Coupled-Worlds Privacy: Exploiting Adversarial Uncertainty in Statistical Data Privacy

Raef Bassily*, Adam Groce†, Jonathan Katz†, and Adam Smith*
*Computer Science and Engineering Department
Pennsylvania State University, State College, Pennsylvania
Email: rbb20@psu.edu, asmith@cse.psu.edu
†Department of Computer Science
University of Maryland, College Park, Maryland
Email: {agroce, jkatz}@cs.umd.edu

***Abstract*—We propose a new framework for defining privacy in statistical databases that enables reasoning about and exploiting *adversarial uncertainty* about the data. Roughly, our framework requires indistinguishability of the real world in which a mechanism is computed over the real dataset, and an ideal world in which a simulator outputs some function of a "scrubbed" version of the dataset (e.g., one in which an individual user's data is removed). In each world, the underlying dataset is drawn from the same distribution in some class (specified as part of the definition), which models the adversary's uncertainty about the dataset.**

We argue that our framework provides meaningful guarantees in a broader range of settings as compared to previous efforts to model privacy in the presence of adversarial uncertainty. We also show that several natural, "noiseless" mechanisms satisfy our definitional framework under realistic assumptions on the distribution of the underlying data.

***Index Terms*—data privacy**

I. Introduction

Suppose Facebook were to release the average income of its users—not a noisy version of the average, but its exact value. Or, suppose an Internet dating service were to release exact aggregate statistics about its users' romantic preferences and sexual habits, as does OkCupid [17]. Such disclosures violate formal privacy definitions such as *differential privacy* [7], [5] but do not appear to constitute serious privacy breaches since an adversary cannot use the released information to learn anything sensitive about an individual user (or even a small set of users) without unrealistically precise knowledge about the millions of users of those sites. Differential privacy appears to be overkill in these settings: it provides strong privacy guarantees for an individual user *even if* an adversary knows everything about the dataset besides that user's data, but in the scenarios just considered such omniscience is implausible.

The goal of this paper is to develop rigorous definitions of privacy for statistical databases that allow us to reason about and exploit existing *adversarial uncertainty* about the underlying data. We are driven by several motivations:

- *Better mechanisms:* Relaxing definitions of privacy potentially allows for mechanisms achieving greater accuracy while still meeting satisfactory notions of privacy.
- *Analyzing existing mechanisms:* A broader goal is to understand what privacy guarantees are achieved by methods in use today (e.g., disclosure-control methods currently used by statistical agencies, or releases that are mandated by law) that were not designed with specific privacy definitions in mind. In some cases, our definitions provide a starting point for making rigorous statements about such methods.
- *Better understanding of the "semantics" of privacy:* Any definitional effort involves translating from natural-language descriptions of privacy to mathematical formulations of the same. We seek to understand the implications of different definitional approaches for the possible *inferences* about sensitive data that an adversary can make based on statistical releases.

The framework we introduce in this paper is flexible and admits several instantiations—including one that is equivalent to differential privacy—and we thus view it as a starting point for future work. We also explore a specific instantiation of the framework that we call *distributional differential privacy*, and illustrate its applicability by studying several appealing "noiseless" mechanisms satisfying the resulting definition.

Some previous works have also looked at modeling and exploiting adversarial uncertainty in private data analysis [4], [1], [16], [12], with the most relevant being the work on noiseless privacy [1] and the Pufferfish framework [16]. (Noiseless privacy can be viewed as one instantiation of the Pufferfish framework.) Both can be viewed as attempts to formalize Dalenius's characterization [3] of a mechanism as private only if it is "[im]possible to determine the value [of sensitive information] more accurately than is possible without access to" the output of that mechanism. Dalenius's definition, however, is unreasonably strong [5], [8], [15] and, in particular, it rules out learning certain global information about a dataset. For example, it would rule out learning a link between smoking and cancer, since given this result one can determine that a known smoker is more likely to have cancer.

In contrast, we start with the premise that learning global information about some population (e.g., a link between smoking and cancer) is *not* a privacy violation. This is, in part, because learning such global information is the main goal of many statistical studies, and in part because it seems counter-intuitive to speak of a violation of a user's privacy that

occurs whether or not that user participates in a study (as in the smoking example). This perspective motivates us to define privacy, as in the case of differential privacy, by comparing the effects of a real-world disclosure to a disclosure computed on a "scrubbed" dataset with, e.g., a user's individual data removed. As we discuss further in Section II-E, this results in definitions very different from those of [1], [16].

A. Our Contributions

We now describe our contributions in more detail.

Definitional framework. We give a framework, *coupled-worlds privacy*, for specifying privacy definitions. As an important example instantiation, we consider *distributional differential privacy*, which generalizes differential privacy. Let x be a dataset containing records $x_1, \ldots, x_n$, where each record corresponds to an individual. At a high level, differential privacy of a mechanism F requires that for any x, and for each user i, the result $F(x)$ reveals nothing more about x_i beyond what would be revealed by $F(x_{-i})$ (where x_{-i} denotes the dataset with x_i removed). Roughly speaking, distributional differential privacy relaxes differential privacy by treating x as a random variable from some distribution in a pre-specified class of distributions Δ, rather than as a fixed value. This means that x_i can be masked by the randomness of the other rows of the database, rather than just by the randomness introduced by the mechanism. (If Δ is taken to be the class of all distributions, this definition is equivalent to differential privacy.)

A bit more formally, our definition requires indistinguishability of the real world in which $F(x)$ is released, and an ideal world in which a simulator releases some function of the "scrubbed" dataset x_{-i}. In each case, the dataset x is drawn from the same distribution in some class Δ specified as part of the definition. Indistinguishability implies, in particular, that the real-world mechanism "leaks" little more than could be inferred from the "scrubbed" dataset in the ideal world, at least under the assumption that one of the distributions in Δ adequately models the true distribution of x given the attacker's auxiliary knowledge (if any).

We prove various properties of definitions within our framework. Although composition does not automatically hold, we show a condition under which it does. We also show that the class of distributions for which a given mechanism satisfies our framework is *convex*. This is a desirable feature (not shared by some previous definitions) since it implies that if a mechanism is private under distributions (i.e., beliefs) $\mathcal{D}$ and $\mathcal{D}'$, then it is also private under a belief that assigns non-zero probability to each of those distributions. Our framework can be instantiated in several ways to yield different definitions. In particular, as in Pufferfish [16], one can tailor the information considered sensitive by appropriate choice of the "scrubbing" operation applied to the dataset given to the simulator.

In addition to the Pufferfish framework, we are aware of at least two concurrent efforts to generalize differential privacy that share some broad ideas, one by Bhowmick and Dwork [2] and one by Dwork, Reingold, Rothblum, and Vadhan [19].

Inference-based semantics. As a way of justifying our definitional framework, we formalize an intuitive, "inference-based" notion of privacy in terms of a Bayesian attacker who updates her belief about the dataset x given the output of some mechanism. We show that if a mechanism is private within our framework then, with high probability, an adversary's posterior beliefs in the real world and the ideal world are close. This generalizes analogous statements shown to hold for differential privacy [14].

The inference-based version of our definition provides a more transparent view of the key difference between our approach and that of previous work taking adversarial uncertainty into account [1], [16]. Previous approaches can be seen as requiring an attacker's posterior belief (in the real world) to be close to its prior belief. Here, in contrast, we compare an attacker's posterior belief in the real world to its posterior belief in a hypothetical (ideal) world involving a "scrubbed" version of the dataset. This results in a more relaxed definition that is arguably more natural; see further discussion in Section II-E.

Analyses of specific mechanisms. On an intuitive level, there are two different ways to exploit the fact that our definition considers datasets drawn from some distribution rather than a "worst-case" dataset as in differential privacy. The first is to leverage the uncertainty of the database to avoid adding noise to the output. The second is to argue that the database sampled will, with high probability, satisfy some condition under which privacy holds. We use these ideas to design several natural, "noiseless" mechanisms—e.g., releasing exact sums, "truncated" histograms, or certain discrete statistical estimators—satisfying distributional differential privacy under reasonable assumptions about the distribution of the underlying data.

II. Coupled-Worlds Privacy

A. Background

We assume datasets are ordered lists of data points from some universe $\mathcal{U}$; that is, a dataset lies in $\mathcal{U}^*$. Each entry in a dataset is associated with some individual i. A *mechanism* F is a randomized algorithm taking inputs in $\mathcal{U}^*$. We generally use upper-case letters to denote random variables, and lower-case letters to denote specific realizations.

We use the following notion extensively:

Definition 1. *Two random variables A and B are (ϵ, δ)-**indistinguishable** (denoted $A \approx_{\epsilon,\delta} B$) if, for all events S, we have*

$$\Pr[A \in S] \leq e^{\epsilon} \cdot \Pr[B \in S] + \delta \quad \textit{and}$$
$$\Pr[B \in S] \leq e^{\epsilon} \cdot \Pr[A \in S] + \delta.$$

When $\delta = 0$ we often omit it and write $A \approx_{\epsilon} B$.

Differential privacy. Recall that a mechanism F is (ϵ, δ)-differentially private if, for every pair of "neighboring" datasets x and y, the random variables $F(x)$ and $F(y)$ are (ϵ, δ)-indistinguishable [7], [6]. As our default notion, we say

that equal-length datasets x and y are neighbors if they differ on a single data point. Thus, for F to be differentially private the output of F must be distributed nearly identically regardless of whether a particular individual's data were used in the dataset or someone else's data were used instead. An alternate approach is to say that x and y are neighbors if they differ by insertion or deletion of a single data point. Formally, given a dataset $x = (x_1, x_2, \ldots)$ let x_{-i} denote the dataset obtained by removing x_i. With respect to this notion of neighboring, differential privacy requires that $F(x) \approx_{\epsilon,\delta} F(x_{-i})$ for all $x \in \mathcal{U}^*$ and all i. This second definition is strictly stronger than our default notion of differential privacy (though with some loss of parameters), but the default notion suffices to analyze most natural mechanisms.

Gehrke et al. [11] reformulate differential privacy by asserting that F is differentially private if there is a simulator Sim such that $F(x) \approx_{\epsilon,\delta} \mathsf{Sim}(x_{-i})$ for all $x \in \mathcal{U}^*$ and all i. This is easily seen to be equivalent to differential privacy: If $F(x) \approx_{\epsilon,\delta} F(x)$ for all x, y differing in one entry, then a valid simulator is given by the algorithm that inserts an arbitrary entry into x_{-i} and then applies F to the result. Conversely, if a suitable Sim exists then for any two datasets x, y that differ in the i-th element we have $x_{-i} = y_{-i}$ and hence $F(x) \approx_{\epsilon,\delta} \mathsf{Sim}(x_{-i}) = \mathsf{Sim}(y_{-i}) \approx_{\epsilon,\delta} F(y)$.

An "inference-based" perspective. A common interpretation of differential privacy, due to Dwork and McSherry (see [5]), is that "no matter what an attacker knows ahead of time, the attacker learns the same information about any individual i from the mechanism whether or not that individual's data were used." This natural-language interpretation was formalized by [9], [14] in terms of a Bayesian attacker who starts with some prior distribution on the dataset and, based on the output of the mechanism, draws inferences about x_i. Specifically, differential privacy implies that *for all* prior distributions $\mathcal{D}$ over the random dataset X, and all indices i, with high probability over $t = F(X)$ we have $X_i\big|_{F(X)=t} \approx_{\epsilon',\delta'} X_i\big|_{F(X_{-i})=t}$ (see Section 1 for discussion and precise parameters). Alternatively, using the simulation-based formulation of Gehrke et al. [11], we can require the existence of a simulator Sim such that for all distributions on X and indices i, and with high probability over $t = F(X)$, we have

$$X_i\big|_{F(X)=t} \approx_{\epsilon,\delta} X_i\big|_{\mathsf{Sim}(X_{-i})=t}. \tag{1}$$

B. A Distributional Version of Differential Privacy

As a warm-up to our general framework, we first describe a particular instantiation that we dub *distributional differential privacy* (DDP). The main idea is that rather than require indistinguishability to hold for *all* distributions over the dataset, we require it to hold only for some specified set of "candidate" distributions Δ. (One can view the set of candidate distributions as representing the possibilities for the "true" distribution of the dataset, or as representing the adversary's possible uncertainty about the dataset.) We present two variants of the definition. The first, which we view as more intuitively appealing, is obtained by relaxing the inference-based definition discussed in the previous section. The second, which can be viewed as a relaxation of the simulation-based definition of Gehrke et al. [11], is somewhat easier to work with and is strictly stronger than our inference-based formulation.

We obtain a distributional variant of the inference-based definition from the previous section by requiring Equation (1) to hold only for some set of candidate distributions Δ rather than for all possible distributions. Fix some class Δ of probability distributions over random variables $(X, Z) \in \mathcal{U}^* \times \{0,1\}^*$, where X represents the dataset and Z denotes auxiliary information known to the adversary. We then have:

Definition 2. *A mechanism F satisfies $(\epsilon, \delta, \Delta)$-**inference-based distributional differential privacy** if there is a simulator Sim such that for all distributions $\mathcal{D} \in \Delta$ on (X, Z), with probability at least $1 - \delta$ over choice of $(t, z) = (F(X), Z)$ the following holds for all i:*

$$X_i\big|_{F(X)=t,\, Z=z} \approx_{\epsilon,\delta} X_i\big|_{\mathsf{Sim}(X_{-i})=t,\, Z=z}.$$

A variant is obtained by generalizing the simulation-based definition of Gehrke et al. [11].

Definition 3. *A mechanism F satisfies $(\epsilon, \delta, \Delta)$-*distributional differential privacy *if there is a simulator Sim such that for all distributions $\mathcal{D} \in \Delta$ on (X, Z), all i, and all $(x_i, z) \in \mathsf{Supp}(X_i, Z)$:*

$$F(X)\big|_{X_i=x_i,\, Z=z} \approx_{\epsilon,\delta} \mathsf{Sim}(X_{-i})\big|_{X_i=x_i,\, Z=z}.$$

In Section III we will show several example DDP mechanisms. In Definitions 2 and 3 taking Δ to be the set of all distributions (or simply all point distributions) gives differential privacy. However, in general DDP is stronger and implies inference-based DDP.

Theorem 1. *Say F satisfies $(\epsilon, \delta, \Delta)$-DDP where distributions in Δ have support only on datasets of size at most n, and $2\sqrt{\delta n} \leq \epsilon e^{\epsilon}$. Then F satisfies $(3\epsilon, 2\sqrt{\delta n})$-inference-based DDP.*

Theorem 1 is a special case of Theorem 2, which we prove in the next section. Theorems 1 and 2 are both generalizations of a result of [14], who proved the same statement for the usual notion of differential privacy.

The converse of Theorem 1 does not hold in general, as the following example shows.

Example 1 (Separation of inference-based and indistinguishability-based DDP)**.** Let Δ contain a single distribution on (X, Z), where $X = (X_1, \ldots, X_n)$ is a tuple of n uniformly distributed bits and $Z = \oplus_{i=1}^{n} X_i$. Say $F(X)$ outputs the parity of its input. Note that for any $x_i, z \in \{0,1\}$, the distribution $F(X)\big|_{X_i=x_i, Z=z}$ is just a point distribution on the value z. However, X_{-i} (and hence $\mathsf{Sim}(X_{-i})$) is independent of $F(X) = Z$, and so the distribution of $\mathsf{Sim}(X_{-i})$ cannot equal Z with probability better than 1/2. Thus, conditioned on Z, the distributions of $F(X)$ and $\mathsf{Sim}(X_{-i})$ are very different

in general, and so F cannot satisfy DDP for any reasonable parameters.

On the other hand, for any t, z the distribution $X_i\big|_{F(X)=t,Z=z}$ is uniform. If Sim outputs a uniform bit, then $X_i\big|_{F(X)=t,Z=z} = X_i\big|_{\mathsf{Sim}(X_{-i})=t,Z=z}$ and so F does satisfy inference-based DDP.

C. General Framework

Distributional differential privacy is just one possible instantiation of a general framework we call *coupled-worlds (CW) privacy*. At a high level, definitions within our framework are specified by two functions[1] alt and priv; if a mechanism F satisfies the definition then, intuitively, "$F(X)$ reveals no more information about $\mathsf{priv}(X)$ than is revealed by $\mathsf{alt}(X)$." That is, priv allows one to specify what information should be kept private, while alt defines a "scrubbed" version of the dataset that is available in some ideal world. For the specific case of DDP, we are interested in the privacy of an individual record X_i (so $\mathsf{priv}(X) = X_i$), and want to ensure that $F(X)$ reveals no more information about X_i than would be revealed if user i had not participated in the study at all (so $\mathsf{alt}(X) = X_{-i}$).

We start with an inference-based version of our framework that we find intuitively compelling.

Definition 4. *A mechanism F satisfies $(\epsilon, \delta, \Delta, \Gamma)$-*inference-based coupled-worlds privacy *if there is a simulator Sim such that for all distributions $\mathcal{D} \in \Delta$ on (X, Z), with probability at least $1 - \delta$ over choice of $(t, z) = (F(X), Z)$ the following holds for all $(\mathsf{alt}, \mathsf{priv}) \in \Gamma$:*

$$\mathsf{priv}(X)\big|_{F(X)=t,\ Z=z} \approx_{\epsilon,\delta} \mathsf{priv}(X)\big|_{\mathsf{Sim}(\mathsf{alt}(X))=t,\ Z=z}.$$

As with DDP, it is convenient to use an alternate, indistinguishability-based definition which implies the inference-based version.

Definition 5. *A mechanism F satisfies $(\epsilon, \delta, \Delta, \Gamma)$-*coupled worlds privacy *if there is a simulator Sim such that for all distributions $\mathcal{D} \in \Delta$ on (X, Z), all $(\mathsf{alt}, \mathsf{priv}) \in \Gamma$, and all $(v, z) \in \mathsf{Supp}(\mathsf{priv}(X), Z)$:*

$$F(X)\big|_{\mathsf{priv}(X)=v,\ Z=z} \approx_{\epsilon,\delta} \mathsf{Sim}(\mathsf{alt}(X))\big|_{\mathsf{priv}(X)=v,\ Z=z}.$$

Theorem 2. *Say F satisfies $(\epsilon, \delta, \Delta, \Gamma)$-CW privacy, where $2\sqrt{\delta|\Gamma|} \le \epsilon e^{\epsilon}$. Then F satisfies $(3\epsilon, 2\sqrt{\delta|\Gamma|}, \Delta, \Gamma)$-inference-based CW privacy.*

The proof of Theorem 2 relies on the following generalization of [14, Lemma 4.1]:

Lemma 1. *Suppose $(A, B) \approx_{\epsilon,\delta} (A', B')$. Then, for every $\delta_2 > 0$ and $\delta_1 = \frac{2\delta}{\delta_2} + \frac{2\delta}{\epsilon e^{\epsilon}}$, the following holds: with probability at least $1 - \delta_1$ over t chosen according to B, the random variables $A|_{B=t}$ and $A'|_{B'=t}$ are $(3\epsilon, \delta_2)$-indistinguishable.*

Proof of Theorem 2. Fix a mechanism F with simulator Sim, a distribution $\mathcal{D}$ in Δ, and a pair $(\mathsf{alt}, \mathsf{priv}) \in \Gamma$. CW privacy implies that:

$$(F(X), \mathsf{priv}(X), Z) \approx_{\epsilon,\delta} (\mathsf{Sim}(\mathsf{alt}(X)), \mathsf{priv}(X), Z).$$

Take $\delta_2 = 2\sqrt{\delta|\Gamma|}$ and $\delta_1 = \frac{2\delta}{\delta_2} + \frac{2\delta}{\epsilon e^{\epsilon}}$. We can apply Lemma 1 with $A = A' = \mathsf{priv}(X)$, $B = (F(X), Z)$, and $B' = (\mathsf{Sim}(\mathsf{alt}(X)), Z)$ to get that with probability $1 - \delta_1$ over (t, z), we have

$$\mathsf{priv}(X)\big|_{F(X)=t,Z=z} \approx_{3\epsilon,\delta_2} \mathsf{priv}(X)\big|_{\mathsf{Sim}(\mathsf{alt}(X))=t,Z=z}.$$

Taking a union bound over all function pairs in Γ, we see that the above holds for all $(\mathsf{alt}, \mathsf{priv}) \in \Gamma$ with probability at least

$$1 - |\Gamma| \cdot \delta_1 = 1 - |\Gamma| \cdot \left(\frac{2\delta}{2\sqrt{\delta|\Gamma|}} + \frac{2\delta}{\epsilon e^{\epsilon}}\right) = 1 - \sqrt{\delta|\Gamma|} - \frac{2\delta|\Gamma|}{\epsilon e^{\epsilon}} \ge 1 - 2\sqrt{\delta|\Gamma|},$$

where the final inequality follows because $\epsilon e^{\epsilon} \ge 2\sqrt{\delta|\Gamma|}$. □

As noted earlier for the specific case of DDP, the implication in Theorem 2 is strict.

Other instantiations. We have already discussed one instantiation of CW privacy (namely, distributional differential privacy) in the previous section. We briefly mention some other interesting instantiations.

- Consider a database representing a social network. Here, the database is a graph and private data is associated with each node or edge. We can define a version of node-level privacy by taking pairs $(\mathsf{alt}, \mathsf{priv})$ in which priv outputs information associated with a given node and its incident edges, and alt removes that node and its incident edges.
- Frequently some data (say, demographic information like gender and age) is public and need not be protected. To model this we can consider pairs $(\mathsf{alt}_i, \mathsf{priv}_i)$ in which priv_i outputs only the private data in record X_i and alt_i removes only the private information.

In all the examples we have discussed so far, alt and priv are complementary. This need not always be the case:

- Imagine a database in which several schools contribute data of their students. In this situation each school might want to make sure that no more can be learned about each of its students than if the *entire school* had chosen not to participate in the study. To model this we can consider pairs $(\mathsf{alt}, \mathsf{priv})$ in which priv still outputs an individual student's record, but alt removes all records associated with that student's school.
- Suppose a study involves a database of assets of several financial firms. Having alt remove all the data of any single firm might be too limiting. Instead we might only require that a certain amount of ambiguity about each firm's data remains. This could be achieved by letting alt add noise to the asset distribution of a firm.

[1] Formally, they are specified by a *set* $\Gamma = \{(\mathsf{alt}_i, \mathsf{priv}_i)\}$ of function pairs.

D. Properties of the Framework

We now explore several properties of the CW privacy framework. We first show that CW privacy is preserved under post-processing. (Proofs of all theorems in this section appear the full version of this work.)

Theorem 3. *Coupled-worlds privacy is preserved under post-processing. Formally, if mechanism F satisfies $(\epsilon, \delta, \Delta, \Gamma)$-CW privacy, then so does $G \circ F$ for any (randomized) function G.*

The next two results show that if a mechanism F satisfies CW privacy with respect to some class of distributions Δ, then it also satisfies CW privacy with respect to a (potentially) larger class Δ'. In the first case, we show that one can take Δ' to include all distributions that are convex combinations of distributions in Δ. Besides being a desirable property in its own right, it also serves as a technically convenient tool.

Theorem 4. *If F satisfies $(\epsilon, \delta, \Delta, \Gamma)$-CW privacy, then it also satisfies $(\epsilon, \delta, \Delta', \Gamma)$-CW privacy for Δ' the convex hull of Δ. That is, Δ' is the set of all convex combinations of distributions in Δ.*

Next, we show that CW privacy continues to hold if the attacker's auxiliary information is reduced. Again, besides being a desirable property in its own right, it is also technically useful since it then suffices to prove CW privacy of some mechanism only with respect to some realistic *upper bound* on the auxiliary information available to an adversary.

Theorem 5. *If F satisfies $(\epsilon, \delta, \Delta, \Gamma)$-CW privacy, then it satisfies $(\epsilon, \delta, \Delta', \Gamma)$-CW privacy for Δ', where $\mathcal{D}' \in \Delta'$ first samples (X, Z) from some $\mathcal{D} \in \Delta$, then outputs (X, Z') with $Z' = f(Z)$ for some (randomized) function f.*

Finally, we turn to the question of the *composition* of two private mechanisms F and G. Here, both $F(X)$ and $G(X)$ are released. Although we are not able to prove as general a composition theorem as we would like, we can show that the composition satisfies CW privacy as long as G is private even when given $F(X)$ as auxiliary information, and F is private when given $\mathsf{Sim}_G(\mathsf{alt}(X))$ as auxiliary information.

Theorem 6. *Let F and G be two mechanisms, Δ a class of distributions, and Γ a family of $(\mathsf{priv}, \mathsf{alt})$ pairs. Say G is $(\epsilon_G, \delta_G, \Delta_G, \Gamma)$-CW private with respect to a simulator Sim_G, where*

$$\Delta_G = \{(X, (Z, F(X))) : \ (X, Z) \in \Delta\}$$

and F is $(\epsilon_F, \delta_F, \Delta_F, \Gamma)$-CW private with respect to a simulator Sim_F, where

$$\Delta_F = \left\{ \left(X, \left(Z, \mathsf{Sim}_G\left(\mathsf{alt}(X)\right)\right)\right) : \ (X, Z) \in \Delta, \right.$$
$$\left. \mathsf{alt} \text{ is the first element of some pair in } \Gamma \right\}.$$

Then the mechanism $H = (F, G)$ is $(\epsilon_H, \delta_H, \Delta, \Gamma)$-CW private where

$$\epsilon_H = \epsilon_F + \epsilon_G$$
$$\delta_H = \max\left(\delta_F e^{\epsilon_G} + \delta_G, \ \delta_F + \delta_G e^{\epsilon_F}\right) = O(\delta_F + \delta_G).$$

E. Relation to Other Definitions

We conclude our definitional treatment by comparing our definition to two other recent proposals: noiseless privacy [1] and Pufferfish [16].

Noiseless privacy was introduced with a similar motivation as our own; the idea was to use adversarial uncertainty about the dataset to eliminate the need for noise in the mechanism itself. The high-level idea is to require that $F(X)$ "looks similar" for any two values of a given record:

Definition 6. *F satisfies $(\epsilon, \delta, \mathcal{D})$-*noiseless privacy *if for all i, x_i, x_i', and z:*

$$F(X)\big|_{X_i = x_i, Z = z} \approx_{\epsilon, \delta} F(X)\big|_{X_i = x_i', Z = z},$$

where (X, Z) is chosen according to distribution $\mathcal{D}$.

When $\delta > 0$ this definition is slightly different from the version in [1]. In particular, we require (ϵ, δ)-indistinguishability to hold for all choices of x_i and x_i', whereas the definition in [1] requires ϵ-indistinguishability to hold except for x_i, x_i' that occur with probability at most δ.

Pufferfish provides a framework for defining privacy. Noiseless privacy can be viewed as one specific instantiation,[2] but others are possible. Pufferfish allows for customization of what information will be kept private by appropriate choice of a function sec, which takes as input a dataset X and outputs an element of $\{0, 1, \perp\}$. Thus, sec defines two disjoint classes of datasets, the preimages of 0 and 1, with the $\perp$ output allowing the function to be indecisive. Roughly, Pufferfish defines a mechanism F to be private if the distribution of $F(X)$ is similar regardless of which value of $\mathsf{sec}(X)$ we condition on.

Definition 7. *A mechanism F satisfies $(\epsilon, \delta, \Delta, \mathcal{S})$-*Pufferfish privacy *if for all $\mathsf{sec} \in \mathcal{S}$, all z, and all distributions (X, Z) in Δ it holds that:*

$$F(X)\big|_{\mathsf{sec}(X) = 0, \ Z = z} \approx_{\epsilon, \delta} F(X)\big|_{\mathsf{sec}(X) = 1, \ Z = z}$$

(This definition differs in some non-essential ways from the definition in [16]. We highlight that we allow $\delta > 0$, something not done in [16].)

In both noiseless privacy (and, by extension, Pufferfish) and our notion of distributional definition privacy, the requirement is that $F(X)$ should be "roughly the same" in each of two possible worlds. The difference between the definitions is in which two worlds are compared. In noiseless privacy and Pufferfish the comparison is between a world in which X_i (resp., $\mathsf{sec}(X)$) takes on one value and a world in which it takes on some other value. In DDP, in contrast, the comparison is between a world in which X_i is included in the dataset and a world in which it is not. This has significant implications. Consider an example in which there is a global parameter μ which is either $+1$ or -1 (with half probability each), and every record is normally distributed with mean μ and standard deviation much smaller than 1. Note that the records

[2] This is true for the definitions as given here, which differ slightly from the definitions given in the original works.

are dependent because they all depend on the value of μ. (They are, however, independent conditioned on μ.) The mechanism F that computes the sample mean $\bar{X}$ of the dataset and then outputs ± 1 depending on which is closer to $\bar{X}$ does *not* satisfy noiseless privacy: the distribution of $F(X)$ conditioned on $X_i \approx -1$ is very different from the distribution of $F(X)$ conditioned on $X_i \approx +1$. On the other hand, F *does* satisfy DDP (with the obvious simulator that simply runs F) since the distributions of $F(X)$ and $F(X_{-i})$ are close for X sampled according to the stated distribution.

To see the difference between our definitions and prior ones, it may help to consider an inference-based version of Pufferfish.

Definition 8. *A mechanism F satisfies $(\epsilon,\ \delta,\ \Delta,\ \mathcal{S})$-*inference-based Pufferfish privacy *if for all $\mathsf{sec} \in \mathcal{S}$, all z, and all distributions (X, Z) in Δ, with probability $1-\delta$ over choice of $t \leftarrow F(X)|_{Z=z}$ we have*

$$\mathsf{sec}(X)|_{F(X)=t,Z=z} \approx_{\epsilon,\delta} \mathsf{sec}(X)|_{Z=z}.$$

We have the following theorem.

Theorem 7. *Say F satisfies $(\epsilon, \delta, \Delta, \mathcal{S})$-Pufferfish privacy, where all $\mathsf{sec} \in S$ have output in $\{0,1\}$ and $2\sqrt{\delta} < \epsilon e^{\epsilon}$. Then it also satisfies $(3\epsilon, 2\sqrt{\delta}, \Delta, \mathcal{S})$-inference-based Pufferfish privacy.*

(A proof is given in the full version. A similar statement was proven in [16] for the $\delta = 0$ case.) Thus, the inference-based version of Pufferfish privacy is implied by the standard version, as long as sec never outputs $\perp$.

Returning to Definition 8, one may interpret Pufferfish as requiring that the distribution of any sensitive information be roughly identical both before and after the release of $F(X)$.[3] This means that releasing estimates of general population parameters (for instance, whether smoking and cancer are correlated) is a privacy violation because it implies something about the information of any individual. In fact, Pufferfish considers the privacy of every individual to be violated in such a setting, even if their data is not used at all. In contrast, inference-based coupled-worlds privacy (cf. Definition 4) only requires that the distribution of any private information be roughly identical whether F is computed over the entire dataset or over a "scrubbed" version of the dataset.

We note also that Pufferfish and noiseless privacy do not satisfy analogues of Theorem 5. In particular, returning to the motivating example before Definition 8, if the parameter μ is included in the auxiliary information then the mechanism in question *is* noiselessly private.

III. Analyses of Specific Mechanisms

In this section we present several noiseless mechanisms that satisfy distributional differential privacy. We first show that for a broad class of distributions the sum of all database rows can be released with no noise added whatsoever. We then show that when the database is sampled randomly from some larger population, a truncated histogram (meaning a histogram with near-empty bins removed) can be released exactly. Finally, we present sufficient conditions for distributional differential privacy and apply this to the computation of maximum a posterior-probability (MAP) estimators. Besides being interesting and useful in their own right, these mechanisms illustrate that distributional differential privacy is a meaningful relaxation of differential privacy.

Recall that for a mechanism F to satisfy distributional differential privacy, there must exist a simulator Sim meeting the requirements of Definition 3. If Sim is identical to F, we call the simulator *canonical*. Throughout this section, all simulators are canonical.

A. Privacy of the Exact Sum of a Real-valued Database

In this section, we consider a basic mechanism that releases the sum (or equivalently, average) the entries of a real-valued database without using any form of randomization (e.g., adding noise). The class of distributions over the pair (X, Z) for which our results apply is natural and contains a wide variety of distributions. In order to simplify the exposition and clearly describe this class, we will first consider a setting where there is no auxiliary information involved and introduce a basic class of distributions on the database under which privacy is guaranteed. Then, we consider a natural setting of auxiliary information and, by invoking the useful convexity property of our privacy framework (Theorem 4), we extend our result to the convex-hull of the former class.

Since we consider here the sum of real data, we assume that the database X is a list of $n+1$ real-valued random variables. (We use $n+1$ for the database size rather than n only because it simplifies the expression of our results). First, let us describe in simple words this basic class of distributions. This class contains all the continuous distributions $\mathcal{D}$ on the database domain, i.e., $\mathbb{R}^{n+1}$, that have the following properties:

- $\mathcal{D}$ is a product distribution. That is, under $\mathcal{D}$, all the database entries are independent.
- Under $\mathcal{D}$, each database entry has a density function whose support (i) is the *same* for all the entries, and (ii) is some bounded interval[4] $[a, b] \subset \mathbb{R}$.
- Over the support, the density function of each entry is bounded from below by some non-zero constant (that does not depend on n).

We say that these distributions have a *uniform component*, because they can be written as the sum of two continuous distributions, one of which is the uniform distribution over $[a, b]$. It is easy to show that DDP of the sum mechanism remains the same (i.e., ϵ and δ of the DDP definition remain unchanged) if every database entry is translated and scaled

[3] The inference-based guarantee provided by differential privacy is stronger, but this does not contradict the fact that differential privacy is a special case of Pufferfish nor does it imply that all Pufferfish mechanisms give the same guarantee. Instead, it is merely a special property of differential privacy itself.

[4] This interval can also be open or half-open. In fact, one can show using some standard tools from measure theory that the result still holds when the support is a bounded measurable (Borel) subset of $\mathbb{R}$, but for clarity we limit ourselves to the simpler case.

(by a non-zero constant) in the same manner. Hence, we will assume, w.l.o.g., that the common support of the database entries is the interval $[-1, 1]$ and that all density functions are bounded from below by some strictly positive constant $\eta > 0$ (that does not depend on n). We denote this class of distributions by $\Delta_{\mathsf{u.c.}}(\eta)$.

When there is no auxiliary information and X comes from the class $\Delta_{\mathsf{u.c.}}(\eta)$ discussed above the exact (noiseless) sum is $(\epsilon,\ \delta,\ \Delta_{\mathsf{u.c.}}(\eta))$-DDP where for every $\epsilon > 0$, δ is exponentially decaying in n.

Theorem 8. (Privacy of the Exact Sum) *Let X denote $\mathbb{R}^{n+1}$-valued database. Let $\eta > 0$. Then, for all $\epsilon > 0$ and $\delta = e^{-\Omega\left(\frac{2}{3}\eta \min\left(\frac{2}{3}\eta, \epsilon^2\right) n\right)}$, the sum of the database entries, $\sum_{i=1}^{n+1} X_i$, is $(\epsilon,\ \delta,\ \Delta_{\mathsf{u.c.}}(\eta))$-DDP.*

Next, we generalize our result, using Theorem 4, to include all the distributions in the convex hull of the class $\Delta_{\mathsf{u.c.}}(\eta)$ given in Theorem 8. We will also take into account a specific form of the auxiliary information Z. Namely, we consider the case where Z is a subset of the database entries denoted by $X_{\mathcal{L}} \triangleq \{X_\ell,\ \ell \in \mathcal{L}\}$ for some set $\mathcal{L} \subset [n]$.

Theorem 9. (Privacy of the Exact Sum - Generalized) *Let $\eta > 0$ and $\mathsf{conv}\left(\Delta_{\mathsf{u.c.}}(\eta)\right)$ be the convex hull of $\Delta_{\mathsf{u.c.}}(\eta)$. Let $\mathcal{L} \subset [n]$ and let L denote $|\mathcal{L}|$. Let X be $\mathbb{R}^{n+1}$-valued database and $Z = X_{\mathcal{L}}$ be the auxiliary information where X is drawn from some distribution in $\mathsf{conv}\left(\Delta_{\mathsf{u.c.}}(\eta)\right)$. Then, for all $\epsilon > 0$ and $\delta = e^{-\Omega\left(\frac{2}{3}\eta \min\left(\frac{2}{3}\eta, \epsilon^2\right)(n-L)\right)}$, the sum of the database entries, $\sum_{i=1}^{n+1} X_i$, is $(\epsilon,\ \delta,\ \mathsf{conv}\left(\Delta_{\mathsf{u.c.}}(\eta)\right))$-DDP.*

Proof. The proof is straightforward. First, we combine the result of Theorem 8 with the fact that, for any independent random variables U, U', V, if $U \approx_{\epsilon,\delta} U'$ then, for all $v \in \mathsf{Supp}(V)$, $(U+V)|_{V=v} \approx_{\epsilon,\delta} (U'+V)|_{V=v}$. Then, we invoke Theorem 4. □

We emphasize that here the database rows are no longer independent. For example, $\mathsf{conv}\left(\Delta_{\mathsf{u.c.}}(\eta)\right)$ includes a setting where it is known that the rows are independently distributed around some mean, but that mean is not known (and hence needs to be estimated with an average query). Also, Theorem 5 means that any auxiliary information that is a function of only some subset of the database is covered by the above theorem.

B. Releasing Exact Histograms under Sampling Priors

Sampling distributions, in which the data are drawn randomly from a fixed underlying population, form a natural class of distributions on data sets. We argue that a *truncated histogram*, which releases a histogram (or contingency table) from which small cell counts have been redacted, is DDP for a large subclass of sampling distributions (and their convex combinations).

The model here is that the random sample is the input to the mechanism. In this context, distributional differential privacy ensures that an adversary cannot determine if a given individual's data was used, *even if the attacker knows that the individual was in the sample*. Consequently, the adversary cannot determine if a given individual was in the sample to begin with. Our results strengthen results of Gehrke et al. [10] on truncated histograms; we explain the relationships between the results further below.

The only condition we require on the sampling distribution is that the size of the sample, denoted N, have some uncertainty (to the adversary).

Definition 9 (Sampling Priors)**.** *Given a finite multiset P (the "population"), and a distribution p_N on nonnegative integers, the* sampling distribution $\mathcal{D}_{P,p_N}$ *picks N according to p_N and obtains X by selecting N individuals uniformly at random from the population P.*

The class $\Delta_{(\epsilon,\delta)\text{-Samp}}$ is the convex closure of the set of sampling distributions for which the random variable N satisfies

$$\Pr_{n\sim N}\left(\frac{\Pr(N=n)}{\Pr(N=n-1)} \in \exp(\pm\epsilon)\right) \geq 1-\delta\,. \qquad (2)$$

The condition on the randomness of the sample size holds in a variety of settings. It is slightly stronger than requiring $N \approx_{\epsilon,\delta} N+1$—it corresponds to requiring that N and $N+1$ be "pointwise" (ϵ,δ)-indistinguishable in the terminology of [14]. Nevertheless, N satisfies the condition when N is either binomial or Poisson[5] (as long as the expectation is sufficiently large, see below) or when $N = const + \mathsf{Lap}(1/\epsilon)$ where Lap is the Laplace distribution. The following lemma is useful in both the discussions of priors and the proof of our main result.

Lemma 2. *For every $\epsilon, \lambda, p, n > 0$, we have (1) $\mathsf{Po}(\lambda)$ satisfies Eq. (2) when $\delta = \exp(-\lambda\epsilon^2)$, and (2) $Bin(n,p)$ satisfies Eq. (2) where $\delta = \exp(-\Omega(np\epsilon^2))$.*

Some examples of sampling priors that fall in the class $\Delta_{(\epsilon,\delta)\text{-Samp}}$:

- Suppose the input is obtained by sampling each element in P independently with probability $p \in (0,1)$. The size N of the sample is binomial $Bin(|P|, p)$, and satisfies our condition when $|P| \cdot p$ is $\Omega(\frac{\log(1/\delta}{\epsilon^2})$ (see Lemma 2).
- Suppose the input to the mechanism is data set is a sample of some known, fixed size n_0. One can enforce the randomness condition by discarding only a few data points at random: set $N = n_0 - [\mathsf{Lap}(1/\epsilon) + \frac{\log(1/\delta)}{\epsilon}]_+$ points and discard $n_0 - N$ data points. Here Lap denotes the Laplace distribution and $[x]_+$ denotes $\max\{x, 0\}$. Note that $N \leq n_0$.
  This results in a "slightly" randomized mechanism that alters at most $2\frac{\log(1/\delta)}{\epsilon}$ bin counts in the histogram, far less than the number required to ensure differential privacy (which requires altering the counts of all bins with some probability).
- *Poisson priors:* In Poisson sampling, the sample size N follows a Poisson distribution. It satisfies our condition when λ is $\Omega(\frac{\log(1/\delta)}{\epsilon^2})$.

[5]Recall that for any nonnegative real number λ, $\mathsf{Po}(\lambda)$ is the distribution over nonnegative integers such that $P(N=n) = e^{-\lambda}\lambda^n/n!$.

- The definition is phrased in terms of a fixed population P, but i.i.d. sampling also falls into this class (one obtains i.i.d. sampling in the limit as $|P|$ goes to infinity).

Given a partition of the data domain $\mathbb{D}$ into "bins", a histogram reports the number of data points in each bin. The k-truncated histogram reports the set of counts with value at least k (and reports "0" for counts less than k).

Theorem 10 (Privacy via Sampling Priors)**.** *There is a constant $C > 0$ such that, for $k > \frac{C \log(1/\delta)}{\epsilon^2}$, the k-truncated histogram is $(3\epsilon, 3\delta)$-DDP for the class $\Delta_{(\epsilon,\delta)\text{-Samp}}$.*

The main difficulty of the proof is that the histogram counts – that is, the entries of the vector $F(X_i)$ – are not independent. For example, when $N = const + Lap(1/\epsilon)$, then the sum of the counts is much more concentrated than it would be if the entries were truly independent (or even if every single count were close to independent from the remaining ones). Nonetheless, we can use the randomness in N to limit the information about the jth entry of $F(X_i)$ that is contained in the remaining entries. The proof can be found in the full version.

Relation to the work of Gehrke et al. Gehrke et al. [10] prove a result which appears, at first glance, very similar: namely, that a mechanism which samples each input record with probability p and computes a histogram on the resulting sample is differentially private.

There are two principal differences between the results. First, we assume the sample *is* the input, and so we ask that the adversary not be able to determine whether or not a particular individual *in the sample* was used. This corresponds very closely, for example, to preventing the type of attack carried out by Homer et al. [13] on genome-wide association study data. In contrast, Gehrke et al. show only that the adversary cannot determine if an individual was present in the underlying population.

Second, the parameters of the two results are incomparable: Gehrke et al. assume the sample itself is very small—approximately a ϵ fraction of the population—whereas our results apply to populations that are very close in size to N (subject to the population always being larger than N). On the other hand, Gehrke et al. require only that k be approximately $\log(1/\delta)/\epsilon$, instead of $\log(1/\delta)/\epsilon^2$. The bound on k is tight for our definition, unfortunately.

Finally, we mention that Gehrke et al. analyze a class of mechanisms, called *crowd-blending private*, that generalize truncated histograms. It seems likely that Theorem 10 also generalizes to this larger class of mechanisms, but we did not verify this.

C. Stable Functions are DDP

We now consider deterministic functions that are "stable," by which we mean that the removal of one record has a low probability of changing the output. (I.e., with high probability $F(X) = F(X_{-i})$.) This is sufficient to guarantee $(0, \delta, \Delta)$-DDP. Furthermore, if we require the existence of non-zero lower bound on the set of all conditional probabilities of the output of $F(X)$ given $X_i = x_i$ and $Z = z$ then the mechanism is $(\epsilon, 0, \Delta)$-DDP. Formally, this gives us two sufficient conditions to prove that a mechanism is DDP.

Theorem 11. *Let $n \in \mathbb{N}$. Consider a deterministic database mechanism $F : \mathbb{D}^n \to \mathcal{S}$ and a class of distributions Δ for the pair (X, Z). Suppose $\exists\ \delta > 0$ such that $\forall i \in [n],\ \forall (x_i, z) \in$ Supp(X_i, Z),*

$$\Pr\left[F(X) \neq F(X_{-i}) | X_i = x_i, Z = z\right] < \delta$$

Then, F is $(0,\ \delta,\ \Delta)$-DDP.

Theorem 12. *Let $n \in \mathbb{N}$. Consider a deterministic function $F : \mathbb{D}^n \to \mathcal{S}$ where $\mathcal{S}$ is a finite set that does not depend on n and a distribution class Δ_n for the pair (X, Z). If there exist $c > 0$ and $\mu_n \in (0, c)$ such that, for all $i \in [n]$, all $t \in \mathcal{S}$, and all$(x_i, z) \in$ Supp(X_i, Z), the following conditions hold simultaneously*

$$\Pr\left[F(X) = t | X_i = x_i, Z = z\right] \geq c$$
$$\Pr\left[F(X) \neq F(X_{-i}) | X_i = x_i, Z = z\right] \leq \mu_n$$

then, F is $(\epsilon_n,\ 0,\ \Delta_n)$-DDP where $\epsilon_n = \ln\left(\frac{c+\mu_n}{c-\mu_n}\right)$. Moreover, if c does not depend on n and $\mu_n \to 0$, then $\epsilon_n \to 0$.

The two above theorems are proved in the full version.

MAP estimators. The sufficient conditions shown above are simple and they cover some very practical functions. As an example, we consider a wide class of estimators known in the literature as "maximum a posteriori probability" (MAP) estimators [18]. At a high level, the scenario we are interested in is one where the database entries are sampled i.i.d. from one of several distributions. Which distribution is used is not known. The MAP estimator calculates, based on provided prior probabilities of each distribution being used, the distribution from which the database entries are most likely sampled. The MAP estimator appears a lot in applications involving parameter estimation and multiple hypothesis testing. We show that the MAP estimator achieves the notions of $(0,\ \delta,\ \Delta)$ and $(\epsilon,\ 0,\ \Delta')$-DDP for two (slightly different) large classes of priors Δ and Δ'. As the MAP estimator is deterministic, this is not possible with differential privacy.

Formally, we consider database entries to come from a set $\mathcal{B}$, and we have a family of probability functions $(f_1, \ldots, f_k)$ that each represents a distribution over $\mathcal{B}$. A distribution $\mathcal{D} \in \Delta$ generates the database X as follows. First, $\mathcal{D}$ picks one of the distributions in the family $(f_1, \ldots, f_k)$, with each f_i chosen with probability p_i (for some probability mass function $(p_1, \ldots, p_k)$). Once that choice has been made, the entries of X are chosen i.i.d. from the chosen f_i. A given $\mathcal{D}$ is defined by the choice of $(p_1, \ldots, p_k)$, and we take Δ to be the union of all distributions $\mathcal{D}$ defined in this manner, where the union is taken over all legitimate probability mass funtions $(p_1, \ldots, p_k)$.

A MAP estimator with respect to $(f_1, \ldots, f_k)$ takes as input a set of prior probabilities $(\pi_1, \ldots, \pi_k)$ (summing to 1) that the

user assigns to the distributions $(f_1, \ldots, f_k)$ and outputs the index of the distribution which is most likely to have generated the given database x. Formally, the output is the value of i that maximizes $\pi_i \prod_{j=1}^{n} f_i(x_j)$ (with ties broken arbitrarily). We emphasize that while intuitively the user is trying to match $(\pi_1, \ldots, \pi_k)$ to the actual priors $(p_1, \ldots, p_k)$, we assume no relationship between them when proving privacy. That is, our results cover the case where the actual priors are unknown to the user.

In the following theorem, proved in the full version, we show that the MAP estimator F, as defined above, is $(0,\ \delta,\ \Delta)$-DDP, where δ decays exponentially to zero in n, if the family $(f_1, \ldots, f_k)$ satisfies some additional regularity conditions.

Theorem 13. *Consider a MAP estimator $F : \mathcal{B}^{n+1} \to [k]$ for a given distribution family $(f_1, \ldots, f_k)$ and a set of strictly positive user-defined weights $(\pi_1, \ldots, \pi_k)$. Suppose that the distribution family satisfies the following condition:*

$$\exists\ M > 1 \quad \textit{s.t.} \quad \forall i, i' \in [k], \forall a \in \textit{Supp}(f_i) \cup \textit{Supp}(f_{i'}), \\ f_i(a) \leq M f_{i'}(a)$$

Then the MAP estimator F is $(0,\ \delta,\ \Delta)$-DDP where Δ is defined as above (with the same choice of distribution family $(f_1, \ldots, f_k)$) and

$$\delta = (k-1)\,\tau\,(M+1)\,e^{-nu} \tag{3}$$

where

$$\tau = \max_{i \neq i'} \frac{\pi_i}{\pi_{i'}} \tag{4}$$

$$u = \min_{i \neq i'} \left(-\ln \left(E\left[\left(\frac{f_i(X_1)}{f_{i'}(X_1)} \right)^s \right] \right) \right) \tag{5}$$

for any fixed $s \in (0, 1)$.

Next, we consider another class of priors Δ_λ indexed by some $\lambda > 0$. This class is defined equivalently to Δ, with the added condition that $p_i \geq \lambda$ for all i. For this class of priors, we give the following result that asserts that the MAP estimator F, as defined above, is $(\epsilon,\ 0,\ \Delta_\lambda)$-DDP, where ϵ decays to zero in n, if the family $(f_1, \ldots, f_k)$ satisfies the same regularity conditions as in the previous theorem.

Theorem 14. *Consider a MAP estimator $F : \mathcal{B}^{n+1} \to [k]$ for a given distribution family $(f_1, \ldots, f_k)$ and a set of strictly positive user-defined weights $(\pi_1, \ldots, \pi_k)$. Suppose that the distribution family satisfies the following condition:*

$$\exists\ M > 1 \quad \textit{s.t.} \quad \forall i, i' \in [k], \forall a \in \textit{Supp}(f_i) \cup \textit{Supp}(f_{i'}), \\ f_i(a) \leq M f_{i'}(a)$$

Then the MAP estimator F is $(\epsilon,\ 0,\ \Delta_\lambda)$-DDP where

$$\epsilon = \ln \left(\frac{1 + (M/\lambda - 1)\,\tau M e^{-nu}}{1 - (k-1)\,\tau M e^{-nu}} \right) \tag{6}$$

where τ is given by (4) and u is given by (5) for any fixed $s \in (0, 1)$.

This theorem is proved in the full version.

The above results hold for a distribution class with no auxiliary information, but they can be extended to the case where the auxiliary information Z is given by a proper subset of the database entries $X_{\mathcal{L}} \triangleq \{X_j,\ j \in \mathcal{L} \subset [n]\}$. Because this auxiliary information can be interpreted as an upper bound, this result covers all cases where the auxiliary information is some function of only a subset of the database. The proof follows exactly the same lines of the proofs of Theorems 13 and 14.

Theorem 15. *Let the auxiliary information be given by $Z = X_{\mathcal{L}}$ for any subset $\mathcal{L} \subset [n]$. The results of Theorems 13 and 14 still hold with n in (3) and (6) being replaced with $n - L$ where $L = |\mathcal{L}|$.*

We believe that both the sufficient conditions and the MAP estimator mechanism itself are of interest independent of our privacy definition. Because the databases' inherent randomness here is only used to avoid problematic situations, rather than as a substitute for added noise, we believe a similar (though possibly slightly less utile) mechanism could be shown to be private under other privacy definitions as well. Mainly, one can show that with a little added noise, the MAP mechanism can be made ϵ-differentially private.

D. Implications for Other Privacy Definitions

In certain settings, privacy under the CW framework and/or its DDP instantiation can imply privacy under the Pufferfish framework and/or noiseless privacy. We first give a general condition that specifies a case in which CW privacy implies Pufferfish privacy. (All results in this section are proved in the full version of this work.)

Theorem 16. *Let F be a database mechanism. Let $\mathcal{S}$ be a set of secret functions* sec *as in Definition 7, and require that each* $\mathsf{sec} \in \mathcal{S}$ *appears at least once as the* priv *function for some* $(\mathsf{alt}, \mathsf{priv}) \in \Gamma$. *Let $\epsilon, \delta > 0$. Suppose that F is $(\epsilon, \delta, \Delta, \Gamma)$-CW private and that* Sim *is the simulator that satisfies Definition 5 in this case. If* $\mathsf{Supp}(Z|_{\mathsf{sec}(X)=0}) = \mathsf{Supp}(Z|_{\mathsf{sec}(X)=1})$ *and there are some $\epsilon_1 > 0$ and $\delta_1 > 0$ such that*

$$\mathsf{Sim}\left(\mathsf{alt}(X)\right)|_{\mathsf{sec}(X)=0, Z=z} \approx_{\epsilon_1, \delta_1} \mathsf{Sim}\left(\mathsf{alt}(X)\right)|_{\mathsf{sec}(X)=1, Z=z}$$

for all $(\mathsf{alt}, \mathsf{sec}) \in \Gamma$ *and all* $z \in \mathsf{Supp}(Z|_{\mathsf{sec}(X)} = 0)$, *then F is $(2\epsilon + \epsilon_1,\ \delta(1 + e^{\epsilon + \epsilon_1}) + \delta_1 e^{\epsilon},\ \Delta,\ \mathcal{S})$-Pufferfish private.*

We can easily apply the choices by which noiseless privacy is an instantiation of Pufferfish to obtain a similar result specifically showing a conversion from DDP to noiseless privacy.

Corollary 1. *Let F be a database mechanism. Let n be the size of the database X. Suppose that F is $(\epsilon, \delta, \Delta)$-DDP for some $\epsilon, \delta > 0$. Let* Sim *be the simulator that satisfies Definition 3 in this case. If, for all $i \in [n]$ and all distinct $x_i,\ x_i' \in$ Range(X_i),* $\mathsf{Supp}(Z|_{X_i = x_i}) = \mathsf{Supp}(Z|_{X_i = x_i'})$ *and*

$$\mathsf{Sim}(X_{-i})|_{X_i = x_i, Z = z} \approx_{\epsilon_1, \delta_1} \mathsf{Sim}(X_{-i})|_{X_i = x_i', Z = z}$$

for all $z \in \mathsf{Supp}(Z|_{X_i=x_i})$*, then* F *is* $(2\epsilon+\epsilon_1,\ \delta(1+e^{\epsilon+\epsilon_1})+\delta_1 e^{\epsilon},\ \mathcal{D})$*-noiselessly private (with respect to Definition 6) for all* $\mathcal{D} \in \Delta$.

Next, we consider a special case where the rows of the database are independent, i.e. Δ is a class of product distributions. Hence, we show that all the DDP mechanisms given in Section 4 are also noiselessly private (with the loss of at most a constant factor in the privacy parameters) whenever the prior distribution is in the product form. In Section II-E we pointed out that noiseless privacy was arguably too strong in that it rules out learning summary statistics because those statistics have implications for an individual. This corollary shows that in some sense this is the only way in which it is stronger than DDP. When such summary statistics are fixed (or included in the auxiliary information) and the only remaining uncertainty is idiosyncratic to each individual noiseless privacy is implied by DDP.

Corollary 2. *Let* Δ *be a class of distributions over* (X, Z) *under each of which the database entries, conditioned on* $Z = z$, $\{X_i|_{Z=z},\ i \in [n]\}$*, are independent for all* $z \in \mathsf{Supp}(Z)$*. Let* F *be a database mechanism. If* F *is* $(\epsilon, \delta, \Delta)$*-DDP for some* $\epsilon, \delta > 0$*, then* F *is* $(2\epsilon,\ \delta(1+e^{\epsilon}),\ \mathcal{D})$*-noiselessly private (with respect to Definition 6) for all* $\mathcal{D} \in \Delta$.

Finally, the following theorem uses Corollary 2 together with the DDP results that were derived for mechanisms in Sections III-A and III-C to show that such mechanisms are also noiselessly private in the case where database entries are chosen independently. For that matter, we need to modify the setting of Theorem 13 a bit. Clearly, for any non-degenerate set of actual priors $(p_1, \ldots, p_k)$, the database entries are not independent. So, in the next theorem, we will assume a setting where one of the distributions in the family $(f_1, \ldots, f_k)$ is picked in a *deterministic* fashion but such distribution is *unknown*, then the database entries are drawn i.i.d. according to this unknown distribution. That is tantamount to say that the set of actual prior is degenerate but unknown, i.e., $p_i = 1$ for some unknown $i \in [k]$.

Theorem 17. *If we replace DDP with Noiseless Privacy (Definition 6) as a definition for privacy, then the results of Theorems 8 and 13 (with the class* Δ *in Theorem 13 is replaced here with* $\{f_1, \ldots, f_k\}$*)* [6] *still hold after replacing each respective pair* $(\epsilon,\ \delta)$ *in each of these theorems with* $(2\epsilon,\ \delta(1+e^{\epsilon}))$.

ACKNOWLEDGMENTS

The work of Jonathan Katz was partially supported by NSF award #0964541. This paper has benefited from discussions with many colleagues over the years, in particular Raghav Bhaskar, Cynthia Dwork, Shiva Kasivswanathan, Dan Kifer, Ashwin Machanavajjhala, Frank McSherry, Sofya Raskhodnikova, and Abhradeep Thakurta.

[6]This represents the choice of f_i being fixed, rather than randomized.

REFERENCES

[1] R. Bhaskar, A. Bhowmick, V. Goyal, S. Laxman, and A. Thakurta. Noiseless database privacy. In D. H. Lee and X. Wang, editors, *ASIACRYPT*, volume 7073 of *Lecture Notes in Computer Science*, pages 215–232. Springer, 2011.

[2] A. Bhowmick and C. Dwork. Natural differential privacy. In *DIMACS Workshop on Recent Work on Differential Privacy across Computer Science*, 2012.

[3] T. Dalenius. Towards a methodology for statistical disclosure control. *Statistik Tidskrift*, (5):35–64, 1977.

[4] Y. Duan. Privacy without noise. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 1517–1520. ACM, 2009.

[5] C. Dwork. Differential Privacy. In *ICALP*, LNCS, pages 1–12. Springer, 2006.

[6] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our Data, Ourselves: Privacy Via Distributed Noise Generation. In *EUROCRYPT*, LNCS, pages 486–503. Springer, 2006.

[7] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, LNCS, pages 265–284. Springer, 2006.

[8] C. Dwork and M. Naor. On the difficulties of disclosure prevention in statistical databases or the case for differential privacy. *J. Privacy and Confidentiality*, 2(1), 2010.

[9] S. R. Ganta, S. P. Kasiviswanathan, and A. Smith. Composition attacks and auxiliary information in data privacy. In *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 265–273. ACM, 2008.

[10] J. Gehrke, M. Hay, E. Lui, and R. Pass. Crowd-blending privacy. In R. Safavi-Naini and R. Canetti, editors, *CRYPTO*, volume 7417 of *Lecture Notes in Computer Science*, pages 479–496. Springer, 2012.

[11] J. Gehrke, E. Lui, and R. Pass. Towards privacy for social networks: A zero-knowledge based definition of privacy. In Y. Ishai, editor, *TCC*, volume 6597 of *Lecture Notes in Computer Science*, pages 432–449. Springer, 2011.

[12] R. Hall, A. Rinaldo, and L. Wasserman. Random differential privacy. *Journal of Privacy and Confidentiality*, 4(2):43–59, 2012.

[13] N. Homer, S. Szelinger, M. Redman, D. Duggan, W. Tembe, J. Muehling, J. V. Pearson, D. A. Stephan, , S. F. Nelson, and D. W. Craig. Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays. *PLOS Genetics*, 4(8), 2008.

[14] S. P. Kasiviswanathan and A. Smith. A note on differential privacy: Defining resistance to arbitrary side information. *CoRR*, arXiv:0803.39461 [cs.CR], 2008.

[15] D. Kifer and A. Machanavajjhala. No Free Lunch in Data Privacy. In *SIGMOD*, pages 193–204, 2011.

[16] D. Kifer and A. Machanavajjhala. A rigorous and customizable framework for privacy. In M. Benedikt, M. Krötzsch, and M. Lenzerini, editors, *PODS*, pages 77–88. ACM, 2012.

[17] OkCupid. OkTrends blog.

[18] H. V. Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 1994.

[19] O. Reingold. Occupy database — privacy is a social choice. http://windowsontheory.org/2012/02/28/occupy-database-privacy-is-a-social-choice/, 2012.