Local Obfuscation Mechanisms for Hiding Probability Distributions *

Yusuke Kawamoto[1] [0000-0002-2151-9560] and
Takao Murakami[2] [0000-0002-5110-1261]

[1] AIST, Tsukuba, Japan
[2] AIST, Tokyo, Japan

Abstract. We introduce a formal model for the information leakage of probability distributions and define a notion called distribution privacy as the local differential privacy for probability distributions. Roughly, the distribution privacy of a local obfuscation mechanism means that the attacker cannot significantly gain any information on the distribution of the mechanism's input by observing its output. Then we show that existing local mechanisms can hide input distributions in terms of distribution privacy, while deteriorating the utility by adding too much noise. For example, we prove that the Laplace mechanism needs to add a large amount of noise proportionally to the infinite Wasserstein distance between the two distributions we want to make indistinguishable. To improve the tradeoff between distribution privacy and utility, we introduce a local obfuscation mechanism, called a tupling mechanism, that adds random dummy data to the output. Then we apply this mechanism to the protection of user attributes in location based services. By experiments, we demonstrate that the tupling mechanism outperforms popular local mechanisms in terms of attribute obfuscation and service quality.

Keywords: local differential privacy · obfuscation mechanism · location privacy · attribute privacy · Wasserstein metric · compositionality

1 Introduction

Differential privacy [1] is a quantitative notion of privacy that has been applied to a wide range of areas, including databases, geo-locations, and social network. The protection of differential privacy can be achieved by adding controlled noise to given data that we wish to hide or obfuscate. In particular, a number of recent studies have proposed *local obfuscation mechanisms* [2–4], namely, randomized algorithms that perturb each single "point" data (e.g., a geo-location point) by adding certain probabilistic noise before sending it out to a data collector. However, the obfuscation of a probability distribution of points (e.g., a distribution of locations of users at home/outside home) still remains to be investigated in terms of differential privacy.

* This work was partially supported by JSPS KAKENHI Grant JP17K12667, JP19H04113, and Inria LOGIS project.

For example, a location-based service (LBS) provider collects each user's geolocation data to provide a service (e.g., navigation or point-of-interest search), and has been widely studied in terms of the privacy of user locations. As shown in [3, 5], users can hide their accurate locations by sending to the LBS provider only approximate location information calculated by an obfuscation mechanism.

Nevertheless, a user's location information can be used for an attacker to infer the user's attributes (e.g., age, gender, social status, and residence area) or activities (e.g., working, sleeping, and shopping) [6–9]. For example, when an attacker knows the distribution of residence locations, he may detect whether given users are *at home* or *outside home* after observing their obfuscated locations. For another example, an attacker may learn whether users are *rich* or *poor* by observing their obfuscated behaviors. These attributes can be used by robbers hence should be protected from them. Privacy issues of such attribute inference are also known in other applications, including recommender systems [10, 11] and online social networks [12, 13]. However, to our knowledge, no literature has addressed the protection of attributes in terms of local differential privacy.

To illustrate the privacy of attributes in an LBS, let us consider a running example where users try to prevent an attacker from inferring whether they are at home or not. Let λ_{home} and λ_{out} be the probability distributions of locations of the users at home and outside home, respectively. Then the privacy of this attribute means that the attacker cannot learn from an obfuscated location whether the actual location follows the distribution λ_{home} or λ_{out}.

This can be formalized using differential privacy. For each $t \in \{home, out\}$, we denote by $p(y \mid \lambda_t)$ the probability of observing an obfuscated location y when an actual location is distributed over λ_t. Then the privacy of t is defined by:

$$\frac{p(y \mid \lambda_{home})}{p(y \mid \lambda_{out})} \leq e^{\varepsilon},$$

which represents that the attacker cannot distinguish whether the users follow the distribution λ_{home} or λ_{out} with degree of ε.

To generalize this, we define a notion, called *distribution privacy* (DistP), as the differential privacy for probability distributions. Roughly, we say that a mechanism A provides DistP w.r.t. λ_{home} and λ_{out} if no attacker can detect whether the actual location (input to A) is sampled from λ_{home} or λ_{out} after he observed an obfuscated location y (output by A)[3]. Here we note that each user applies the mechanism A locally by herself, hence can customize the amount of noise added to y according to the attributes she wants to hide.

Although existing local differential privacy mechanisms are designed to protect point data, they also hide the distribution that the point data follow. However, we demonstrate that they need to add a large amount of noise to obfuscate distributions, and thus deteriorate the utility of the mechanisms.

To achieve both high utility and strong privacy of attributes, we introduce a mechanism, called the *tupling mechanism*, that not only perturbs an actual

[3] In our setting, the attacker observes only a sampled output of A, and not the exact histogram of A's output distribution. See Section 3.5 for more details.

input, but also adds random dummy data to the output. Then we prove that this mechanism provides DistP. Since the random dummy data obfuscate the shape of the distribution, users can instead reduce the amount of noise added to the actual input, hence they get better utility (e.g., quality of a POI service).

This implies that DistP is a relaxation of differential privacy that guarantees the privacy of attributes while achieving higher utility by weakening the differentially private protection of point data. For example, suppose that users do not mind revealing their actual locations outside home, but want to hide (e.g., from robbers) the fact that they are outside home. When the users employ the tupling mechanism, they output both their (slightly perturbed) actual locations and random dummy locations. Since their outputs include their (roughly) actual locations, they obtain high utility (e.g., learning shops near their locations), while their actual location points are protected only weakly by differential privacy. However, their attributes *at home*/*outside home* are hidden among the dummy locations, hence protected by DistP. By experiments, we demonstrate that the tupling mechanism is useful to protect the privacy of attributes, and outperforms popular existing mechanisms (the randomized response [14], the planar Laplace [3] and Gaussian mechanisms) in terms of DistP and service quality.

Our contributions. The main contributions of this work are given as follows:

- We propose a formal model for the privacy of probability distributions in terms of differential privacy. Specifically, we define the notion of distribution privacy (DistP) to represent that the attacker cannot significantly gain information on the distribution of a mechanism's input by observing its output.
- We provide theoretical foundation of DistP, including its useful properties (e.g., compositionality) and its interpretation (e.g., in terms of Bayes factor).
- We quantify the effect of distribution obfuscation by existing local mechanisms. In particular, we show that (extended) differential privacy mechanisms are able to make any two distributions less distinguishable, while deteriorating the utility by adding too much noise to protect all point data.
- For instance, we prove that extended differential privacy mechanisms (e.g., the Laplace mechanism) need to add a large amount of noise proportionally to the ∞-Wasserstein distance $W_{\infty,d}(\lambda_0, \lambda_1)$ between the two distributions λ_0 and λ_1 that we want to make indistinguishable.
- We show that DistP is a useful relaxation of differential privacy when users want to hide their attributes, but not necessarily to protect all point data.
- To improve the tradeoff between DistP and utility, we introduce the *tupling mechanism*, which locally adds random dummies to the output. Then we show that this mechanism provides DistP and hight utility for users.
- We apply local mechanisms to the obfuscation of attributes in location based services (LBSs). Then we show that the tupling mechanism outperforms popular existing mechanisms in terms of DistP and service quality.

All proofs of technical results can be found in Appendix.

2 Preliminaries

In this section we recall some notions of privacy and metrics used in this paper. Let $\mathbb{N}^{>0}$ be the set of positive integers, and $\mathbb{R}^{>0}$ (resp. $\mathbb{R}^{\geq 0}$) be the set of positive (resp. non-negative) real numbers. Let $[0, 1]$ be the set of non-negative real numbers not grater than 1. Let $\varepsilon, \varepsilon_0, \varepsilon_1 \in \mathbb{R}^{\geq 0}$ and $\delta, \delta_0, \delta_1 \in [0, 1]$.

2.1 Notations for Probability Distributions

We denote by $\mathbb{D}\mathcal{X}$ the *set of all probability distributions* over a set $\mathcal{X}$, and by $|\mathcal{X}|$ the number of elements in a finite set $\mathcal{X}$.

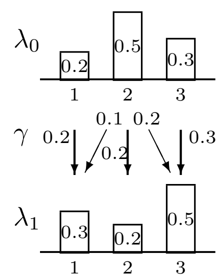


Fig. 1: Coupling γ.

Given a finite set $\mathcal{X}$ and a distribution $\lambda \in \mathbb{D}\mathcal{X}$, the probability of drawing a value x from λ is denoted by $\lambda[x]$. For a finite subset $\mathcal{X}' \subseteq \mathcal{X}$ we define $\lambda[\mathcal{X}']$ by: $\lambda[\mathcal{X}'] = \sum_{x' \in \mathcal{X}'} \lambda[x']$. For a distribution λ over a finite set $\mathcal{X}$, its *support* $\mathsf{supp}(\lambda)$ is defined by $\mathsf{supp}(\lambda) = \{x \in \mathcal{X} \colon \lambda[x] > 0\}$. Given a $\lambda \in \mathbb{D}\mathcal{X}$ and a $f : \mathcal{X} \to \mathbb{R}$, the expected value of f over λ is: $\mathbb{E}_{x\sim\lambda}[f(x)] \stackrel{\text{def}}{=} \sum_{x\in\mathcal{X}} \lambda[x]f(x)$.

For a randomized algorithm $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ and a set $R \subseteq \mathcal{Y}$ we denote by $A(x)[R]$ the probability that given input x, A outputs one of the elements of R. Given a randomized algorithm $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ and a distribution λ over $\mathcal{X}$, we define $A^{\#}(\lambda)$ as the distribution of the output of A. Formally, for a finite set $\mathcal{X}$, the *lifting* of A w.r.t. $\mathcal{X}$ is the function $A^{\#} : \mathbb{D}\mathcal{X} \to \mathbb{D}\mathcal{Y}$ such that $A^{\#}(\lambda)[R] \stackrel{\text{def}}{=} \sum_{x\in\mathcal{X}} \lambda[x]A(x)[R]$.

2.2 Differential Privacy (DP)

Differential privacy [1] captures the idea that given two "adjacent" inputs x and x' (from a set $\mathcal{X}$ of data with an adjacency relation Φ), a randomized algorithm A cannot distinguish x from x' (with degree of ε and up to exceptions δ).

Definition 1 (Differential privacy). Let e be the base of natural logarithm. A randomized algorithm $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, δ)-*differential privacy (DP)* w.r.t. an adjacency relation $\Phi \subseteq \mathcal{X} \times \mathcal{X}$ if for any $(x, x') \in \Phi$ and any $R \subseteq \mathcal{Y}$,

$$\Pr[A(x) \in R] \leq e^{\varepsilon} \Pr[A(x') \in R] + \delta$$

where the probability is taken over the random choices in A.

2.3 Differential Privacy Mechanisms and Sensitivity

Differential privacy can be achieved by a *privacy mechanism*, namely a randomized algorithm that adds probabilistic noise to a given input that we want to protect. The amount of noise added by some popular mechanisms (e.g., the exponential mechanism) depends on a *utility function* $u : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ that maps a pair of input and output to a utility score. More precisely, the noise is added according to the "sensitivity" of u, which we define as follows.

Definition 2 (Utility distance). The *utility distance* w.r.t a utility function $u : (\mathcal{X} \times \mathcal{Y}) \to \mathbb{R}$ is the function d given by: $d(x, x') \stackrel{\text{def}}{=} \max_{y \in \mathcal{Y}} \big| u(x, y) - u(x', y) \big|$.

Note that d is a pseudometric. Hereafter we assume that for all x, y, $u(x, y) = 0$ is logically equivalent to $x = y$. Then the utility distance d is a metric.

Definition 3 (Sensitivity w.r.t. an adjacency relation). The *sensitivity* of a utility function u w.r.t. an adjacency relation $\Phi \subseteq \mathcal{X} \times \mathcal{X}$ is defined as:

$$\Delta_{\Phi,d} \stackrel{\text{def}}{=} \max_{(x,x') \in \Phi} d(x, x') = \max_{(x,x') \in \Phi} \max_{y \in \mathcal{Y}} \big| u(x, y) - u(x', y) \big|.$$

2.4 Extended Differential Privacy (XDP)

We review the notion of *extended differential privacy* [15], which relaxes DP by incorporating a metric d. Intuitively, this notion guarantees that when two inputs x and x' are closer in terms of d, the output distributions are less distinguishable.

Definition 4 (Extended differential privacy). For a metric $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, we say that a randomized algorithm $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, δ, d)-*extended differential privacy (XDP)* if for all $x, x' \in \mathcal{X}$ and for any $R \subseteq \mathcal{Y}$,

$$\Pr[A(x) \in R] \le e^{\varepsilon d(x,x')} \Pr[A(x') \in R] + \delta.$$

2.5 Wasserstein Metric

We recall the notion of probability coupling as follows.

Definition 5 (Coupling). Given $\lambda_0 \in \mathbb{D}\mathcal{X}_0$ and $\lambda_1 \in \mathbb{D}\mathcal{X}_1$, a *coupling* of λ_0 and λ_1 is a $\gamma \in \mathbb{D}(\mathcal{X}_0 \times \mathcal{X}_1)$ such that λ_0 and λ_1 are γ's marginal distributions, i.e., for each $x_0 \in \mathcal{X}_0$, $\lambda_0[x_0] = \sum_{x'_1 \in \mathcal{X}_1} \gamma[x_0, x'_1]$ and for each $x_1 \in \mathcal{X}_1$, $\lambda_1[x_1] = \sum_{x'_0 \in \mathcal{X}_0} \gamma[x'_0, x_1]$. We denote by $\mathsf{cp}(\lambda_0, \lambda_1)$ the set of all couplings of λ_0 and λ_1.

Example 1 (Coupling as transformation of distributions). Let us consider two distributions λ_0 and λ_1 shown in Fig. 1. A coupling γ of λ_0 and λ_1 shows a way of transforming λ_0 to λ_1. For example, $\gamma[2, 1] = 0.1$ moves from $\lambda_0[2]$ to $\lambda_1[1]$.

We then recall the ∞-Wasserstein metric [16] between two distributions.

Definition 6 (∞-Wasserstein metric). Let d be a metric over $\mathcal{X}$. The ∞-*Wasserstein metric* $W_{\infty,d}$ w.r.t. d is defined by: for any $\lambda_0, \lambda_1 \in \mathbb{D}\mathcal{X}$,

$$W_{\infty,d}(\lambda_0, \lambda_1) = \min_{\gamma \in \mathsf{cp}(\lambda_0,\lambda_1)} \max_{(x_0,x_1) \in \mathsf{supp}(\gamma)} d(x_0, x_1).$$

The ∞-Wasserstein metric $W_{\infty,d}(\lambda_0, \lambda_1)$ represents the minimum largest move between points in a transportation from λ_0 to λ_1. Specifically, in a transportation γ, $\max_{(x_0,x_1) \in \mathsf{supp}(\gamma)} d(x_0, x_1)$ represents the largest move from a point in λ_0 to another in λ_1. For instance, in the coupling γ in Example 1, the largest move is 1 (from $\lambda_0[2]$ to $\lambda_1[1]$, and from $\lambda_0[2]$ to $\lambda_1[3]$). Such a largest move is minimized by a coupling that achieves the ∞-Wasserstein metric. We denote by $\Gamma_{\infty,d}$ the set of all couplings that achieve the ∞-Wasserstein metric.

Finally, we recall the notion of the lifting of relations.

Definition 7 (Lifting of relations). Given a relation $\Phi \subseteq \mathcal{X} \times \mathcal{X}$, the *lifting* of Φ is the maximum relation $\Phi^{\#} \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$ such that for any $(\lambda_0, \lambda_1) \in \Phi^{\#}$, there exists a coupling $\gamma \in \mathsf{cp}(\lambda_0, \lambda_1)$ satisfying $\mathsf{supp}(\gamma) \subseteq \Phi$.

Note that by Definition 5, the coupling γ is a probability distribution over Φ whose marginal distributions are λ_0 and λ_1. If $\Phi = \mathcal{X} \times \mathcal{X}$, then $\Phi^{\#} = \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$.

3 Privacy Notions for Probability Distributions

In this section we introduce a formal model for the privacy of user attributes, which is motivated in Section 1.

3.1 Modeling the Privacy of User Attributes in Terms of DP

As a running example, we consider an LBS (location based service) in which each user queries an LBS provider for a list of shops nearby. To hide a user's exact location x from the provider, the user applies a randomized algorithm $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$, called a *local obfuscation mechanism*, to her location x, and obtains an approximate information y with the probability $A(x)[y]$.

To illustrate the privacy of attributes, let us consider an example in which users try to prevent an attacker from inferring whether they are *male* or *female* by obfuscating their own exact locations using a mechanism A. For each $t \in \{male, female\}$, let $\lambda_t \in \mathbb{D}\mathcal{X}$ be the prior distribution of the location of the users who have the attribute t. Intuitively, λ_{male} (resp. λ_{female}) represents an attacker's belief on the location of the male (resp. female) users before the attacker observes an output of the mechanism A. Then the privacy of t can be modeled as a property that the attacker has no idea on whether the actual location x follows the distribution λ_{male} or λ_{female} after observing an output y of A.

This can be formalized in terms of ε-local DP. For each $t \in \{male, female\}$, we denote by $p(y \mid \lambda_t)$ the probability of observing an obfuscated location y when an actual location x is distributed over λ_t, i.e., $p(y \mid \lambda_t) = \sum_{x \in \mathcal{X}} \lambda_t[x] A(x)[y]$. Then we can define the privacy of t by:

$$\frac{p(y \mid \lambda_{male})}{p(y \mid \lambda_{female})} \leq e^{\varepsilon}.$$

3.2 Distribution Privacy and Extended Distribution Privacy

We generalize the privacy of attributes (in Section 3.1) and define the notion of *distribution privacy* (DistP) as the differential privacy where the input is a probability distribution of data rather than a value of data. This notion models a level of obfuscation that hides which distribution a data value is drawn from. Intuitively, we say a randomized algorithm A provides DistP if, by observing an output of A, we cannot detect from which distribution an input to A is generated.

Definition 8 (Distribution privacy). Let $\varepsilon \in \mathbb{R}^{\geq 0}$ and $\delta \in [0,1]$. We say that a randomized algorithm $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, δ)-*distribution privacy (DistP) w.r.t.* an adjacency relation $\Psi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$ if its lifting $A^{\#} : \mathbb{D}\mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, δ)-DP w.r.t. Ψ, i.e., for all pairs $(\lambda, \lambda') \in \Psi$ and all $R \subseteq \mathcal{Y}$, we have:

$$A^{\#}(\lambda)[R] \leq e^{\varepsilon} \cdot A^{\#}(\lambda')[R] + \delta.$$

We say A provides (ε, δ)-*DistP w.r.t.* $\Lambda \subseteq \mathbb{D}\mathcal{X}$ if it provides (ε, δ)-DistP w.r.t. Λ^2.

For example, the privacy of a user attribute $t \in \{male, female\}$ described in Section 3.1 can be formalized as $(\varepsilon, 0)$-DistP w.r.t. $\{\lambda_{male}, \lambda_{female}\}$.

Mathematically, DistP is not a new notion but the DP for distributions. To contrast with DistP, we refer to the DP for data values as *point privacy*.

Next we introduce an extended form of distribution privacy to a metric. Intuitively, extended distribution privacy guarantees that when two input distributions are closer, then the output distributions must be less distinguishable.

Definition 9 (Extended distribution privacy). Let $d : (\mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}) \to \mathbb{R}$ be a utility distance, and $\Psi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$. We say that a mechanism $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, d, δ)-*extended distribution privacy (XDistP) w.r.t.* Ψ if the lifting $A^{\#}$ provides (ε, d, δ)-XDP w.r.t. Ψ, i.e., for all $(\lambda, \lambda') \in \Psi$ and all $R \subseteq \mathcal{Y}$, we have:

$$A^{\#}(\lambda)[R] \leq e^{\varepsilon d(\lambda, \lambda')} \cdot A^{\#}(\lambda')[R] + \delta.$$

3.3 Interpretation by Bayes Factor

The interpretation of DP has been explored in previous work [17, 15] using the notion of *Bayes factor*. Similarly, the meaning of DistP can also be explained in terms of Bayes factor, which compares the attacker's prior and posterior beliefs.

Assume that an attacker has some belief on the input distribution before observing the output values of an obfuscater A. We denote by $p(\lambda)$ the prior probability that a distribution λ is chosen as the input distribution. By observing an output y of A, the attacker updates his belief on the input distribution. We denote by $p(\lambda|y)$ the posterior probability of λ being chosen, given an output y.

For two distributions λ_0, λ_1, the Bayes factor $K(\lambda_0, \lambda_1, y)$ is defined as the ratio of the two posteriors divided by that of the two priors: $K(\lambda_0, \lambda_1, y) = \frac{p(\lambda_0|y)}{p(\lambda_1|y)} \big/ \frac{p(\lambda_0)}{p(\lambda_1)}$. If the Bayes factor is far from 1 the attacker significantly updates his belief on the distribution by observing a perturbed output y of A.

Assume that A provides $(\varepsilon, 0)$-DistP. By Bayes' theorem, we obtain:

$$K(\lambda_0, \lambda_1, y) = \frac{p(\lambda_0|y)}{p(\lambda_1|y)} \cdot \frac{p(\lambda_1)}{p(\lambda_0)} = \frac{p(y|\lambda_0)}{p(y|\lambda_1)} = \frac{A^{\#}(\lambda_0)[y]}{A^{\#}(\lambda_1)[y]} \leq e^{\varepsilon}.$$

Intuitively, if the attacker believes that λ_0 is k times more likely than λ_1 before the observation, then he believes that λ_0 is $k \cdot e^{\varepsilon}$ times more likely than λ_1 after the observation. This means that for a small value of ε, DistP guarantees that the attacker does not gain information on the distribution by observing y.

In the case of XDistP, the Bayes factor $K(\lambda_0, \lambda_1, y)$ is bounded above by $e^{\varepsilon d(\lambda_0, \lambda_1)}$. Hence the attacker gains more information for a larger distance $d(\lambda_0, \lambda_1)$.

Table 1: Summary of basic properties of DistP.

Sequential composition $\odot$	A_b is $(\varepsilon_b, \delta_b)$-DistP $\Rightarrow A_1 \odot A_0$ is $(\varepsilon_0 + \varepsilon_1,\ (\delta_0 + \delta_1) \cdot \lvert\Phi\rvert)$-DistP
Sequential composition $\bullet$	A_b is $(\varepsilon_b, \delta_b)$-DistP $\Rightarrow A_1 \bullet A_0$ is $(\varepsilon_0 + \varepsilon_1, \delta_0 + \delta_1)$-DistP
Post-processing	A_0 is (ε, δ)-DistP $\Rightarrow A_1 \circ A_0$ is (ε, δ)-DistP
Pre-processing (by c-stable T)	A is (ε, δ)-DistP $\Rightarrow A \circ T$ is $(c\,\varepsilon, \delta)$-DistP

3.4 Privacy Guarantee for Attackers with Close Beliefs

In the previous sections, we assume that we know the distance between two actual input distributions, and can determine the amount of noise required for distribution obfuscation. However, an attacker may have different beliefs on the distributions that are closer to the actual ones, e.g., more accurate distributions obtained by more observations and specific situations (e.g., daytime/nighttime).

To see this, for each $\lambda \in \mathbb{D}\mathcal{X}$, let $\tilde{\lambda}$ be an attacker's belief on λ. We say that an attacker has (c, d)-*close beliefs* if each distribution λ satisfies $d(\lambda, \tilde{\lambda}) \leq c$. Then extended distribution privacy in the presence of an attacker is given by:

Proposition 1 (XDistP with close beliefs) *Let $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provide $(\varepsilon, d, 0)$-XDistP w.r.t. some $\Psi \subseteq \mathcal{X} \times \mathcal{X}$. If an attacker has (c, d)-close beliefs, then for all $(\lambda_0, \lambda_1) \in \Psi$ and all $R \subseteq \mathcal{Y}$, we have $A^{\#}(\tilde{\lambda_0})[R] \leq e^{\varepsilon(d(\lambda_0,\lambda_1)+2c)} \cdot A^{\#}(\tilde{\lambda_1})[R]$.*

When the attacker's beliefs are closer to ours, then c is smaller, hence a stronger distribution privacy is guaranteed. See Appendix B.2 for DistP. Note that assuming some attacker's beliefs are inevitable also in many previous studies, e.g., when we want to protect the privacy of correlated data [18–20].

3.5 Difference from the Histogram Privacy

Finally, we present a brief remark on the difference between DistP and the *differential privacy of histogram publication* (e.g., [21]). Roughly, a histogram publication mechanism is a *central* mechanism that aims at *hiding a single record* $x \in \mathcal{X}$ and outputs an obfuscated histogram, e.g., a distribution $\mu \in \mathbb{D}\mathcal{Y}$, whereas a DistP mechanism is a *local* mechanism that aims at *hiding an input distribution* $\lambda \in \mathbb{D}\mathcal{X}$ and outputs a single perturbed value $y \in \mathcal{Y}$.

Note that neither of these implies the other. The ε-DP of a histogram publication mechanism means that for any two adjacent inputs $x, x' \in \mathcal{X}$ and any histogram $\mu \in \mathbb{D}\mathcal{Y}$, $\frac{p(\mu|x)}{p(\mu|x')} \leq e^{\varepsilon}$. However, this does not derive ε-DistP, i.e., for any adjacent input distributions $\lambda, \lambda' \in \mathbb{D}\mathcal{X}$ and any output $y \in \mathcal{Y}$, $\frac{p(y|\lambda)}{p(y|\lambda')} \leq e^{\varepsilon}$.

4 Basic Properties of Distribution Privacy

In Table 1, we show basic properties of DistP.(See Appendices B.4 and B.5 for the details.)

The composition $A_1 \odot A_0$ means that an identical input x is given to two DistP mechanisms A_0 and A_1, whereas the composition $A_1 \bullet A_0$ means that independent inputs x_b are provided to mechanisms A_b [22]. The compositionality can be used to quantify the attribute privacy against an attacker who obtains multiple released data each obfuscated for the purpose of protecting a different attribute. For example, let $\Psi = \{(\lambda_{male}, \lambda_{female}), (\lambda_{home}, \lambda_{out})\}$, and A_0 (resp. A_1) be a mechanism providing ε_0-DistP (resp. ε_1-DistP) w.r.t. Ψ. When A_0 (resp. A_1) obfuscates a location x_0 for the sake of protecting *male*/*female* (resp. *home*/*out*), then both *male*/*female* and *home*/*out* are protected with $(\varepsilon_0 + \varepsilon_1)$-DistP.

As for pre-processing, the stability notion is different from that for DP:

Definition 10 (Stability). Let $c \in \mathbb{N}^{>0}$, $\Psi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$, and W be a metric over $\mathbb{D}\mathcal{X}$. A transformation $T : \mathbb{D}\mathcal{X} \to \mathbb{D}\mathcal{X}$ is (c, Ψ)-*stable* if for any $(\lambda_0, \lambda_1) \in \Psi$, $T(\lambda_0)$ can be reached from $T(\lambda_1)$ at most c-steps over Ψ. Analogously, $T : \mathbb{D}\mathcal{X} \to \mathbb{D}\mathcal{X}$ is (c, W)-*stable* if for any $\lambda_0, \lambda_1 \in \mathbb{D}\mathcal{X}$, $W(T(\lambda_0), T(\lambda_1)) \leq c\,W(\lambda_0, \lambda_1)$.

We present relationships among privacy notions in Appendices B.3 and B.7 An important property is that when the relation $\Psi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$ includes pairs of point distributions (i.e., distributions having single points with probability 1), DistP (resp. XDistP) implies DP (resp. XDP). In contrast, if Ψ does not include pairs of point distributions, DistP (resp. XDistP) may not imply DP (resp. XDP), as in Section 6.

5 Distribution Obfuscation by Point Obfuscation

In this section we present how the point obfuscation mechanisms (including DP and XDP mechanisms) contribute to the obfuscation of probability distributions. (See Appendix B.1 for the proofs.)

5.1 Distribution Obfuscation by DP Mechanisms

We first show every DP mechanism provides DistP. (See Definition 7 for $\Phi^{\#}$.)

Theorem 1 ((ε, δ)-DP $\Rightarrow$ $(\varepsilon, \delta \cdot |\Phi|)$-DistP) *Let $\Phi \subseteq \mathcal{X} \times \mathcal{X}$. If $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, δ)-DP w.r.t. Φ, then it provides $(\varepsilon, \delta \cdot |\Phi|)$-DistP w.r.t. $\Phi^{\#}$.*

This means that the mechanism A makes any pair $(\lambda_0, \lambda_1) \in \Phi^{\#}$ indistinguishable up to the threshold ε and with exceptions $\delta \cdot |\Phi|$. Intuitively, when λ_0 and λ_1 are adjacent w.r.t. the relation $\Phi^{\#}$, we can construct λ_1 from λ_0 only by moving mass from $\lambda_0[x_0]$ to $\lambda_1[x_1]$ where $(x_0, x_1) \in \Phi$ (i.e., x_0 is adjacent to x_1).

Example 2 (Randomized response). By Theorem 1, the $(\varepsilon, 0)$-DP randomized response [14] and RAPPOR [4] provide $(\varepsilon, 0)$-DistP. When we use these mechanisms, the estimation of the input distribution is harder for a smaller ε. However, these DP mechanisms tend to have small utility, because they add much noise to hide not only the input distributions, but everything about inputs.

5.2 Distribution Obfuscation by XDP Mechanisms

Compared to DP mechanisms, XDP mechanisms are known to provide better utility. Alvim et al. [23] show the planar Laplace mechanism [3] adds less noise than the randomized response, since XDP hides only closer locations. However, we show XDP mechanisms still need to add much noise proportionally to the ∞-Wasserstein distance between the distributions we want make indistinguishable.

The ∞-Wasserstein distance $W_{\infty,d}$ as Utility Distance We first observe how much ε' is sufficient for an ε'-XDP mechanism (e.g., the Laplace mechanism) to make two distribution λ_0 and λ_1 indistinguishable in terms of ε-DistP.

Suppose that λ_0 and λ_1 are point distributions such that $\lambda_0[x_0] = \lambda_1[x_1] = 1$ for some $x_0, x_1 \in \mathcal{X}$. Then an ε'-XDP mechanism A satisfies:

$$D_\infty(A^\#(\lambda_0) \,\|\, A^\#(\lambda_1)) = D_\infty(A(x_0) \,\|\, A(x_1)) \leq \varepsilon' d(x_0, x_1).$$

In order for A to provide ε-DistP, ε' should be defined as $\frac{\varepsilon}{d(x_0,x_1)}$. That is, the noise added by A should be proportional to the distance between x_0 and x_1.

To extend this to arbitrary distributions, we need to define a utility metric between distributions. A natural possible definition would be the largest distance between values of λ_0 and λ_1, i.e., the *diameter* over the supports defined by:

$$\mathsf{diam}(\lambda_0, \lambda_1) = \max_{x_0 \in \mathsf{supp}(\lambda_0), x_1 \in \mathsf{supp}(\lambda_1)} d(x_0, x_1).$$

However, when there is an outlier in λ_0 or λ_1 that is far from other values in the supports, then the diameter $\mathsf{diam}(\lambda_0, \lambda_1)$ is large. Hence the mechanisms that add noise proportionally to the diameter would lose utility too much.

To have better utility, we employ the ∞-Wasserstein metric $W_{\infty,d}$. The idea is that given two distributions λ_0 and λ_1 over $\mathcal{X}$, we consider the cost of a transportation of weights from λ_0 to λ_1. The transportation is formalized as a coupling γ of λ_0 and λ_1 (see Definition 5), and the cost of the largest move is $\varDelta_{\mathsf{supp}(\gamma),d} = \max_{(x_0,x_1) \in \mathsf{supp}(\gamma)} d(x_0, x_1)$, i.e., the sensitivity w.r.t. the adjacency relation $\mathsf{supp}(\gamma) \subseteq \mathcal{X} \times \mathcal{X}$ (Definition 3). The minimum cost of the largest move is given by the ∞-Wasserstein metric: $W_{\infty,d}(\lambda_0, \lambda_1) = \min_{\gamma \in \mathsf{cp}(\lambda_0,\lambda_1)} \varDelta_{\mathsf{supp}(\gamma),d}$.

XDP implies XDistP We show every XDP mechanism provides XDistP with the metric $W_{\infty,d}$. To formalize this, we define a lifted relation $\varPhi^\#_{W_\infty}$ as the maximum relation over $\mathbb{D}\mathcal{X}$ s.t. for any $(\lambda_0, \lambda_1) \in \varPhi^\#_{W_\infty}$, there is a coupling $\gamma \in \mathsf{cp}(\lambda_0, \lambda_1)$ satisfying $\mathsf{supp}(\gamma) \subseteq \varPhi$ and $\gamma \in \varGamma_{\infty,d}(\lambda_0, \lambda_1)$. Then $\varPhi^\#_{W_\infty} \subseteq \varPhi^\#$ holds.

Theorem 2 (**(ε, d, δ)-XDP$\Rightarrow$($\varepsilon, W_{\infty,d}, \delta\cdot|\varPhi|$)-XDistP**) *If $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, d, δ)-XDP w.r.t. $\varPhi \subseteq \mathcal{X} \times \mathcal{X}$, it provides $(\varepsilon, W_{\infty,d}, \delta\cdot|\varPhi|)$-XDistP w.r.t. $\varPhi^\#_{W_\infty}$.*

Algorithm 1 Tupling mechanism $Q^{\mathsf{tp}}_{k,\nu,A}$

Input: x: input, k: #dummies, ν: distribution of dummies, A: randomized algorithm
Output: $y = (r_1, \ldots, r_i, s, r_{i+1}, \ldots, r_k)$: the output value of the tupling mechanism

$s \xleftarrow{\$} A(x)$; // Draw an obfuscated value s of an input x
$r_1, r_2, \ldots, r_k \xleftarrow{\$} \nu$; // Draw k dummies from a given distribution ν
$i \xleftarrow{\$} \{1, 2, \ldots, k+1\}$; // Draw i to decide the order of the outputs
return $(r_1, \ldots, r_i, s, r_{i+1}, \ldots, r_k)$;

By Theorem 2, when $\delta > 0$, the noise required for obfuscation is proportional to $|\varPhi|$, which is at most the domain size squared $|\mathcal{X}|^2$. This implies that for a larger domain $\mathcal{X}$, the Gaussian mechanism is not suited for distribution obfuscation. We will demonstrate this by experiments in Section 7.4.

In contrast, the Laplace/exponential mechanisms provide $(\varepsilon, W_{\infty,d}, 0)$-$\mathsf{DistP}$. Since $W_{\infty,d}(\lambda_0, \lambda_1) \le \mathsf{diam}(\lambda_0, \lambda_1)$, the noise added proportionally to $W_{\infty,d}$ can be smaller than diam. This implies that obfuscating a distribution requires less noise than obfuscating a set of data. However, the required noise can still be very large when we want to make two distant distributions indistinguishable.

6 Distribution Obfuscation by Random Dummies

In this section we introduce a local mechanism called a *tupling mechanism* to improve the tradeoff between DistP and utility, as motivated in Section 1.

6.1 Tupling Mechanism

We first define the *tupling mechanism* as a local mechanism that obfuscates a given input x by using a point perturbation mechanism A (not necessarily in terms of DP or XDP), and that also adds k random dummies $r_1, r_2, \ldots, r_k$ to the output to obfuscate the input distribution (Algorithm 1). The probability that given an input x, the mechanism $Q^{\mathsf{tp}}_{k,\nu,A}$ outputs $\bar{y}$ is given by $Q^{\mathsf{tp}}_{k,\nu,A}(x)[\bar{y}]$.

6.2 Privacy of the Tupling Mechanism

Next we show that the tupling mechanism provides DistP w.r.t. the following class of distributions. Given $\beta, \eta \in [0,1]$ and $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$, we define $\varLambda_{\beta,\eta,A}$ by:

$$\varLambda_{\beta,\eta,A} = \left\{ \lambda \in \mathbb{D}\mathcal{X} \mid \Pr\left[y \xleftarrow{\$} \mathcal{Y} : A^{\#}(\lambda)[y] \le \beta \right] \ge 1 - \eta \right\}.$$

For instance, a distribution λ satisfying $\max_x \lambda[x] \le \beta$ belongs to $\varLambda_{\beta,0,A}$. (See Proposition 12 in Appendix B.8.)

Theorem 3 (DistP of the tupling mechanism) *Let $k \in \mathbb{N}^{>0}$, ν be the uniform distribution over $\mathcal{Y}$, $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$, and $\beta, \eta \in [0,1]$. Given an $0 < \alpha < \frac{k}{|\mathcal{Y}|}$,* *let* $\varepsilon_\alpha = \ln \frac{k+(\alpha+\beta)\cdot|\mathcal{Y}|}{k-\alpha\cdot|\mathcal{Y}|}$ *and* $\delta_\alpha = 2e^{-\frac{2\alpha^2}{k\beta^2}} + \eta$. *Then the* (k,ν,A)*-tupling mechanism provides* $(\varepsilon_\alpha,\delta_\alpha)$*-*$\mathsf{DistP}$ *w.r.t.* $\Lambda^2_{\beta,\eta,A}$.

This claim states that just adding random dummies achieves DistP without any assumption on A (e.g., A does not have to provide DP). For a smaller range size $|\mathcal{Y}|$ and a larger number k of dummies, we obtain a stronger DistP.

Note that the distributions protected by $Q^{\mathsf{tp}}_{k,\nu,A}$ belong to the set $\Lambda_{\beta,\eta,A}$.

- When $\beta = 1$, $\Lambda_{\beta,\eta,A}$ is the set of all distributions (i.e., $\Lambda_{1,\eta,A} = \mathbb{D}\mathcal{X}$) while ε_α and δ_α tend to be large.
- For a smaller β, the set $\Lambda_{\beta,\eta,A}$ is smaller while ε_α and δ_α are smaller; that is, the mechanism provides a stronger DistP for a smaller set of distributions.
- If A provides ε_A-DP, $\Lambda_{\beta,\eta,A}$ goes to $\mathbb{D}\mathcal{X}$ for $\varepsilon_A \to 0$. More generally, $\Lambda_{\beta,\eta,A}$ is larger when the maximum output probability $\max_y A^{\#}(\lambda)[y]$ is smaller.

In practice, even when ε_A is relatively large, a small number of dummies enables us to provide a strong DistP, as shown by experiments in Section 7.

We note that Theorem 3 may not imply DP of the tupling mechanism, depending on A. For example, suppose that A is the identity function. For small ε_α and δ_α, we have $\beta \ll 1$, hence no point distribution λ (where $\lambda[x] = 1$ for some x) belongs to $\Lambda_{\beta,\eta,A}$, namely, the tupling mechanism does not provide $(\varepsilon_\alpha,\delta_\alpha)$-$\mathsf{DP}$.

6.3 Service Quality Loss and Cost of the Tupling Mechanism

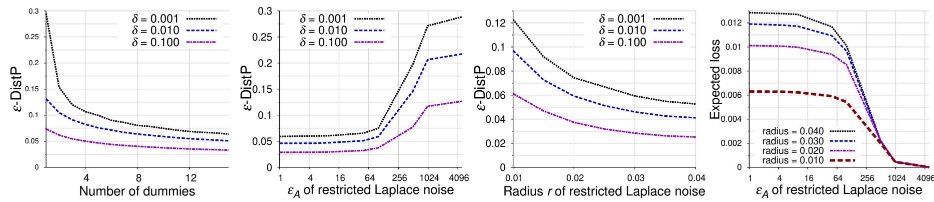
When a mechanism outputs a value y closer to the original input x, she obtains a larger *utility*, or equivalently, a smaller *service quality loss* $d(x,y)$. For example, in an LBS (location based service), if a user located at x submits an obfuscated location y, the LBS provider returns the shops near y, hence the service quality loss can be expressed as the Euclidean distance $d(x,y) \stackrel{\text{def}}{=} \|x-y\|$.

Since each output of the tupling mechanism consists of $k+1$ elements, the quality loss of submitting a tuple $\bar{y} = (y_1, y_2, \ldots, y_{k+1})$ amounts to $d(x,\bar{y})$:=$\min_i d(x,y_i)$. Then the expected quality loss of the mechanism is defined as follows.

Definition 11 (Expected quality loss of the tupling mechanism). For a $\lambda \in \mathbb{D}\mathcal{X}$ and a metric $d : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$, the *expected quality loss* of $Q^{\mathsf{tp}}_{k,\nu,A}$ is:

$$L\big(Q^{\mathsf{tp}}_{k,\nu,A}\big) = \textstyle\sum_{x\in\mathcal{X}} \sum_{\bar{y}\in\mathcal{Y}^{k+1}} \lambda[x]\; Q^{\mathsf{tp}}_{k,\nu,A}(x)[\bar{y}]\; \min_i d(x,y_i).$$

For a larger number k of random dummies, $\min_i d(x,y_i)$ is smaller on average, hence $L\big(Q^{\mathsf{tp}}_{k,\nu,A}\big)$ is also smaller. Furthermore, thanks to the distribution obfuscation by random dummies, we can instead reduce the perturbation noise added to the actual input x to obtain the same level of DistP. Therefore, the service quality is much higher than existing mechanisms, as shown in Section 7.



(a) #dummies and ε-DistP (when using $(100, 0.020)$-RL mechanism).

(b) ε_A of $(\varepsilon_A, 0.020)$-RL mechanism and ε-DistP (with 10 dummies).

(c) A radius r of $(100, r)$-RL mechanism and ε-DistP (with 10 dummies).

(d) ε_A of (ε_A, r)- RL mechanism and the expected loss (with 5 dummies).

Fig. 2: Empirical DistP and quality loss of $Q^{\mathsf{tp}}_{k,\nu,A}$ for the attribute *male/female*.

6.4 Improving the Worst-Case Quality Loss

As a point obfuscation mechanism A used in the tupling mechanism $Q^{\mathsf{tp}}_{k,\nu,A}$, we define the *restricted Laplace (RL) mechanism* below. Intuitively, (ε_A, r)-RL mechanism adds ε_A-XDP Laplace noise only within a radius r of the original location x. This ensures that the worst-case quality loss of the tupling mechanisms is bounded above by the radius r, whereas the standard Laplace mechanism reports a location y that is arbitrarily distant from x with a small probability.

Definition 12 (RL mechanism). Let $\mathcal{Y}_{x,r} = \{y' \in \mathcal{Y} \mid d(x, y') \leq r\}$. We define (ε_A, r)-*restricted Laplace (RL)* mechanism as the $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ defined by: $A(x)[y] = \frac{e^{-\varepsilon d(x,y)}}{\sum_{y' \in \mathcal{Y}_{x,r}} e^{-\varepsilon d(x,y')}}$ if $y \in \mathcal{Y}_{x,r}$, and $A(x)[y] = 0$ otherwise.

Since the support of A is limited to $\mathcal{Y}_{x,r}$, A provides better service quality but does not provide DP. Nevertheless, as shown in Theorem 3, $Q^{\mathsf{tp}}_{k,\nu,A}$ provides DistP, due to dummies in $\mathcal{Y} \setminus \mathcal{Y}_{x,r}$. This implies that DistP is a relaxation of DP that guarantees the privacy of attributes while achieving higher utility by weakening the DP protection of point data. In other words, DistP mechanisms are useful when users want both to keep high utility and to protect the attribute privacy more strongly than what a DP mechanism can guarantee (e.g., when users do not mind revealing their actual locations outside home, but want to hide from robbers the fact that they are outside home, as motivated in Section 1).

7 Application to Attribute Privacy in LBSs

In this section we apply local mechanisms to the protection of the attribute privacy in location based services (LBSs) where each user submits her own location x to an LBS provider to obtain information relevant to x (e.g., shops near x).

7.1 Experimental Setup

We perform experiments on location privacy in Manhattan by using the Foursquare dataset (Global-scale Check-in Dataset) [24]. We first divide Manhattan into

11×10 regions with 1.0km intervals. To provide more useful information to users in crowded regions, we further re-divide these regions to 276 regions by recursively partitioning each crowded region into four until each resulting region has roughly similar population density.[4] Let $\mathcal{Y}$ be the set of those 276 regions, and $\mathcal{X}$ be the set of the 228 regions inside the central 10km $\times$ 9km area in $\mathcal{Y}$.

As an obfuscation mechanism Q, we use the tupling mechanism $Q^{\mathsf{tp}}_{k,\nu,A}$ that uses an (ε_A, r)-RL mechanism A and the uniform distribution ν over $\mathcal{Y}$ to generate dummy locations. Note that ν is close to the population density distribution over $\mathcal{Y}$, because each region in $\mathcal{Y}$ is constructed to have roughly similar population density. In the definitions of the RL mechanism and the quality loss, we use the Euclidean distance $\|\cdot\|$ between the central points of the regions.

In the experiments, we measure the privacy of user attributes, formalized as DistP. For example, let us consider the attribute *male*/*female*. For each $t \in \{\mathit{male}, \mathit{female}\}$, let $\lambda_t \in \mathbb{D}\mathcal{X}$ be the prior distribution of the location of the users having the attribute t. Then, λ_{male} (resp. λ_{female}) represents an attacker's belief on the location of the male (resp. female) users. We define these as the empirical distributions that the attacker can calculate from the above Foursquare dataset.

7.2 Evaluation of the Tupling Mechanism

Distribution privacy We demonstrate by experiments that the *male* users cannot be recognized as which of *male* or *female* in terms of DistP. In Fig. 2, we show the experimental results on the DistP of the tupling mechanism $Q^{\mathsf{tp}}_{k,\nu,A}$. For a larger number k of dummy locations, we have a stronger DistP (Fig. 2a). For a larger ε_A, $(\varepsilon_A, 0.020)$-RL mechanism A adds less noise, hence the tupling mechanism provides a weaker DistP (Fig. 2b)[5]. For a larger radius r, the RL mechanism A spreads the original distribution λ_{male} and thus provides a strong DistP (Fig. 2c). We also show the relationship between k and DistP in the eastern/western Tokyo and London, which have different levels of privacy (Fig. 3).

These results imply that if we add more dummies, we can decrease the noise level/radius of A to have better utility, while keeping the same level ε of DistP. Conversely, if A adds more noise, we can decrease the number k of dummies.

Expected quality loss In Fig. 2d, we show the experimental results on the expected quality loss of the tupling mechanism. For a larger ε_A, A adds less noise, hence the loss is smaller. We confirm that for more dummy data, the expected quality loss is smaller. Unlike the planar Laplace mechanism (PL), A ensures that the worst quality loss is bounded above by the radius r. Furthermore, for a smaller radius r, the expected loss is also smaller as shown in Fig. 2d.

[4] This partition may be useful to achieve smaller values (ε, δ) of DistP, because β tends to be smaller when the population density is closer to the uniform distribution.

[5] In Fig. 2b, for $\varepsilon_A \rightarrow 0$, ε does not converge to 0, since the radius $r = 0.020$ of RL does not cover the whole $\mathcal{Y}$. However, if $r \geq \max_{x,y} \|x - y\|$, ε converges to 0.

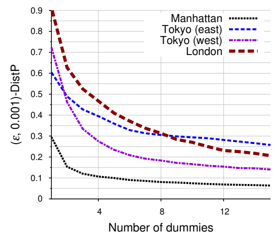


Fig. 3: k and DistP for *male*/*female* in different cities.

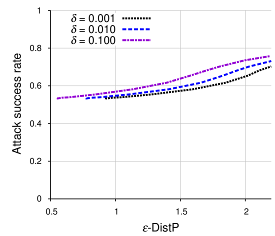


Fig. 4: DistP and ASR of the tupling ($k = 10$, $r = 0.020$).

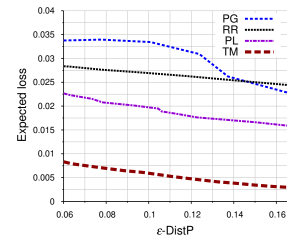


Fig. 5: $(\varepsilon, 0.001)$-DistP and expected loss for *male*/*female* and TM using $k = 10$, $r = 0.020$.

7.3 Appropriate Parameters

We define the *attack success rate (ASR)* as the ratio that the attacker succeeds to infer a user has an attribute when she does actually. We use an inference algorithm based on the Bayes decision rule [25] to minimize the identification error probability when the estimated posterior probability is accurate [25].

In Fig. 4, we show the relationships between DistP and ASR in Manhattan for the attribute *home*, meaning the users located at their home. In theory, ASR $= 0.5$ represents the attacker learns nothing about the attribute, whereas the empirical ASR in our experiments fluctuates around 0.5. This seems to be caused by the fact that the dataset and the number of locations are finite. From Fig. 4, we conclude that $\varepsilon = 1$ is an appropriate parameter for $(\varepsilon, 0.001)$-DistP to achieve ASR $= 0.5$ in our setting, and we confirm this for other attributes. However, we note that this is an empirical criterion possibly depending on our setting, and the choice of ε for DistP can be as controversial as that for DP and should also be investigated using approaches for DP (e.g., [26]) in future work.

In Appendices A.1, A.2, and A.3, we show that different levels of DistP are achieved for different attributes in different cities. In particular, we present appropriate parameters for the tupling mechanism for various attributes in a couple of cities.

7.4 Comparison of Obfuscation Mechanisms

We demonstrate that the tupling mechanism (TM) outperforms the popular mechanisms: the randomized response (RR), the planar Laplace (PL), and the planar Gaussian (PG). In Fig. 5 we compare these concerning the relationship between ε-DistP and expected quality loss. Since PG always has some δ, it provides a weaker DistP than PL for the same quality loss. We also confirm that PL has smaller loss than RR, since it adds noise proportionally to the distance. Comparison using other attributes can be found in Appendix A.1.

Finally, we briefly discuss the computational cost of the tupling mechanism $Q^{\mathsf{tp}}_{k,\nu,A}$, compared to PL. In the implementation, for a larger domain $\mathcal{X}$, PL deals with a larger size $|\mathcal{X}| \times |\mathcal{Y}|$ of the mechanism's matrix, since it outputs each region with a non-zero probability. In contrast, since the RL mechanism A used in $Q^{\mathsf{tp}}_{k,\nu,A}$ maps each location x to a region within a radius r of x, the size of A's matrix is $|\mathcal{X}| \times |\mathcal{Y}_{x,r}|$, requiring much smaller memory space than PL.

Furthermore, the users of TM can simply ignore the responses to dummy queries, whereas the users of PL need to select relevant POIs (point of interests) from a large radius of x, which could cost computationally for many POIs. Therefore, TM is more suited to be used in mobile environments than PL.

8 Related Work

Differential privacy. Since the seminal work of Dwork [1] on DP, a number of its variants have been studied to provide different privacy guarantees; e.g., f-divergence privacy [27], d-privacy [15], Pufferfish privacy [19], local DP [2], and utility-optimized local DP [28]. All of these are intended to protect the input data rather than the input distributions. Note that distribution*al* privacy [29] is different from DistP and does not aim at protecting the privacy of distributions.

To our knowledge, this is the first work that investigates the *differential privacy* of probability distributions lying behind the input. However, a few studies have proposed related notions. Jelasity et al. [30] propose *distributional differential privacy* w.r.t. parameters θ and θ' of two distributions, which aims at protecting the privacy of the distribution parameters but is defined in a Bayesian style (unlike DP and DistP) to satisfy that for any output sequence y, $p(\theta|y) \leq e^{\varepsilon} p(\theta'|y)$. After a preliminary version of this paper appeared in arXiv [31], a notion generalizing DistP, called *profile based privacy*, is proposed in [32].

Some studies are technically related to our work. Song *et al.* [20] propose the Wasserstein mechanism to provide Pufferfish privacy, which protects correlated inputs. Fernandes *et al.* [33] introduce Earth mover's privacy, which is technically different from DistP in that their mechanism obfuscates a vector (a bag-of-words) instead of a distribution, and perturbs each element of the vector. Sei *et al.* [34] propose a variant of the randomized response to protect individual data and provide high utility of database. However, we emphasize again that our work differs from these studies in that we aim at protecting input distributions.

Location privacy. Location privacy has been widely studied in the literature, and its survey can be found in [35]. A number of location obfuscation methods have been proposed so far, and they can be broadly divided into the following four types: perturbation (adding noise) [36, 3, 5], location generalization (merging regions) [37, 38], and location hiding (deleting) [37, 39], and adding dummy locations [40–42]. Location obfuscation based on DP (or its variant) have also been widely studied, and they can be categorized into the ones in the centralized model [43, 44] and the ones in the local model [3, 5]. However, these methods aim at protecting locations, and neither at protecting users' attributes (e.g., age, gender) nor activities (e.g., working, shopping) in a DP manner. Despite the fact that users' attributes and activities can be inferred from their locations [6–8], to our knowledge, no studies have proposed obfuscation mechanisms to provide rigorous DP guarantee for such attributes and activities.

9 Conclusion

We have proposed a formal model for the privacy of probability distributions and introduced the notion of distribution privacy (DistP). Then we have shown that existing local mechanisms deteriorate the utility by adding too much noise to provide DistP. To improve the tradeoff between DistP and utility, we have introduced the tupling mechanism and applied it to the protection of user attributes in LBSs. Then we have demonstrated that the tupling mechanism outperforms popular local mechanisms in terms of attribute obfuscation and service quality.

As future work, we will improve the theoretical bound on ε for the tupling mechanism. We also plan to design optimal mechanisms that minimize the quality loss and computational costs, while providing DistP, in various applications.

Acknowledgment We thank the reviewers, Catuscia Palamidessi, Gilles Barthe, and Frank D. Valencia for their helpful comments on preliminary drafts.

References

1. C. Dwork, "Differential privacy," in *Proc. ICALP*, 2006, pp. 1–12.
2. J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *Proc. FOCS*, 2013, pp. 429–438.
3. M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: differential privacy for location-based systems," in *Proc. CCS*. ACM, 2013, pp. 901–914.
4. Úlfar Erlingsson, V. Pihur, and A. Korolova, "RAPPOR: Randomized aggregatable privacy-preserving ordinal response," in *Proc. CCS*, 2014, pp. 1054–1067.
5. N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Optimal geo-indistinguishable mechanisms for location privacy," in *Proc. CCS*, 2014, pp. 251–262.
6. L. Liao, D. Fox, and henry Kautz, "Extracting places and activities from GPS traces using hierarchical conditional random fields," *Int. J. Robotics Res.*, vol. 1, no. 26, pp. 119–134, 2007.
7. V. W. Zheng, Y. Zheng, and Q. Yang, "Joint learning user's activities and profiles from GPS data," in *Proc. LBSN*, 2009, pp. 17–20.
8. Y. Matsuo, N. Okazaki, K. Izumi, Y. Nakamura, T. Nishimura, and K. Hasida, "Inferring long-term user properties based on users' location history," in *Proc. IJCAI*, 2007, pp. 2159–2165.
9. D. Yang, B. Qu, and P. Cudré-Mauroux, "Privacy-preserving social media data publishing for personalized ranking-based recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 3, pp. 507–520, 2019.
10. J. Otterbacher, "Inferring gender of movie reviewers: Exploiting writing style, content and metadata," in *Proc. CIKM*, 2010, pp. 369–378.
11. U. Weinsberg, S. Bhagat, S. Ioannidis, and N. Taft, "BlurMe: Inferring and obfuscating user gender based on ratings," in *Proc. RecSys*, 2012, pp. 195–202.
12. N. Z. Gong and B. Liu, "Attribute inference attacks in online social networks," *ACM Trans. Privacy and Security*, vol. 21, no. 1, pp. 3:1–3:30, 2018.
13. A. Mislove, B. Viswanath, P. K. Gummadi, and P. Druschel, "You are who you know: inferring user profiles in online social networks," in *Proc. WSDM*, 2010, pp. 251–260.

14. P. Kairouz, K. Bonawitz, and D. Ramage, "Discrete distribution estimation under local privacy," in *Proc. ICML*, 2016, pp. 2436–2444.
15. K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe, and C. Palamidessi, "Broadening the scope of Differential Privacy using metrics," in *Proc. PETS*, 2013, pp. 82–102.
16. L. Vaserstein, "Markovian processes on countable space product describing large systems of automata." *Probl. Peredachi Inf.*, vol. 5, no. 3, pp. 64–72, 1969.
17. C. Dwork, F. Mcsherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. TCC*, 2006, pp. 265–284.
18. D. Kifer and A. Machanavajjhala, "No free lunch in data privacy," in *Proc. SIGMOD*, 2011, pp. 193–204.
19. ——, "A rigorous and customizable framework for privacy," in *Proc. PODS*, 2012, pp. 77–88.
20. S. Song, Y. Wang, and K. Chaudhuri, "Pufferfish privacy mechanisms for correlated data," in *Proc. SIGMOD*, 2017, pp. 1291–1306.
21. J. Xu, Z. Zhang, X. Xiao, Y. Yang, G. Yu, and M. Winslett, "Differentially private histogram publication," *VLDB J.*, vol. 22, no. 6, pp. 797–822, 2013.
22. Y. Kawamoto, K. Chatzikokolakis, and C. Palamidessi, "On the compositionality of quantitative information flow," *Log. Methods Comput. Sci.*, vol. 13, no. 3, 2017.
23. M. S. Alvim, K. Chatzikokolakis, C. Palamidessi, and A. Pazii, "Invited paper: Local differential privacy on metric spaces: Optimizing the trade-off with utility," in *Proc. CSF*, 2018, pp. 262–267.
24. D. Yang, D. Zhang, and B. Qu, "Participatory cultural mapping based on collective behavior data in location based social networks," *ACM Transactions on Intelligent Systems and Technology*, vol. 7, no. 3, pp. 30:1–30:23, 2015.
25. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification.* Wiley-Interscience, 2000.
26. J. Hsu, M. Gaboardi, A. Haeberlen, S. Khanna, A. Narayan, B. C. Pierce, and A. Roth, "Differential privacy: An economic method for choosing epsilon," in *Proc. CSF*, 2014, pp. 398–410.
27. G. Barthe and F. Olmedo, "Beyond differential privacy: Composition theorems and relational logic for f-divergences between probabilistic programs," in *Proc. ICALP*, ser. LNCS, vol. 7966, 2013, pp. 49–60.
28. T. Murakami and Y. Kawamoto, "Utility-optimized local differential privacy mechanisms for distribution estimation," in *Proc. USENIX Security*, 2019, to appear.
29. A. Blum, K. Ligett, and A. Roth, "A learning theory approach to noninteractive database privacy," *J. ACM*, vol. 60, no. 2, pp. 12:1–12:25, 2013.
30. M. Jelasity and K. P. Birman, "Distributional differential privacy for large-scale smart metering," in *Proc. IH&MMSec*, 2014, pp. 141–146.
31. Y. Kawamoto and T. Murakami, "Local obfuscation mechanisms for hiding probability distributions," *CoRR*, vol. abs/1812.00939, 2018. [Online]. Available: http://arxiv.org/abs/1812.00939
32. J. Geumlek and K. Chaudhuri, "Profile-based privacy for locally private computations," *CoRR*, vol. abs/1903.09084, 2019.
33. N. Fernandes, M. Dras, and A. McIver, "Generalised differential privacy for text document processing," in *Proc. POST*, 2019, pp. 123–148.
34. Y. Sei and A. Ohsuga, "Differential private data collection and analysis based on randomized multiple dummies for untrusted mobile crowdsensing," *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 4, pp. 926–939, 2017.

35. K. Chatzikokolakis, E. ElSalamouny, C. Palamidessi, and P. Anna, "Methods for location privacy: A comparative overview," *Foundations and Trends®in Privacy and Security*, vol. 1, no. 4, pp. 199–257, 2017.
36. R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. L. Boudec, "Protecting location privacy: optimal strategy against localization attacks," in *Proc. CCS*. ACM, 2012, pp. 617–627.
37. R. Shokri, G. Theodorakopoulos, J.-Y. L. Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *Proc. S&P*. IEEE, 2011, pp. 247–262.
38. M. Xue, P. Kalnis, and H. Pung, "Location diversity: Enhanced privacy protection in location based services," in *Proc. LoCA*. Springer, 2009, pp. 70–87.
39. B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Preserving privacy in gps traces via uncertainty-aware path cloaking," in *Proc. CCS*. ACM, 2007, pp. 161–171.
40. V. Bindschaedler and R. Shokri, "Synthesizing plausible privacy-preserving location traces," in *Proc. S&P*, 2016, pp. 546–563.
41. R. Chow and P. Golle, "Faking contextual data for fun, profit, and privacy," in *Proc. PES*. ACM, 2009, pp. 105–108.
42. H. Kido, Y. Yanagisawa, and T. Satoh, "Protection of location privacy using dummies for location-based services," in *Proc. ICDE Workshops*, 2005, p. 1248.
43. A. Machanavajjhala, D. Kifer, J. M. Abowd, J. Gehrke, and L. Vilhuber, "Privacy: Theory meets practice on the map," in *Proc. ICDE*. IEEE, 2008, pp. 277–286.
44. S.-S. Ho and S. Ruan, "Differential privacy for location pattern mining," in *Proc. SPRINGL*. ACM, 2011, pp. 17–24.
45. Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui, "Exploring millions of footprints in location sharing services," in *Proc. ICWSM*, 2011.
46. Please Rob Me - Raising Awareness about Over-sharing, http://pleaserobme.com/.
47. Y. Kawamoto and T. Murakami, "Local distribution obfuscation via probability coupling," in *Proc. Allerton*, 2019, to appear.
48. C. Dwork, G. N. Rothblum, and S. P. Vadhan, "Boosting and differential privacy," in *Proc. FOCS*, 2010, pp. 51–60.

A Experimental Results

A.1 Experimental Comparison among Attributes in Manhattan

In this section we present the experimental results on more attributes of users in the location data in Manhattan. We investigate the following four attributes:

- *social/less-social* (Fig. 9) represent whether a user's social status (defined in [45] as the number of followers divided by the number of followings) is greater than 5 or not. In [9], the social status is regarded as a private attribute that can be leaked by the user's behaviour and should be protected. We set 5 to be a threshold, because mobility patterns of people are substantially different when their social status is greater than 5 [45].
- *workplace/non-workplace* (Fig. 10) represent whether a user is at office or not. This attribute can be sensitive when it implies users are unemployed.
- *home/out* (Fig. 11) represent whether a user is at home or not. This attribute should be protected, for instance, from robbers [46].
- *north/south* (Fig. 12) represent whether a user's home is located in the northern or southern Manhattan. This attribute needs to be protected, for instance, from stalkers. However, the residential area is highly correlated with visited places, hence can be inferred by the current location relatively easily.

Note that these figures also show the results of KL-DistP, which is not defined in this paper but is introduced in [47].

Compared to the attribute of *male/female* (Fig. 2), the attributes *home/out* (Fig. 11) and *north/south* (Fig. 12) require more noise for distribution obfuscation. This is because the distances between the two distributions of users of these attributes are larger than that of λ_{male} and λ_{female}. The histograms for *male/female* and for *north/south* are shown in Fig. 7 and Fig. 8 respectively.

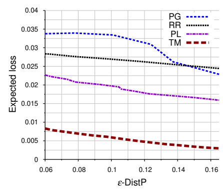
We also show the comparison of different obfuscation mechanisms for various attributes in Fig. 6.

A.2 Experimental Comparison among Time Periods in Manhattan

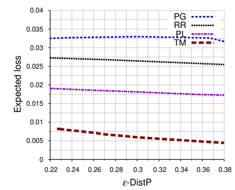
In this section we present the experimental results on different periods of time: 00h-05h, 06h-11h, 12h-17h, and 18h-23h in Manhattan in Figs. 13 to 16. It can be seen that ε depends on a period of time. For example, hiding the attribute *male/female* and *social/less-social* requires more noise in 06h-11h. This might be related to the fact that more social people tend to attend social events in the morning. On the other hand, the attribute *home/out* requires similar amount of noise in any period of time. This implies that residential areas are distant from other areas in each city, and this fact does not change over time.

A.3 Experimental Comparison among Various Cities

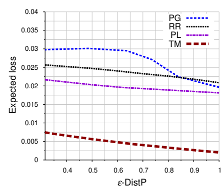
In this section we compare the experimental results on five cities: Manhattan, eastern Tokyo, western Tokyo, London, and Paris. In Table 3 we show examples



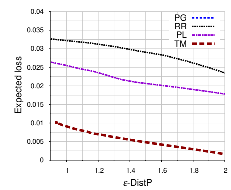
(a) Relationship between $(\varepsilon, 0.001)$-DistP and expected loss for *male/female*.



(b) Relationship between $(\varepsilon, 0.001)$-DistP and expected loss for *social/less-social*.



(c) Relationship between $(\varepsilon, 0.001)$-DistP and loss for *workplace/non-workplace*.



(d) Relationship between $(\varepsilon, 0.001)$-DistP and expected loss for *home/out*.

Fig. 6: Comparison of the randomized response (RR), the planar Laplace mechanism (PL), the planar Gaussian mechanism (PG), and the tupling mechanism (TM) $Q^{\mathsf{tp}}_{k,\nu,A}$ with $k = 10$ dummies and a radius $r = 0.020$. The experiments are performed for the location data in Manhattan.

of parameters that achieve the same levels of DistP in different cities. More detailed comparison among those cities is shown in Fig. 17 (*male/female*), Fig. 18 (*social/less-social*), Fig. 19 (*workplace/non-workplace*), and Fig. 20 (*home/out*).

When we use the same parameters in the tupling mechanism, the levels ε of DistP differ among those cities. In western Tokyo, for instance, the attributes *social/less-social* and *workplace/non-workplace* are more difficult to hide. This implies that areas for social events and workplace might be more separated from the other areas in western Tokyo. Since each attribute in each city may require different levels of noise for distribution obfuscation, the reference parameters (e.g., those shown in Table 3) would be useful to select appropriate parameters for the tupling mechanism to protect the privacy of attributes in different cities.

A.4 Theoretical/Empirical Values of ε-DistP

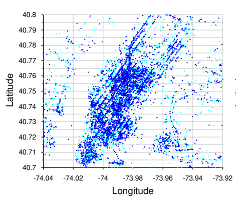
In Table 2, we show the theoretical values of ε calculated by Theorem 3 for $\delta = 0.001, 0.01, 0.1$. Compared to the experimental results, those theoretical values can only give loose upper bounds on ε. This is because the concentration inequality used to derive Theorem 3 give loose bounds.

Table 2: Theoretical/empirical ε-DistP of $Q^{\mathsf{tp}}_{k,\nu,A}$ ($k = 10$, $\varepsilon_A = 10$, $r = 0.020$).

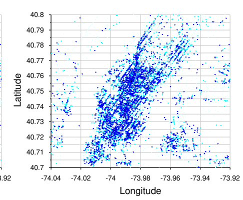
	$\delta = 0.001$	$\delta = 0.01$	$\delta = 0.1$
Theoretical bounds	2.170	1.625	1.140
Empirical values	0.04450	0.03534	0.02295

Table 3: The number k of dummies required for achieving DistP in different cities (MH = Manhattan, TKE = Tokyo (east), TKW = Tokyo (west), LD = London, PR = Paris) when $\varepsilon_A = 100$ and $r = 0.020$. Note that the data of Paris for *male*/*female* are excluded because of the insufficient sample size.

	MH	**TKE**	**TKW**	**LD**	**PR**
$(0.25, 0.001)$-DistP for **male / female**	2	>20	5	10	—
$(0.50, 0.001)$-DistP for **social / less social**	2	3	>20	2	3
$(1.00, 0.001)$-DistP for **work / non-work**	2	2	>20	1	2
$(1.50, 0.001)$-DistP for **home / outside**	3	5	>20	>20	4

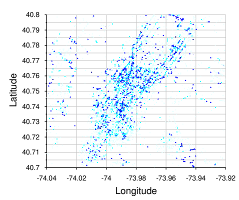


(e) The histogram of locations of the male users in Manhattan.

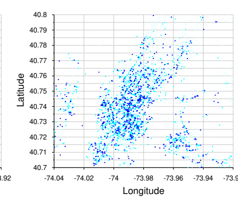


(f) The histogram of locations of the female users in Manhattan.

Fig. 7: Histograms of user locations used in experiments. Darker colors represent larger population.

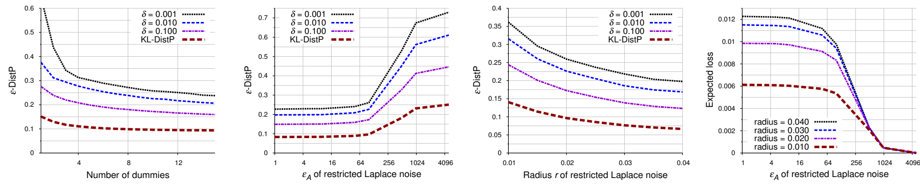


(a) The histogram of locations of the users living in northern Manhattan.



(b) The histogram of locations of the users living in southern Manhattan.

Fig. 8: Histograms of user locations used in experiments. Darker colors represent larger population.



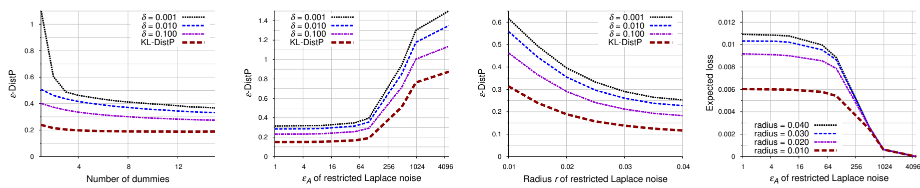
(a) Relationship between ε-DistP and #dummies (when using $(100, 0.020)$-RL mechanism).

(b) Relationship between ε-DistP and ε_A of $(\varepsilon_A, 0.020)$-RL mechanism (with 10 dummies).

(c) Relationship between ε-DistP and a radius r of $(100, r)$-RL mechanism (with 10 dummies).

(d) Relationship between the expected loss and ε_A of (ε_A, r)-RL mechanism (with 5 dummies).

Fig. 9: Empirical DistP and loss for attribute *social*/*less-social* in Manhattan.



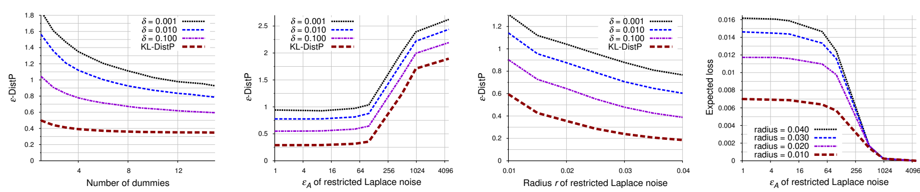
(a) Relationship between ε-DistP and #dummies (when using $(100, 0.020)$-RL mechanism).

(b) Relationship between ε-DistP and ε_A of $(\varepsilon_A, 0.020)$-RL mechanism (with 10 dummies).

(c) Relationship between ε-DistP and a radius r of $(100, r)$-RL mechanism (with 10 dummies).

(d) Relationship between the expected loss and ε_A of (ε_A, r)-RL mechanism (with 5 dummies).

Fig. 10: Empirical DistP and loss for *workplace*/*non-workplace* in Manhattan.



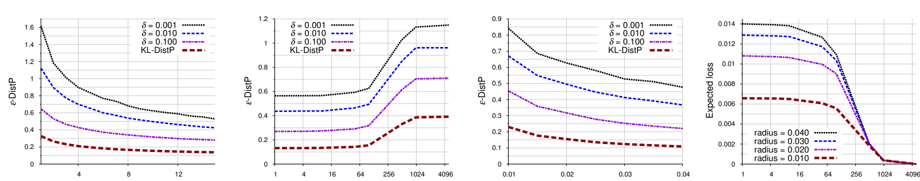
(a) Relationship between ε-DistP and #dummies (when using $(100, 0.020)$-RL mechanism).

(b) Relationship between ε-DistP and ε_A of $(\varepsilon_A, 0.020)$-RL mechanism (with 10 dummies).

(c) Relationship between ε-DistP and a radius r of $(100, r)$-RL mechanism (with 10 dummies).

(d) Relationship between the expected loss and ε_A of (ε_A, r)-RL mechanism (with 5 dummies).

Fig. 11: Empirical DistP and loss for the attribute *home*/*out* in Manhattan.



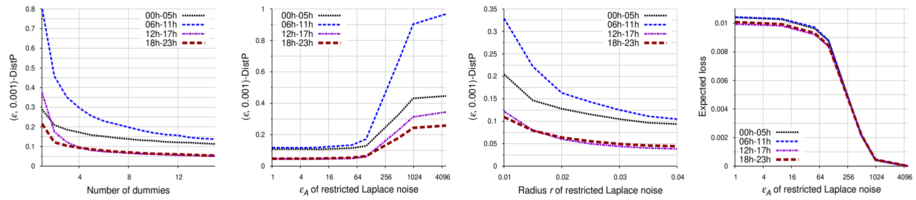
(a) Relationship between ε-DistP and #dummies (when using $(100, 0.020)$-RL mechanism).

(b) Relationship between ε-DistP and ε_A of $(\varepsilon_A, 0.020)$-RL mechanism (with 10 dummies).

(c) Relationship between ε-DistP and a radius r of $(100, r)$-RL mechanism (with 10 dummies).

(d) Relationship between the expected loss and ε_A of (ε_A, r)-RL mechanism (with 5 dummies).

Fig. 12: Empirical DistP and loss for the attribute *north*/*south* in Manhattan.



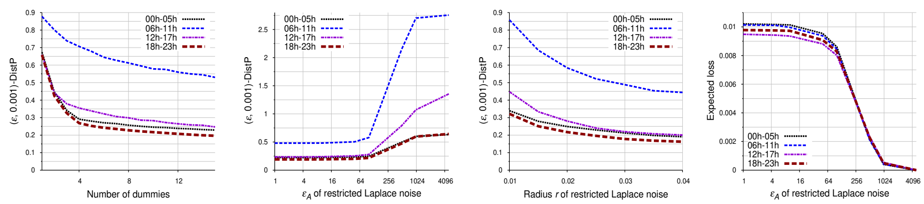
(a) Relationship between ε-DistP and #dummies (when using $(100, 0.020)$-RL mechanism).

(b) Relationship between ε-DistP and ε_A of $(\varepsilon_A, 0.020)$-RL mechanism (with 10 dummies).

(c) Relationship between ε-DistP and a radius r of $(100, r)$-RL mechanism (with 10 dummies).

(d) Relationship between the expected loss and ε_A of (ε_A, r)-RL mechanism (with 5 dummies).

Fig. 13: Empirical DistP and loss for *male*/*female* in different hours.



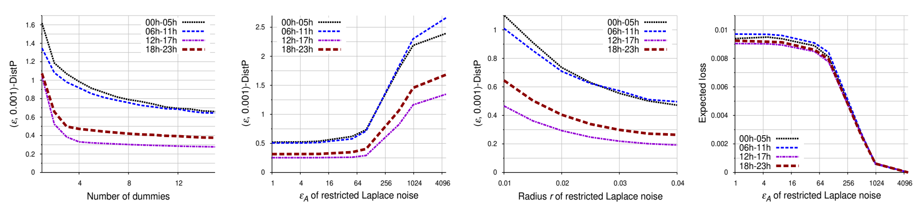
(a) Relationship between ε-DistP and #dummies (when using $(100, 0.020)$-RL mechanism).

(b) Relationship between ε-DistP and ε_A of $(\varepsilon_A, 0.020)$-RL mechanism (with 10 dummies).

(c) Relationship between ε-DistP and a radius r of $(100, r)$-RL mechanism (with 10 dummies).

(d) Relationship between the expected loss and ε_A of (ε_A, r)-RL mechanism (with 5 dummies).

Fig. 14: Empirical DistP and loss for *social*/*less-social* in different hours.



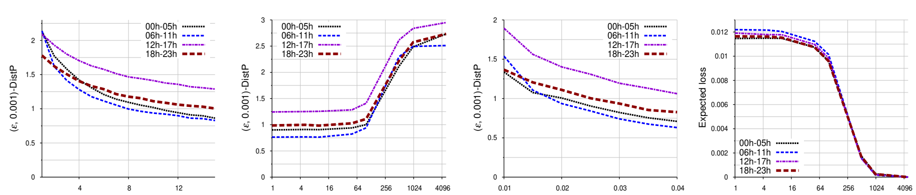
(a) Relationship between ε-DistP and #dummies (when using $(100, 0.020)$-RL mechanism).

(b) Relationship between ε-DistP and ε_A of $(\varepsilon_A, 0.020)$-RL mechanism (with 10 dummies).

(c) Relationship between ε-DistP and a radius r of $(100, r)$-RL mechanism (with 10 dummies).

(d) Relationship between the expected loss and ε_A of (ε_A, r)-RL mechanism (with 5 dummies).

Fig. 15: Empirical DistP and loss for *workplace*/*non-workplace* in different hours.



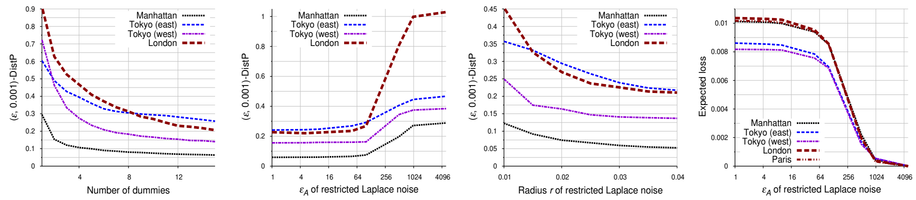
(a) Relationship between ε-DistP and #dummies (when using $(100, 0.020)$-RL mechanism).

(b) Relationship between ε-DistP and ε_A of $(\varepsilon_A, 0.020)$-RL mechanism (with 10 dummies).

(c) Relationship between ε-DistP and a radius r of $(100, r)$-RL mechanism (with 10 dummies).

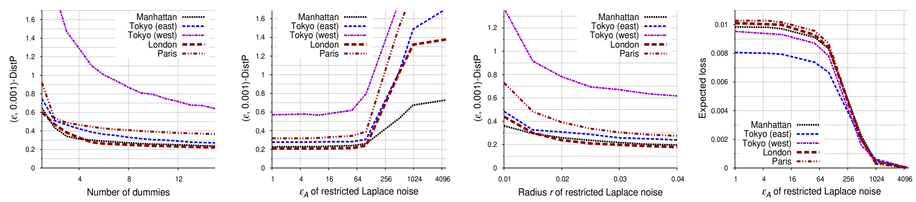
(d) Relationship between the expected loss and ε_A of (ε_A, r)-RL mechanism (with 5 dummies).

Fig. 16: Empirical DistP and loss for *home*/*out* in different hours.



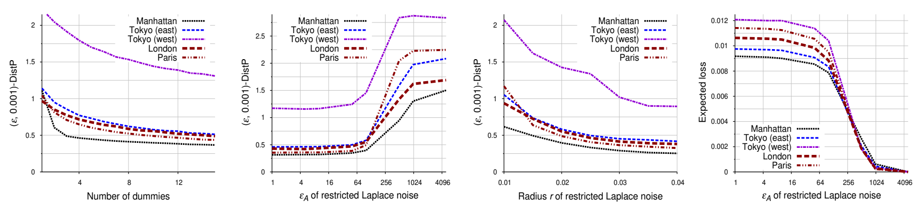
(a) Relationship between ε-DistP and #dummies (when using $(100, 0.020)$-RL mechanism).
(b) Relationship between ε-DistP and ε_A of $(\varepsilon_A, 0.020)$-RL mechanism (with 10 dummies).
(c) Relationship between ε-DistP and a radius r of $(100, r)$-RL mechanism (with 10 dummies).
(d) Relationship between the expected loss and ε_A of (ε_A, r)-RL mechanism (with 5 dummies).

Fig. 17: Empirical DistP and loss for *male*/*female* in different cities.



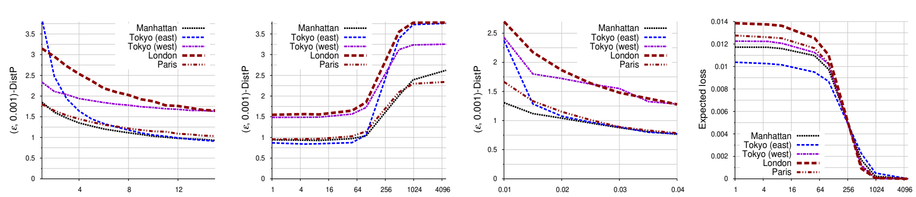
(a) Relationship between ε-DistP and #dummies (when using $(100, 0.020)$-RL mechanism).
(b) Relationship between ε-DistP and ε_A of $(\varepsilon_A, 0.020)$-RL mechanism (with 10 dummies).
(c) Relationship between ε-DistP and a radius r of $(100, r)$-RL mechanism (with 10 dummies).
(d) Relationship between the expected loss and ε_A of (ε_A, r)-RL mechanism (with 5 dummies).

Fig. 18: Empirical DistP and loss for *social*/*less-social* in different cities.



(a) Relationship between ε-DistP and #dummies (when using $(100, 0.020)$-RL mechanism).
(b) Relationship between ε-DistP and ε_A of $(\varepsilon_A, 0.020)$-RL mechanism (with 10 dummies).
(c) Relationship between ε-DistP and a radius r of $(100, r)$-RL mechanism (with 10 dummies).
(d) Relationship between the expected loss and ε_A of (ε_A, r)-RL mechanism (with 5 dummies).

Fig. 19: Empirical DistP and loss for *workplace*/*non-workplace* in different cities.



(a) Relationship between ε-DistP and #dummies (when using $(100, 0.020)$-RL mechanism).
(b) Relationship between ε-DistP and ε_A of $(\varepsilon_A, 0.020)$-RL mechanism (with 10 dummies).
(c) Relationship between ε-DistP and a radius r of $(100, r)$-RL mechanism (with 10 dummies).
(d) Relationship between the expected loss and ε_A of (ε_A, r)-RL mechanism (with 5 dummies).

Fig. 20: Empirical DistP and loss for *home*/*out* in different cities.

B Proofs of Technical Results

B.1 Distribution Obfuscation by Point Obfuscation

In this section we show the results on how point obfuscation mechanisms provide distribution privacy.

Theorem 1 **((ε, δ)-DP $\Rightarrow$ $(\varepsilon, \delta \cdot |\Phi|)$-DistP)** *Let $\Phi \subseteq \mathcal{X} \times \mathcal{X}$. If $A : \mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}$ provides (ε, δ)-DP w.r.t. Φ, then it provides $(\varepsilon, \delta \cdot |\Phi|)$-DistP w.r.t. $\Phi^{\#}$.*

Proof. Assume A provides (ε, δ)-DP w.r.t. Φ. Let $\delta' = \delta \cdot |\Phi|$, and $(\lambda_0, \lambda_1) \in \Phi^{\#}$. By Definition 7, there is a coupling $\gamma \in \mathsf{cp}(\lambda_0, \lambda_1)$ satisfying $\mathsf{supp}(\gamma) \subseteq \Phi$. By Definition 5, we have:

$$\text{for each } x_0 \in \mathcal{X}, \quad \lambda_0[x_0] = \textstyle\sum_{x_1 \in \mathcal{X}} \gamma[x_0, x_1] \tag{1}$$

$$\text{for each } x_1 \in \mathcal{X}, \quad \lambda_1[x_1] = \textstyle\sum_{x_0 \in \mathcal{X}} \gamma[x_0, x_1]. \tag{2}$$

Let $R \subseteq \mathcal{Y}$ such that $A^{\#}(\lambda_0)[R] > \delta'$. We first show that $A^{\#}(\lambda_1)[R] > 0$ as follows. To derive a contradiction, we assume $A^{\#}(\lambda_1)[R] = 0$. Then for any $x_1 \in \mathsf{supp}(\lambda_1)$, $A(x_1)[R] = 0$. Let $S_{x_1} = \{x_0 \in \mathcal{X} \mid (x_0, x_1) \in \Phi\}$. For any $x_0 \in S_{x_1}$, we have:

$$\begin{aligned} A(x_0)[R] &\leq e^{\varepsilon} A(x_1)[R] + \delta && \text{(by the } (\varepsilon, \delta)\text{-DP of } A \text{ w.r.t. } \Phi) \\ &\leq \delta && \text{(by } A(x_1)[R] = 0) \end{aligned}$$

Thus we obtain:

$$\begin{aligned} A^{\#}(\lambda_0)[R] &= \sum_{x_0 \in \mathsf{supp}(\lambda_0)} \lambda_0[x_0] A(x_0)[R] \\ &\leq |\mathsf{supp}(\lambda_0)| \cdot \delta \\ &\leq |\Phi| \cdot \delta \qquad \text{(by } |\mathsf{supp}(\lambda_0)| \leq |\mathsf{supp}(\gamma)| \leq |\Phi|) \\ &= \delta' \end{aligned}$$

This contradicts the definition of R. Hence $A^{\#}(\lambda_1)[R] > 0$.

Then we calculate the ratio of the probability that $A^{\#}$ outputs an element of R given input λ_0 to that given input λ_1:

$$\begin{aligned} &\frac{A^{\#}(\lambda_0)[R] - \delta'}{A^{\#}(\lambda_1)[R]} \\ =&\frac{\left(\sum_{x_0 \in \mathcal{X}} \lambda_0[x_0] \cdot A(x_0)[R]\right) - \delta'}{\sum_{x_1 \in \mathcal{X}} \lambda_1[x_1] \cdot A(x_1)[R]} \\ =&\frac{\sum_{(x_0, x_1) \in \mathsf{supp}(\gamma)} \left(\gamma[x_0, x_1] A(x_0)[R] - \frac{\delta'}{|\mathsf{supp}(\gamma)|}\right)}{\sum_{(x_0, x_1) \in \mathsf{supp}(\gamma)} \gamma[x_0, x_1] A(x_1)[R]} \quad \text{(by (1), (2))} \\ \leq& \max_{(x_0, x_1) \in \mathsf{supp}(\gamma)} \frac{\gamma[x_0, x_1] \cdot A(x_0)[R] - \frac{\delta'}{|\mathsf{supp}(\gamma)|}}{\gamma[x_0, x_1] \cdot A(x_1)[R]} \end{aligned}$$

$$\leq \max_{(x_0,x_1)\in\mathsf{supp}(\gamma)} \frac{A(x_0)[R]-\delta}{A(x_1)[R]} \qquad \left(\text{by } -\tfrac{\delta\cdot|\Phi|}{\gamma[x_0,x_1]\cdot|\mathsf{supp}(\gamma)|}\leq -\delta\right)$$

$$\leq e^{\varepsilon}. \qquad (\text{by } (x_0,x_1)\in\mathsf{supp}(\gamma)\subseteq\Phi \text{ and } (\varepsilon,\delta)\text{-}\mathsf{DP} \text{ w.r.t. } \Phi)$$

Therefore A provides $(\varepsilon,\, \delta\cdot|\Phi|)$-DistP w.r.t. $\Phi^{\#}$. □

Theorem 2 ((ε,d,δ)-XDP$\Rightarrow$$(\varepsilon,\, W_{\infty,d},\delta\cdot|\Phi|)$-XDistP) *If $A:\mathcal{X}\to\mathbb{D}\mathcal{Y}$ provides (ε,d,δ)-XDP w.r.t. $\Phi\subseteq\mathcal{X}\times\mathcal{X}$, it provides $(\varepsilon,\, W_{\infty,d},\delta\cdot|\Phi|)$-XDistP w.r.t. $\Phi^{\#}_{W_\infty}$.*

Proof. Assume that A provides (ε,d,δ)-XDP w.r.t. Φ. Let $(\lambda_0,\lambda_1)\in\Phi^{\#}_{W_\infty}$. By definition, there exists a coupling $\gamma\in\mathsf{cp}(\lambda_0,\lambda_1)$ that satisfies $\mathsf{supp}(\gamma)\subseteq\Phi$ and $\gamma\in\Gamma_{\infty,d}(\lambda_0,\lambda_1)$. Then γ minimizes the sensitivity $\Delta_{\mathsf{supp}(\gamma)}$; i.e., $\gamma\in\operatorname{argmin}_{\gamma'\in\mathsf{cp}(\lambda_0,\lambda_1)}\Delta_{\mathsf{supp}(\gamma')}$.

Let $\delta'=\delta\cdot|\Phi|$. Let $R\subseteq\mathcal{Y}$ such that $A^{\#}(\lambda_0)[R]>\delta'$. Analogously to the proof for Theorem 1, we obtain $A^{\#}(\lambda_1)[R]>0$. Then it follows from $\mathsf{supp}(\gamma)\subseteq\Phi$ and (ε,d,δ)-XDP w.r.t. Φ that:

$$\begin{aligned}\frac{A^{\#}(\lambda_0)[R]-\delta'}{A^{\#}(\lambda_1)[R]} &\leq \max_{(x_0,x_1)\in\mathsf{supp}(\gamma)}\frac{A(x_0)[R]-\delta}{A(x_1)[R]}\\ &\leq \max_{(x_0,x_1)\in\mathsf{supp}(\gamma)} e^{\varepsilon\, d(x_0,x_1)}\\ &= e^{\varepsilon\, W_{\infty,d}(\lambda_0,\lambda_1)}. \quad (\text{by } \gamma\in\Gamma_{\infty,d}(\lambda_0,\lambda_1))\end{aligned}$$

Therefore A provides $(\varepsilon,\, W_{\infty,d},\delta')$-XDistP w.r.t. $\Phi^{\#}_{W_\infty}$. □

B.2 Distribution Privacy with Attacker's Close Beliefs

Next we show the propositions on distribution privacy with an attacker's beliefs.

Proposition 1 (XDistP with close beliefs) *Let $A:\mathcal{X}\to\mathbb{D}\mathcal{Y}$ provide $(\varepsilon,d,0)$-XDistP w.r.t. some $\Psi\subseteq\mathcal{X}\times\mathcal{X}$. If an attacker has (c,d)-close beliefs, then for all $(\lambda_0,\lambda_1)\in\Psi$ and all $R\subseteq\mathcal{Y}$, we have $A^{\#}(\tilde{\lambda_0})[R]\leq e^{\varepsilon(d(\lambda_0,\lambda_1)+2c)}\cdot A^{\#}(\tilde{\lambda_1})[R]$.*

Proof. We obtain the proposition by:

$$\begin{aligned}A^{\#}(\tilde{\lambda_0})[R] &\leq e^{\varepsilon c}\cdot A^{\#}(\lambda_0)[R] && (\text{by } d(\lambda_0,\tilde{\lambda_0})\leq c)\\ &\leq e^{\varepsilon(d(\lambda_0,\lambda_1)+c)}\cdot A^{\#}(\lambda_1)[R]\\ &\leq e^{\varepsilon(d(\lambda_0,\lambda_1)+2c)}\cdot A^{\#}(\tilde{\lambda_1})[R] && (\text{by } d(\lambda_1,\tilde{\lambda_1})\leq c).\end{aligned}$$

□

Similarly, we say that the attacker has *close beliefs* w.r.t an adjacency relation Ψ if for each $(\lambda_0,\lambda_1)\in\Psi$, we have $(\tilde{\lambda_0},\lambda_0)\in\Psi$ and $(\lambda_1,\tilde{\lambda_1})\in\Psi$. Then we obtain:

Proposition 2 (DistP with close beliefs) *Let $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provide (ε, δ)-DistP w.r.t. an adjacency relation $\Psi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$. Let $\delta' = (1 + e^{\varepsilon} + e^{2\varepsilon})\delta$. If an attacker has close beliefs w.r.t. Ψ, then we obtain $(3\varepsilon, \delta')$-DistP against the attacker, i.e., for all $(\lambda_0, \lambda_1) \in \Psi$ and all $R \subseteq \mathcal{Y}$, we have:*

$$A^{\#}(\tilde{\lambda_0})[R] \leq e^{3\varepsilon} \cdot A^{\#}(\tilde{\lambda_1})[R] + \delta'.$$

Proof. We obtain the proposition by:

$$\begin{aligned}
A^{\#}(\tilde{\lambda_0})[R] &\leq e^{\varepsilon} \cdot A^{\#}(\lambda_0)[R] + \delta && \text{(by } (\tilde{\lambda_0}, \lambda_0) \in \Psi) \\
&\leq e^{\varepsilon}(e^{\varepsilon} \cdot A^{\#}(\lambda_1)[R] + \delta) + \delta && \text{(by } (\lambda_0, \lambda_1) \in \Psi) \\
&\leq e^{\varepsilon}(e^{2\varepsilon} \cdot A^{\#}(\tilde{\lambda_1})[R] + (1 + e^{\varepsilon})\delta) + \delta && \text{(by } (\lambda_1, \tilde{\lambda_1}) \in \Psi) \\
&\leq e^{3\varepsilon} \cdot A^{\#}(\tilde{\lambda_1})[R] + (1 + e^{\varepsilon} + e^{2\varepsilon})\delta. && \square
\end{aligned}$$

B.3 Point Obfuscation by Distribution Obfuscation

Next we show that DP is an instance of DistP if an adjacency relation includes pairs of point distributions.

Definition 13 (Point distribution). For each $x \in \mathcal{X}$, the *point distribution* η_x of x is the distribution over $\mathcal{X}$ such that: $\eta_x[x'] = 1$ if $x' = x$, and $\eta_x[x'] = 0$ otherwise.

Lemma 1. *Let $\Phi \subseteq \mathcal{X} \times \mathcal{X}$. For any $(x_0, x_1) \in \Phi$, we have $(\eta_{x_0}, \eta_{x_1}) \in \Phi^{\#}_{W_\infty}$.*

Proof. Let $(x_0, x_1) \in \Phi$. We define $\gamma \in \mathbb{D}(\mathcal{X} \times \mathcal{X})$ by: $\gamma[x, x'] = 1$ if $x = x_0$ and $x' = x_1$, and $\gamma[x, x'] = 0$ otherwise. Then γ is the only coupling between η_{x_0} and η_{x_1}, hence $\gamma \in \Gamma_{\infty,d}(\eta_{x_0}, \eta_{x_1})$. Also $\mathsf{supp}(\gamma) = \{(x_0, x_1)\} \subseteq \Phi$. Therefore by definition, we obtain $(\eta_{x_0}, \eta_{x_1}) \in \Phi^{\#}_{W_\infty}$. $\square$

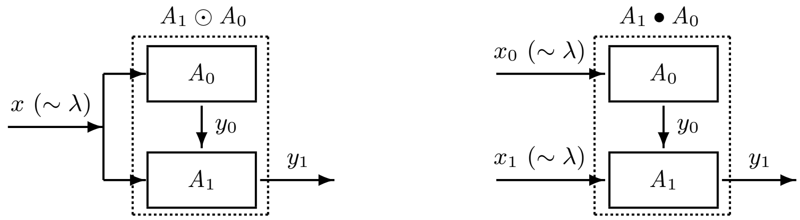
Theorem 4 (DistP $\Rightarrow$ DP and XDistP $\Rightarrow$ XDP) *Let $\varepsilon \in \mathbb{R}^{\geq 0}$, $\Phi \subseteq \mathcal{X} \times \mathcal{X}$, and $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ be a randomized algorithm.*

1. *If A provides (ε, δ)-DistP w.r.t. $\Phi^{\#}$, it provides (ε, δ)-DP w.r.t. Φ.*
2. *If A provides $(\varepsilon, W_{\infty,d}, \delta)$-XDistP w.r.t. $\Phi^{\#}_{W_\infty}$, it provides (ε, d, δ)-XDP w.r.t. Φ.*

Proof. We prove the first claim as follows. Assume that A provides (ε, δ)-DistP w.r.t. $\Phi^{\#}$. Let $(x_0, x_1) \in \Phi$, and η_{x_0} and η_{x_1} be the point distributions, defined in Definition 13. By Lemma 1 and $\Phi^{\#}_{W_\infty} \subseteq \Phi^{\#}$, we have $(\eta_{x_0}, \eta_{x_1}) \in \Phi^{\#}$. It follows from (ε, δ)-DistP that for any $R \subseteq \mathcal{Y}$, we obtain:

$$A(x_0)[R] = A^{\#}(\eta_{x_0})[R] \leq e^{\varepsilon} \cdot A^{\#}(\eta_{x_1})[R] + \delta = e^{\varepsilon} \cdot A(x_1)[R] + \delta.$$

Hence A provides (ε, δ)-DP w.r.t. Φ.



(a) Composition $\odot$ with shared input. (b) Composition $\bullet$ with independent inputs.

Fig. 21: Two kinds of sequential compositions $\odot$ and $\bullet$.

Next we show the second claim. Assume that A provides $(\varepsilon, W_{\infty,d}, \delta)$-XDistP w.r.t. $\Phi^{\#}_{W_\infty}$. Let $(x_0, x_1) \in \Phi$, and η_{x_0} and η_{x_1} be the point distributions. By Lemma 1, we have $(\eta_{x_0}, \eta_{x_1}) \in \Phi^{\#}_{W_\infty}$. Then for any $R \subseteq \mathcal{Y}$, we obtain:

$$\begin{aligned} A(x_0)[R] &= A^{\#}(\eta_{x_0})[R] \\ &\le e^{\varepsilon W_{\infty,d}(\eta_{x_0}, \eta_{x_1})} \cdot A^{\#}(\eta_{x_1})[R] + \delta \quad \text{(by XDistP of } A) \\ &= e^{\varepsilon d(x_0, x_1)} \cdot A(x_1)[R] + \delta \end{aligned}$$

where the last equality follows from the definition of $W_{\infty,d}$. Hence A provides (ε, d, δ)-XDP w.r.t. Φ. □

B.4 Sequential Compositions $\odot$ and $\bullet$

In this section we show two kinds of compositionality.

Sequential Composition $\odot$ with Shared Input We first present the definition of the sequential composition with shared input (Fig. 21a).

Definition 14 (Sequential composition $\odot$). Given two randomized algorithms $A_0 : \mathcal{X} \to \mathbb{D}\mathcal{Y}_0$ and $A_1 : \mathcal{Y}_0 \times \mathcal{X} \to \mathbb{D}\mathcal{Y}_1$, we define the *sequential composition* of A_0 and A_1 as the randomized algorithm $A_1 \odot A_0 : \mathcal{X} \to \mathbb{D}\mathcal{Y}_1$ such that: for any $x \in \mathcal{X}$, $(A_1 \odot A_0)(x) = A_1(A_0(x), x))$.

Then we show that it is harder to obfuscate distributions when an identical input is applied to the mechanism multiple times.

Proposition 3 (Sequential composition $\odot$ of (ε, δ)-DistP) *Let $\Phi \subseteq \mathcal{X} \times \mathcal{X}$. If $A_0 : \mathcal{X} \to \mathbb{D}\mathcal{Y}_0$ provides $(\varepsilon_0, \delta_0)$-DistP w.r.t. $\Phi^{\#}$ and for each $y_0 \in \mathcal{Y}_0$, $A_1(y_0) : \mathcal{X} \to \mathbb{D}\mathcal{Y}_1$ provides $(\varepsilon_1, \delta_1)$-DistP w.r.t. $\Phi^{\#}$ then the sequential composition $A_1 \odot A_0$ provides $(\varepsilon_0 + \varepsilon_1, (\delta_0 + \delta_1) \cdot |\Phi|)$-DistP w.r.t. $\Phi^{\#}$.*

Proof. By Theorem 4, A_0 provides $(\varepsilon_0, \delta_0)$-DP w.r.t. Φ, and for each $y_0 \in \mathcal{Y}_0$, $A_1(y_0)$ provides $(\varepsilon_1, \delta_1)$-DP w.r.t. Φ. By the sequential composition theorem for DP mechanisms, $A_1 \odot A_0$ provides $(\varepsilon_0 + \varepsilon_1, \delta_0 + \delta_1)$-DP w.r.t. Φ. By Theorem 1, $A_1 \odot A_0$ provides $(\varepsilon_0 + \varepsilon_1, (\delta_0 + \delta_1) \cdot |\Phi|)$-DistP w.r.t. $\Phi^{\#}$.

The compositionality for XDistP can be shown analogously to Proposition 3.

Proposition 4 (Sequential composition $\odot$ of (ε, δ)-XDistP) *Let $\Phi \subseteq \mathcal{X} \times \mathcal{X}$. If $A_0 : \mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}_0$ provides $(\varepsilon_0, W_{\infty,d}, \delta_0)$-XDistP w.r.t. $\Phi^{\#}_{W_\infty}$ and for each $y_0 \in \mathcal{Y}_0$, $A_1(y_0) : \mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}_1$ provides $(\varepsilon_1, W_{\infty,d}, \delta_1)$-XDistP w.r.t. $\Phi^{\#}_{W_\infty}$ then the sequential composition $A_1 \odot A_0$ provides $(\varepsilon_0 + \varepsilon_1, W_{\infty,d}, (\delta_0 + \delta_1) \cdot |\Phi|)$-XDistP w.r.t. $\Phi^{\#}_{W_\infty}$.*

Proof. By Theorem 4, A_0 provides $(\varepsilon_0, d, \delta_0)$-XDP w.r.t. Φ, and for each $y_0 \in \mathcal{Y}_0$, $A_1(y_0)$ provides $(\varepsilon_1, d, \delta_1)$-XDP w.r.t. Φ. By the sequential composition theorem for XDP mechanisms, $A_1 \odot A_0$ provides $(\varepsilon_0 + \varepsilon_1, d, \delta_0 + \delta_1)$-XDP w.r.t. Φ. By Theorem 2, $A_1 \odot A_0$ provides $(\varepsilon_0 + \varepsilon_1, W_{\infty,d}, (\delta_0 + \delta_1) \cdot |\Phi|)$-XDistP w.r.t. $\Phi^{\#}_{W_\infty}$.

Sequential Composition $\bullet$ with Independent Sampling We first note that the lifting of the sequential composition $(A_1 \odot A_0)^{\#}$ does not coincide with the sequential composition of the liftings ${A_1}^{\#} \odot {A_1}^{\#}$. Then the latter is a randomized algorithm ${A_1}^{\#} \odot {A_0}^{\#} : \mathbb{D}\mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}_1$ such that for any $\lambda \in \mathbb{D}\mathcal{X}$, $({A_1}^{\#} \odot {A_0}^{\#})(\lambda) = {A_1}^{\#}({A_0}^{\#}(\lambda), \lambda)$. Then each of A_0 and A_1 receives an input independently sampled from λ. This is different from $A_1 \odot A_0$, in which A_0 and A_1 share an identical input drawn from λ, as shown in Figure 21a.

To see this difference in detail, we deal with another definition of sequential composition.

Definition 15 (Sequential composition $\bullet$). Given two randomized algorithms $A_0 : \mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}_0$ and $A_1 : \mathcal{Y}_0 \times \mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}_1$, we define the *sequential composition* of A_0 and A_1 as the randomized algorithm $A_1 \bullet A_0 : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}_1$ such that: for any $x_0, x_1 \in \mathcal{X}$, $(A_1 \bullet A_0)(x_0, x_1) = A_1(A_0(x_0), x_1))$.

Then the lifting of the sequential composition $(A_1 \bullet A_0)^{\#}$ coincides with the sequential composition of the liftings ${A_1}^{\#} \bullet {A_1}^{\#}$ in the sense that for any $\lambda_0, \lambda_1 \in \mathbb{D}\mathcal{X}$,

$$(A_1 \bullet A_0)^{\#}(\lambda_0 \times \lambda_1) = {A_1}^{\#}({A_0}^{\#}(\lambda_0), \lambda_1) = ({A_1}^{\#} \bullet {A_1}^{\#})(\lambda_0, \lambda_1), \tag{3}$$

where $\lambda_0 \times \lambda_1$ is the probability distribution over $\mathcal{X} \times \mathcal{X}$ such that for all $x_0, x_1 \in \mathcal{X}$, $(\lambda_0 \times \lambda_1)[x_0, x_1] = \lambda_0[x_0]\lambda_1[x_1]$.

To show the compositionality for distribution privacy, we introduce an operator $\diamond$ between binary relations Ψ_0 and Ψ_1 by:

$$\Psi_0 \diamond \Psi_1 = \{(\lambda_0 \times \lambda_1, \lambda'_0 \times \lambda'_1) \mid (\lambda_0, \lambda'_0) \in \Psi_0, (\lambda_1, \lambda'_1) \in \Psi_1\}.$$

Proposition 5 (Sequential composition $\bullet$ of (ε, δ)-DistP) *Let $\Psi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$. If $A_0 : \mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}_0$ provides $(\varepsilon_0, \delta_0)$-DistP w.r.t. Ψ and for each $y_0 \in \mathcal{Y}_0$, $A_1(y_0) : \mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}_1$ provides $(\varepsilon_1, \delta_1)$-DistP w.r.t. Ψ then the sequential composition $A_1 \bullet A_0$ provides $(\varepsilon_0 + \varepsilon_1, \delta_0 + \delta_1)$-DistP w.r.t. $\Psi \diamond \Psi$.*

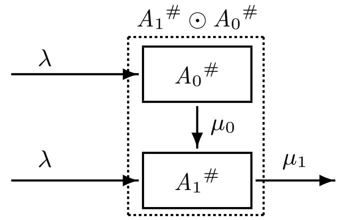


Fig. 22: Sequential composition $A_1{}^{\#} \odot A_0{}^{\#}$ involves two independent samplings from λ.

Proof. By the definition of DistP, we have that $A_0{}^{\#}$ provides $(\varepsilon_0, \delta_0)$-DP w.r.t. Ψ and for each $y_0 \in \mathcal{Y}_0$, $A_1(y_0)^{\#}$ provides $(\varepsilon_1, \delta_1)$-DP w.r.t. Ψ. By the sequential composition theorem for DP mechanisms and Equation (3), $(A_1 \bullet A_0)^{\#}$ provides $(\varepsilon_0+\varepsilon_1, \delta_0+\delta_1)$-DP w.r.t. $\Psi \diamond \Psi$. Therefore $A_1 \bullet A_0$ provides $(\varepsilon_0+\varepsilon_1, \delta_0+\delta_1)$-DistP w.r.t. $\Psi \diamond \Psi$.

For brevity, we omit the case of XDistP.

Finally, we remark that the comparison of the compositions with shared input and with independent input is also discussed from the viewpoint of quantitative information flow in [22].

B.5 Post-processing and Pre-processing

Next we show that distribution privacy is immune to the post-processing. For $A_0 : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ and $A_1 : \mathcal{Y} \to \mathbb{D}\mathcal{Z}$, we define $A_1 \circ A_0$ by: $(A_1 \circ A_0)(x) = A_1(A_0(x))$.

Proposition 6 (Post-processing) *Let $\Psi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$, and $W : \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X} \to \mathbb{R}^{\geq 0}$ be a metric. Let $A_0 : \mathcal{X} \to \mathbb{D}\mathcal{Y}$, and $A_1 : \mathcal{Y} \to \mathbb{D}\mathcal{Z}$.*

1. *If A_0 provides (ε, δ)-DistP w.r.t. Ψ then so does the composite function $A_1 \circ A_0$.*
2. *If A_0 provides (ε, W, δ)-XDistP w.r.t. Ψ then so does the composite function $A_1 \circ A_0$.*

Proof. Let $(\lambda_0, \lambda_1) \in \Psi$. Since every randomized algorithm is a convex combination of deterministic algorithms, there are a distribution μ over an index set $\mathcal{I}$ and deterministic algorithms $A_{1,i} : U_i \to R$ such that $A_1 = \sum_{i \in \mathcal{I}} \mu[i] A_{1,i}$. Then we obtain:

$$\begin{aligned}
&(A_1 \circ A_0)^{\#}(\lambda_0)[R] \\
&= \textstyle\sum_{x_0 \in \mathcal{X}} \lambda_0[x_0] \cdot (A_1 \circ A_0)(x_0)[R] \\
&= \textstyle\sum_{i \in \mathcal{I}} \mu[i] \sum_{x_0 \in \mathcal{X}} \lambda_0[x_0] \cdot (A_{1,i} \circ A_0)(x_0)[R] \\
&= \textstyle\sum_{i \in \mathcal{I}} \mu[i] \sum_{x_0 \in \mathcal{X}} \lambda_0[x_0] \cdot A_0(x_0)[U_i] \\
&= \textstyle\sum_{i \in \mathcal{I}} \mu[i] \cdot A_0{}^{\#}(\lambda_0)[U_i].
\end{aligned} \tag{4}$$

Then we show the first claim as follows. Assume that A_0 provides (ε,δ)-DistP w.r.t. Ψ.

$$\begin{aligned}
&(A_1 \circ A_0)^{\#}(\lambda_0)[R] \\
&= \textstyle\sum_{i\in\mathcal{I}} \mu[i] \cdot A_0{}^{\#}(\lambda_0)[U_i] && \text{(by (4))} \\
&\le \textstyle\sum_{i\in\mathcal{I}} \mu[i] \cdot \big(e^{\varepsilon} \cdot A_0{}^{\#}(\lambda_1)[U_i] + \delta\big) \\
& && \text{(by } (\varepsilon,\delta)\text{-}\mathsf{DistP} \text{ of } A_0{}^{\#}) \\
&= \textstyle\sum_{i\in\mathcal{I}} \mu[i] \cdot \big(e^{\varepsilon} \cdot \sum_{x_1} \lambda_1[x_1] A_0(x_1)[U_i] + \delta\big) \\
&= \textstyle\sum_{i\in\mathcal{I}} \mu[i] \cdot \big(e^{\varepsilon} \cdot \sum_{x_1} \lambda_1[x_1] (A_{1,i} \circ A_0)(x_1)[R] + \delta\big) \\
&= e^{\varepsilon} \cdot \Big(\textstyle\sum_{i\in\mathcal{I}} \mu[i] \cdot (A_{1,i} \circ A_0)^{\#}(\lambda_1)[R]\Big) + \delta \\
&= e^{\varepsilon} \cdot (A_1 \circ A_0)^{\#}(\lambda_1))[R] + \delta.
\end{aligned}$$

Therefore $A_1 \circ A_0$ provides (ε,δ)-DistP w.r.t. Ψ.

Analogously, we show the second claim. Assume that A_0 provides (ε, W, δ)-XDistP w.r.t. Ψ.

$$\begin{aligned}
&(A_1 \circ A_0)^{\#}(\lambda_0)[R] \\
&= \sum_{i\in\mathcal{I}} \mu[i] \cdot A_0{}^{\#}(\lambda_0)[U_i] && \text{(by (4))} \\
&\le \sum_{i\in\mathcal{I}} \mu[i] \cdot \big(e^{\varepsilon\cdot W(\lambda_0,\lambda_1)} \cdot A_0{}^{\#}(\lambda_1)[U_i] + \delta\big) \\
& && \text{(by } (\varepsilon, W,\delta)\text{-}\mathsf{XDistP} \text{ of } A_0{}^{\#}) \\
&= e^{\varepsilon\cdot W(\lambda_0,\lambda_1)} \cdot \Big(\sum_{i\in\mathcal{I}} \mu[i] \cdot (A_{1,i} \circ A_0)^{\#}(\lambda_1)[R]\Big) + \delta \\
&= e^{\varepsilon\cdot W(\lambda_0,\lambda_1)} \cdot (A_1 \circ A_0)^{\#}(\lambda_1))[R] + \delta.
\end{aligned}$$

Therefore $A_1 \circ A_0$ provides (ε, W, δ)-XDistP w.r.t. Ψ. □

We then show a property on pre-processing.

Proposition 7 (Pre-processing) *Let $c \in \mathbb{R}^{\ge 0}$, $\Psi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$, and $W : \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X} \to \mathbb{R}^{\ge 0}$ be a metric.*

1. *If $T : \mathbb{D}\mathcal{X} \to \mathbb{D}\mathcal{X}$ is a (c,Ψ)-*stable *transformation and $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε,δ)-DistP w.r.t Ψ, then $A \circ T$ provides $(c\,\varepsilon,\delta)$-DistP w.r.t Ψ.*
2. *If $T : \mathbb{D}\mathcal{X} \to \mathbb{D}\mathcal{X}$ is a (c, W)-*stable *transformation and $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, W, δ)-XDistP, then $A \circ T$ provides $(c\,\varepsilon, W, \delta)$-XDistP.*

Proof. We show the first claim as follows. Assume that A provides (ε,δ)-DistP w.r.t. Ψ. Let $(\lambda,\lambda') \in \Psi$, and $R \subseteq \mathcal{Y}$. Then we have:

$$\begin{aligned}
(A \circ T)^{\#}(\lambda)[R] &= A^{\#}(T^{\#}(\lambda))[R] \\
&\le e^{c\varepsilon} A^{\#}(T^{\#}(\lambda'))[R] \\
&= e^{c\varepsilon} (A \circ T)^{\#}(\lambda')[R].
\end{aligned}$$

Therefore $A \circ T$ provides $(c\,\varepsilon, \delta)$-DistP.

Next we show the second claim. Assume that A provides (ε, W, δ)-XDistP. Let $\lambda, \lambda' \in \mathbb{D}\mathcal{X}$, and $R \subseteq \mathcal{Y}$. Then we obtain:

$$\begin{aligned}(A \circ T)^{\#}(\lambda)[R] &= A^{\#}(T^{\#}(\lambda))[R] \\ &\le e^{\varepsilon W(T^{\#}(\lambda), T^{\#}(\lambda'))} A^{\#}(T^{\#}(\lambda'))[R] \\ &\le e^{c\varepsilon W(\lambda,\lambda')} (A \circ T)^{\#}(\lambda')[R].\end{aligned}$$

Therefore $A \circ T$ provides $(c\,\varepsilon, W, \delta)$-XDistP. □

B.6 Probabilistic Distribution Privacy (PDistP)

We next introduce an approximate notion of distribution privacy analogously to the notion of probabilistic differential privacy (PDP) [48]. Intuitively, a randomized algorithm provides (ε, δ)-*probabilistic distribution privacy* if it provides ε-distribution privacy with probability at least $(1 - \delta)$.

Definition 16 (Probabilistic distribution privacy). Let $\Psi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$. We say that a randomized algorithm $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, δ)-*probabilistic distribution privacy (PDistP) w.r.t.* Ψ if the lifting $A^{\#}$ provides (ε, δ)-probabilistic differential privacy w.r.t. Ψ , i.e., for all $(\lambda, \lambda') \in \Psi$, there exists an $R' \subseteq \mathcal{Y}$ such that $A^{\#}(\lambda)[R'] \le \delta$, $A^{\#}(\lambda')[R'] \le \delta$, and that for all $R \subseteq \mathcal{Y}$, we have:

$$\begin{aligned} A^{\#}(\lambda)[R \setminus R'] &\le e^{\varepsilon} \cdot A^{\#}(\lambda')[R \setminus R'] \\ A^{\#}(\lambda')[R \setminus R'] &\le e^{\varepsilon} \cdot A^{\#}(\lambda)[R \setminus R'], \end{aligned}$$

where the probability space is taken over the choices of randomness in A.

B.7 Relationships between PDistP and DistP

In this section we show the relationships between PDistP and DistP. By definition, $(\varepsilon, 0)$-DistP is equivalent to $(\varepsilon, 0)$-PDistP. In general, however, (ε, δ)-DistP does not imply (ε, δ)-PDistP, while (ε, δ)-PDistP implies (ε, δ)-DistP.

Proposition 8 ($(\varepsilon, 0)$-PDistP $\Leftrightarrow$ $(\varepsilon, 0)$-DistP) *A randomized algorithm A provides $(\varepsilon, 0)$-PDistP w.r.t. an adjacency relation Ψ iff it provides $(\varepsilon, 0)$-DistP w.r.t. Ψ.*

Proof. Immediate from the definitions. □

Proposition 9 ((ε, δ)-PDistP $\Rightarrow$ (ε, δ)-DistP) *If a randomized algorithm A provides (ε, δ)-PDistP w.r.t. an adjacency relation Ψ, then it provides (ε, δ)-DistP w.r.t. Ψ.*

Proof. Let $\Psi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$, $(\lambda, \lambda') \in \Psi$, $R \subseteq \mathcal{Y}$, and $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ be any randomized algorithm that provides (ε, δ)-PDistP w.r.t. Ψ. Then there exists an $R' \subseteq \mathcal{Y}$ such that $A^{\#}(\lambda)[R'] \le \delta$, and $A^{\#}(\lambda)[R \setminus R'] \le e^{\varepsilon} \cdot A^{\#}(\lambda')[R \setminus R']$. Hence we obtain:

$$\begin{aligned} A^{\#}(\lambda)[R] &= A^{\#}(\lambda)[R \setminus R'] + A^{\#}(\lambda)[R'] \\ &\le e^{\varepsilon} \cdot A^{\#}(\lambda')[R \setminus R'] + \delta \\ &\le e^{\varepsilon} \cdot A^{\#}(\lambda')[R] + \delta. \end{aligned}$$

Therefore A provides (ε, δ)-DistP w.r.t. Ψ. □

B.8 Properties of the Tupling Mechanism

In this section we show properties of the tupling mechanism.

We first present a minor result with $\delta = 0$.

Proposition 10 ($(\varepsilon_A, 0)$-DistP of the tupling mechanism) *If A provides $(\varepsilon_A, 0)$-DP w.r.t. Φ, then the (k, ν, A)-tupling mechanism $Q^{\mathsf{tp}}_{k,\nu,A}$ provides $(\varepsilon_A, 0)$-DistP w.r.t. $\Phi^{\#}$.*

Proof. Let $\Phi \subseteq \mathcal{X} \times \mathcal{X}$. Since A provides $(\varepsilon_A, 0)$-DP w.r.t. Φ, Theorem 1 implies that A provides $(\varepsilon_A, 0)$-DistP w.r.t. $\Phi^{\#}$.

Let $(\lambda_0, \lambda_1) \in \Phi^{\#}$ and $R \subseteq \mathcal{Y}^{k+1}$ such that $Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_0)[R] > 0$. Since the dummies are uniformly distributed over $\mathcal{Y}$, we have $Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_1)[R] > 0$. Let $\bar{y} = (y_1, y_2, \ldots, y_{k+1}) \in R$ be an output of $Q^{\mathsf{tp}}_{k,\nu,A}$. Since i is uniformly drawn from $\{1, 2, \ldots, k+1\}$ in the mechanism $Q^{\mathsf{tp}}_{k,\nu,A}$, the output of A appears as the i-th element y_i of the tuple $\bar{y}$ with probability $\frac{1}{k+1}$. For each $b = 0, 1$, when an input x is drawn from λ_b, the probability that A outputs y_i is:

$$A^{\#}(\lambda_b)[y_i] = \textstyle\sum_{x \in \mathcal{X}} \lambda_b[x] A(x)[y_i].$$

On the other hand, for each $j \ne i$, the probability that y_j is drawn from the dummy distribution ν is given by $\nu[y_j]$. Therefore, the probability that the mechanism $Q^{\mathsf{tp}}_{k,\nu,A}$ outputs the tuple $\bar{y}$ is:

$$Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_b)[\bar{y}] = \tfrac{1}{k+1} \textstyle\sum_{i=1}^{k+1} A^{\#}(\lambda_b)[y_i] \prod_{j \ne i} \nu[y_j].$$

Hence we obtain:

$$\begin{aligned}
&\frac{Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_0)[R]}{Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_1)[R]} \\
&= \frac{\sum_{\bar{y}\in R} Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_0)[\bar{y}]}{\sum_{\bar{y}\in R} Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_1)[\bar{y}]} \\
&= \frac{\sum_{\bar{y}\in R} \frac{1}{k+1} \sum_{i=1}^{k+1} A^{\#}(\lambda_0)[y_i] \prod_{j\neq i} \nu[y_j]}{\sum_{\bar{y}\in R} \frac{1}{k+1} \sum_{i=1}^{k+1} A^{\#}(\lambda_1)[y_i] \prod_{j\neq i} \nu[y_j]} \\
&= \frac{\sum_{\bar{y}\in R} \sum_{i=1}^{k+1} A^{\#}(\lambda_0)[y_i]}{\sum_{\bar{y}\in R} \sum_{i=1}^{k+1} A^{\#}(\lambda_1)[y_i]} && \text{(since ν is uniform)} \\
&\leq \max_{\bar{y}\in R} \max_{i\in\{1,2,\ldots,k+1\}} \frac{A^{\#}(\lambda_0)[y_i]}{A^{\#}(\lambda_1)[y_i]} \\
&\leq e^{\varepsilon_A} && \text{(by $(\varepsilon_A, 0)$-DistP of A w.r.t. $\Phi^{\#}$)}
\end{aligned}$$

Therefore $Q^{\mathsf{tp}}_{k,\nu,A}$ provides $(\varepsilon_A, 0)$-DistP w.r.t. $\Phi^{\#}$.

Next we show that the tupling mechanism provides PDistP without any restriction on A, i.e., A does not have to provide DP in order for the tupling mechanism to provides PDistP.

Proposition 11 ($(\varepsilon_\alpha, \delta_\alpha)$-PDistP of the tupling mechanism) *Let $k \in \mathbb{N}^{>0}$, ν be the uniform distribution over $\mathcal{Y}$, $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$, and $\beta, \eta \in [0, 1]$. For an $\alpha \in \mathbb{R}^{>0}$, let $\varepsilon_\alpha = \ln \frac{k+(\alpha+\beta)\cdot|\mathcal{Y}|}{k-\alpha\cdot|\mathcal{Y}|}$ and $\delta_\alpha = 2\exp\bigl(-\frac{2\alpha^2}{k\beta^2}\bigr) + \eta$. Then for any $0 < \alpha < \frac{k}{|\mathcal{Y}|}$, the tupling mechanism $Q^{\mathsf{tp}}_{k,\nu,A}$ provides $(\varepsilon_\alpha, \delta_\alpha)$-PDistP w.r.t $\Lambda^2_{\beta,\eta,A}$.*

Proof. Let $\lambda_0, \lambda_1 \in \Lambda_{\beta,\eta,A}$, and $R \subseteq \mathcal{Y}^{k+1}$ such that $Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_0)[R] > 0$. Since the dummies are uniformly distributed over $\mathcal{Y}$, we have $Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_1)[R] > 0$. As in the proof for Proposition 10, for any $\bar{y} = (y_1, y_2, \ldots, y_{k+1}) \in R$:

$$\frac{Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_0)[\bar{y}]}{Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_1)[\bar{y}]} = \frac{\sum_{i=1}^{k+1} A^{\#}(\lambda_0)[y_i]}{\sum_{i=1}^{k+1} A^{\#}(\lambda_1)[y_i]}. \tag{5}$$

Recall that in the definition of the tupling mechanism, y_i is the output of A while for each $j \neq i$, y_j is generated from the uniform distribution ν over $\mathcal{Y}$. Hence the expected value of $\sum_{j\neq i} A^{\#}(\lambda_0)[y_j]$ is given by $\frac{k}{|\mathcal{Y}|}$. Therefore it follows from the Hoeffding's inequality that for any $\alpha > 0$, we have:

$$\begin{aligned}
\Pr\bigl[\textstyle\sum_{j\neq i} A^{\#}(\lambda_0)[y_j] \geq \frac{k}{|\mathcal{Y}|} + \alpha\bigr] &\leq \exp\bigl(-\tfrac{2\alpha^2}{k\beta^2}\bigr) \\
\Pr\bigl[\textstyle\sum_{j\neq i} A^{\#}(\lambda_1)[y_j] \leq \frac{k}{|\mathcal{Y}|} - \alpha\bigr] &\leq \exp\bigl(-\tfrac{2\alpha^2}{k\beta^2}\bigr)
\end{aligned}$$

where each y_j is independently drawn from ν, hence each of $A^{\#}(\lambda_0)[y_j]$ and $A^{\#}(\lambda_1)[y_j]$ is independent. By $\delta_\alpha = 2\exp\bigl(-\frac{2\alpha^2}{k\beta^2}\bigr)$, we obtain:

$$\Pr\Bigl[\sum_{j\neq i} A^{\#}(\lambda_0)[y_j] \geq \tfrac{k}{|\mathcal{Y}|} + \alpha \text{ and } \sum_{j\neq i} A^{\#}(\lambda_1)[y_j] \leq \tfrac{k}{|\mathcal{Y}|} - \alpha\Bigr] \leq \delta_\alpha. \tag{6}$$

When $\bar{y}$ is drawn from $Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_0)$, let y_i be the output of $A^{\#}(\lambda_0)$. Then we obtain:

$$
\begin{aligned}
&\Pr\left[\frac{Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_0)[\bar{y}]}{Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_1)[\bar{y}]} \ge e^{\varepsilon}\right] \\
&= \Pr\left[\frac{\sum_{j=1}^{k+1} A^{\#}(\lambda_0)[y_j]}{\sum_{j=1}^{k+1} A^{\#}(\lambda_1)[y_j]} \ge e^{\varepsilon}\right] && \text{(by (5))} \\
&= \Pr\left[\frac{\sum_{j=1}^{k+1} A^{\#}(\lambda_0)[y_j]}{\sum_{j=1}^{k+1} A^{\#}(\lambda_1)[y_j]} \ge \frac{\frac{k}{|\mathcal{Y}|} + \alpha + \beta}{\frac{k}{|\mathcal{Y}|} - \alpha}\right] && \text{(by def. of } \varepsilon) \\
&\le \Pr\left[\frac{\sum_{j=1}^{k+1} A^{\#}(\lambda_0)[y_j]}{\sum_{j=1}^{k+1} A^{\#}(\lambda_1)[y_j]} \ge \frac{\frac{k}{|\mathcal{Y}|} + \alpha + A^{\#}(\lambda_0)[y_i]}{\frac{k}{|\mathcal{Y}|} - \alpha + A^{\#}(\lambda_1)[y_i]}\right] + \eta \\
&\qquad \bigl(\text{by } \Pr\bigl[A^{\#}(\lambda_0)[y_i] \le \beta\bigr] \ge 1-\eta \ \text{ and } A^{\#}(\lambda_1)[y_i] \ge 0\bigr) \\
&= \Pr\left[\frac{\sum_{j\neq i} A^{\#}(\lambda_0)[y_j]}{\sum_{j\neq i} A^{\#}(\lambda_1)[y_j]} \ge \frac{\frac{k}{|\mathcal{Y}|} + \alpha}{\frac{k}{|\mathcal{Y}|} - \alpha}\right] + \eta \\
&\le \delta_\alpha. && \text{(by (6))}
\end{aligned}
$$

Then there is an $R' \subseteq \mathcal{Y}^{k+1}$ such that $Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_0)[R'] \le \delta_\alpha$, and that for any $\bar{y} \in R$, $\bar{y} \in R'$ iff $\frac{Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_0)[\bar{y}]}{Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_1)[\bar{y}]} \ge e^{\varepsilon}$. Then:

$$
\begin{aligned}
Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_1)[R'] &= \sum_{\bar{y}\in R'} Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_1)[\bar{y}] \\
&\le \sum_{\bar{y}\in R'} e^{-\varepsilon} \cdot Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_0)[\bar{y}] \\
&= e^{-\varepsilon} \cdot Q^{\mathsf{tp}}_{k,\nu,A}(\lambda_0)[R'] \\
&\le \delta_\alpha.
\end{aligned}
$$

Therefore $Q^{\mathsf{tp}}_{k,\nu,A}$ provides $(\varepsilon_\alpha, \delta_\alpha)$-$\mathsf{PDistP}$ w.r.t. $\varLambda^2_{\beta,\eta,A}$.

Then, we obtain the DistP of the tupling mechanism from Propositions 11 and 9 as follows.

Theorem 3 (DistP of the tupling mechanism) *Let $k \in \mathbb{N}^{>0}$, ν be the uniform distribution over $\mathcal{Y}$, $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$, and $\beta, \eta \in [0,1]$. Given an $0 < \alpha < \frac{k}{|\mathcal{Y}|}$, let $\varepsilon_\alpha = \ln \frac{k+(\alpha+\beta)\cdot|\mathcal{Y}|}{k-\alpha\cdot|\mathcal{Y}|}$ and $\delta_\alpha = 2e^{-\frac{2\alpha^2}{k\beta^2}} + \eta$. Then the (k,ν,A)-tupling mechanism provides $(\varepsilon_\alpha, \delta_\alpha)$-$\mathsf{DistP}$ w.r.t. $\varLambda^2_{\beta,\eta,A}$.*

Proof. By Proposition 11, $Q^{\mathsf{tp}}_{k,\nu,A}$ provides $(\varepsilon_\alpha, \delta_\alpha)$-$\mathsf{PDistP}$ w.r.t. $\varLambda^2_{\beta,\eta,A}$. Hence by Proposition 9, $Q^{\mathsf{tp}}_{k,\nu,A}$ provides $(\varepsilon_\alpha, \delta_\alpha)$-$\mathsf{DistP}$ w.r.t. $\varLambda^2_{\beta,\eta,A}$.

Finally, we note that if A spreads the input distribution (e.g., A provides ε-DP for a smaller ε), then β becomes smaller (by Proposition 12), hence smaller values of ε_α and δ_α as shown in experimental results in Section 7.

Proposition 12 *Let* $\lambda \in \mathbb{D}\mathcal{X}$, $A : \mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}$, *and* $y \in \mathcal{Y}$. *Then we have* $\min_{x \in \mathcal{X}} \lambda[x] \le A^{\#}(\lambda)[y] \le \max_{x \in \mathcal{X}} \lambda[x]$.

Proof. By $A^{\#}(\lambda)[y] = \sum_{x' \in \mathcal{X}} \lambda[x']A(x')[y]$, we obtain:

$$
\begin{aligned}
A^{\#}(\lambda)[y] &\le \big(\max_{x \in \mathcal{X}} \lambda[x]\big) \cdot \textstyle\sum_{x' \in \mathcal{X}} A(x')[y] = \max_{x \in \mathcal{X}} \lambda[x] \\
A^{\#}(\lambda)[y] &\ge \big(\min_{x \in \mathcal{X}} \lambda[x]\big) \cdot \textstyle\sum_{x' \in \mathcal{X}} A(x')[y] = \min_{x \in \mathcal{X}} \lambda[x].
\end{aligned}
$$