filtering, we effectively average in time to smooth the signal, and ultimately reduce the variance after subsampling. As detailed in Section 2 we represent a filter as a matrix

$A$ , which, after multiplying with the signal, produces the smoother signal.

However, from the DP point of view, filtering spreads the content of  $t$ -th time step across all time steps; thus, the effects of subsampling are now less obvious. As an extreme example, consider a filter that simply averages across all time points, i.e.,  $A_{ij} = 1/T$  for any  $i$  and  $j$ . Then, any individual's contribution scatters equally along all time steps. Thus, the intuition above does not hold here, but still, we can apply a similar way of thinking if  $A$  is somewhat similar to the identity matrix—there exist time steps that  $i$ -th individual's influence is large, and the probability of choosing almost all such time steps is low.

We define the function as  $f_{\text{fs}}(\cdot, \cdot) = \text{Subsample}(A \text{ count}(\cdot, \cdot))$ . By letting  $\delta_i = 1$  if  $i \in J$  and 0 otherwise and  $B = \text{diag}(\delta_1, \dots, \delta_T)A$ , we can express  $f_{\text{fs}}(x, J) = Bx$ . Fix two signals  $x, x'$  induced by neighboring datasets. Then, it holds that

$$\begin{aligned} \Delta_{2, f_{\text{fs}}}^2 &= \|Bx - Bx'\|_2^2 \\ &= (x - x')^\top B^\top B(x - x') \\ &\leq \lambda_{\max}(B^\top B) \|x - x'\|_2^2 \leq \sigma_{\max}^2(B)I, \end{aligned}$$

where the last inequality follows from Assumption 1. It remains to bound the maximum singular value of  $B$  to complete bounding the sensitivity.

To bound how large the maximum singular value is likely to be, we use the following simplified version of the theorem by Tropp (2015).

**Theorem 3** (Simplified version, (Tropp, 2015)). *For  $\varepsilon \geq 0$ ,*

$$\begin{aligned} &\mathbb{P} \{ \sigma_{\max}^2(B) \geq (1 + \varepsilon) p \sigma_{\max}^2(A) \} \\ &\leq 2 \text{srnk}(A) \left( \frac{e^\varepsilon}{(1 + \varepsilon)^{1 + \varepsilon}} \right)^{p \sigma_{\max}^2(A)/L}, \end{aligned}$$

where  $\text{srnk}(A) = \frac{\|A\|_F^2}{\|A\|_2^2}$  is the stable rank of  $A$ , the ratio between the squared Frobenius norm of  $A$  and the squared spectral norm of  $A$ , and  $L = \max_t \|A_t\|_2^2$ .

As we assume in Section 2 that  $A$  is a circular matrix and  $\|A_t\|_1 = 1$  for all  $t \in [T]$ ,  $\sigma_{\max}(A) = 1$  and  $L = \|A_1\|_2^2$ . Combining this theorem with the sensitivity bound with the maximum singular value, we are able to bound the reduction in sensitivity from *both* subsampling *and* filtering simultaneously. Our final sensitivity reduction theorem for filtered and subsampled signals is given below.

**Theorem 4.** *Fix  $\alpha \in (0, 1]$ . Let  $\delta' = 2 \text{srnk}(A) \left( \frac{e^{\alpha^2/p} - 1}{(\alpha^2/p)^{\alpha^2/p}} \right)^{p/L}$ . Then, with probability at least  $1 - \delta'$ ,  $\Delta_{2, f_{\text{fs}}} = \alpha\sqrt{T}$ .*

---

**Algorithm 1** Our  $(\epsilon, \delta)$ -DP Algorithm

---

**Input:** count signal  $x \in \mathbb{R}^T$ , filter  $A \in \mathbb{R}^{T \times T}$ , parameters  $p, \alpha, \epsilon, \delta$

**Output:**  $\hat{z} \in \mathbb{R}^T$

$y = Ax$

Sample  $J$  uniform randomly with a parameter  $p$ .

$y_s = \text{Subsample}(y, J)$

$z = y_s + (\xi_1, \dots, \xi_{|J|})^\top$ , where  $\xi_i \sim \mathcal{N}\left(0, \frac{2 \ln(1.25/\delta)}{(\epsilon/(\alpha\sqrt{T}))^2}\right)$

Interpolate  $z$  to get  $\hat{z}$  with  $J$ .

---

Note that the theorem above does not fully make use of Assumption 1; thus, the statement holds even if we replace  $I$  with  $T$ . Also, we note that by taking  $A$  to be an identity matrix, the function  $f_{\text{fs}}$  is exactly the same as  $f_s$ , but the bound is looser than Theorem 2 in general. Furthermore, it is hard to obtain the explicit upper bound for  $\alpha$  in general, but we give the relationship between  $\alpha$  and  $\delta'$  in Supplementary Material C.1.

### 3.3 Algorithm

Algorithm 1 gives our  $(\epsilon, \delta)$ -DP algorithm for a filtered and subsampled count signal. The central idea is to utilize the sensitivity analysis in the previous section and add less variance noise than the standard Gaussian mechanism. One note is that the interpolation in the last line of Algorithm 1 does not incur a privacy breach by using  $J$  as long as the interpolation is bijective given  $J$ , which is usually the case.

From Theorems 2 and 4, we see that Algorithm 1 satisfies  $(\epsilon, \delta)$ -DP. In short, our privacy guarantee allows smaller  $\epsilon$  in exchange for a little larger  $\delta$  under the same noise variance as the standard Gaussian mechanism.

**Corollary 2.** *Let  $\delta'$  be the one in Theorem 4. Algorithm 1 satisfies  $(\epsilon, \delta + \delta'(\exp(\epsilon/\alpha) - \exp(\epsilon)))$ -DP.*

The corollary follows directly from Theorem 4. We provide the detailed proof in Supplementary Material A.4.

The guarantee implicitly requires  $\delta'$  to be small enough. In other words, if  $\delta'(\exp(\epsilon/\alpha) - \exp(\epsilon))$  is smaller or equal to  $\delta$ , the compensation for the  $\delta$  term is negligible and we obtain the better privacy. Also,  $\delta'$  is a monotonously decreasing function with respect to  $\alpha$ ; thus, one can easily find appropriate  $\alpha$  to satisfy desired  $\delta'$  numerically.

When  $A$  is an identity matrix, i.e., not filtering before subsampling, we have a better privacy guarantee.

**Corollary 3.** *Let  $A$  be an identity matrix,  $I' = \lceil \alpha^2 I \rceil$  and  $\delta'$  be the one in Theorem 2. Algorithm 1 satisfies*

$(\epsilon, \delta + \delta'(\exp(\sqrt{I/I'}\epsilon) - \exp(\epsilon)))-DP$ .

The proof is almost identical to the one for Corollary 2. Note that the above guarantee heavily relies on Assumption 1 as opposed to Corollary 2.

### 3.3.1 Graceful Privacy Degradation When Assumption 1 Fails

Although we succeed in amplifying privacy by subsampling, one would cast doubt on the validity of Assumption 1, i.e., any individual participates at most  $I$  time steps. Here, we show even if the assumption does not hold precisely, our algorithm still satisfies DP and the privacy level does not degrade abruptly.

**Proposition 1.** *If  $\sup_i \sum_t d_{i,t} = cI$  for some  $c > 1$ , then Algorithm 1 satisfies  $(\sqrt{c}\epsilon, \delta + \delta'(\exp(\sqrt{c}\epsilon/\alpha) - \exp(\sqrt{c}\epsilon)))-DP$ .*

In short, the above proposition claims that if the *actual* participation limit is  $cI$  instead of  $I$ , privacy parameters increase only slightly; in particular,  $\epsilon$  only increases by a factor of  $\sqrt{c}$ .

## 4 EXPERIMENTS

We next investigate how well our algorithm works in practice by empirically evaluating it on real and synthetic data <sup>2</sup>. Specifically, we ask the following questions:

1. How does our algorithm compare with existing baselines such as the Gaussian mechanism and DFT in terms of accuracy?
2. How does filtering before subsampling improve accuracy?
3. How does the sampling frequency of the time-series data affect accuracy?

We answer the first question by experimenting with three real datasets in Section 4.2. We then answer the rest of the questions with a synthetic dataset in Section 4.3.

### 4.1 Setup

**Datasets.** We consider four datasets – three real, and one synthetic.

**PeMS** ( $T = 1800$ ) (Chen et al., 2001) is a dataset collected by California Transportation Agencies (CalTrans) Performance Measurement System (PeMS). We

select a sensor from a highway in the Bay area, California, and sample the flow data, the number of cars that passed the sensor, every five minutes from Jan. 2017. **Gowalla** ( $T = 725$ ) (Cho et al., 2011) is a dataset of check-in information from a location-based social networking website called Gowalla. We use the most visited location and aggregate the check-in count every 24 hours from Feb. 2009. **Foursquare** ( $T = 2126$ ) (Yang et al., 2015, 2016) is a similar dataset with check-in information from the location data platform Foursquare. We pick the most checked-in place and aggregate the count every 6 hours from April 2012. Since raw Gowalla and Foursquare datasets contain user IDs, we use them to calculate the empirical  $I$  in Table 1. We set  $I = \lceil T/10 \rceil$  as confirmed by Table 1.

Since we do not have the data generating mechanisms for real data, to understand how the sampling frequency affects the accuracy and the conditions under which filtering helps, we also experiment with a synthetic dataset **Synth**. We generate the signal value given time step  $t$  as  $x_t = a \sin(\omega t) + b + ct$ , where we set  $a = 200$ ,  $b = 500$  and  $c = 0.1$ . The resulting signal has two overall trends: periodicity and linear growth. We expect a good algorithm to return a randomized version that maintains *both* trends. We control the sampling frequency to examine how the extent of oversampling affects the utility. We do this by, given (relative) sampling frequency  $f$ , generating the signal  $x^{(f)} = (x_0, x_{1/f}, x_{2/f}, \dots)$ . Thus, as the relative frequency increases, the signal is oversampled more and the length gets larger. We set  $T = 10000$  for  $f = 1$  (thus,  $T = 10000 \cdot f$  given  $f$ ) and  $I = 100$  for any  $f$ . We generate several signals with varying  $f$ 's ( $f \in \{1, 1/2, 1/4, 1/8, 1/16, 1/32, 1/64\}$ ). We provide the reason why we fix  $I$  for the Synth data in Supplementary Material B.2. Furthermore, we consider the case where there is observation noise: unbiased randomness added to each time step that is typical in real-world time-series. To do so, we add mean-zero Gaussian noise to each time step, yielding a more realistic, noisy version of the signal:  $\tilde{x}_t = x_t + d\hat{\sigma}_t$ , where  $\hat{\sigma}_t$ 's are drawn i.i.d. from  $\mathcal{N}(0, 1)$  and  $d = 100$ .

**Algorithms.** The simplest baseline for our problem is the standard Gaussian mechanism. The Gaussian mechanism for time-series data adds noise corresponding to the sensitivity  $\sqrt{I}$  to each of  $T$  time steps by considering time-series data as a  $T$ -dimensional vector. The more involved baseline is the algorithm by Rastogi and Nath (2010) (DFT). It adds noise to the first  $k$  discrete Fourier transform coefficients and ignores higher frequency contents of the signal. For DFT, we use the hyperparameters reported as being the best in their paper. While DFT is originally an  $\epsilon$ -DP algorithm, we extend it to be  $(\epsilon, \delta)$ -DP by changing the

<sup>2</sup>All codes to reproduce the results are available at <https://github.com/tacchan7412/Privacy-Amplification-by-Subsampling-in-Time-Domain>

Table 2: Mean MAEs and standard deviations of Gaussian mechanism, DFT and our algorithms (with and without filter) on PeMS, Gowalla, and Foursquare dataset for  $(\epsilon, \delta) = (0.5, 10^{-4})$ . We see that our algorithm without filtering performs the best on these real world datasets. We explore when filtering is optimal on the Synth dataset.

	PeMS	Gowalla	Foursquare
Gaussian	$93.0 \pm 1.6$	$33.0 \pm 1.5$	$57.4 \pm 1.5$
DFT	$55.7 \pm 2.8$	$39.3 \pm 3.5$	$42.9 \pm 3.4$
<b>Ours w/o Filter</b>	$42.8 \pm 2.9$	$16.7 \pm 2.7$	$29.5 \pm 2.4$
<b>Ours w/ Filter</b>	$60.9 \pm 4.0$	$21.2 \pm 3.9$	$39.6 \pm 3.6$

noise distribution to Gaussian from Laplace for direct comparison with our algorithm. For our algorithm, we consider two variations—with and without filtering.

**Experiment Setup.** We set the privacy parameters and  $\alpha$  so that each mechanism guarantees  $(\epsilon, \delta) = (0.5, 10^{-4})$ -DP throughout the experiments. Accuracy is measured by the mean absolute error (MAE) between the raw signal and the output. For the noisy Synth data, we measure the MAE between the true signal (before adding observation noise) and the output. For each time-series, we repeat the algorithms 1000 times and report the mean and standard deviation of MAE; this is because the outputs depend heavily on the randomness of the mechanisms. We choose a subsampling parameter for our algorithm  $p$  to be 0.1, i.e., the subsampled signal length would be around  $T/10$  for the real datasets. For the synthetic signal,  $p = 0.1 \cdot (1 - \log(f)/4)$ . We present further details of the experiments in Supplementary Material B.

## 4.2 Results on Real Data

Table 2 summarizes the MAEs on the real datasets. We see that for all the datasets, our algorithm without filtering performs the best, indicating the superiority of this algorithm. Ours with filtering performs better than the baselines on the Gowalla and Foursquare dataset, but slightly worse than DFT on the PeMS dataset. This might be due to the periodic nature of the PeMS signal—DFT is suitable for periodic signals in general. Although the overall performance of our algorithms is better than the baselines, it seems that the variance of MAE is slightly larger than the standard Gaussian mechanism. We anticipate that the main reason is the interpolation in Algorithm 1.

## 4.3 Results on Synthetic Data

To better understand when filtering works well, we compare the results with and without observation noise. We anticipate that the averaging effect of filtering offers better performance when observation noise is present. Figure 1 shows in orange examples of outputs of four mechanisms: the two baseline mechanisms

and our mechanism with and without filtering, each operating on the Synth time-series with observation noise. For reference the Synth signal is plotted in blue (without obs. noise for clarity). We show the outputs operating on the *noiseless* Synth signal in Supplementary Material B.3. Figures 2a and 2b show the MAEs for Synth dataset with and without observation noise while sweeping relative sampling frequency  $f$ . We see that our mechanism with filtering generally outperforms the one without filtering when there is observation noise, confirming our hypothesis. Filtering before subsampling has an averaging effect in time, canceling out the observation noise and thus accurately conveying the underlying trends. We can observe this effect directly in Figure 1. The output from our algorithm with filtering concentrates around the raw Synth signal the most, yielding a lower MAE and better capturing underlying trends. On the other hand, when the signal is noiseless, filtering does not help achieve better utility. As seen in Section 3.2, the sensitivity reduction is larger for subsampling without filtering; thus, ours without filtering is more helpful for less noisy signals.

To understand the impact of oversampling, i.e., how sampling frequency matters to utility, we observe the results under several (relative) sampling frequencies ( $f$ ). The larger  $f$  is, the more oversampled the signal is. For both Figures 2a and 2b, we see a similar trend that MAEs get smaller as  $f$  becomes larger for our algorithms (with and without filtering). This trend validates the hypothesis that almost no information gets lost by subsampling for oversampled signals. In such a case, we gain utility due to the reduced sensitivity achieved by subsampling. By contrast, for less frequently sampled, fast-moving signals, our algorithms perform worse than the standard Gaussian mechanism. We expect that this is because the error caused by the interpolation dominates the error by adding noise. The linear interpolation does not supplement the information lost by subsampling. Note that MAEs for the Gaussian mechanism and DFT are almost constant because, for experiments on Synth, we fix  $I$  for all  $f$ 's.

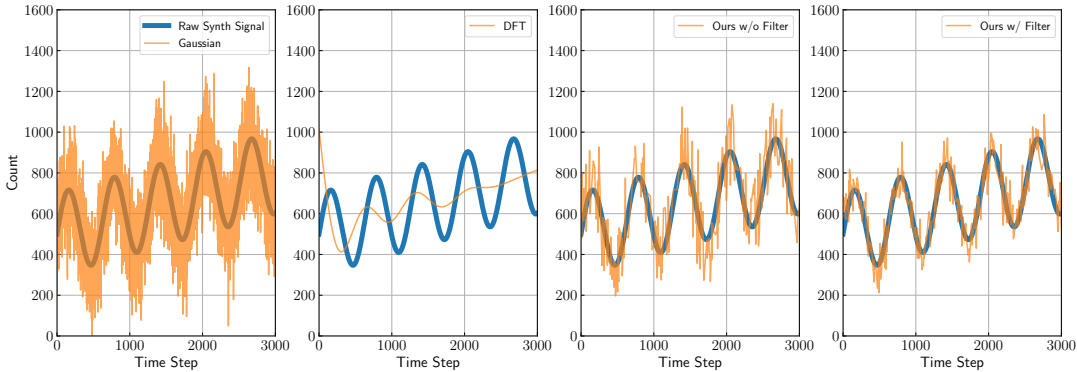


Figure 1: Two baseline mechanisms (left) and our mechanism with and without filtering (right), each operating on Synth time-series with observation noise. We plot the noiseless Synth time-series in blue for reference. Output by our algorithm with filtering concentrates around raw signal the most, demonstrating value of our filter + subsample algorithm when operating on noisy time-series.

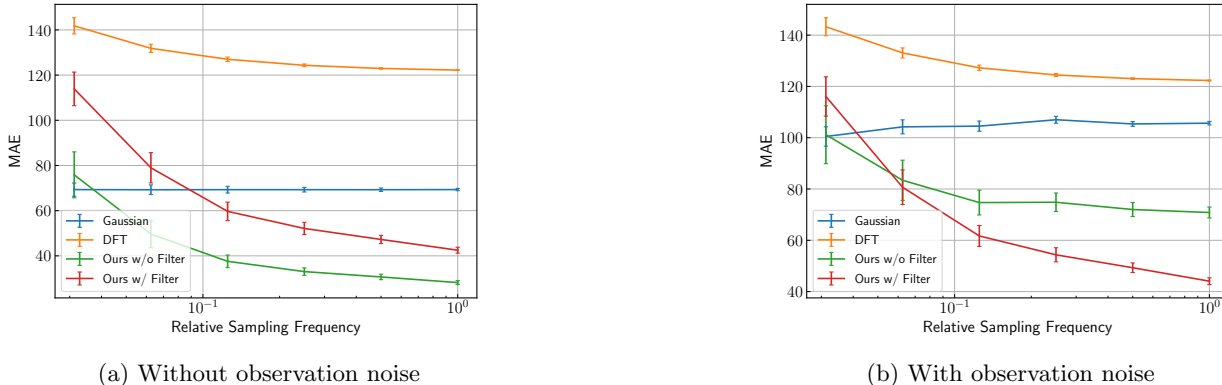


Figure 2: Mean MAEs (smaller is better) and standard deviations of Gaussian mechanism, DFT, and our algorithms (with and without filter) on Synth data under varying relative sampling frequencies for  $(\epsilon, \delta) = (0.5, 10^{-4})$ . Signal is sampled more frequently, i.e., more oversampled, as relative sampling frequency gets larger. We see that (1) our algorithms (with and without filter) perform better as signal gets more oversampled, and (2) our algorithm with filtering performs best when the input signal is noisy (right).

4.4 Discussion

Throughout the experiments, we have made the three questions clear.

Our experimental results on real data reveal that our algorithm, especially without filtering, is superior to the baseline mechanisms, the Gaussian mechanism and DFT, in terms of MAE. This is mainly because the sensitivity reduction by subsampling reduces the noise variance and the error incurred by the interpolation is small due to the oversampled nature of real data.

Filtering improves utility when signals contain observation noise as confirmed by the result on the Synth data. It is helpful when we want to capture underlying trends rather than to reproduce original signals because applying a filter has an averaging effect and

thus lessens the effect of observation noise. On the contrary, for noiseless signals, we should use our algorithm without filtering since the sensitivity reduction is larger for subsampling but not filtering in general.

The result on the Synth data also indicates that our algorithm works better as data gets oversampled. This result verifies our hypothesis that subsampling preserves overall trends for oversampled signals. The sensitivity reduction is also significant since our algorithm outperforms the baseline mechanisms for such signals.

5 CONCLUSION AND FUTURE WORK

This paper investigates how subsampling and filtering help achieve a better privacy-utility tradeoff for



the longstanding problem of publishing sanitized aggregate time-series data with a DP guarantee. Using a novel concentration analysis, we show subsampling and filtering reduce the sensitivity with high probability under reasonable assumptions for real time-series data. We then propose a class of DP mechanisms exploiting subsampling and/or filtering, which empirically outperform baselines. Going forward, we plan to develop theoretical utility guarantees for our mechanisms, and explore how alternative interpolation schemes impact utility. Finally, we believe feature-wise subsampling is beneficial beyond time-series data, and are investigating other applications for our techniques.

### Acknowledgements

TK, CM, and KC would like to thank ONR under N00014-20-1-2334 and UC Lab Fees under LFR 18-548554 for research support. Also, TK is supported in part by Funai Overseas Fellowship. We would also like to thank our reviewers for their insightful feedback.

### References

- Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. (2016). Deep Learning with Differential Privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, CCS '16*, pages 308–318, New York, NY, USA. Association for Computing Machinery.
- Balle, B., Barthe, G., and Gaboardi, M. (2018). Privacy Amplification by Subsampling: Tight Analyses via Couplings and Divergences. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.
- Bassily, R., Smith, A., and Thakurta, A. (2014). Private Empirical Risk Minimization: Efficient Algorithms and Tight Error Bounds. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 464–473.
- Beimel, A., Brenner, H., Kasiviswanathan, S. P., and Nissim, K. (2014). Bounds on the sample complexity for private learning and private data release. *Mach Learn*, 94(3):401–437.
- Beimel, A., Nissim, K., and Stemmer, U. (2013). Characterizing the sample complexity of private learners. In *Proceedings of the 4th Conference on Innovations in Theoretical Computer Science, ITCS '13*, pages 97–110, New York, NY, USA. Association for Computing Machinery.
- Cao, Y., Xiao, Y., Xiong, L., and Bai, L. (2019). PriSTE: From Location Privacy to Spatiotemporal Event Privacy. In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, pages 1606–1609, Macao, Macao. IEEE.
- Cao, Y., Yoshikawa, M., Xiao, Y., and Xiong, L. (2017). Quantifying Differential Privacy under Temporal Correlations. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pages 821–832.
- Chan, T.-H. H., Shi, E., and Song, D. (2011). Private and Continual Release of Statistics. *ACM Trans. Inf. Syst. Secur.*, 14(3):26:1–26:24.
- Chen, C., Petty, K., Skabardonis, A., Varaiya, P., and Jia, Z. (2001). Freeway Performance Measurement System: Mining Loop Detector Data. *Transportation Research Record*, 1748(1):96–102.
- Cho, E., Myers, S. A., and Leskovec, J. (2011). Friendship and mobility: User movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '11*, pages 1082–1090, New York, NY, USA. Association for Computing Machinery.
- Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., and Naor, M. (2006a). Our Data, Ourselves: Privacy Via Distributed Noise Generation. In Vaudey, S., editor, *Advances in Cryptology - EUROCRYPT 2006*, Lecture Notes in Computer Science, pages 486–503, Berlin, Heidelberg. Springer.
- Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006b). Calibrating Noise to Sensitivity in Private Data Analysis. In Halevi, S. and Rabin, T., editors, *Theory of Cryptography*, Lecture Notes in Computer Science, pages 265–284, Berlin, Heidelberg. Springer.
- Dwork, C., Naor, M., Pitassi, T., and Rothblum, G. N. (2010). Differential privacy under continual observation. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing, STOC '10*, pages 715–724, New York, NY, USA. Association for Computing Machinery.
- Dwork, C. and Roth, A. (2014). The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.*, 9(3–4):211–407.
- Fan, L. and Xiong, L. (2012). Real-time aggregate monitoring with differential privacy. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management, CIKM '12*, pages 2169–2173, New York, NY, USA. Association for Computing Machinery.
- Kasiviswanathan, S. P., Lee, H. K., Nissim, K., Raskhodnikova, S., and Smith, A. (2011). What Can We Learn Privately? *SIAM J. Comput.*, 40(3):793–826.

- Meehan, C. and Chaudhuri, K. (2021). Location Trace Privacy Under Conditional Priors. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, pages 2881–2889. PMLR.
- Ny, J. L. and Pappas, G. J. (2014). Differentially Private Filtering. *IEEE Transactions on Automatic Control*, 59(2):341–354.
- Rastogi, V. and Nath, S. (2010). Differentially private aggregation of distributed time-series with transformation and encryption. In *Proceedings of the 2010 International Conference on Management of Data - SIGMOD '10*, page 735, Indianapolis, Indiana, USA. ACM Press.
- Tropp, J. A. (2015). An Introduction to Matrix Concentration Inequalities. *MAL*, 8(1-2):1–230.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., and van Mulbregt, P. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272.
- Xiao, Y. and Xiong, L. (2015). Protecting Locations with Differential Privacy under Temporal Correlations. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, CCS '15*, pages 1298–1309, New York, NY, USA. Association for Computing Machinery.
- Yang, D., Zhang, D., Chen, L., and Qu, B. (2015). NationTelescope: Monitoring and visualizing large-scale collective behavior in LBSNs. *Journal of Network and Computer Applications*, 55:170–180.
- Yang, D., Zhang, D., and Qu, B. (2016). Participatory Cultural Mapping Based on Collective Behavior Data in Location-Based Social Networks. *ACM Trans. Intell. Syst. Technol.*, 7(3):30:1–30:23.

---

## Supplementary Material: Privacy Amplification by Subsampling in Time Domain

---

### A PROOF OF RESULTS

#### A.1 Proof of Theorem 2

*Proof.* For fixed  $I' \leq I$  ( $I' \in \mathbb{N}$ ), it holds that  $\Pr\{\Delta_{2,f_s} > \sqrt{I'}\} = \Pr\{|J \cap j_i| > I'\}$ , where the probability is taken over the randomness of  $J$ . Since  $|J \cap j_i|$  follows the Binomial distribution with parameters  $|j_i|$  and  $p$ , the tail probability is bounded as below:

$$\begin{aligned} \Pr\{|J \cap j_i| > I'\} &= \sum_{t=I'+1}^{|j_i|} \binom{|j_i|}{t} p^t (1-p)^{|j_i|-t} \\ &\leq \sum_{t=I'+1}^I \binom{I}{t} p^t (1-p)^{I-t}, \end{aligned}$$

where the last inequality follows from  $|j_i| \leq I$  and the monotonicity in a parameter  $n$  of the tail probability of the Binomial distribution. Setting the last term to be  $\delta'$  concludes the proof.  $\square$

#### A.2 Proof of Corollary 1

*Proof.* From Theorem 2, for fixed  $\delta'$ , if  $I'$  satisfies  $\Pr\{|J \cap j_i| > I'\} = \delta'$ , then with probability  $1 - \delta'$ ,  $\Delta_{2,f_s} = \sqrt{I'}$ . Here, applying Hoeffding's inequality to  $I$  i.i.d. Bernoulli random variables with a parameter  $p$  yields that

$$\begin{aligned} \Pr\{|J \cap j_i| > I'\} &\leq \Pr\{|J \cap j_i| \geq I'\} \\ &\leq \exp\left(-2I\left(p - \frac{I'}{I}\right)^2\right), \end{aligned}$$

for  $I' \geq Ip$ . Rearranging the equality  $\delta' = \exp\left(-2I\left(p - \frac{I'}{I}\right)^2\right)$  implies that for  $I'$  s.t.  $I' = Ip + \sqrt{\frac{I}{2} \log 1/\delta'}$ , it holds that with probability  $1 - \delta'$ ,  $\Delta_{2,f_s} = \sqrt{I'}$ .  $\square$

#### A.3 Proof of Theorem 4

*Proof.* As argued in the main paper,  $\Delta_{2,f_s}^2 \leq \sigma_{\max}^2(B)I$ . Furthermore, from Theorem 3, we know

$$\Pr\{\sigma_{\max}^2(B) \geq (1 + \varepsilon)p\sigma_{\max}^2(A)\} \leq 2\text{srank}(A) \left(\frac{e^\varepsilon}{(1 + \varepsilon)^{1+\varepsilon}}\right)^{p\sigma_{\max}^2(A)/L}.$$

By using the fact that  $\sigma_{\max}(A) = 1$  and letting  $\alpha = \sqrt{(1 + \varepsilon)p}$ , we have

$$\Pr\{\sigma_{\max}^2(B) \geq \alpha^2\} \leq 2\text{srank}(A) \left(\frac{e^{\alpha^2/p-1}}{(\alpha^2/p)^{\alpha^2/p}}\right)^{p/L}.$$

Thus, it holds that  $\Pr\{\sigma_{\max}(B) \leq \alpha\} \geq 1 - \delta'$ , where  $\delta' = 2\text{srank}(A) \left(\frac{e^{\alpha^2/p-1}}{(\alpha^2/p)^{\alpha^2/p}}\right)^{p/L}$ . Therefore, we have  $\Pr\{\Delta_{2,f_s} \leq \alpha\sqrt{I}\} \geq 1 - \delta'$ .  $\square$

#### A.4 Proof of Corollary 2

*Proof.* Let  $M' : \mathbb{R}^T \rightarrow \mathbb{R}^T$  be Algorithm 1 and  $M : \mathbb{R}^T \rightarrow \mathbb{R}^{T'}$  be Algorithm 1 without the interpolation in the last step. We define the interpolation map given  $j \in \mathcal{J}$  as  $\text{Interpolate}_j : \mathbb{R}^{|j|} \rightarrow \mathbb{R}^T$ . Note that this map is bijective.

Let  $S' \subseteq \text{range}(M')$  be arbitrary. Also, we denote the corresponding set by the inverse of  $\text{Interpolate}_j$  for  $M$  as  $S \subseteq \text{range}(M)$ . Furthermore, we let  $F \subseteq \mathcal{J}$  be the failure indices set. More formally,  $F = \{j \in \mathcal{J} : \sigma_{\max}(B) > \alpha\}$ , where  $B = \text{diag}(\delta_1, \dots, \delta_T)A$ . Note that under  $j \in F$ , the sensitivity is  $\sqrt{I}$ , but under  $j \in F^c$ , it is  $\alpha\sqrt{I}$ . Then, it holds that

$$\begin{aligned}
 \Pr\{M'(x) \in S'\} &= \sum_{j \in \mathcal{J}} \Pr\{M'(x) \in S', J = j\} \\
 &= \sum_{j \in \mathcal{J}} \Pr\{M(x) \in S, J = j\} \\
 &= \sum_{j \in F} \Pr\{M(x) \in S | J = j\} \Pr\{J = j\} + \sum_{j \in F^c} \Pr\{M(x) \in S | J = j\} \Pr\{J = j\} \\
 &\leq \sum_{j \in F} (\exp(\epsilon/\alpha) \Pr\{M(x') \in S | J = j\} + \delta) \Pr\{J = j\} \\
 &\quad + \sum_{j \in F^c} (\exp(\epsilon) \Pr\{M(x') \in S | J = j\} + \delta) \Pr\{J = j\} \\
 &= \exp(\epsilon/\alpha) \sum_{j \in F} \Pr\{M(x') \in S | J = j\} \Pr\{J = j\} + \delta \Pr\{J \in F\} \\
 &\quad + \exp(\epsilon) \sum_{j \in F^c} \Pr\{M(x') \in S | J = j\} \Pr\{J = j\} + \delta \Pr\{J \in F^c\} \\
 &= \exp(\epsilon) \sum_{j \in \mathcal{J}} \Pr\{M(x') \in S, J = j\} + \delta \\
 &\quad + (\exp(\epsilon/\alpha) - \exp(\epsilon)) \sum_{j \in F} \Pr\{M(x') \in S | J = j\} \Pr\{J = j\} \\
 &= \exp(\epsilon) \sum_{j \in \mathcal{J}} \Pr\{M'(x') \in S', J = j\} + \delta \\
 &\quad + (\exp(\epsilon/\alpha) - \exp(\epsilon)) \sum_{j \in F} \Pr\{M(x') \in S | J = j\} \Pr\{J = j\} \\
 &\leq \exp(\epsilon) \Pr\{M'(x') \in S'\} + \delta + (\exp(\epsilon/\alpha) - \exp(\epsilon)) \Pr\{J \in F\} \\
 &\leq \exp(\epsilon) \Pr\{M'(x') \in S'\} + \delta + \delta' (\exp(\epsilon/\alpha) - \exp(\epsilon))
 \end{aligned}$$

where the first inequality follows from the use of the Gaussian mechanism, the second inequality holds due to  $\Pr[M(x') \in S | J = j] \leq 1$ , and the last inequality follows by Theorem 4.  $\square$

#### A.5 Proof of Corollary 3

*Proof.* If  $A$  is an identity matrix, we are able to use the sensitivity bound in Theorem 2; thus,  $\delta'$  is the one in Theorem 2 with  $I' = \lceil \alpha^2 I \rceil$ . Note that  $I' \in \mathbb{N}$ . Let  $F = \{j \in \mathcal{J} : |j \cap j_i| > I'\}$ . Then, under  $j \in F^c$ , the sensitivity is  $\sqrt{I'}$ . Finally, replacing  $\alpha$  with  $\sqrt{I'}/I$  and proceeding the same transformation as the proof of Corollary 2 yield the final claim, where we use Theorem 2 to bound  $\Pr\{J \in F\}$  in the final step.  $\square$

#### A.6 Proof of Proposition 1

*Proof.* By the assumption that  $\sup_i \sum_t d_{i,t} = cI$  for some  $c > 1$ , the sensitivity becomes  $\sqrt{cI}$  under  $j \in F$  and  $\alpha\sqrt{cI}$  under  $j \in F^c$ , where  $F$  is the failure indices set defined in the proof of Corollary 2. Therefore, if we replace  $\epsilon$  in Algorithm 1 with  $\sqrt{c}\epsilon$ , then we can proceed the identical transformation as the proof of Corollary 2. Thus,

we have the privacy guarantee that replaces  $\epsilon$  in Corollary 2 with  $\sqrt{c}\epsilon$  for the case where  $\sup_i \sum_t d_{i,t} = cI$  for some  $c > 1$ .  $\square$

## B DETAILS OF EXPERIMENTS

### B.1 Instantiation of Our Algorithm

We instantiate the filter with the gaussian filter. More formally, we define the filtering vector  $h$  that constitutes  $A$  by

$$\hat{h}_t = \exp\left(-\frac{1}{2}\left(\frac{T/2 - |t - T/2|}{\sigma_g}\right)^2\right) \quad \forall t \in [T]$$

$$h = \hat{h}/\|\hat{h}\|_1.$$

We set  $\sigma_g = 10$  for the real datasets and  $\sigma_g = 10/(1 - \log(f)/4)$  for the synthetic dataset.

We implement the interpolation in Algorithm 1 by SciPy’s (Virtanen et al., 2020) `scipy.interpolate.interp1d` method with `kind` parameter set to `linear`. There’s a case that  $J$  in Algorithm 1 does not contain the first and last indices; thus, we need to extrapolate the signal. We deal with it by copying the nearest value. For example, if  $\min J = 2$ , we publish  $\hat{z} = (z_2, z_2, \dots)^\top$ .

### B.2 Why We Fix $I$ for Synthetic Data

In our experiment with Synth data, we fix  $I$  for any (relative) sampling frequency  $f$  even though different sampling frequencies  $f$  yield different numbers of time steps  $T$ : the larger  $f$  becomes, the larger  $T$  becomes. However, for e.g., the PeMS data (traffic flow),  $I$  stays constant and does not scale with  $f$ . For example, if we assume a commuter passes a given sensor at most once every hour, then as long as the sampling frequency remains above once an hour,  $I$  will stay fixed regardless of  $T$ . In other words, we can consider the number of time steps,  $T$ , as  $T = (\text{Duration}) \times (\text{Sampling Frequency})$ , and (for a high enough sampling frequency  $f$ )  $I$  only depends on signal duration. In fact, in the instance that sampling frequency descends below once per hour, sensitivity still remains unchanged given the above assumption. For instance, if the sampling frequency drops to once every two hours, an individual may participate *twice* in a time-step. While this may seem to invalidate our assumption of single participation per time-step ( $M = 1$ ), the sensitivity stays constant since the product  $MI$  (number of participation by an individual in a fixed duration of the time-series) remains constant.

On the other hand, for real signals, the sampling frequency tends to be fixed; thus,  $I$  scales with  $T$  since larger  $T$  means longer observation duration. Therefore, it is indeed necessary to reduce the sensitivity scaling with  $T$  or  $I$  as an analyst wants to use a signal observed for a longer period with a given frequency  $f$ .

### B.3 Outputs Operating on the Noiseless Synth Signal

Figure 3 shows in orange examples of outputs of four mechanisms: the two baseline mechanisms and our mechanism with and without filtering, each taking the Synth signal without observation noise as input. For reference the noiseless Synth signal is plotted in blue. Similar to the case where each mechanism takes the noisy Synth signal as input, outputs from our mechanism with and without filtering concentrate around the raw signal. Comparing the outputs by our mechanism with and without filtering, we see that subsampling alone helps produce less noisy output and capture underlying trends better. This supports the fact that the sensitivity reduction is larger for subsampling without filtering in general.

## C MISCELLANY

### C.1 Relationship between $\alpha$ and $\delta'$ in Theorem 4

Figure 4 illustrates the tail probability  $\delta'$  of  $\Delta_{2,f_{\text{fs}}}$  exceeding  $\alpha\sqrt{T}$  given  $\alpha \in [0, 1]$ , i.e.,  $\delta' = \Pr\{\Delta_{2,f_{\text{fs}}} > \alpha\sqrt{T}\}$ . We set other parameters as  $T = 10000$  and  $\sigma_g = 10$ , yielding  $\text{srnk}(A) \approx 280$  and  $L \approx 0.028$  in Theorem 4. As expected, we get more sensitivity reduction as subsampling parameter  $p$  gets smaller. More importantly, we

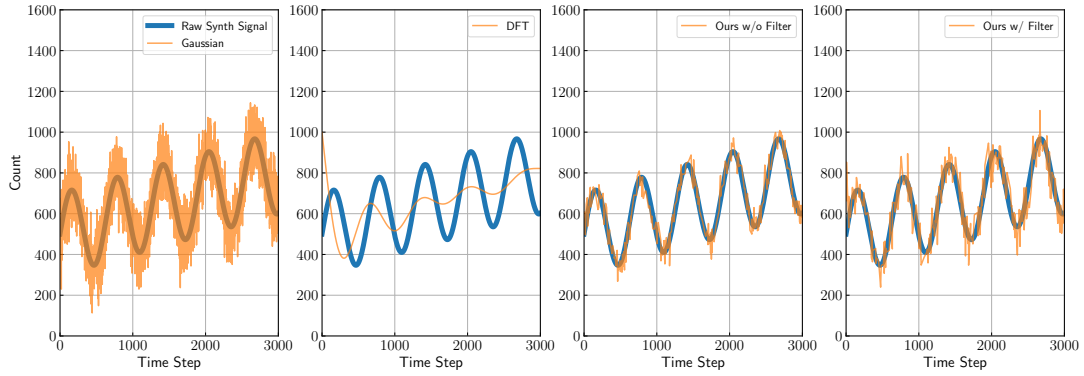


Figure 3: Two baseline mechanisms (left) and our mechanism with and without filtering (right), each operating on Synth time-series without observation noise. We plot the noiseless Synth time-series in blue for reference. Output by our algorithm without filtering concentrates around raw signal the most, indicating superiority of our algorithm with subsampling alone for smooth time-series signals.

observe that  $\delta'$  decreases at least exponentially. We believe the reduction is still notable though it is smaller for filtering and subsampling compared with subsampling alone.

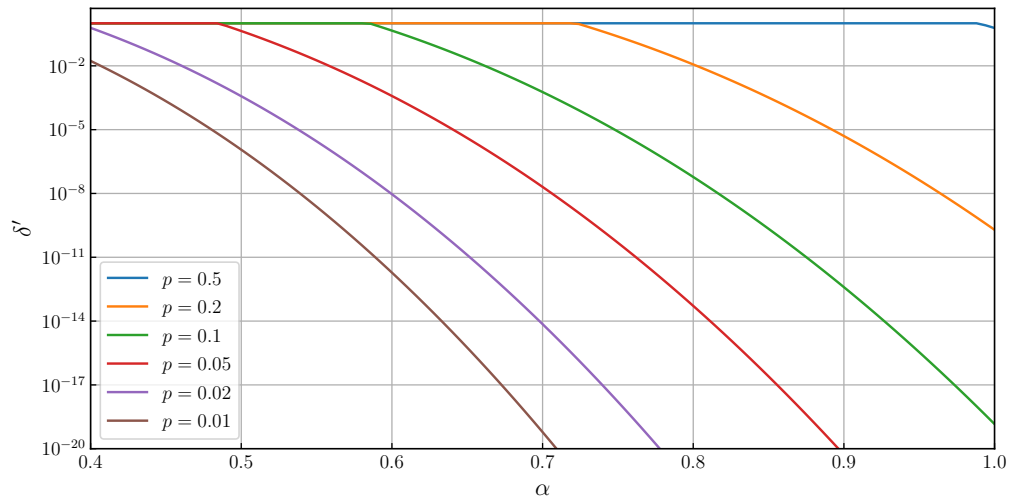


Figure 4: Tail probability,  $\delta'$ , of  $\Delta_{2, f_{\text{ts}}}$  exceeding  $\alpha\sqrt{I}$  given  $\alpha$  in Theorem 4 for varying subsampling parameter  $p$ . We observe that decrease rate of  $\delta'$  is at least exponentially fast. We also see that as  $p$  becomes smaller we get more sensitivity reduction.