Temporal Differential Privacy

kyunghee KIM

Department of Statistics

Sungkyunkwan University

May 12, 2022

Overview

1. Last week question
2. Sum up
3. Time Series DP

Last week question

- $\epsilon \propto$ Noise?
- Applications?
- Performance Verification?

$\epsilon \propto$ Noise?

Let noise follows laplace distribution (can be something else)

- Laplace distribution

$$f(x|\sigma) = \frac{1}{2\sigma} exp(-\frac{|x|}{\sigma})$$

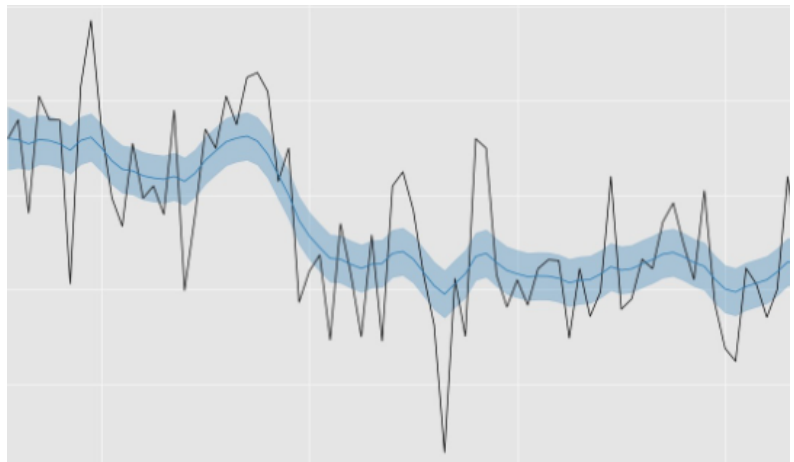
- Add noise

$$K(D) = f(D) + Lap(\sigma),\ \sigma = \Delta f/\epsilon$$

$$\begin{aligned}\frac{P(K(D_1) = y)}{P(K(D_2) = y)} &= \frac{P(f(D_1) + Lap(\Delta f/\epsilon) = y}{P(f(D_2) + Lap(\Delta f/\epsilon) = y} \\ &= \frac{exp(-|y - f(D_1)|\epsilon/\Delta f)}{exp(-|y - f(D_2)|\epsilon/\Delta f)} \leq exp(\epsilon)\end{aligned}$$

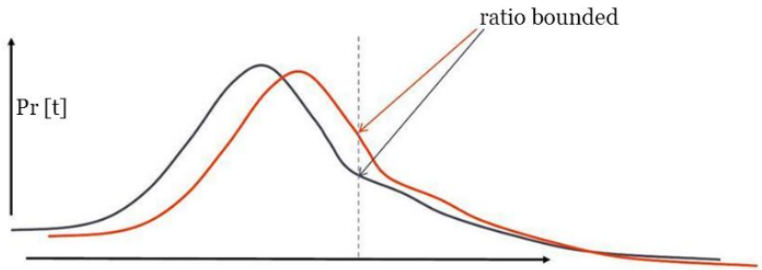
Application?

The noise addition method is not applicable to all queries. For example, if the result of a query is a string, a picture, etc., you cannot add noise and you must use another method.



Performance verification?

- Too accurate



  For utility measure, regression estimates are fine not 99% quantile, etc (Park, 2016).

- Cost measures

  A paper to be introduced later, they quantify the utility in terms of the missing, repetition, empty and delay cost.

Statistical disclosure control for public microdata: present and future, Min-Jeong Park, 2016.

Sum up

LDP; Local Differential Privacy

Data owners add noise to provide processed data.

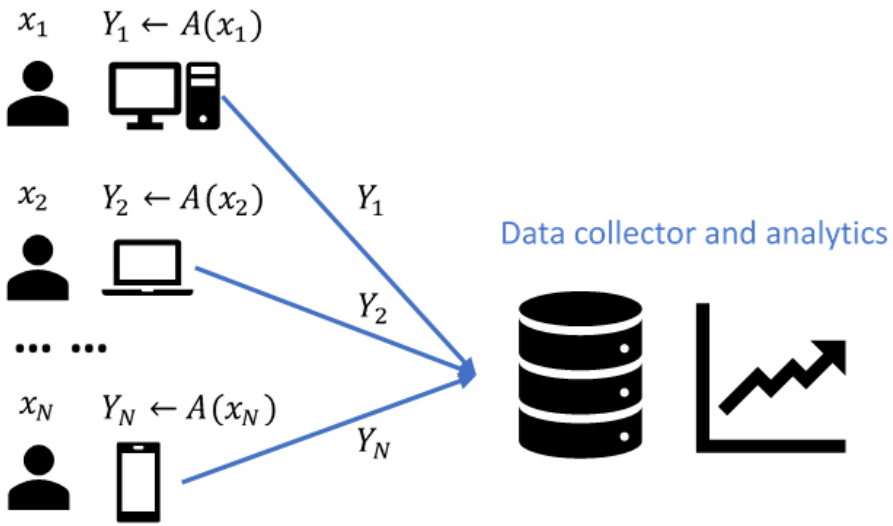


Figure 1. Data collection and analytics under the LDP model

Sum up

For any two tuples t and t', and any possible output t*:

$$P(A(t) = t*) \leq e^{\epsilon} P(A(t') = t*)$$

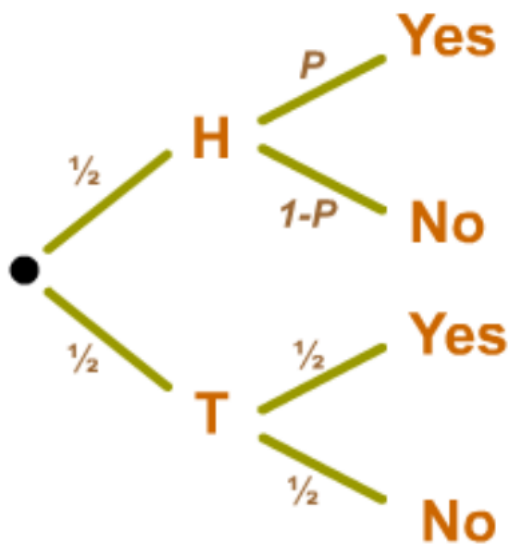
To satisfy $\epsilon-$LDP, both 1) Laplace mechanism and 2) Generalized Randomized Response are predominant perturbation mechanisms.

Sum up

Noise type: Randomized Response, Laplace dist,$\cdots$.

- Randomized Response

  Coin flipping example.



Temporal DP

The challenge is the time-series data, which has a strong correlation among successive values in the series (Bodhi Chakraborty, 2019). So, there is a need for preserving the privacy of the event of interest and the sensor nodes even when the adversary has a rough estimate of occurrence of an event with time.

Temporal Differential Privacy in Wireless Sensor Networks, Bodhi Chakraborty, 2019.

TLDP vs VLDP

Time series has both values and timestamps.

1. Value setting (VLDP)
2. Temporal setting (TLDP)

VLDP is an extension of LDP, it can be satisfied by existing value perturbation mechanism such as Laplace mechanism or Randomized Response. In this presentation, we focus on TLDP. *VLDP: Value Local DP. *TLDP: Temporal Local DP.

Temporal DP

For this, delay at the intermediate time points has been suggested. The data are buffered for random time durations at the intermediate time.

- T_0: Actual time of occurrence of an event.
- T_m = Apply DP at T_0

  Perturbs the original transmission time T_0 by appling DP which makes modified time of transmission T_m (Bodhi, 2019).

TLDP Applications

- Example 1. Biosensors in Health monitoring

  Monitor vital metrics including heart rate, ECG, and blood pressure.

- Example 2. Mobility Tracking

  Moving speed such as stationary, walking, jogging or driving.

- Example 3. Sensor Readings in Smart Home

  Monitor temperature, humidity, air quality.

Mechanisms

We will compare several temporal perturbation mechanisms

- Forward mechanism (baseline 1)
- Backward mechanism (baseline 2)
- Threshold mechanism

The paper proposes Threshold mechanism, which is not only free of missing, repetition and empty costs, but also has the lowest delay costs.

Cost measures

Cost(Utility) of Released TS

1. Missing Cost (C_M)
2. Repetition Cost (C_N)
3. Empty Cost (C_E)
4. Delay Cost (C_D)

They depend on time series manipulations and mechanisms (FC, SMA, ...)

Backward Perturbation Mechanism

- Backward Perturbation Mechanism

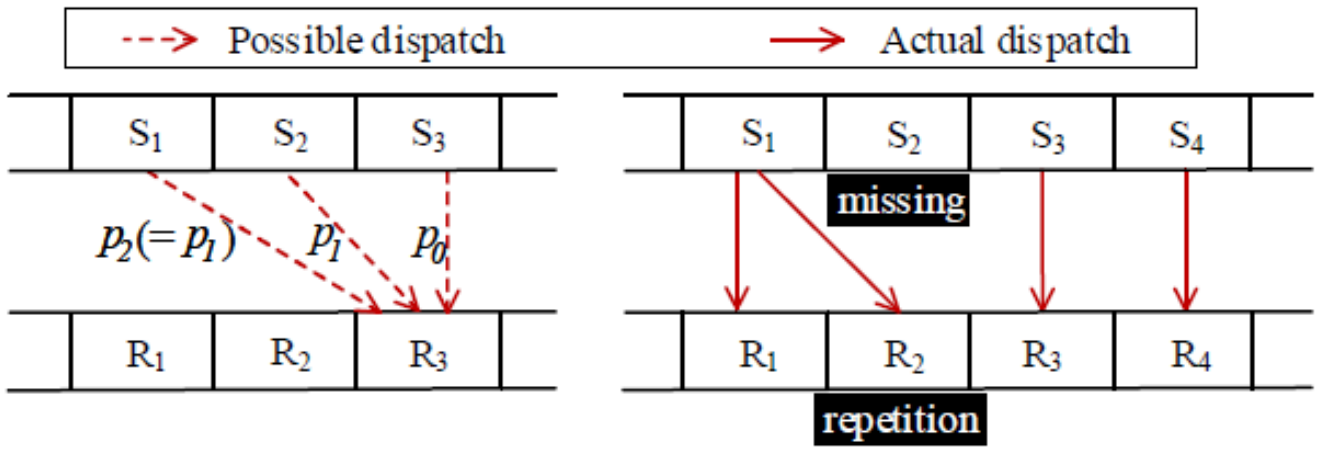
  It probabilistically selects a value from previous k timestamps in the original time series to release at the current timestamp.

  1) Perturbation protocol

$$\Pr(R_i = S_{i-j}) := p_j = \begin{cases} \frac{e^{\epsilon/2}}{k-1+e^{\epsilon/2}}, & j = 0 \\ \frac{1}{k-1+e^{\epsilon/2}}, & j \in \{1, 2, ..., k-1\} \end{cases}$$

  2) Cost Analysis

BPM (Cont'd)



(a) Perturbation protocol (b) Dispatching example

Fig. 1. Backward Perturbation mechanism

Forward Perturbation Mechanism

- Forward Perturbation Mechanism

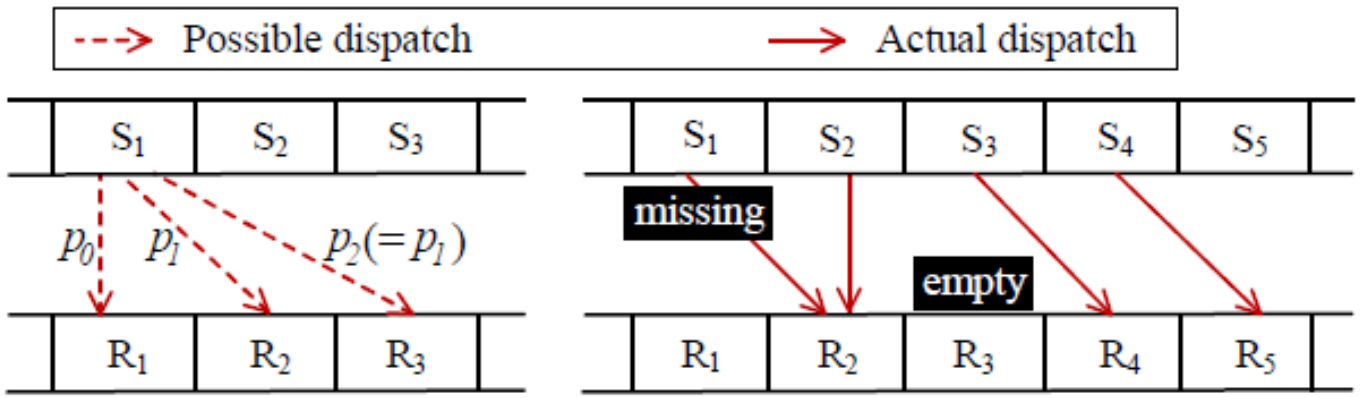
  It probabilistically dispatches the value at the current timestamp to on one of the subsequent k timestamps to release.

  1) Perturbation protocol

  As oppose to Backward Perturbation mechanism which finds for each R_i a previous S_{i-j} to dispatch to, the Forward dispatches each S_i to one of the R_{i+j}'s in the next k timestamps.

  2) Cost Analysis

FPM (Cont'd)



(a) Perturbation protocol (b) Dispatching example

Fig. 2. Forward Perturbation mechanism

BPM and FPM Problem

Two baseline mechanisms essentially cause a "collision". However, the counter example in Figure shows that the collision-free variant of Forward Perturbation mechaism no longer satisfies DP.

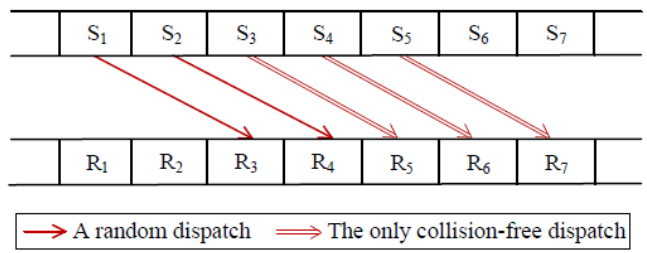


Fig. 3. Collision-free Forward Perturbation violates ϵ-TLDP.

An adversary can infer, with almost 100% confidence.

Compare Mechanisms

Comparison of Backward and Forward Perturbation

COST ANALYSIS OF DIFFERENT MECHANISMS

Cost	Backward Pert. mechanism	Forward Pert. mechanism	Threshold mechanism
Missing	$(1-p_0)(1-p_1)^{k-1}M$	$(1-p_0)(1-p_1)^{k-1}M$	0
Repetition	$(1-p_0)(1-p_1)^{k-1}N$	0	0
Empty	0	$(1-p_0)(1-p_1)^{k-1}E$	0
Delay	$p_1(1-p_0)\cdot$ $\sum_{j=1}^{k-1} j(1-p_1)^{j-1}D$	$p_1(1-p_0)\cdot$ $\sum_{j=1}^{k-1} j(1-p_1)^{j-1}D$	$(k-c_0)D$

Neither mechanism can avoid value missing, which is very costly or even unacceptable in many time series applications.

Threshold Mechanism

Threshold mechanism

Above or equal to a threshold c_0 in the range of $[2, k-1]$, it always satisfy DP.

Threshold mechanism is to add the following rule to the collision-free Forward Perturbation. The core of this algorithm is to find c_0* the optimal threshold of this mechanism using a binary search.

Threshold Mechanism

Algorithm 1 Perturbation protocol: $Perturb(\cdot)$

Input: Original time series $S = \{S_1, S_2, ..., S_n, ...\}$
Time window length k
Threshold c_0
Output: $R = Perturb(S, k, c_0)$ is the released time series
Procedure:

Initialize $x = \mathbf{0}$, and $R = \emptyset$
for each value $S_i \in S$ **do**
  Count the number of "0"s in $\{x_i, x_{i+1}, ..., x_{i+k-1}\}$, denoted by c
  if $c > c_0$ **then**
    Randomly select an index l from $\mathbb{X} = \{j | x_j = 0, i \leq j \leq i+k-1\}$
    Dispatch S_i to R_l, and set $x_l = 1$
  else
    if $x_i = 0$ **then**
      Dispatch S_i to R_i, and set $x_i = 1$
    else
      Randomly select an index l from $\mathbb{X} = \{j | x_j = 0, i < j \leq i+k-1\}$
      Dispatch S_i to R_l, and set $x_l = 1$
  Release R_i
return $R = \{R_1, R_2, ..., R_n, ...\}$

Experimental Evaluation

Datasets

1. U.S. Stocks

   14058 trading days

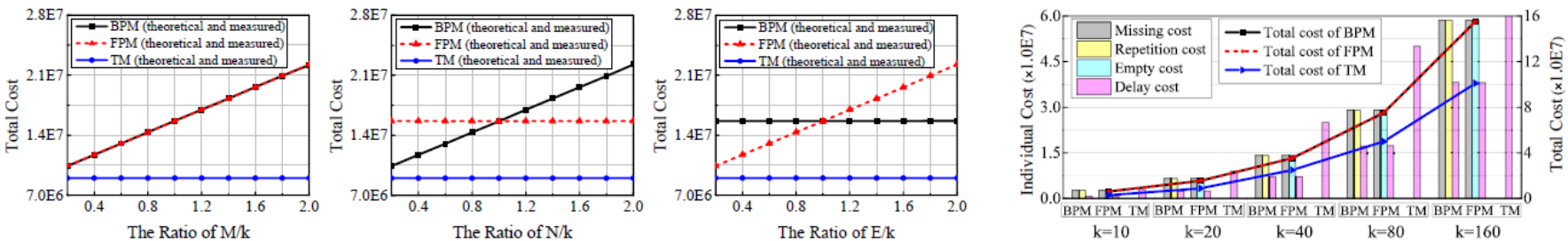
2. Taxi Trajectories

   6357 taxi trajectories, each of which has GPS coordinate in a

   15-second interval and has at least 300 timestamps

Compare Cost

The first set evaluates the overall cost of the three TLDP perturbation mechanisms. Backward Perturbation mechanism(BPM), Forward (FPM), Threshold (TM), Extended Threshold (ETM).

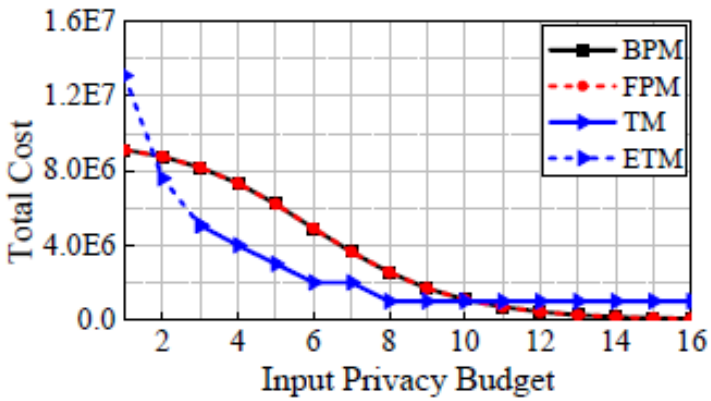


(a) Total cost varying M (b) Total cost varying N (c) Total cost varying E (d) Impact of k on individual and total costs

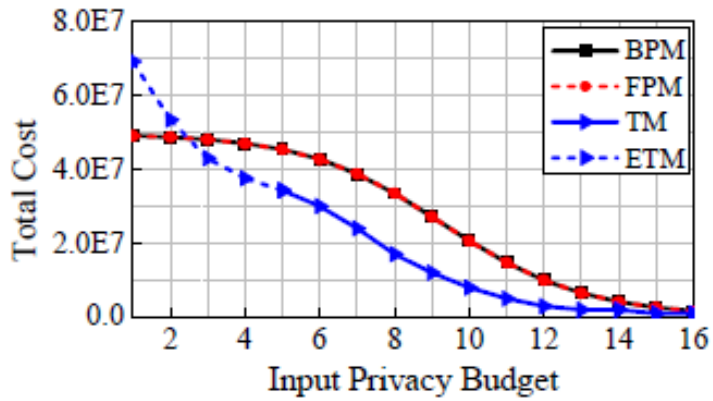
Fig. 5. Impact of relative cost and sliding window length on total cost

Compare Total Cost vary with ϵ

Compare total cost vary with ϵ.



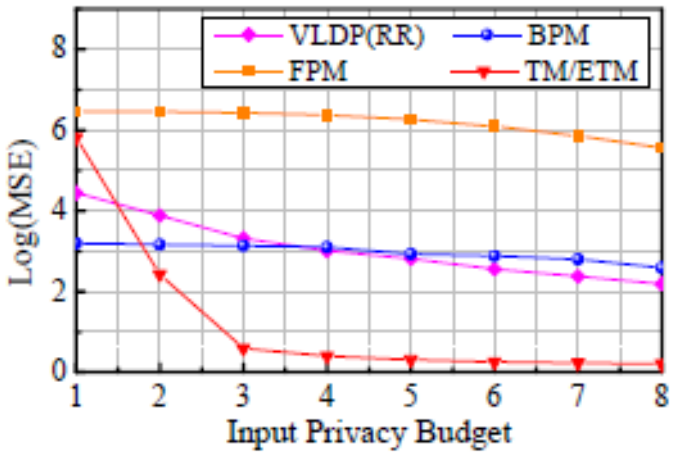
(a) $k = 10$



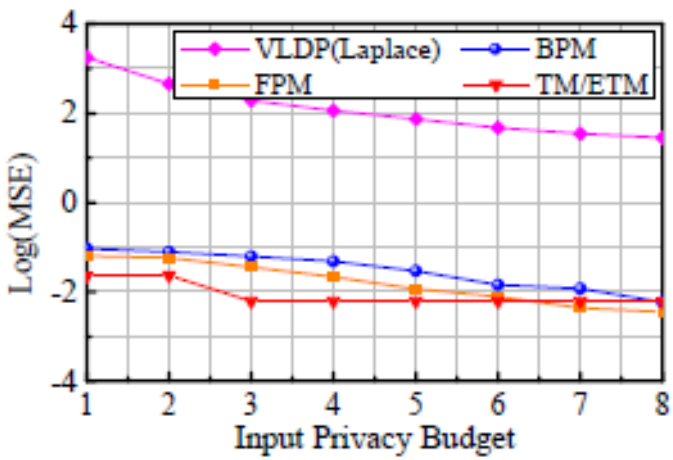
(b) $k = 50$

TLDP vs VLDP

Compares real utility of TLDP (BPM, FPM, TM, ETM) against VLDP perturbation (Randomized Response or Laplace Mechanism) in three real-world applications (e.g., Frequency counting, simple moving average and trajectory clustering).



(a) FC, VLDP vs. TLDP



(c) SMA, VLDP vs. TLDP

Accuracy

TABLE II
NMI ON TRAJECTORY CLUSTERING

Input ϵ	1	2	3	4	5	6	7	8
Laplace	0.003	0.003	0.004	0.005	0.006	0.005	0.006	0.007
TM/ETM	0.572^*	0.590^*	0.610^*	0.616	0.699	0.705	0.706	0.768

*: Extended Threshold Mechanism (ETM) is used.

The End