Local Distribution Obfuscation via Probability Coupling*

Yusuke Kawamoto[1] Takao Murakami[2]

***Abstract*— We introduce a general model for the local obfuscation of probability distributions by probabilistic perturbation, e.g., by adding differentially private noise, and investigate its theoretical properties. Specifically, we relax a notion of distribution privacy (DistP) by generalizing it to divergence, and propose local obfuscation mechanisms that provide divergence distribution privacy. To provide f-divergence distribution privacy, we prove that probabilistic perturbation noise should be added proportionally to the Earth mover's distance between the probability distributions that we want to make indistinguishable. Furthermore, we introduce a local obfuscation mechanism, which we call a *coupling mechanism*, that provides divergence distribution privacy while optimizing the utility of obfuscated data by using exact/approximate auxiliary information on the input distributions we want to protect.**

I. Introduction

Differential privacy (DP) [1] is one of the most popular privacy notions that have been studied in various areas, including databases, machine learning, geo-locations, and social networks. The protection of DP can be achieved by adding probabilistic noise to the data we want to obfuscate. In particular, many studies have proposed *local obfuscation mechanisms* [2], [3], [4] that perturb each single "point" datum (e.g., a geo-location point) by adding controlled probabilistic noise before sending it out to a data collector.

Recent researches [5], [6], [7] show that local obfuscation mechanisms can be used to hide the probability distributions that lie behind such point data and implicitly represent sensitive attributes (e.g., age, gender, social status). In particular, [6] proposes the notion of *distribution privacy* (DistP) as the local DP of probability distributions. Roughly, DistP of a local obfuscation mechanism A represents that the adversary cannot significantly gain information on the distribution of A's input by observing A's output. However, since DistP assumes the worst case risk in the sense of DP, it imposes strong requirement and might unnecessarily lose the utility of obfuscated data.

In this paper, we relax the notion of DistP by generalizing it to an arbitrary divergence. The basic idea is similar to point privacy notions that relax DP and improve utility by relying on some divergence (e.g., total variation privacy [8], Kullback-Leibler divergence privacy [8], [9], and Rényi differential privacy [10]). We define the notion of *divergence distribution privacy* by replacing the DP-style with an arbitrary divergence D. This relaxation allows us to formalize "on-average" DistP, and to explore privacy notions against an adversary performing the statistical hypothesis test corresponding to the divergence [8].

Furthermore, we propose and investigate local obfuscation mechanisms that provide divergence DistP. Specifically, we consider the following two scenarios:

(i) when we have no idea on the input distributions;
(ii) when we know exact or approximate information on the input distributions (e.g., when we can use public datasets [11], [12] to learn approximate distributions of locations of male/female users if we want to obfuscate the attribute male/female).

For the scenario (i), we clarify how much perturbation noise should be added to provide f-divergence DistP when we use an existing mechanism for obfuscating point data. For the scenario (ii), we introduce a local obfuscation mechanism that provides divergence DistP while optimizing the utility of obfuscated data by using the auxiliary information. Here it should be noted that probability coupling techniques are crucial in constructing divergence DistP mechanisms in both the scenarios.

Our contributions. The main contributions are as follows:

- We introduce notions of divergence DistP and investigate theoretical properties of distribution obfuscation, especially the relationships between local distribution obfuscation and probability coupling.
- We investigate the relationships among various notions of DistP based on f-divergences, such as Kullback-Leibler divergence, which models "on-average" risk.
- In the scenario (i), we present how much divergence DistP can be achieved by local obfuscation. In particular, by using probability coupling techniques, we prove that perturbation noise should be added proportionally to the Earth mover's distance between the input distributions that we want to make indistinguishable.
- In the scenario (ii), we propose a local obfuscation mechanism, called a *(utility-optimal) coupling mechanism*, that provides divergence DistP while minimizing utility loss. The construction of the mechanism relies on solving an optimal transportation problem using probability coupling.
- We theoretically evaluate the divergence DistP and utility loss of coupling mechanisms that can use exact/approximate knowledge on the input distributions.

Paper organization. The rest of this paper is organized as follows. Section II presents background knowledge. Section III introduces notions of divergence DistP. Section IV investigates important properties of divergence DistP, and relationships among privacy notions. Section V shows that

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

[1]Yusuke Kawamoto is with AIST, Tsukuba, Japan.
[2]Takao Murakami is with AIST, Tokyo, Japan.

in the scenario (i), an f-privacy mechanism can provide f-divergence DistP. Section VI generalizes DistP to use exact/approximate information on the input distribution in the scenario (ii), and proposes a local mechanism for providing DistP while optimizing utility. Section VII discusses related work and Section VIII concludes.

II. Preliminaries

In this section we recall some notions of privacy, divergence, and metrics used in this paper.

A. Notations

Let $\mathbb{R}^{\geq 0}$ be the set of non-negative real numbers, and $[0, 1] \stackrel{\text{def}}{=} \{r \in \mathbb{R}^{\geq 0} \mid r \leq 1\}$. Let $\varepsilon, \varepsilon_0, \varepsilon_1 \in \mathbb{R}^{\geq 0}$, $\delta, \delta_0, \delta_1 \in [0, 1]$, and e be the base of natural logarithm.

We denote by $|\mathcal{X}|$ the number of elements in a finite set $\mathcal{X}$, and by $\mathbb{D}\mathcal{X}$ the *set of all probability distributions* over a set $\mathcal{X}$. Given a probability distribution λ over a finite set $\mathcal{X}$, 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* is $\mathsf{supp}(\lambda) = \{x \in \mathcal{X} \colon \lambda[x] > 0\}$.

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

B. Differential Privacy

Differential privacy [1] is a notion of privacy guaranteeing that we cannot learn which of two "adjacent" inputs x and x' is used to generate an output of a randomized algorithm. This notion is parameterized by a degree ε of indistinguishability, a ratio δ of exception, and some adjacency relation $\varPhi$ over a set $\mathcal{X}$ of data. The formal definition is given as follows.

Definition 1 (Differential privacy): A randomized algorithm $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, δ)-*differential privacy (DP)* w.r.t. an adjacency relation $\varPhi \subseteq \mathcal{X} \times \mathcal{X}$ if for any $(x, x') \in \varPhi$ 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.

Then the protection of DP is stronger for smaller ε and δ.

DP can be achieved by a *local obfuscation mechanism* or *privacy mechanism* (illustrated in Fig. 1), namely a randomized algorithm that adds controlled noise probabilistically to given inputs that we want to protect.

C. Extended Differential Privacy (XDP)

The notion of DP can be relaxed by incorporating a metric d over the set $\mathcal{X}$ of input data. In [13] Chatzikokolakis *et al.* propose the notion of "d-privacy", an extension of $(\varepsilon, 0)$-DP to a metric d on input data. Intuitively, this notion guarantees that when two inputs x and x' are closer in terms of d, the output distributions are less distinguishable[1]. Here we show the definition of this extended DP equipped with δ.

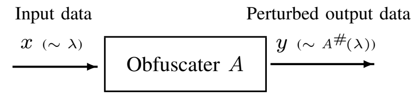


Fig. 1: A local obfuscation mechanism A perturbs input data x and returns output data y. Then the underlying probability distribution λ can also be seen to be obfuscated.

Definition 2 (Extended differential privacy): Let $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a metric. 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 $R \subseteq \mathcal{Y}$,

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

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

To achieve XDP, obfuscation mechanisms should add noise proportionally to the distance $d(x, x')$ between the two inputs x and x' that we want to make indistinguishable, hence more noise is require for a larger $d(x, x')$.

D. Distribution Privacy and Extended Distribution Privacy

Distribution privacy (DistP) [6] is a privacy notion that measures how much information on the input distribution is leaked by an output of a randomized algorithm. For example, let λ_{male} (resp. λ_{female}) be a (prior) probability distribution of the locations of the male (resp. female) users. When we observe an output of an obfuscation mechanism A and cannot learn whether the input to A is drawn from λ_{male} or λ_{female}, then we say that A provides (ε, δ)-DistP w.r.t. $(\lambda_{male}, \lambda_{female})$. Formally, DistP is defined as follows.

Definition 3 (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 $\varPsi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$ if its lifting $A^{\#} : \mathbb{D}\mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, δ)-DP w.r.t. $\varPsi$, i.e., for all pairs $(\lambda, \lambda') \in \varPsi$ and $R \subseteq \mathcal{Y}$, we have $A^{\#}(\lambda)[R] \leq e^{\varepsilon} \cdot A^{\#}(\lambda')[R] + \delta$.

Next we recall an extension [6] of DistP with a metric d as follows. Intuitively, this extended notion guarantees that when two input distributions are closer, then the output distributions must be less distinguishable.

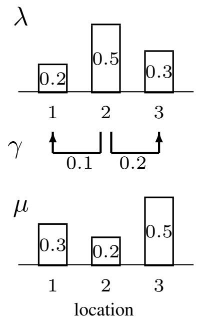
Definition 4 (Extended distribution privacy): Let $d : (\mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}) \to \mathbb{R}$ be a metric, and $\varPsi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$. We say that a mechanism $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, d, δ)-*extended distribution privacy (XDistP) w.r.t.* $\varPsi$ if the lifting $A^{\#}$ provides (ε, d, δ)-XDP w.r.t. $\varPsi$, i.e., for all $(\lambda, \lambda') \in \varPsi$ and $R \subseteq \mathcal{Y}$, we have $A^{\#}(\lambda)[R] \leq e^{\varepsilon d(\lambda, \lambda')} \cdot A^{\#}(\lambda')[R] + \delta$.

Analogously to XDP, noise should be added proportionally to the distance $d(\lambda, \lambda')$.

[1]Compared to DP, XDP provides weaker privacy and higher utility, as it obfuscates closer points. E.g., [14] shows the planar Laplace mechanism [3] (with XDP) adds less noise than the randomized response (with DP).

TABLE I: Instances of f-divergence

Divergence	$f(t)$
KL-divergence	$t \log t$
Reverse KL-divergence	$-t \log t$
Total variation	$\frac{1}{2}\lvert t-1\rvert$
χ^2-divergence	$(t-1)^2$
Hellinger distance	$\frac{1}{2}(\sqrt{t}-1)^2$



(a) An original distribution λ and a target distribution μ.

γ		μ 1	2	3
	1	0.2		
λ	2	0.1	0.2	0.2
	3			0.3

(b) A coupling γ of its two marginals λ and μ can be interpreted as a transportation that transforms λ to μ. E.g., to construct μ from λ, 0.1 moves from 2 to 1, and 0.2 moves from 2 to 3.

Fig. 2: A coupling γ that transforms λ to μ.

E. *Divergence*

A *divergence* over a non-empty set $\mathcal{Y}$ is a function $D(\cdot \parallel \cdot)$: $\mathbb{D}\mathcal{Y} \times \mathbb{D}\mathcal{Y} \to \mathbb{R}^{\geq 0}$ such that for all $\mu, \mu' \in \mathbb{D}\mathcal{Y}$, (i) $D(\mu \parallel \mu') \geq 0$ and (ii) $D(\mu \parallel \mu') = 0$ iff $\mu = \mu'$. Note that a divergence may not be symmetric or subadditive. We denote by $\mathsf{Div}(\mathcal{Y})$ the set of all divergences over $\mathcal{Y}$.

Next we recall the notion of (approximate) max divergence, which can be used to define DP.

Definition 5 (Max divergence): Let $\delta \in [0,1]$ and $\mu, \mu' \in \mathbb{D}\mathcal{Y}$. Then δ-*approximate max divergence* between μ, μ' is:

$$D^{\delta}_{\infty}(\mu \parallel \mu') = \max_{R \subseteq \mathsf{supp}(\mu), \mu[R] \geq \delta} \ln \tfrac{\mu[R]-\delta}{\mu'[R]}.$$

We recall the notion of the f-divergences [15]. As shown in Table I, many divergence notions (e.g. Kullback-Leibler-divergence [16]) are instances of f-divergence.

Definition 6 (f-divergence): Let $\mathcal{F}$ be the collection of functions defined by:

$$\mathcal{F} = \{ f : \mathbb{R}^+ \to \mathbb{R}^+ \mid f \text{ is convex and } f(1) = 0 \}.$$

Let $\mathcal{Y}$ be a finite set, and $\mu, \mu' \in \mathbb{D}\mathcal{Y}$ such that for every $y \in \mathcal{Y}$, $\mu'[y] = 0$ implies $\mu[y] = 0$. Then for an $f \in \mathcal{F}$, the f-*divergence* of μ from μ' is defined as:

$$D_f(\mu \parallel \mu') = \sum_{y \in \mathsf{supp}(\mu')} \mu'[y]\, f\big(\tfrac{\mu[y]}{\mu'[y]}\big).$$

F. *Probability Coupling*

We recall the notion of probability coupling as follows.

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

Formally, a coupling is defined as follows.

Definition 7 (Coupling): Given $\lambda \in \mathbb{D}\mathcal{X}_0$ and $\mu \in \mathbb{D}\mathcal{X}_1$, a *coupling* of λ and μ is a joint distribution $\gamma \in \mathbb{D}(\mathcal{X}_0 \times \mathcal{X}_1)$ such that λ and μ are γ's marginal distributions, i.e., for each $x_0 \in \mathcal{X}_0$, $\lambda[x_0] = \sum_{x'_1 \in \mathcal{X}_1} \gamma[x_0, x'_1]$ and for each $x_1 \in \mathcal{X}_1$, $\mu[x_1] = \sum_{x'_0 \in \mathcal{X}_0} \gamma[x'_0, x_1]$. We denote by $\mathsf{cp}(\lambda, \mu)$ the set of all couplings of λ and μ.

G. *p-Wasserstein Metric*

Then we recall the p-Wasserstein metric [17] between two distributions, which is defined using a coupling as follows.

Definition 8 (p-Wasserstein metric): Let d be a metric over $\mathcal{X}$, and $p \in \mathbb{R}^{\geq 1} \cup \{\infty\}$. The *$p$-Wasserstein metric* $W_{p,d}$ w.r.t. d is defined by: for any two distributions $\lambda, \mu \in \mathbb{D}\mathcal{X}$,

$$W_{p,d}(\lambda, \mu) = \min_{\gamma \in \mathsf{cp}(\lambda,\mu)} \Big(\sum_{(x_0,x_1) \in \mathsf{supp}(\gamma)} d(x_0, x_1)^p \gamma[x_0, x_1] \Big)^{\frac{1}{p}}.$$

$W_{1,d}$ is also called the *Earth mover's distance*.

The intuitive meaning of $W_{1,d}(\lambda, \mu)$ is the minimum cost of transportation from λ to μ in transportation theory. As illustrated in Fig. 2, we regard the distribution λ (resp. μ) as the set of points where each point x has weight $\lambda[x]$ (resp. $\mu[x]$), and we move some weight in λ from a point x_0 to another x_1 to construct μ. We represent by $\gamma[x_0, x_1]$ the amount of weight moved from x_0 to x_1.[2] We denote by $d(x_0, x_1)$ the cost (i.e., distance) of move from x_0 to x_1. Then the minimum cost of the whole transportation is:

$$W_{1,d}(\lambda, \mu) = \min_{\gamma \in \mathsf{cp}(\lambda,\mu)} \sum_{(x_0,x_1) \in \mathsf{supp}(\gamma)} d(x_0, x_1)\, \gamma[x_0, x_1].$$

E.g., in Fig. 2, when the cost function d is the Euclid distance over $\mathcal{X}$ (e.g., $d(2,1) = |2-1| = 1$), the transportation γ achieves the minimum cost $0.1 \cdot 1 + 0.2 \cdot 1 = 0.3$.

Let $\Gamma_{p,d}$ the set of all couplings achieving $W_{p,d}$; i.e.,

$$\Gamma_{p,d}(\lambda, \mu) = \operatorname*{argmin}_{\gamma \in \mathsf{cp}(\lambda,\mu)} \Big(\sum_{(x_0,x_1) \in \mathsf{supp}(\gamma)} d(x_0, x_1)^p \gamma[x_0, x_1] \Big)^{\frac{1}{p}}.$$

Then $\gamma \in \Gamma_{1,d}(\lambda, \mu)$ can be efficiently computed by the *North-West corner rule* [18] when d is *submodular* [3].

III. Divergence Distribution Privacy

In this section we introduce new definitions of distribution privacy generalized to an arbitrary divergence D. The main motivation is to discuss distribution privacy based on f-divergences, especially Kullback-Leibler divergence, which models "on-average" risk.

A. *Divergence DP and Divergence XDP*

To generalize distribution privacy notions, we first present a generalized formulation of point privacy parameterized with a divergence D. Intuitively, we say that a randomized algorithm A provides (ε, D)-DP if a divergence D cannot distinguish the input to A by observing an output of A.

[2] The amount of weight moved from a point x_0 in λ is given by $\lambda[x_0] = \sum_{x'_1 \in \mathcal{X}} \gamma[x_0, x'_1]$, while the amount moved into x_1 in μ is given by $\mu[x_1] = \sum_{x'_0 \in \mathcal{X}} \gamma[x'_0, x_1]$. Hence γ is a coupling of λ and μ.

[3] d is submodular if $d(x_0, x_1) + d(x'_0, x'_1) \leq d(x'_0, x_1) + d(x_0, x'_1)$.

Definition 9 (Divergence DP w.r.t. adjacency relation): For an adjacency relation $\Phi \subseteq \mathcal{X} \times \mathcal{X}$ and a divergence $D \in \mathsf{Div}(\mathcal{Y})$, we say that a randomized algorithm $A : \mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}$ provides (ε, D)-*DP w.r.t.* Φ if for all $(x, x') \in \Phi$, we have $D(A(x) \,\|\, A(x')) \le \varepsilon$ and $D(A(x') \,\|\, A(x)) \le \varepsilon$.

Note that some instances of divergence DP are known in the literature. In [8], (ε, D_f)-DP is called *ε-f-divergence privacy*, $(\varepsilon, D_{\mathrm{KL}})$-DP (KLP) is called *$\varepsilon$-KL-privacy*, and $(\varepsilon, D_{\mathrm{TV}})$-DP is called *$\varepsilon$-total variation privacy*. Furthermore, $(\varepsilon, D^\delta_\infty)$-DP is equivalent to (ε, δ)-DP, since it is known that (ε, δ)-DP can be defined using the approximate max divergence D^δ_∞ as follows:

Proposition 1: A randomized algorithm $A : \mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}$ provides (ε, δ)-DP w.r.t. $\Phi \subseteq \mathcal{X} \times \mathcal{X}$ iff for any $(x, x') \in \Phi$, $D^\delta_\infty(A(x) \parallel A(x')) \le \varepsilon$ and $D^\delta_\infty(A(x') \parallel A(x)) \le \varepsilon$.

Next we generalize the notion of extended differential privacy (XDP) to an arbitrary divergence D as follows.

Definition 10 (Divergence XDP): Let $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ be a metric, $\Phi \subseteq \mathcal{X} \times \mathcal{X}$, and $D \in \mathsf{Div}(\mathcal{Y})$. We say that a randomized algorithm $A : \mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}$ provides (ε, d, D)-*XDP w.r.t.* Φ if for all $(x, x') \in \Phi$, $D(A(x) \parallel A(x')) \le \varepsilon d(x, x')$.

These notions will be used to define (extended) divergence distribution privacy in the next section.

B. Divergence DistP and Divergence XDistP

In this section we generalize the notion of (extended) distribution privacy to an arbitrary divergence D. The main aim of generalization is to present theoretical properties of distribution privacy in a more general form, and also to discuss distribution privacy based on the f-divergences.

Intuitively, we say that a randomized algorithm A provides (ε, D)-distribution privacy w.r.t. a set Ψ of pairs of distributions if for each pair $(\lambda_0, \lambda_1) \in \Psi$, a divergence D cannot distinguish which distribution (of λ_0 and λ_1) is used to generate A's input value.

Definition 11 (Divergence DistP): Let $D \in \mathsf{Div}(\mathcal{Y})$, and $\Psi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$. We say that a randomized algorithm $A : \mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}$ provides (ε, D)-*distribution privacy (DistP) w.r.t.* Ψ if the lifting $A^\#$ provides (ε, D)-DP w.r.t. Ψ, i.e., for all $(\lambda, \lambda') \in \Psi$,

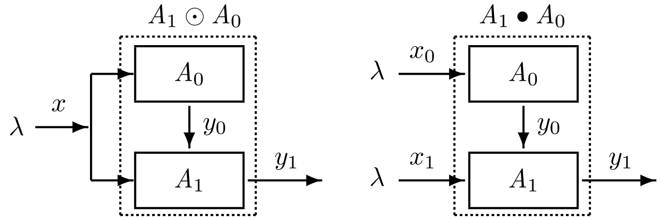
$$D(A^\#(\lambda) \parallel A^\#(\lambda')) \le \varepsilon.$$

As with the generalization of DP to f-divergence [8], D-DistP expresses privacy against an adversary performing the hypothesis test corresponding to the divergence D. When D involves averaging (e.g., $D = D_{\mathrm{KL}}$), D-DistP formalizes "on-average" privacy, which relaxes the original DistP.

Next we introduce XDistP parameterized with a divergence D. Intuitively, XDistP with a divergence D guarantees that when two input distributions λ and λ' are closer (in terms of a metric d), then the output distributions $A^\#(\lambda)$ and $A^\#(\lambda')$ must be less distinguishable (in terms of D).

Definition 12 (Divergence XDistP): Let d be a metric over $\mathbb{D}\mathcal{X}$, $D \in \mathsf{Div}(\mathcal{Y})$, and $\Psi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$. We say that a randomized algorithm $A : \mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}$ provides (ε, d, D)-*extended distribution privacy (XDistP) w.r.t.* Ψ if the lifting $A^\#$ provides (ε, d, D)-XDP w.r.t. Ψ, i.e., for all $(\lambda, \lambda') \in \Psi$,

$$D(A^\#(\lambda) \parallel A^\#(\lambda')) \le \varepsilon d(\lambda, \lambda').$$



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

Fig. 3: Two kinds of sequential compositions $\odot$ and $\bullet$.

IV. Properties of Divergence Distribution Privacy

In this section we show useful properties of divergence distribution privacy, such as compositionality and relationships among distribution privacy notions.

A. Basic Properties of Divergence Distribution Privacy

In Tables II and III we summarize the results on two kinds of sequential compositions $\odot$ (Fig. 3a) and $\bullet$ (Fig. 3b), post-processing, and pre-processing for divergence DistP and for divergence XDistP, respectively. We present the details and proofs for these results in Appendices D, E, and F.

The two kinds of composition have been studies in previous work (e.g., [19], [6]). For two mechanisms A_0 and A_1, the composition $A_1 \odot A_0$ means that an identical input value x is given to two DistP mechanisms A_0 and A_1, whereas$A_1 \bullet A_0$ means that independent inputs x_b are provided to mechanisms A_b. Note that this kind of composition is adaptive in the sense that the output of A_1 can be dependent on that of A_0. Hence the compositonality does not hold in general for f-divergence, whereas we show the compositionality for KL-divergence in Tables II and III. For non-adaptive sequential composition, the compositionality of divergence DistP/XDistP is straightforward from [20], which show the compositionality of popular f-divergences, including total variation and Hellinger distance.

As for pre-processing, we use the following definition of stability [6], which is analogous to the stability for DP.

Definition 13 (Stability): Let $c \in \mathbb{N}$, $\Psi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$, and W be a metric over $\mathbb{D}\mathcal{X}$. A transformation $T : \mathbb{D}\mathcal{X} \rightarrow \mathbb{D}\mathcal{X}$ 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, $T : \mathbb{D}\mathcal{X} \rightarrow \mathbb{D}\mathcal{X}$ is (c, W)-*stable* if for any $\lambda_0, \lambda_1 \in \mathbb{D}\mathcal{X}$, $W(T(\lambda_0), T(\lambda_1)) \le c\, W(\lambda_0, \lambda_1)$.

B. Relationships among Distribution Privacy Notions

In Fig. 4 we show the summary of the relationships among notions of divergence XDP and divergence XDistP. See Appendices B and G for details and proofs.

V. Local Mechanisms for Divergence Distribution Privacy

In this section we present how much degree of divergence DistP/XDistP can be achieved by local obfuscation. Specifically, we show how f-divergence privacy contribute

TABLE II: Summary of basic properties of divergence DistP.

Sequential composition $\odot$ (D_{KL})	A_b is $(\varepsilon_b, D_{\mathrm{KL}})$-DistP $\Rightarrow A_1 \odot A_0$ is $(\varepsilon_0+\varepsilon_1, D_{\mathrm{KL}})$-DistP
Sequential composition $\bullet$ (D_{KL})	A_b is $(\varepsilon_b, D_{\mathrm{KL}})$-DistP $\Rightarrow A_1 \bullet A_0$ is $(\varepsilon_0+\varepsilon_1, D_{\mathrm{KL}})$-DistP
Post-processing	A_0 is (ε, D_f)-DistP $\Rightarrow A_1 \circ A_0$ is (ε, D_f)-DistP
Pre-processing (by c-stable T)	A is (ε, D)-DistP $\Rightarrow A \circ T$ is $(c\,\varepsilon, D)$-DistP

TABLE III: Summary of basic properties of divergence XDistP.

Sequential composition $\odot$ (D_{KL})	A_b is $(\varepsilon_b, W_{1,d}, D_{\mathrm{KL}})$-XDistP $\Rightarrow A_1 \odot A_0$ is $(\varepsilon_0+\varepsilon_1, W_{1,d}, D_{\mathrm{KL}})$-XDistP
Sequential composition $\bullet$ (D_{KL})	A_b is $(\varepsilon_b, W_{1,d}, D_{\mathrm{KL}})$-XDistP $\Rightarrow A_1 \bullet A_0$ is $(\varepsilon_0+\varepsilon_1, W_{1,d}, D_{\mathrm{KL}})$-XDistP
Post-processing	A_0 is (ε, W, D_f)-XDistP $\Rightarrow A_1 \circ A_0$ is (ε, W, D_f)-XDistP
Pre-processing (by c-stable T)	A is (ε, W, D)-XDistP $\Rightarrow A \circ T$ is $(c\,\varepsilon, W, D)$-XDistP

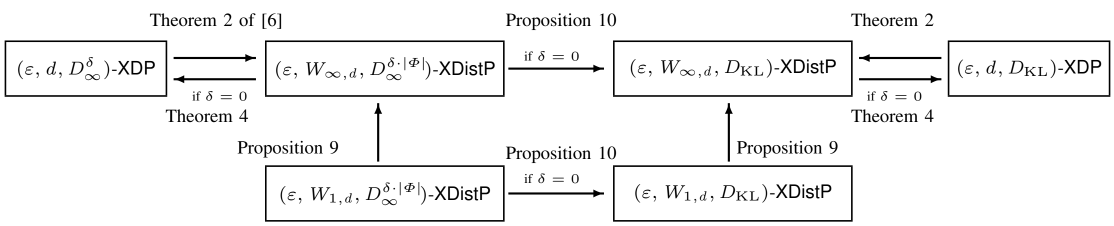


Fig. 4: Relationships among divergence XDistP notions.

to the obfuscation of probability distributions. To prove those results, we use the notion of probability coupling.

A. Divergence DistP by Local Obfuscation

We first show that f-divergence privacy mechanisms provide D_f-DistP. To present this formally, we recall the notion of the lifting of relations as follows.

Definition 14 (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 $\mathsf{supp}(\gamma) \subseteq \Phi$.

Intuitively, when λ_0 and λ_1 are adjacent w.r.t. the lifted relation $\Phi^{\#}$, then we can construct λ_1 from λ_0 according to the coupling γ, that is, 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). Note that by Definition 7, 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 $\Phi^{\#} = \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$.

Now we show that every f-divergence privacy mechanism provides D_f-DistP as follows. (See Appendix A for the proof.)

Theorem 1 ((ε, D_f)-DP $\Rightarrow$ (ε, D_f)-DistP): Let $\Phi \subseteq \mathcal{X} \times \mathcal{X}$. If a randomized algorithm $A : \mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}$ provides (ε, D_f)-DP w.r.t. Φ, then it provides (ε, D_f)-DistP w.r.t. $\Phi^{\#}$.

Intuitively, the f-divergence privacy mechanism A makes any pair (λ_0, λ_1) of input distributions in $\Phi^{\#}$ indistinguishable in terms of D_f up to the threshold ε.

B. Divergence XDistP by Local Obfuscation

Next we investigate how much noise should be added for local obfuscation mechanisms to provide divergence XDistP.

We first consider two point distributions λ_0 at x_0 and λ_1 at x_1, i.e., $\lambda_0[x_0] = \lambda_1[x_1] = 1$. Then an (ε, d, D_f)-XDP mechanism A satisfies:

$$D_f(A^{\#}(\lambda_0) \,\|\, A^{\#}(\lambda_1)) = D_f(A(x_0) \,\|\, A(x_1)) \le \varepsilon d(x_0, x_1).$$

Hence the noise added by A should be proportional to the distance $d(x_0, x_1)$ between x_0 and x_1.

To generalize this observation on point distributions to arbitrary distributions, we need to employ some metric between distributions. As the metric, we could use the *diameter* over the supports, which is defined by:

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

or the ∞-Wasserstein metric $W_{\infty,d}$, which is used for XDistP [6]. However, when there is an outlier in λ_0 or λ_1, then $\mathsf{diam}(\lambda_0, \lambda_1)$ and $W_{\infty,d}(\lambda_0, \lambda_1)$ tend to be large. Since the mechanism needs to add noise proportionally to the distance $\mathsf{diam}(\lambda_0, \lambda_1)$ or $W_{\infty,d}(\lambda_0, \lambda_1)$ to achieve XDistP, it needs to add large amount of noise and thus loses utility significantly.

To have better utility, we employ the Earth mover's distance (1-Wasserstein metric) $W_{1,d}$ as a metric for D_f-XDistP mechanisms. Given two distributions λ_0 and λ_1 over $\mathcal{X}$, we consider a transportation γ from λ_0 to λ_1 that minimizes the expected cost of the transportation. Then the minimum of the expected cost is given by the Earth mover's distance $W_{1,d}(\lambda_0, \lambda_1)$.

Now we show that, to achieve D_f-XDistP, we only have to add noise proportionally to the Earth mover's distance $W_{1,d}$ between the input distributions. To formalize this, we define a lifted relation $\Phi^{\#}_{W_p}$ as the maximum relation over $\mathbb{D}\mathcal{X}$ s.t. for any $(\lambda_0, \lambda_1) \in \Phi^{\#}_{W_p}$, there is a coupling $\gamma \in \mathsf{cp}(\lambda_0, \lambda_1)$ satisfying $\mathsf{supp}(\gamma) \subseteq \Phi$ and $\gamma \in \Gamma_{p,d}(\lambda_0, \lambda_1)$.

Theorem 2 ((ε, d, D_f)-XDP $\Rightarrow$ $(\varepsilon, W_{1,d}, D_f)$-XDistP): Let $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a metric. If a randomized algorithm $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, d, D_f)-XDP w.r.t. Φ then it provides $(\varepsilon, W_{1,d}, D_f)$-XDistP w.r.t. $\Phi^{\#}_{W_1}$.

See Appendix A for the proof. Since the Earth mover's distance is not grater than the diameter or ∞-Wasserstein distance, D_f-XDistP may require less noise than D_∞-XDistP.

VI. Local Distribution Obfuscation with Auxiliary Inputs

In this section we introduce a local obfuscation mechanism which we call a *coupling mechanism* in order to provide distribution privacy while optimizing utility. Specifically, a coupling mechanism uses (full or approximate) knowledge on the input probability distributions to perturb each single input value so that the output distribution gets indistinguishable from some target probability distribution. To define the mechanism, we calculate the probability coupling of each input distribution and the target distribution.

A. Privacy Definitions with Auxiliary Inputs

We first extend the definition of divergence DistP so that a local obfuscation mechanism A can receive some *auxiliary input* (e.g. context information) ranging over a set $\mathcal{S}$, which might be used for A to apply different randomized algorithms in different situations or to different input distributions.

Definition 15 (Divergence DistP with auxiliary inputs): Let $\varepsilon \in \mathbb{R}^{\geq 0}$, $D \in \mathsf{Div}(\mathcal{Y})$, and $\Psi \subseteq (\mathcal{S} \times \mathbb{D}\mathcal{X}) \times (\mathcal{S} \times \mathbb{D}\mathcal{X})$. We say that a randomized algorithm $A : \mathcal{S} \times \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, D)-*distribution privacy w.r.t.* Ψ if for all pairs $((s, \lambda), (s', \lambda')) \in \Psi$,

$$D(A^{\#}(s, \lambda) \parallel A^{\#}(s', \lambda')) \leq \varepsilon.$$

In this definition, the auxiliary input over $\mathcal{S}$ typically represents contextual information about where the obfuscation mechanism A is used or what distribution an input is sampled from. Such information may be useful to customize A to improve utility while providing distribution privacy in specific situations. For example, assume that each auxiliary input s represents the fact that an input x is sampled from a distribution λ_s. If a local mechanism A uses this auxiliary information to always produce a distribution μ of outputs[4], it can prevent the leakage of information on the input distribution λ_s. We elaborate on this in the next sections.

[4]If A can use no auxiliary information but wants to produce μ, then the output value needs to be independent of the input, hence very poor utility.

B. Coupling Mechanisms

In this section we introduce a new local obfuscation mechanism, which we call a *coupling mechanism*. The aim of the new mechanism is to improve the utility while protecting distribution privacy when we know the input distribution fully or approximately. Intuitively, a coupling mechanism uses (full or partial) information on the input distribution $\lambda \in \mathbb{D}\mathcal{X}$ and produces an output value following some identical distribution $\mu \in \mathbb{D}\mathcal{Y}$, which we call a *target distribution*. More specifically, given some auxiliary information s about λ, a coupling mechanism $A : \mathcal{S} \times \mathcal{X} \to \mathbb{D}\mathcal{Y}$ probabilistically maps each input value x to some output value y so that y is distributed over the target distribution μ.

The simplest construction of a coupling mechanism would be to randomly sample a value y from μ independently of the input x. However, this mechanism provides very poor utility, since the output y loses all information on x.

Instead, we construct a mechanism by calculating a coupling $\gamma \in \mathbb{D}(\mathcal{X}\times\mathcal{Y})$ that transforms λ to μ with the minimum loss. We explain this using a simple example below.

Example 2 (Coupling mechanism): A coupling γ of two distributions λ and μ (Fig. 2b) shows a way of transforming λ to μ by probabilistically adding noise to each single input value drawn from λ. More specifically, $\gamma[2, 1] = 0.1$ means that 0.1 (out of $\lambda[2] = 0.5$) moves from 2 to 1, and $\gamma[2, 3] = 0.2$ means that 0.2 moves from 2 to 3. Based on this coupling γ, we construct the coupling mechanism C that maps the input 2 to the output 1 with probability $20\%(= 0.1/0.5)$, and to the output 3 with probability $40\%(= 0.2/0.5)$. By applying this mechanism C to the input distribution λ, the resulting output distribution $C^{\#}(\lambda)$ is identical to μ.

Formally, we assume that for each auxiliary input $s \in \mathcal{S}$, we learn that the input distribution is approximately $\widehat{\lambda}_s \in \mathbb{D}\mathcal{X}$ while the actual distribution is $\lambda_s \in \mathbb{D}\mathcal{X}$. Then we define the coupling mechanism C as follows.

Definition 16 (Coupling mechanism): Let $\mu \in \mathbb{D}\mathcal{Y}$. For each $s \in \mathcal{S}$, let $\widehat{\lambda}_s \in \mathbb{D}\mathcal{X}$ be an approximate input distribution, and $\gamma_s \in \mathsf{cp}(\widehat{\lambda}_s, \mu)$ be a coupling of $\widehat{\lambda}_s$ and μ. Then a *coupling mechanism* w.r.t. μ is defined as a randomized algorithm $C : \mathcal{S} \times \mathcal{X} \to \mathbb{D}\mathcal{Y}$ such that given $s \in \mathcal{S}$ and $x \in \mathcal{X}$, outputs $y \in \mathcal{Y}$ with the probability:

$$C(s, x)[y] = \tfrac{\gamma_s[x,y]}{\widehat{\lambda}_s[x]}.$$

When C can access the exact information on λ_s (i.e., $\widehat{\lambda}_s$ is identical to the actual distribution λ_s from which inputs are sampled), then C provides $(0, D)$-DistP for any divergence D, i.e., no information on the input distribution is leaked by the output of C. However, we often obtain only approximate information on the input distribution. In this case, C still provides strong privacy as shown in the next section.

C. Distribution Privacy of Coupling Mechanisms

In this section we evaluate the DistP and utility of coupling mechanisms. (See Appendix C for the proof.)

Theorem 3 (DistP of the coupling mechanism): Let $\Psi \subseteq (\mathcal{S} \times \mathbb{D}\mathcal{X}) \times (\mathcal{S} \times \mathbb{D}\mathcal{X})$ such that each element of Ψ is of the form (s, λ_s) for some $s \in \mathcal{S}$. Let C be a coupling mechanism w.r.t. a target distribution μ. Assume that for each $s \in \mathcal{S}$, the approximate knowledge $\widehat{\lambda}_s$ is close to the actual distribution λ_s in the sense that $D_\infty(\widehat{\lambda}_s \parallel \lambda_s) \leq \varepsilon$ and $D_\infty(\lambda_s \parallel \widehat{\lambda}_s) \leq \varepsilon$. Then C provides:

1) $(2\varepsilon, D_\infty)$-DistP w.r.t. Ψ;
2) $(2\varepsilon\, e^\varepsilon, D_{\mathrm{KL}})$-DistP w.r.t. Ψ;
3) $(e^\varepsilon f(e^{2\varepsilon}), D_f)$-DistP w.r.t. Ψ.

This theorem implies that when the mechanism C learns the exact distribution, i.e., $\widehat{\lambda}_s = \lambda_s$, then by $\varepsilon = 0$ it provides $(0, D_\infty)$-DistP, hence there is no leaked information on the input distributions. For $\varepsilon \approx 0$, we have $\varepsilon\, e^\varepsilon \approx \varepsilon(1+\varepsilon) \approx \varepsilon$, hence C provides approximately $(2\varepsilon, D_{\mathrm{KL}})$-DistP .

D. Utility-Optimal Coupling Mechanisms

In this section we introduce a utility-optimal coupling mechanism. Here we assume that there is some metric d over $\mathcal{X} \cup \mathcal{Y}$. Then the notion of utility loss of a local obfuscation mechanism is defined as follows.

Definition 17 (Expected utility loss): Given an input distribution $\lambda \in \mathbb{D}\mathcal{X}$ and a metric d over $\mathcal{X} \cup \mathcal{Y}$, the *expected utility loss* of a randomized algorithm $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ is:

$$\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \lambda[x] A(x)[y]\, d(x, y).$$

The utility loss of a coupling mechanism depends on the choice of the coupling used in the mechanism. Given an Euclid distance d and an input distribution $\widehat{\lambda}_s$, the expected utility loss of a coupling mechanism w.r.t. a target distribution μ using a coupling γ_s is represented by $\sum_{(x_0,x_1) \in \mathsf{supp}(\gamma_s)} d(x_0, x_1)\gamma_s[x_0, x_1]$.

Now we define the coupling mechanism that minimizes the expected utility loss as follows.

Definition 18 (Utility-optimal coupling mechanism): Let $\mu \in \mathbb{D}\mathcal{Y}$. A *utility-optimal coupling mechanism* w.r.t. μ is a coupling mechanism w.r.t. μ that uses a coupling $\gamma_s \in \mathit{\Gamma}_{1,d}(\lambda_s, \mu)$ for each $s \in \mathcal{S}$.

Proposition 2 (Loss of the coupling mechanism): For each $s \in \mathcal{S}$, the expected utility loss of a utility-optimal coupling mechanism w.r.t. a target distribution $\mu \in \mathbb{D}\mathcal{Y}$ is given by the Earth mover's distance $W_{1,d}(\widehat{\lambda}_s, \mu)$.

The proof is straightforward from the definition of the Earth mover's distance. Note that as mentioned in Section II-G, the coupling $\gamma_s \in \mathit{\Gamma}_{1,d}(\lambda_s, \mu)$ can be efficiently calculated by the North-West corner rule when d is submodular.

Analogously, we could define a coupling mechanism that minimizes the maximum loss by using a coupling $\gamma_s \in \mathit{\Gamma}_{\infty,d}(\lambda_s, \mu)$ for each $s \in \mathcal{S}$. Then the worst-case utility loss is given by the ∞-Wasserstein metric $W_{\infty,d}(\widehat{\lambda}_s, \mu)$.

VII. Related work

Since the seminal work of Dwork [1] on differential privacy (DP), a lot of its variants have been studied to provide different types of privacy guarantees [21]; e.g., d-privacy [13], f-divergence privacy [20], [8], mutual-information DP [9], concentrated DP [22], Rényi DP [10], Pufferfish privacy [23], Bayesian DP [24], local DP [2], personalized DP [25], and utility-optimized local DP [26]. All of these are intended to protect single input values instead of input distributions.

A few researches have explored the privacy of distributions. Jelasity *et al*. [5] propose distributional DP to protect the privacy of distribution parameters θ in a Bayesian style (unlike DP and DistP). Kawamoto *et al*. [6] propose the DistP notion in a DP style. Geumlek *et al*. [7] propose profile-based privacy, a variant of DistP that allows the mechanisms to depend on the perfect knowledge of input distributions. However, these studies deal only with the worst-case risk, and neither relax them to the average-case risk (with divergence) nor allow them to use arbitrary auxiliary information (in spite that available information on input distributions is often approximate only).

There have been many studies (e.g., [27]) on the *DP of histogram publishing*, which is different from DistP as follows. Histogram publishing is a *central* mechanism that *hides a single record* $x \in \mathcal{X}$ and outputs an obfuscated histogram, e.g., $\mu \in \mathbb{D}\mathcal{Y}$, 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 $y \in \mathcal{Y}$. As explained in [6], neither of these implies the other.

VIII. Conclusion

We introduced the notions of divergence DistP and presented their useful theoretical properties in a general form. By using probability coupling techniques, we presented how much divergence DistP can be achieved by local obfuscation. In particular, we proved that the perturbation noise should be added proportionally to the Earth mover's distance between the input distributions. We also proposed a local mechanism called a (utility-optimal) coupling mechanism and theoretically evaluated their DistP and utility loss in the presence of (exact or approximate) knowledge on the input distributions.

As for future work, we are planning to develop various kinds of coupling mechanisms for specific applications, such as location privacy.

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] M. Jelasity and K. P. Birman, "Distributional differential privacy for large-scale smart metering," in *Proc. IH&MMSec*, 2014, pp. 141–146.
[6] Y. Kawamoto and T. Murakami, "Local obfuscation mechanisms for hiding probability distributions," in *Proc. ESORICS*, 2019, pp. 128–148.
[7] J. Geumlek and K. Chaudhuri, "Profile-based privacy for locally private computations," *CoRR*, vol. abs/1903.09084, 2019.
[8] R. F. Barber and J. C. Duchi, "Privacy and statistical risk: Formalisms and minimax bounds," *CoRR*, vol. abs/1412.4451, 2014.
[9] P. Cuff and L. Yu, "Differential privacy as a mutual information constraint," in *Proc. CCS*, 2016, pp. 43–54.
[10] I. Mironov, "Rényi differential privacy," in *Proc. CSF*, 2017, pp. 263–275.

[11] D. Yang, D. Zhang, and B. Qu, "Participatory cultural mapping based on collective behavior data in location based social networks," *ACM TIST*, vol. 7, no. 3, pp. 30:1–30:23, 2015.

[12] 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.

[13] 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.

[14] 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.

[15] I. Csiszar, "Information measures of difference of probability distributions and indirect observations," *Studia Sci. Math. Hungar.*, vol. 2, pp. 299–318, 1967.

[16] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 03 1951.

[17] L. Vaserstein, "Markovian processes on countable space product describing large systems of automata." *Probl. Peredachi Inf.*, vol. 5, no. 3, pp. 64–72, 1969.

[18] A. J. Hoffman, "On simple linear programming problems," in *Convexity: Proceedings of the Seventh Symposium in Pure Mathematics of the American Mathematical Society*, 1963, vol. 7, p. 317.

[19] Y. Kawamoto, K. Chatzikokolakis, and C. Palamidessi, "On the compositionality of quantitative information flow," *Log. Methods Comput. Sci.*, vol. 13, no. 3, 2017.

[20] 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.

[21] D. Desfontaines and B. Pejó, "Sok: Differential privacies," *CoRR*, vol. abs/1906.01337, 2019. [Online]. Available: http://arxiv.org/abs/1906.01337

[22] C. Dwork and G. N. Rothblum, "Concentrated differential privacy," *CoRR*, vol. abs/1603.01887, 2016.

[23] D. Kifer and A. Machanavajjhala, "A rigorous and customizable framework for privacy," in *Proc. PODS*, 2012, pp. 77–88.

[24] B. Yang, I. Sato, and H. Nakagawa, "Bayesian differential privacy on correlated data," in *Proc. SIGMOD*, 2015, pp. 747–762.

[25] Z. Jorgensen, T. Yu, and G. Cormode, "Conservative or liberal? Personalized differential privacy," in *Proc. ICDE'15*, 2015, pp. 1023–1034.

[26] T. Murakami and Y. Kawamoto, "Utility-optimized local differential privacy mechanisms for distribution estimation," in *Proc. USENIX Security*, 2019, pp. 1877–1894.

[27] 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.

APPENDIX

A. Local Mechanisms for D_f-DistP/XDistP

We first show the proofs for the D_f-DistP/XDistP achieved by local obfuscation mechanisms.

Theorem 1 ((ε, D_f)-DP $\Rightarrow$ (ε, D_f)-DistP): Let $\Phi \subseteq \mathcal{X} \times \mathcal{X}$. If a randomized algorithm $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, D_f)-DP w.r.t. Φ, then it provides (ε, D_f)-DistP w.r.t. $\Phi^\#$.

Proof: Let $(\lambda_0, \lambda_1) \in \Phi^\#$ and $\Gamma \stackrel{\text{def}}{=} \mathsf{cp}(\lambda_0, \lambda_1)$.

$$
\begin{aligned}
& D_f(A^\#(\lambda_0) \parallel A^\#(\lambda_1)) \\
&= \textstyle\sum_y A^\#(\lambda_1)[y] \ f\big(\frac{A^\#(\lambda_0)[y]}{A^\#(\lambda_1)[y]}\big) \\
&= \textstyle\sum_y \sum_{x_1} \lambda_1[x_1]\, A(x_1)[y] \ f\Big(\frac{\sum_{x_0} \lambda_0[x_0]\, A(x_0)[y]}{\sum_{x_1} \lambda_1[x_1]\, A(x_1)[y]}\Big) \\
&= \min_{\gamma\in\Gamma} \textstyle\sum_{y, x_0, x_1} \gamma[x_0, x_1]\, A(x_1)[y] \ f\Big(\frac{\sum_{x_0,x_1} \gamma[x_0,x_1]\, A(x_0)[y]}{\sum_{x_0,x_1} \gamma[x_0,x_1]\, A(x_1)[y]}\Big) \\
&\hspace{12em} \text{(where } (x_0, x_1) \text{ ranges over } \mathsf{supp}(\gamma)) \\
&= \min_{\gamma\in\Gamma} \textstyle\sum_y c \ f\Big(\frac{1}{c} \sum_{x_0,x_1} \gamma[x_0, x_1]\, A(x_0)[y]\Big) \\
&\hspace{12em} \text{(where } c = \textstyle\sum_{x_0,x_1} \gamma[x_0,x_1]\, A(x_1)[y]) \\
&= \min_{\gamma\in\Gamma} \textstyle\sum_y c \ f\big(\sum_{x_0,x_1} \frac{\gamma[x_0,x_1]\, A(x_1)[y]}{c} \cdot \frac{\gamma[x_0,x_1]\, A(x_0)[y]}{\gamma[x_0,x_1]\, A(x_1)[y]}\big) \\
&\le \min_{\gamma\in\Gamma} \textstyle\sum_y c \sum_{x_0,x_1} \frac{\gamma[x_0,x_1]\, A(x_1)[y]}{c} \cdot f\big(\frac{\gamma[x_0,x_1]\, A(x_0)[y]}{\gamma[x_0,x_1]\, A(x_1)[y]}\big) \\
&\hspace{8em} \text{(by Jensen's inequality and the convexity of } f) \\
&= \min_{\gamma\in\Gamma} \textstyle\sum_{x_0,x_1} \gamma[x_0, x_1] \displaystyle\sum_y A(x_1)[y] \ \cdot f\big(\tfrac{A(x_0)[y]}{A(x_1)[y]}\big) \\
&= \min_{\gamma\in\Gamma} \textstyle\sum_{x_0,x_1} \gamma[x_0, x_1]\, D_f(A(x_0) \parallel A(x_1)). \qquad (1)
\end{aligned}
$$

Assume that A provides (ε, D_f)-DP w.r.t. Φ. By Definition 14, there is a coupling $\gamma \in \Gamma$ with $\mathsf{supp}(\gamma) \subseteq \Phi$. Then:

$$
\begin{aligned}
& D_f(A^\#(\lambda_0) \parallel A^\#(\lambda_1)) \\
&= \min_{\gamma\in\Gamma} \textstyle\sum_{x_0,x_1} \gamma[x_0, x_1]\, D_f(A(x_0) \parallel A(x_1)) \qquad \text{(by (1))} \\
&\le \min_{\gamma\in\Gamma} \textstyle\sum_{x_0,x_1} \gamma[x_0, x_1]\, \varepsilon \\
&\hspace{4em} \text{(by } (x_0, x_1) \in \mathsf{supp}(\gamma) \subseteq \Phi \text{ and } (\varepsilon, D_f)\text{-DP)} \\
&= \varepsilon.
\end{aligned}
$$

Hence A provides (ε, D_f)-DistP w.r.t. $\Phi^\#$. ■

Theorem 2 ((ε, d, D_f)-XDP $\Rightarrow$ ($\varepsilon, W_{1,d}, D_f$)-XDistP): Let $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a metric. If a randomized algorithm $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, d, D_f)-XDP w.r.t. Φ then it provides $(\varepsilon, W_{1,d}, D_f)$-XDistP w.r.t. $\Phi^\#_{W_1}$.

Proof: Assume that A provides (ε, d, D_f)-XDP w.r.t. Φ. Let $(\lambda_0, \lambda_1) \in \Phi^\#_{W_1}$. By definition, there exists a coupling $\gamma \in \Gamma$ that satisfies $\mathsf{supp}(\gamma) \subseteq \Phi$ and $\gamma \in \Gamma_{1,d}(\lambda_0, \lambda_1)$. Then it follows from (1) in the proof for Theorem 1 that:

$$
\begin{aligned}
& D_f(A^\#(\lambda_0) \parallel A^\#(\lambda_1)) \\
&= \min_{\gamma\in\Gamma} \textstyle\sum_{x_0,x_1} \gamma[x_0, x_1]\, D_f(A(x_0) \parallel A(x_1)) \qquad \text{(by (1))} \\
&\le \min_{\gamma\in\Gamma} \textstyle\sum_{x_0,x_1} \gamma[x_0, x_1]\, \varepsilon\, d(x_0, x_1) \\
&\hspace{4em} \text{(by } (x_0, x_1) \in \mathsf{supp}(\gamma) \subseteq \Phi \text{ and } (\varepsilon, d, D_f)\text{-XDP)} \\
&= \varepsilon\, W_{1,d}(\lambda_0, \lambda_1). \qquad \text{(by } \gamma \in \Gamma_{1,d}(\lambda_0, \lambda_1))
\end{aligned}
$$

Hence A provides $(\varepsilon, W_{1,d}, D_f)$-XDistP w.r.t. $\Phi^\#_{W_1}$. ■

B. Point Obfuscation by Distribution Obfuscation

Next we show that divergence DP is an instance of divergence DistP if an adjacency relation includes pairs of point distributions (i.e., distributions having single points with probability 1).

Lemma 1: Let $p \in \mathbb{R}^{\ge 1} \cup \{\infty\}$ and $\Phi \subseteq \mathcal{X} \times \mathcal{X}$. For any $(x_0, x_1) \in \Phi$, we have $(\eta_{x_0}, \eta_{x_1}) \in \Phi^\#_{W_p}$.

Theorem 4 (DistP $\Rightarrow$ DP and XDistP $\Rightarrow$ XDP): Let $\varepsilon \in \mathbb{R}^{\ge 0}$, $p \in \mathbb{R}^{\ge 1} \cup \{\infty\}$, $D \in \mathsf{Div}(\mathcal{Y})$, $\Phi \subseteq \mathcal{X} \times \mathcal{X}$, and $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ be a randomized algorithm.

1) If A provides (ε, D)-DistP w.r.t. $\Phi^\#$, then it provides (ε, D)-DP w.r.t. Φ.
2) If A provides $(\varepsilon, W_{p,d}, D)$-XDistP w.r.t. $\Phi^\#_{W_p}$, then it provides (ε, d, D)-XDP w.r.t. Φ.

Proof: We show the first claim as follows. Assume that A provides (ε, D)-DistP w.r.t. $\Phi^\#$. Let $(x_0, x_1) \in \Phi$, and η_{x_0} and η_{x_1} be the point distributions. By Lemma 1 and $\Phi^\#_{W_p} \subseteq \Phi^\#$, we have $(\eta_{x_0}, \eta_{x_1}) \in \Phi^\#$. By (ε, D)-DistP, we obtain $D(A(x_0) \parallel A(x_1)) = D(A^{\#}(\eta_{x_0}) \parallel A^{\#}(\eta_{x_1})) \le \varepsilon$. Hence A provides (ε, D)-DP w.r.t. Φ.

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

$$\begin{aligned} D(A(x_0) \,\|\, A(x_1)) =&\ D(A^{\#}(\eta_{x_0}) \,\|\, A^{\#}(\eta_{x_1})) \\ \le&\ \varepsilon W_{p,d}(\eta_{x_0}, \eta_{x_1}) \quad \text{(by XDistP of } A) \\ =&\ \varepsilon d(x_0, x_1), \end{aligned}$$

where the last equality follows from the definition of $W_{p,d}$. Hence A provides (ε, d, D)-XDP w.r.t. Φ. ■

C. Privacy and Utility of Coupling Mechanisms

Next, we show the privacy of the coupling mechanisms.

Theorem 3 (DistP of the coupling mechanism): Let $\Psi \subseteq (\mathcal{S} \times \mathbb{D}\mathcal{X}) \times (\mathcal{S} \times \mathbb{D}\mathcal{X})$ such that each element of Ψ is of the form (s, λ_s) for some $s \in \mathcal{S}$. Let C be a coupling mechanism w.r.t. a target distribution μ. Assume that for each $s \in \mathcal{S}$, the approximate knowledge $\widehat{\lambda}_s$ is close to the actual distribution λ_s in the sense that $D_\infty(\widehat{\lambda}_s \parallel \lambda_s) \le \varepsilon$ and $D_\infty(\lambda_s \parallel \widehat{\lambda}_s) \le \varepsilon$. Then C provides:

1) $(2\varepsilon, D_\infty)$-DistP w.r.t. Ψ;
2) $(2\varepsilon\, e^{\varepsilon}, D_{\mathrm{KL}})$-DistP w.r.t. Ψ;
3) $(e^{\varepsilon} f(e^{2\varepsilon}), D_f)$-DistP w.r.t. Ψ.

Proof: Let $((s_0, \lambda_{s_0}), (s_1, \lambda_{s_1})) \in \Psi$, and $R \subseteq \mathcal{Y}$. When C is applied to λ_{s_0} the output distribution is given by:

$$\begin{aligned} C^{\#}(s_0, \lambda_{s_0})[R] &= \sum_{x \in \mathcal{X}} \lambda_{s_0}[x] \cdot \tfrac{\gamma_{s_0}[x,R]}{\widehat{\lambda}_{s_0}[x]} \\ &\le e^{\varepsilon} \sum_{x \in \mathcal{X}} \gamma_{s_0}[x, R] \ \text{ (by } D_\infty(\lambda_{s_0} \,\|\, \widehat{\lambda}_{s_0}) \le \varepsilon) \\ &= e^{\varepsilon} \mu[R]. \end{aligned}$$

When C is applied to λ_{s_1} the output distribution is:

$$\begin{aligned} C^{\#}(s_1, \lambda_{s_1})[R] &= \sum_{x \in \mathcal{X}} \lambda_{s_1}[x] \cdot \tfrac{\gamma_{s_1}[x,R]}{\widehat{\lambda}_{s_1}[x]} \\ &\ge e^{-\varepsilon} \sum_{x \in \mathcal{X}} \gamma_{s_1}[x, R] \ \text{ (by } D_\infty(\widehat{\lambda}_{s_1} \,\|\, \lambda_{s_1}) \le \varepsilon) \\ &= e^{-\varepsilon} \mu[R]. \end{aligned}$$

Hence $\frac{C^{\#}(s_0,\lambda_0)[R]}{C^{\#}(s_1,\lambda_1)[R]} \le e^{2\varepsilon}$. Therefore C provides $(2\varepsilon, D_\infty)$-DistP w.r.t. Ψ.

Next the KL-divergence is given by:

$$\begin{aligned} & D_{\mathrm{KL}}(C^{\#}(s_0, \lambda_{s_0}) \,\|\, C^{\#}(s_1, \lambda_{s_1})) \\ &= \sup_y C^{\#}(s_0, \lambda_{s_0})[y] \cdot \ln\left(\tfrac{C^{\#}(s_0,\lambda_{s_0})[y]}{C^{\#}(s_1,\lambda_{s_1})[y]}\right) \\ &\le e^{\varepsilon} \sup_y \mu[y] \ln\left(e^{2\varepsilon}\right) \\ &\le 2\varepsilon\, e^{\varepsilon}. \end{aligned}$$

Therefore C provides $(2\varepsilon\, e^{\varepsilon}, D_f)$-DistP w.r.t. Ψ.

Finally, the f-divergence is given by:

$$\begin{aligned} & D_f(C^{\#}(s_0, \lambda_{s_0}) \,\|\, C^{\#}(s_1, \lambda_{s_1})) \\ &= \sup_y C^{\#}(s_1, \lambda_{s_1})[y] \cdot f\left(\tfrac{C^{\#}(s_0,\lambda_{s_0})[y]}{C^{\#}(s_1,\lambda_{s_1})[y]}\right) \\ &\le e^{\varepsilon} \sup_y \mu[y] f\left(e^{2\varepsilon}\right) \\ &\le e^{\varepsilon} f(e^{2\varepsilon}). \end{aligned}$$

Therefore C provides $(e^{\varepsilon} f(e^{2\varepsilon}), D_f)$-DistP w.r.t. Ψ. ■

D. Sequential Composition $\odot$ with Shared Input

We first recall the definition of the sequential composition $\odot$ with shared input (Fig. 3a) in previous work.

Definition 19 (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 present the compositionality of D_{KL}-DistP. Note that since this composition is adaptive, the compositionality does not hold in general for f-divergence.

Proposition 3 (Sequential composition $\odot$ of D_{KL}-DistP): Let $\Phi \subseteq \mathcal{X} \times \mathcal{X}$. If $A_0 : \mathcal{X} \to \mathbb{D}\mathcal{Y}_0$ provides $(\varepsilon_0, D_{\mathrm{KL}})$-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 $(\varepsilon_1, D_{\mathrm{KL}})$-DistP w.r.t. $\Phi^{\#}$, the sequential composition $A_1 \odot A_0$ provides $(\varepsilon_0 + \varepsilon_1, D_{\mathrm{KL}})$-DistP w.r.t. $\Phi^{\#}$.

Proof: By Theorem 4 in Appendix B, A_0 provides $(\varepsilon_0, D_{\mathrm{KL}})$-DP w.r.t. Φ, and for each $y_0 \in \mathcal{Y}_0$, $A_1(y_0)$ provides $(\varepsilon_1, D_{\mathrm{KL}})$-DP w.r.t. Φ. Let $(x, x') \in \Phi$. Then:

$$\begin{aligned} & D_{\mathrm{KL}}((A_1 \odot A_0)(x) \parallel (A_1 \odot A_0)(x')) \\ &= \sum_{y_1} (A_1 \odot A_0)(x)[y_1] \ln \tfrac{(A_1 \odot A_0)(x)[y_1]}{(A_1 \odot A_0)(x')[y_1]} \\ &= \sum_{y_0, y_1} A_0(x)[y_0] \cdot A_1(y_0, x)[y_1] \ln \tfrac{A_0(x)[y_0] \cdot A_1(y_0,x)[y_1]}{A_0(x')[y_0] \cdot A_1(y_0,x')[y_1]} \\ &= \sum_{y_0} A_0(x)[y_0] \ln \tfrac{A_0(x)[y_0]}{A_0(x')[y_0]} \\ &\quad + \sum_{y_0, y_1} A_0(x)[y_0] A_1(y_0, x)[y_1] \ln \tfrac{A_1(y_0,x)[y_1]}{A_1(y_0,x')[y_1]} \\ &\le D_{\mathrm{KL}}(A_0(x) \,\|\, A_0(x')) \\ &\quad + \max_{y_0} \sum_{y_1} A_1(y_0, x)[y_1] \ln \tfrac{A_1(y_0,x)[y_1]}{A_1(y_0,x')[y_1]} \\ &= D_{\mathrm{KL}}(A_0(x) \,\|\, A_0(x')) + \max_{y_0} D_{\mathrm{KL}}(A_1(y_0, x) \,\|\, A_1(y_0, x')) \\ &\le \varepsilon_0 + \varepsilon_1. \end{aligned}$$

Hence $A_1 \odot A_0$ provides $(\varepsilon_0 + \varepsilon_1, D_{\mathrm{KL}})$-DP w.r.t. Φ. By Theorem 1, $A_1 \odot A_0$ provides $(\varepsilon_0 + \varepsilon_1, D_{\mathrm{KL}})$-DistP w.r.t. $\Phi^{\#}$. ■

Proposition 4 (Sequential composition $\odot$ of D_{KL}-XDistP): Let d be a metric over $\mathcal{X}$, and $\Phi \subseteq \mathcal{X} \times \mathcal{X}$. If $A_0 : \mathcal{X} \to \mathbb{D}\mathcal{Y}_0$ provides $(\varepsilon_0, W_{1,d}, D_{\mathrm{KL}})$-XDistP w.r.t. $\Phi^{\#}_{W_1}$ and for each $y_0 \in \mathcal{Y}_0$, $A_1(y_0) : \mathcal{X} \to \mathbb{D}\mathcal{Y}_1$ provides $(\varepsilon_1, W_{1,d}, D_{\mathrm{KL}})$-XDistP w.r.t. $\Phi^{\#}_{W_1}$ then the sequential composition $A_1 \odot A_0$ provides $(\varepsilon_0 + \varepsilon_1, W_{1,d}, D_{\mathrm{KL}})$-XDistP w.r.t. $\Phi^{\#}_{W_1}$.

Proof: Analogous to the proof for Proposition 3. ■

E. Sequential Composition $\bullet$ with Independent Sampling

In this section we present the compositionality with independent sampling, which is defined as follows.

Definition 20 (Sequential composition $\bullet$): 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 $A_1 \bullet A_0 : \mathcal{X} \times \mathcal{X} \to \mathbb{D}\mathcal{Y}_1$ such that: for any $x_0, x_1 \in \mathcal{X}$, $(A_1 \bullet A_0)(x_0, x_1) = A_1(A_0(x_0), x_1))$.

We define an operator $\diamond$ between binary relations $\varPsi_0$ and $\varPsi_1$:

$$\varPsi_0 \diamond \varPsi_1 = \{(\lambda_0 \times \lambda_1, \lambda'_0 \times \lambda'_1) \mid (\lambda_0, \lambda'_0) \in \varPsi_0, (\lambda_1, \lambda'_1) \in \varPsi_1\}.$$

Now we show the compositionality for D_{KL}-DistP.

Proposition 5 (Sequential composition $\bullet$ of D_{KL}-DistP): Let $\varPsi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$. If $A_0 : \mathcal{X} \to \mathbb{D}\mathcal{Y}_0$ provides $(\varepsilon_0, D_{\mathrm{KL}})$-DistP w.r.t. $\varPsi$ and for each $y_0 \in \mathcal{Y}_0$, $A_1(y_0) : \mathcal{X} \to \mathbb{D}\mathcal{Y}_1$ provides $(\varepsilon_1, D_{\mathrm{KL}})$-DistP w.r.t. $\varPsi$, then the composition $A_1 \bullet A_0$ provides $(\varepsilon_0 + \varepsilon_1, D_{\mathrm{KL}})$-DistP w.r.t. $\varPsi \diamond \varPsi$.

Proof: Let $(\lambda_0, \lambda'_0), (\lambda_1, \lambda'_1) \in \varPsi$.

$$\begin{aligned}
&D_{\mathrm{KL}}((A_1 \bullet A_0)^{\#}(\lambda_0 \times \lambda_1) \parallel (A_1 \bullet A_0)^{\#}(\lambda'_0 \times \lambda'_1)) \\
&= \sum_{y_1} (A_1 \bullet A_0)^{\#}(\lambda_0 \times \lambda_1)[y_1] \ln \tfrac{(A_1 \bullet A_0)^{\#}(\lambda_0 \times \lambda_1)[y_1]}{(A_1 \bullet A_0)^{\#}(\lambda'_0 \times \lambda'_1)[y_1]} \\
&= \sum_{y_0, y_1} {A_0}^{\#}(\lambda_0)[y_0] A_1(y_0)^{\#}(\lambda_1)[y_1] \ln \tfrac{{A_0}^{\#}(\lambda_0)[y_0] A_1(y_0)^{\#}(\lambda_1)[y_1]}{{A_0}^{\#}(\lambda'_0)[y_0] A_1(y_0)^{\#}(\lambda'_1)[y_1]} \\
&= \sum_{y_0} {A_0}^{\#}(\lambda_0)[y_0] \ln \tfrac{{A_0}^{\#}(\lambda_0)[y_0]}{{A_0}^{\#}(\lambda'_0)[y_0]} \\
&\quad + \sum_{y_0, y_1} {A_0}^{\#}(\lambda_0)[y_0] A_1(y_0)^{\#}(\lambda_1)[y_1] \ln \tfrac{A_1(y_0)^{\#}(\lambda_1)[y_1]}{A_1(y_0)^{\#}(\lambda'_1)[y_1]} \\
&\le D_{\mathrm{KL}}({A_0}^{\#}(\lambda_0) \parallel {A_0}^{\#}(\lambda'_0)) \\
&\quad + \max_{y_0} \sum_{y_1} {A_1}^{\#}(y_0)(\lambda_1)[y_1] \ln \tfrac{{A_1}^{\#}(y_0)(\lambda_1)[y_1]}{{A_1}^{\#}(y_0)(\lambda'_1)[y_1]} \\
&= D_{\mathrm{KL}}({A_0}^{\#}(\lambda_0) \parallel {A_0}^{\#}(\lambda'_0)) \\
&\quad + \max_{y_0} D_{\mathrm{KL}}({A_1}^{\#}(y_0)(\lambda_1) \parallel {A_1}^{\#}(y_0)(\lambda'_1)) \\
&\le \varepsilon_0 + \varepsilon_1.
\end{aligned}$$

Hence $A_1 \bullet A_0$ provides $(\varepsilon_0 + \varepsilon_1, D_{\mathrm{KL}})$-DistP w.r.t. $\varPsi \diamond \varPsi$. ■

Proposition 6 (Sequential composition $\bullet$ of D_{KL}-XDistP): Let d be a metric over $\mathcal{X}$, and $\varPsi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$. If $A_0 : \mathcal{X} \to \mathbb{D}\mathcal{Y}_0$ provides $(\varepsilon_0, W_{1,d}, D_{\mathrm{KL}})$-XDistP w.r.t. $\varPsi$ and for each $y_0 \in \mathcal{Y}_0$, $A_1(y_0) : \mathcal{X} \to \mathbb{D}\mathcal{Y}_1$ provides $(\varepsilon_1, W_{1,d}, D_{\mathrm{KL}})$-XDistP w.r.t. $\varPsi$, then the composition $A_1 \bullet A_0$ provides $(\varepsilon_0 + \varepsilon_1, W_{1,d}, D_{\mathrm{KL}})$-XDistP w.r.t. $\varPsi \diamond \varPsi$.

Proof: Analogous to the proof for Proposition 5. ■

F. Post-processing and Pre-processing

Next we show that divergence 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 7 (Post-processing): Let $\varPsi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$, and $W : \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X} \to \mathbb{R}^{\ge 0}$ be a metric. 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, D_f)-DistP w.r.t. $\varPsi$ then so does the composite function $A_1 \circ A_0$.
2) If A_0 provides (ε, W, D_f)-XDistP w.r.t. $\varPsi$ then so does the composite function $A_1 \circ A_0$.

Proof: The claim is immediate from the data processing inequality for the f-divergence. ■

We then show properties of pre-processing as follows.

Proposition 8 (Pre-processing): Let $c \in \mathbb{R}^{\ge 0}$, $\varPsi \subseteq \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}$, $W : \mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X} \to \mathbb{R}^{\ge 0}$ be a metric, and $D \in \mathsf{Div}(\mathcal{Y})$.

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

Proof: We show the first claim as follows. Assume that A provides (ε, D)-DistP w.r.t. $\varPsi$. Let $(\lambda, \lambda') \in \varPsi$. Then $D((A \circ T)^{\#}(\lambda) \parallel (A \circ T)^{\#}(\lambda')) = D(A^{\#}(T^{\#}(\lambda)) \parallel A^{\#}(T^{\#}(\lambda'))) \le c\varepsilon$ by $(c, \varPsi)$-stability. Therefore $A \circ T$ provides $(c\,\varepsilon, D)$-DistP w.r.t. $\varPsi$.

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

$$\begin{aligned}
&D((A \circ T)^{\#}(\lambda) \parallel (A \circ T)^{\#}(\lambda')) \\
&= D(A^{\#}(T^{\#}(\lambda)) \parallel A^{\#}(T^{\#}(\lambda'))) \\
&\le \varepsilon W(T^{\#}(\lambda), T^{\#}(\lambda')) \\
&\le c\,\varepsilon W(\lambda, \lambda') \quad \text{(by } (c, W)\text{-stable).}
\end{aligned}$$

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

G. Relationships among XDistP Notions

Finally, we show relationships among distribution privacy notions with different metric d and divergence D.

Proposition 9 ($W_{1,d}$-XDistP $\Rightarrow$ $W_{\infty,d}$-XDistP): Let $D \in \mathsf{Div}(\mathcal{Y})$. If $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides $(\varepsilon, W_{1,d}, D)$-XDistP, then it provides $(\varepsilon, W_{\infty,d}, D)$-XDistP.

Proof: Assume that A provides $(\varepsilon, W_{1,d}, D)$-XDistP. Let $\lambda_0, \lambda_1 \in \mathbb{D}\mathcal{X}$. By the property of the p-Wasserstein metric, $W_{1,d}(\lambda_0, \lambda_1) \le W_{\infty,d}(\lambda_0, \lambda_1)$. Then $D(\mu_0 \parallel \mu_1) \le W_{1,d}(\lambda_0, \lambda_1) \le W_{\infty,d}(\lambda_0, \lambda_1)$. Hence the claim follows. ■

Proposition 10 ($D \le D'$ & D'-XDistP $\Rightarrow$ D-XDistP): Let $d : (\mathbb{D}\mathcal{X} \times \mathbb{D}\mathcal{X}) \to \mathbb{R}$ be a metric. Let $D, D' \in \mathsf{Div}(\mathcal{Y})$ be two divergences such that for all $\mu_0, \mu_1 \in \mathbb{D}\mathcal{Y}$, $D(\mu_0 \parallel \mu_1) \le D'(\mu_0 \parallel \mu_1)$. If $A : \mathcal{X} \to \mathbb{D}\mathcal{Y}$ provides (ε, d, D')-XDistP, then it provides (ε, d, D)-XDistP.

Then $(\varepsilon, d, D_\infty)$-XDistP implies $(\varepsilon, d, D_{\mathrm{KL}})$-XDistP.

Proof: Assume A provides (ε, d, D')-XDistP. Let $\lambda_0, \lambda_1 \in \mathbb{D}\mathcal{X}$. Then $D'(A^{\#}(\lambda_0) \| A^{\#}(\lambda_1)) \le \varepsilon d(\lambda_0, \lambda_1)$. By definition, $D(A^{\#}(\lambda_0) \| A^{\#}(\lambda_1)) \le D'(A^{\#}(\lambda_0) \| A^{\#}(\lambda_1)) \le \varepsilon d(\lambda_0, \lambda_1)$. Thus A provides (ε, d, D)-XDistP. ■