Limits of Computational Differential Privacy in the Client/Server Setting

Adam Groce, Jonathan Katz, and Arkady Yerukhimovich

Dept. of Computer Science
University of Maryland
{agroce, jkatz, arkady}@cs.umd.edu

Abstract. Differential privacy is a well established definition guaranteeing that queries to a database do not reveal "too much" information about specific individuals who have contributed to the database. The standard definition of differential privacy is information theoretic in nature, but it is natural to consider *computational* relaxations and to explore what can be achieved with respect to such notions. Mironov et al. (Crypto 2009) and McGregor et al. (FOCS 2010) recently introduced and studied several variants of computational differential privacy, and show that in the *two-party* setting (where data is split between two parties) these relaxations can offer significant advantages.

Left open by prior work was the extent, if any, to which computational differential privacy can help in the usual *client/server* setting where the entire database resides at the server, and the client poses queries on this data. We show, for queries with output in $\mathbb{R}^n$ (for constant n) and with respect to a large class of utilities, that any computationally private mechanism can be converted to a statistically private mechanism that is equally efficient and achieves roughly the same utility.

1 Introduction

A statistical database holds data representing some population. It is often desirable to allow clients to query this database to learn properties of the underlying population. However, it is also important to protect the privacy of the individual users whose data is contained in the database. This conflict between utility and privacy has motivated a significant amount of research in recent years, and several definitions of privacy as well as techniques for achieving these definitions have appeared in the literature.

The foundational definition of privacy in this setting is that of *differential privacy* [6, 5, 3]. Very coarsely, this definition can be viewed as limiting the amount of information the answer to some query reveals about any particular user in the database. The standard definition of differential privacy is very strong, requiring unconditional privacy guarantees against computationally unbounded adversaries. Despite this fact, there has been a good amount of success in designing differentially private mechanisms for many types of queries and in various settings [1, 5, 12, 2, 9].

Recently, Mironov et al. [11] introduced various notions of *computational* differential privacy and explored relations between them. There are several reasons to consider such relaxations of differential privacy. In practice a computational notion of security suffices, yet the stringent notion of (statistical) differential privacy rules out some mechanisms that are intuitively secure: e.g., a differentially private mechanism implemented using pseudorandom noise in place of truly random noise, or a differentially private mechanism implemented using secure multi-party computation [4, 11]. One might hope that by considering a relaxed definition we can circumvent limitations or impossibility results that arise in the information-theoretic setting, in the same way that computationally secure notions of encryption allow bypassing known bounds for perfectly secure encryption. Recent results [11, 10] show that this is the case in the *two-party* setting where the database is partitioned between two parties who wish to evaluate some query over their joint data. Specifically, McGregor et al. [10] show a strong separation between the accuracy that can be obtained when using differential privacy as opposed to using *computational* differential privacy.

McGregor et al. [10], however, leave open the analogous question in the more widely studied *client/server* setting where a server holds the entire database on which a client may pose queries. Indeed, they explicitly remark [10, Section 1]:

> *[Our] strong separation between (information-theoretic) differential privacy and computational differential privacy ... stands in sharp contrast with the client-server setting where... there are not even candidates for a separation.*

It is this question we address in this paper.

1.1 Summary of Our Results

There are (at least) two notions of computational privacy that can be considered: IND-CDP and SIM-CDP. These notions are introduced in [11], where it is shown that any SIM-CDP mechanism is also IND-CDP (the other direction is not known); thus, SIM-CDP is a possibly stronger definition. (Mironov et al. also define the notion of $\text{SIM}_{\forall\exists}$-CDPbut this notion is equivalent to IND-CDP.) We review these definitions in Section 2.

There are two measures one could hope to improve upon when moving from the setting of (statistical) differential privacy to the setting of computational differential privacy: the best possible *utility* (or *accuracy*) that can be achieved, and the *efficiency* of implementing a mechanism that achieves some level of utility. With respect to the definitions given by Mironov et al., it is not hard to see that the best achievable utility cannot be improved as long as the utility is an efficiently computable function of the database and the output of the mechanism. (This is an immediate consequence of the SIM-CDP and $\text{SIM}_{\forall\exists}$-CDP definitions, since otherwise the utility function itself serves as a distinguisher.) The interesting question is therefore to look for improvements in the efficiency, e.g., to show that the best possible utility *for polynomial-time mechanisms* is better

in the computational case, or even to show a polynomial factor improvement in the efficiency in moving from one case to the other. Unfortunately, we show two negative results indicating that such improvements are unlikely in certain natural settings:

1. Our first result concerns *black-box* constructions of computationally secure mechanisms from a wide range of cryptographic primitives including trapdoor permutations, collision-resistant hash functions, and/or random oracles. Roughly, we show that for any black-box construction of a computationally private mechanism there exists a corresponding *statistically* private mechanism that performs just as well in terms of both efficiency and utility (with respect to any utility measure).
2. Our main results rules out improvements by *arbitrary* mechanisms, for a specific (but large) class of queries and utility measures. That is, for queries with output in $\mathbb{R}^n$ (for constant n) and a natural class of utilities, we show that *any* computationally private mechanism can be converted to a statistically private mechanism that is roughly as efficient and achieves almost the same utility.

Each result applies to both the IND-CDP and SIM-CDP definitions.

We believe our results represent an important step in understanding the benefits and limitations of computational notions of privacy. Although we show negative results, they may point toward specific situations where computational differential privacy gives some advantage. We leave it as an open question to find utility measures or query classes with respect to which computational differential privacy *can* help in the client/server setting, or to extend our impossibility results to show that no such improvements can be hoped for.

Limitations of our results. There are several types of queries to which our results do not apply. The most important are queries with outputs that cannot naturally be thought of as tuples of real numbers. This includes, e.g., queries that return classifiers (as in [9]), graphs, or synthetic databases.

Our results also do not apply, in general, to queries that return output in $\mathbb{R}^n$ for "large" n (i.e., n that grows with the security parameter k). In particular, this means that our results are somewhat limited when it comes to analyzing differential privacy of multiple queries. (Note that n queries with outputs in $\mathbb{R}$ can be viewed as a single query with output in $\mathbb{R}^n$.) Our results do apply to any *constant* number of queries. In addition, using composition properties of differential privacy, our results apply to the case where arbitrarily many queries are answered, and all queries are answered independently (i.e., the server maintains no state). However, in some cases it is known that answering many queries at the same time can be done with better privacy than would be achieved by answering each query independently; in such cases our results do not apply.

Our results also hold only for restricted classes of utility functions. For example, they do not apply when there is no polynomial bound on the error.

2 (Computational) Differential Privacy

We begin by reviewing definitions for the various notions of differential privacy that will be discussed in this paper. All have roughly the same underlying intuition, but the technical differences are crucial. We begin by defining "adjacent" databases.

Definition 1 *Two databases $D, D' \in \mathcal{D}$ are* adjacent *if they differ in at most 1 entry.*

Differential privacy guarantees that the results of queries on two adjacent databases cannot be distinguished very well. This is a very strong privacy guarantee, that in particular ensures that the presence or absence of any one user in the database cannot affect the results very much.

One way to formalize this notion is to require that no set of answers can be significantly more likely to result from D than from D'. Formalizing this yields the by-now-standard notion of (statistical) differential privacy:

Definition 2 *A randomized function $f : \mathcal{D} \to \mathcal{R}$ is* ϵ-differentially private *(ϵ-DP) if for all adjacent databases $D, D' \in \mathcal{D}$ and all subsets $S \subset \mathcal{R}$:*

$$\Pr[f(D) \in S] \leq e^{\epsilon} \times \Pr[f(D') \in S].$$

This is the strongest definition of differential privacy. It can, in fact, be criticized as too strong. For example, consider a set of responses that are possible outputs when querying D but impossible when querying D'. The existence of such responses violates differential privacy, even if the probability of outputting one of these responses is small. To allow for this sort of situation one can consider a slightly weaker notion of differential privacy, called (ϵ, δ)-differential privacy, that allows a small additive factor in the inequality [4].

Definition 3 *A randomized function $f : \mathcal{D} \to \mathcal{R}$ is* (ϵ, δ)-differentially private *((ϵ, δ)-DP) if for all adjacent databases $D, D' \in \mathcal{D}$ and all subsets $S \subset \mathcal{R}$:*

$$\Pr[f(D) \in S] \leq e^{\epsilon} \times \Pr[f(D') \in S] + \delta.$$

It is worth noting that while for ϵ-DP it is sufficient to require the inequality in the definition to hold pointwise, for (ϵ, δ)-differential privacy it is important to explicitly consider all subsets S.

We say a family of mechanisms $\{f_k\}$ is *efficient* if the running time of $f_k(D)$ is at most $\mathsf{poly}(|D|, k)$. A family $\{f_k\}$ is *uniform* if there is a Turing machine f such that $f(k, D) = f_k(D)$. It is reasonable even in the information-theoretic setting to consider a family of mechanisms $\{f_k\}$ indexed by a security parameter k, and to require that δ become negligible in k.

Definition 4 *Let ϵ be an arbitrary function. A family of randomized functions $\{f_k\}_{k \in \mathbb{N}}$ is $(\epsilon, \mathsf{negl})$-DP if there exists a negligible function δ such that each f_k is $(\epsilon(k), \delta(k))$-DP.*

The above definitions are all information-theoretic in nature, but it is natural to consider computational variants. Mironov et al. [11] propose two definitions of computational differential privacy, SIM-CDP and IND-CDP. Roughly, one can view IND-CDP as an "indistinguishability-based" relaxation whereas SIM-CDP is a "simulation-based" notion. SIM-CDP is at least as strong as IND-CDP [11], but the converse is not known. All the definitions can be presented for either uniform or non-uniform adversaries; for consistency with [11], we give non-uniform definitions here. While we state our results for the case of non-uniform adversaries, our results all carry over to the uniform setting as well.

IND-CDP provides perhaps the most natural relaxation of differential privacy.

Definition 5 (IND-CDP) *Let ϵ be an arbitrary function. A family of functions $\{f_k\}_{k\in\mathbb{N}}$ is ϵ-IND-CDP if for every non-uniform polynomial-time $\mathcal{A}$ and every sequence $\{(D_k, D'_k)\}_{k\in\mathbb{N}}$ of (ordered pairs of) polynomial-size, adjacent databases, there is a negligible function negl such that*

$$\Pr[\mathcal{A}(f_k(D_k)) = 1] \leq e^{\epsilon(k)} \times \Pr[\mathcal{A}(f_k(D'_k)) = 1] + \mathsf{negl}(k).$$

The notion of SIM-CDP requires that there be a statistically private mechanism that is indistinguishable from the mechanism under consideration.

Definition 6 (SIM-CDP) *Let ϵ be an arbitrary function. A family of functions $\{f_k\}_{k\in\mathbb{N}}$ is ϵ-SIM-CDP if there exists a family of functions $\{F_k\}_{k\in\mathbb{N}}$ that is $(\epsilon, \mathsf{negl})$-DP and is computationally indistinguishable from $\{f_k\}$. The latter means there is a negligible function negl such that for any non-uniform polynomial-time $\mathcal{A}$ and any database D:*

$$\Big|\Pr[\mathcal{A}(f_k(D)) = 1] - \Pr[\mathcal{A}(F_k(D)) = 1]\Big| \leq \mathsf{negl}(k).$$

In [11] it is required that $\{F_k\}_{k\in\mathbb{N}}$ be ϵ-DP (rather than $(\epsilon, \mathsf{negl})$-DP). Thus our definition is slightly weaker, which makes our impossibility results stronger.

We also recall the notion of $\text{SIM}_{\forall\exists}$-CDP, which weakens SIM-CDP by reversing the order of quantifiers in the definition: here, the statistically private mechanism F is allowed to be different for each pair of databases (D, D'). Crucially for our purposes, this definition is known to be equivalent to IND-CDP [11].

Definition 7 ($\text{SIM}_{\forall\exists}$-CDP) *Let ϵ be an arbitrary function. A family of functions $\{f_k\}_{k\in\mathbb{N}}$ is ϵ-$\text{SIM}_{\forall\exists}$-CDP if for all sequences of (unordered pairs of) adjacent databases $\{\{D_k, D'_k\}\}_{k\in\mathbb{N}}$ there is a family of functions $\{F_k\}_{k\in\mathbb{N}}$ such that:*

1. *$\{F_k\}$ is ϵ-DP on $\{\{D_k, D'_k\}\}_{k\in\mathbb{N}}$; i.e., for all subsets $S \subset \mathcal{R}$ we have*

$$\Pr[F_k(D_k) \in S] \leq e^{\epsilon(k)} \times \Pr[F_k(D'_k) \in S].$$

2. *$f_k(D_k)$ and $f_k(D'_k)$ are indistinguishable from $F_k(D_k)$ and $F_k(D'_k)$ respectively. Formally, for any non-uniform, polynomial-time adversary $\mathcal{A}$*

$$\Big|\Pr[\mathcal{A}(f_k(D_k)) = 1] - \Pr[\mathcal{A}(F_k(D_k)) = 1]\Big| \leq \mathsf{negl}(k),$$

and similarly for D'_k.

Thus far we have only discussed privacy but have not mentioned *utility*. In general, we assume a utility measure U that takes as input a database D and the output of some mechanism $f(D)$ and returns a real number. In Section 4 we consider a specific class of utilities.

3 Limitations on Black-Box Constructions

Here we show that black-box constructions (of a very general sort) cannot help in the setting of computational differential privacy. (We refer the reader to [13] for further discussion and definitional treatment of black-box constructions.) For concreteness, in the technical discussion we focus on black-box constructions from one-way functions, but at the end of the section we discuss generalizations of the result.

Roughly, a fully black-box construction of an ϵ-IND-CDP mechanism from a one-way function is a family of polynomial-time oracle machines $\{f_k^{(\cdot)}\}_{k\in\mathbb{N}}$ such that for every $\mathcal{A}$ and every $\mathcal{O}$ that is one-way against $\mathcal{A}$ it holds that $\{f_k^{\mathcal{O}}\}_{k\in\mathbb{N}}$ is ϵ-IND-CDP against $\mathcal{A}$. It would make sense also to impose a utility condition on the construction (which could be viewed as a correctness requirement on the constructions), but we do not do so here.

Theorem 1 *If there exists a fully black-box construction $\{f_k\}_{k\in\mathbb{N}}$ of an ϵ-IND-CDP mechanism from one-way functions, then there exists an $(\epsilon, \mathsf{negl})$-DP family $\{f'_k\}_{k\in\mathbb{N}}$ that is roughly as efficient and such that, for all databases D and utility measures U,*

$$\left|\mathbf{E}\left[U(D, f_k^{\mathcal{O}}(D))\right] - \mathbf{E}\left[U(D, f'_k(D))\right]\right| \leq \mathsf{negl}(k),$$

where the expectations are both taken over the randomness of the mechanism, and the expectation on the left is additionally taken over random choice of a function $\mathcal{O}$.

Proof. The key idea behind the proof is as follows: a random function is one-way with overwhelming probability [8, 7]; thus, the mechanism $f_k^{\mathcal{O}}$ with $\mathcal{O}$ chosen at random is also ϵ-IND-CDP. Since the construction is fully black-box (and hence relativizing), one-wayness of $\mathcal{O}$ (and hence indistinguishability of the mechanism) holds even for an unbounded adversary as long as the adversary makes only polynomially many queries to $\mathcal{O}$. We construct f'_k by having it simply run f_k as a subroutine, simulating a random function $\mathcal{O}$ on behalf of f_k. This idea is motivated by analogous techniques used in [7].

Let Func denote the set of length-preserving functions from $\{0,1\}^*$ to $\{0,1\}^*$, and let f'_k be as just described. Then for any adjacent databases D, D' and any (unbounded) $\mathcal{A}$:

$$\Pr[\mathcal{A}(f'_k(D)) = 1] = \Pr_{\mathcal{O}\leftarrow\mathsf{Func}}[\mathcal{A}(f_k^{\mathcal{O}}(D)) = 1]$$

and

$$\Pr[\mathcal{A}(f'_k(D')) = 1] = \Pr_{\mathcal{O}\leftarrow\mathsf{Func}}[\mathcal{A}(f_k^{\mathcal{O}}(D')) = 1].$$

Letting OWF denote the event that $\mathcal{O}$ is one-way, we have

$$\begin{aligned}\Pr[\mathcal{A}(f'_k(D)) = 1] &\leq \Pr\left[\mathcal{A}(f_k^{\mathcal{O}}(D)) = 1 \mid \mathsf{OWF}\right] + \mathsf{negl}(k) \\ &\leq e^{\epsilon(k)} \times \Pr\left[\mathcal{A}(f_k^{\mathcal{O}}(D')) = 1 \mid \mathsf{OWF}\right] + \mathsf{negl}'(k) \\ &\leq e^{\epsilon(k)} \times \Pr[\mathcal{A}(f'_k(D')) = 1] + \mathsf{negl}''(k).\end{aligned}$$

The second inequality holds since $\{f_k\}$ is a fully black-box construction of an ϵ-IND-CDP mechanism from one-way functions. (Note that, above, $\mathcal{A}$ is not given access to $\mathcal{O}$ at all.) But the condition that

$$\Pr[\mathcal{A}(f'_k(D)) = 1] \leq e^{\epsilon(k)} \times \Pr[\mathcal{A}(f'_k(D')) = 1] + \mathsf{negl}''(k)$$

for an unbounded $\mathcal{A}$ is equivalent to $(\epsilon, \mathsf{negl})$-differential privacy.

The claim regarding the utility of $\{f'_k\}$ follows by a similar argument. (Note that we do not require that U be efficiently computable.)

Note that the above proof holds not just for constructions based on one-way functions, but for any black-box construction from a primitive P that can be instantiated with a random object. This includes, e.g., ideal ciphers, collision-resistant hash functions, and trapdoor permutations [7].

4 Limitations for Computational Differential Privacy

In the previous section we ruled out black-box constructions from general assumptions, but with regard to arbitrary measures of utility and arbitrary mechanisms. Here, we focus on *arbitrary* mechanisms with output in $\mathbb{R}^n$ (for constant n), and a large, but specific, class of efficiently computable utilities. Specifically, we look at utilities defined by (a generalization of) the L_p norm.

Definition 8 (L_p-norm) *The L_p-norm of a vector $\mathbf{x} \in \mathbb{R}^n$, denoted $||\mathbf{x}||_p$, is defined as*

$$||\mathbf{x}||_p \stackrel{\text{def}}{=} (|x_1|^p + |x_2|^p + \ldots + |x_n|^p)^{1/p}$$

for $p \in \mathbb{N}^+$, where x_i is the ith coordinate of $\mathbf{x}$. (We do not deal with the L_0 norm in this paper.) We also allow $p = \infty$, where

$$||\mathbf{x}||_\infty \stackrel{\text{def}}{=} \max(|x_1|, |x_2|, \ldots, |x_n|).$$

A natural notion of utility would be to look at the average distance (in some L_p norm) from the true answer to the output of the mechanism. We broaden this to include things like mean-squared error that are commonly used in statistics.

Definition 9 (Average (p, v)-error) *Let $f_k : \mathcal{D} \to \mathbb{R}^n$ be a mechanism for answering a query $q : \mathcal{D} \to \mathbb{R}^n$. The* average (p, v)-error *(also called the v*th moment of the L_p error*) of this mechanism ($p > 0, v \geq 1$) on database D is*

$$\sigma_{p,v}(q, D, f_k) \stackrel{\text{def}}{=} \mathbf{E}\left[||f_k(D) - q(D)||_p^v\right].$$

We often refer to the above as "error" rather than "utility"; lower error values are good, whereas lower utility values are bad. We remark that we can handle utility measures beyond the above, as long as they satisfy a technical requirement that follows from our proof. Since we do not currently have any clean way to state this requirement, we do not discuss it further

Given a mechanism $\{f_k : \mathcal{D} \to \mathbb{R}^n\}_{k\in\mathbb{N}}$ for answering a query $q : \mathcal{D} \to \mathbb{R}^n$, we say *the average (p, v)-error of $\{f_k\}$ is polynomially bounded* if there is a polynomial err such that, for all D and k, we have

$$\sigma_{p,v}(q, D, f_k) \leq \mathsf{err}(k).$$

Theorem 2, below, shows that nothing can be gained by using computational differential privacy rather than statistical differential privacy, as long as we consider mechanisms whose error is polynomially bounded. Before giving the formal theorem statement and proof in the following section, we give an intuitive explanation here.

Let f_k be a polynomial-time ϵ-SIM-CDP mechanism for answering some query $q : \mathcal{D} \to \mathbb{R}^n$, where we assume that f_k also has output in $\mathbb{R}^n$ (and n is independent of k). Let $p > 0, v \geq 1$ be arbitrary, and assume the average (p, v)-error of f_k is polynomially bounded with error bound err. We claim there is an $(\epsilon, \mathsf{negl})$-DP mechanism $\hat{f}_k$ with essentially the same running time[1] as f_k, and such that $\sigma_{p,v}(q, D, \hat{f}_k) < \mathsf{err}(k) + \mathsf{negl}(k)$.

Let $\{F_k\}$ be a mechanism that is $(\epsilon, \mathsf{negl})$-DP and indistinguishable from $\{f_k\}$. Such a mechanism is guaranteed to exist by definition of SIM-CDP. Note that $\{F_k\}$ may be much less efficient than $\{f_k\}$, and may not even be implementable in polynomial time. On the other hand, F_k and f_k must induce distributions over $\mathbb{R}^n$ that are, in some sense, very close. Intuitively, in any "box" in $\mathbb{R}^n$ of noticeable size, the probabilities with which the outputs of F_k or f_k lie in that cell must be roughly equal; if not, the difference in probabilities could be used to distinguish F_k and f_k (since membership in the box can be efficiently tested).

We derive $\hat{f}_k$ by adding a small amount of uniform noise to the output of f_k. Carefully setting the amount of noise to be sufficiently small, we can bound the error introduced in moving from f_k to $\hat{f}_k$. To analyze privacy of the resulting mechanism, we look at the mechanism $\hat{F}_k$ where a small amount of uniform noise is added to F_k. For any particular value x, the probability with which $\hat{f}_k$ (resp., $\hat{F}_k$) outputs x is proportional to the probability that f_k (resp., F_k) outputs a value within a box centered at x. This box is sufficiently big so that $\hat{F}_k$ and $\hat{f}_k$ have similar probabilities of outputting any particular value.

While $\hat{F}_k$ and $\hat{f}_k$ have similar probabilities of outputting any particular value, these small differences could, in principle, compound and become unacceptably large when summed over all values in some set $S \subset \mathbb{R}^n$. To show that such differences do not grow too large, we use the fact that f_k has polynomially bounded error. This allows us to break our analysis into two parts: one focusing on a region S_c "close" to the correct answer $q(D)$, and the other focusing on

[1] Specifically, $\hat{f}_k$ runs f_k and adds a random number to its output.

$S_f = S \setminus S_c$. We show that

$$\left|\Pr[\hat{f}_k(D) \in S_c] - \Pr[\hat{F}_k(D) \in S_c]\right|$$

is small, using the argument discussed above; we also show that

$$\max\{\Pr[\hat{f}_k(D) \in S_f], \Pr[\hat{F}_k(D) \in S_f]\}$$

is small by the polynomial bound on the error. Combined, this shows that for every S, the difference

$$\left|\Pr[\hat{f}_k(D) \in S] - \Pr[\hat{F}_k(D) \in S]\right|$$

is small, as required. Since F_k, and hence $\hat{F}_k$, is *statistically* differentially private, this means that $\hat{f}_k$ is also.

Formal details are given in the following section.

4.1 Statement and Proof of the Main Result

We first present a proof that applies to the (stronger) SIM-CDP definition. We then outline the changes needed to prove the result for the case of IND-CDP.

Theorem 2 *Fix $p > 0, v \geq 1$. Let $\{f_k : \mathcal{D} \to \mathbb{R}^n\}$ be an efficient ϵ-SIM-CDP mechanism whose average (p, v)-error is polynomially bounded by err. Then there is an efficient $(\epsilon, \mathsf{negl})$-DP mechanism $\{\hat{f}_k\}$ with $\sigma_{p,v}(q, D, \hat{f}_k) < \mathsf{err}(k) + \mathsf{negl}(k)$.*

Moreover, $\hat{f}_k$ has essentially the same running time as f_k; specifically, $\hat{f}_k$ only adds uniform noise to f_k.

Proof. Let $\{F_k\}$ be an $(\epsilon, \mathsf{negl})$-DP family of mechanisms that is indistinguishable from $\{f_k\}$. Let negl_1 be a negligible function such that for any non-uniform polynomial-time $\mathcal{A}$ and any database D,

$$\left|\Pr[\mathcal{A}(f_k(D)) = 1] - \Pr[\mathcal{A}(F_k(D)) = 1]\right| \leq \mathsf{negl}_1(k).$$

(Such a function exists by definition of SIM-CDP.)

Since $\{f_k\}$ is efficient, its output must have some polynomial length. We assume that f_k (and hence F_k) give output in fixed-point notation with k bits of precision. Formally, let $\mathbb{R}_k$ be the set

$$\mathbb{R}_k = \{x \in \mathbb{R} \mid \exists j \in \mathbb{Z} : x = j \cdot 2^{-k}\};$$

then we assume that f_k gives output in $\mathbb{R}_k^n$. (More generally, the proof given here works when the precision is any polynomial in k. Moreover, fixed-point notation is not essential; in particular, the proof can be modified for the case when the output of f_k is given in floating-point notation.) For $x \in \mathbb{R}$ and $k \in \mathbb{N}$, define $\lceil x \rceil_k \stackrel{\text{def}}{=} \lceil x \cdot 2^k \rceil \cdot 2^{-k}$ to be the value x "rounded up" so that it lies in $\mathbb{R}_k$.

A set $\mathcal{B} \subset \mathbb{R}^n$ is a *box* if it a Cartesian product of closed intervals in $\mathbb{R}$. Abusing notation, we call a sequence $\{\mathcal{B}_k\}$ of boxes $\mathcal{B}_k \subset \mathbb{R}_k^n$ a box as well. The following is an immediate consequence of the SIM-CDP definition (recall the definition requires indistinguishability against non-uniform adversaries):

Lemma 1 *For any box $\{\mathcal{B}_k\}$ and any database D:*

$$|\Pr[f_k(D) \in \mathcal{B}_k] - \Pr[F_k(D) \in \mathcal{B}_k]| \leq \mathsf{negl}_1(k).$$

We next define two mechanisms $\{\hat{F}_k\}$ and $\{\hat{f}_k\}$ that are "noisy" versions of $F(D)$ and $f(D)$, respectively. Because we are dealing with discrete rather than continuous values, the definition is more complicated than simply adding uniform noise in some range.

Set $c(k) = \left\lceil \sqrt[4n]{\mathsf{negl}_1(k)} \right\rceil_k$. For $\mathbf{x} \in \mathbb{R}_k^n$, let $\mathcal{B}_{c,k}(\mathbf{x})$ denote the box with radius $c(k)$ (in the L_∞ norm) centered at $\mathbf{x}$; that is,

$$\mathcal{B}_{c,k}(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}_k^n \,:\, ||\mathbf{y} - \mathbf{x}||_\infty \leq c(k)\}.$$

Mechanism $\{\hat{f}_k\}$ is defined as follows: $\hat{f}_k(D)$ computes $f_k(D)$, and then outputs a uniform value in $\mathcal{B}_{c,k}(f(D))$. (This is equivalent to adding uniform, independent, discretized noise from $[-c(k), c(k)]$ to each coordinate of $f(D)$.) Mechanism $\{\hat{F}_k\}$ is defined to be the analogous mechanism that adds noise to F instead of f.

$B_{c,k}(\mathbf{x})$ contains $\left(c(k) \cdot 2^{k+1} + 1\right)^n$ points and thus, for any D and $\mathbf{x} \in \mathbb{R}_k^n$:

$$\Pr[\hat{F}_k(D) = \mathbf{x}] = \left(c(k) \cdot 2^{k+1} + 1\right)^{-n} \cdot \Pr[F_k(D) \in B_{c,k}(\mathbf{x})]$$

and

$$\Pr[\hat{f}_k(D) = \mathbf{x}] = \left(c(k) \cdot 2^{k+1} + 1\right)^{-n} \cdot \Pr[f_k(D) \in B_{c,k}(\mathbf{x})].$$

Taking $\mathcal{B}_k = \mathcal{B}_{c,k}(\mathbf{x}_k)$ (for an arbitrary sequence $\{\mathbf{x}_k\}$ with $\mathbf{x}_k \in \mathbb{R}_k^n$) in Lemma 1, we obtain:

$$\begin{aligned}
&\left|\Pr[\hat{F}_k(D) = \mathbf{x}_k] - \Pr[\hat{f}_k(D) = \mathbf{x}_k]\right| \\
&= \left(c(k) \cdot 2^{k+1} + 1\right)^{-n} \cdot \left|\Pr[F_k(D) \in B_{c,k}(\mathbf{x}_k)] - \Pr[f_k(D) \in B_{c,k}(\mathbf{x}_k)]\right| \\
&\leq \left(c(k) \cdot 2^{k+1} + 1\right)^{-n} \cdot \mathsf{negl}_1(k). \qquad (1)
\end{aligned}$$

The above holds for an arbitrary database D, and so it also holds for any adjacent database D'.

$\hat{F}_k$ applies post-processing to the output of F_k, so $\{\hat{F}_k\}$ is also $(\epsilon, \mathsf{negl})$-DP. Let negl_2 be a negligible function such that for all sets S and adjacent databases D and D' it holds that

$$\Pr[\hat{F}_k(D) \in S] \leq e^{\epsilon(k)} \times \Pr[\hat{F}_k(D') \in S] + \mathsf{negl}_2(k). \qquad (2)$$

Our goal is to prove that $\hat{f}_k(D)$ is *statistically* close to $\hat{F}_k(D)$, for any D, which will then imply the theorem. We have already shown (cf. Equation (1)) that the distributions of $\hat{f}_k(D)$ and $\hat{F}_k(D)$ are *pointwise* negligibly close. We need to show that this is true also for arbitrary subsets. To do this, we first use the polynomial error bound on f_k to argue that f_k (and hence $\hat{f}_k$) must put relatively low weight on outputs that are far from the correct output. Formally:

Lemma 2 *There is a polynomial b such that, for any D, we have*

$$\sigma_{p,v}(q, D, \hat{f}_k) \leq \mathsf{err}(k) + c(k) \cdot b(k).$$

The lemma follows from the observation that, for any fixed output $y = f_k(D)$, the output $\hat{y} = \hat{f}_k(D)$ satisfies

$$||\hat{y} - q(D)||_p \leq ||y - q(D)||_p + n \cdot c(k).$$

The proof of the lemma is tedious, and so we defer it to Appendix A.

Fix an arbitrary D. We now show that with high probability the output of $\hat{f}_k(D)$ is close to the true answer $q(D)$. Set $z(k) = \left\lceil \frac{1}{\sqrt[4n]{\mathsf{negl}_1(k)}} \right\rceil_k$, and define

$$\mathsf{Close}_k \stackrel{\text{def}}{=} \{\mathbf{x} \in \mathbb{R}_k^d : ||\mathbf{x} - q(D)||_p^v \leq z(k)\};$$

i.e., these are the points close to $q(D)$. Let $\mathsf{Far}_k \stackrel{\text{def}}{=} \mathbb{R}_k^n \setminus \mathsf{Close}_k$. Because the average error of $\hat{f}_k$ is at most $\mathsf{err}(k) + b(k) \cdot c(k)$, we have

$$\Pr[\hat{f}_k(D) \in \mathsf{Far}_k] \leq \left(\mathsf{err}(k) + b(k) \cdot c(k)\right) / z(k). \tag{3}$$

Indistinguishability of $\{f_k\}$ and $\{F_k\}$, and the manner in which $\{\hat{f}_k\}$ and $\{\hat{F}_k\}$ are constructed, implies that $\{\hat{f}_k\}$ and $\{\hat{F}_k\}$ are indistinguishable as well. As in the proof of Lemma 1, this means that

$$\left|\Pr[\hat{f}_k(D) \in \mathsf{Far}_k] - \Pr[\hat{F}_k(D) \in \mathsf{Far}_k]\right| \leq \mathsf{negl}_1(k).$$

Combining this with Equation (3) yields

$$\Pr[\hat{F}_k(D) \in \mathsf{Far}_k] \leq \left(\mathsf{err}(k) + b(k) \cdot c(k)\right) / z(k) + \mathsf{negl}_1(k).$$

We now use the above results to relate the probabilities that $\hat{F}_k(D)$ or $\hat{f}_k(D)$ lie within some arbitrary set. The number of points in Close_k is bounded from above by $(z(k) \cdot 2^{k+1} + 1)^n$, since its size is largest (for fixed $z(k)$) when $p = \infty$ and $v = 1$. For any $S_k \subset \mathbb{R}_k^n$, we can thus lower-bound $\Pr[\hat{F}_k(D) \in S_k]$ via

$$\begin{aligned}
\Pr[\hat{F}_k(D) \in S_k] &= \sum_{\mathbf{x} \in S_k} \Pr[\hat{F}_k(D) = \mathbf{x}] \\
&\geq \sum_{\mathbf{x} \in S_k \cap \mathsf{Close}_k} \Pr[\hat{F}_k(D) = \mathbf{x}] \\
&\geq \sum_{\mathbf{x} \in S_k \cap \mathsf{Close}_k} \left(\Pr[\hat{f}_k(D) = \mathbf{x}] - \left(c(k) \cdot 2^{k+1} + 1\right)^{-n} \cdot \mathsf{negl}_1(k)\right),
\end{aligned}$$

using Equation (1), which bounds the difference in probabilities between $\hat{f}_k$ and $\hat{F}_k$ pointwise. Continuing, we have

$$\begin{aligned}
&\Pr[\hat{F}_k(D) \in S_k] \\
&\geq \Pr[\hat{f}_k(D) \in S_k \cap \mathsf{Close}_k] - \left(z(k) \cdot 2^{k+1} + 1\right)^n \cdot \left(c(k) \cdot 2^{k+1} + 1\right)^{-n} \cdot \mathsf{negl}_1(k) \\
&\geq \Pr[\hat{f}_k(D) \in S_k \cap \mathsf{Close}_k] - \left(\frac{z(k)+1}{c(k)}\right)^n \cdot \mathsf{negl}_1(k) \\
&\quad + \left(\Pr[\hat{f}_k(D) \in S_k \cap \mathsf{Far}_k] - \left(\mathsf{err}(k) + b(k) \cdot c(k)\right)/z(k)\right) \\
&\geq \Pr[\hat{f}_k(D) \in S_k] - \left(\frac{z(k)+1}{c(k)}\right)^n \cdot \mathsf{negl}_1(k) - \left(\mathsf{err}(k) + b(k) \cdot c(k)\right)/z(k). \quad (4)
\end{aligned}$$

Similarly, we can upper-bound $\Pr[\hat{F}_k(D) \in S_k]$ via

$$\begin{aligned}
&\Pr[\hat{F}_k(D) \in S_k] \\
&\leq \sum_{\mathbf{x} \in S_k \cap \mathsf{Close}_k} \Pr[\hat{F}_k(D) = \mathbf{x}] + \Pr[\hat{F}_k(D') \in \mathsf{Far}_k] \\
&\leq \sum_{\mathbf{x} \in S_k \cap \mathsf{Close}_k} \left(\Pr[\hat{f}_k(D) = \mathbf{x}] + \left(c(k) \cdot 2^{k+1} + 1\right)^{-n} \cdot \mathsf{negl}_1(k)\right) \\
&\quad + \Pr[\hat{F}_k(D) \in \mathsf{Far}_k] \\
&\leq \Pr[\hat{f}_k(D) \in S_k] + \left(\frac{z(k)+1}{c(k)}\right)^n \cdot \mathsf{negl}_1(k) \\
&\quad + \left(\mathsf{err}(k) + b(k) \cdot c(k)\right)/z(k) + \mathsf{negl}_1(k). \quad (5)
\end{aligned}$$

Equations (4) and (5) hold for an arbitrary database D, and thus also hold for any adjacent database D'. Substituting into Equation (2) and simplifying, we obtain

$$\begin{aligned}
&\Pr[\hat{f}_k(D) \in S_k] \\
&\leq e^{\epsilon(k)} \times \Pr[\hat{f}_k(D') \in S_k] \\
&\quad + \left(e^{\epsilon(k)} + 1\right) \times \left(\left(\frac{z(k)+1}{c(k)}\right)^n \mathsf{negl}_1(k) + \left(\mathsf{err}(k) + b(k) \cdot c(k)\right)/z(k)\right) \\
&\quad + e^{\epsilon(k)} \cdot \mathsf{negl}_1(k) + \mathsf{negl}_2(k)\,.
\end{aligned}$$

We show that the additive terms are all negligible. Note first that

$$\begin{aligned}
\left(\frac{z(k)+1}{c(k)}\right)^n \cdot \mathsf{negl}_1(k) &\leq \left(\frac{\frac{1}{\sqrt[4n]{\mathsf{negl}_1(k)}} + 2}{\sqrt[4n]{\mathsf{negl}_1(k)}}\right)^n \cdot \mathsf{negl}_1(k) \\
&\leq \left(\frac{3}{\sqrt[2n]{\mathsf{negl}_1(k)}}\right)^n \mathsf{negl}_1(k) \\
&\leq 3^n \cdot \sqrt{\mathsf{negl}_1(k)},
\end{aligned}$$

which is negligible in k (recall n is constant). To bound $\left(\mathsf{err}(k) + b(k) \cdot c(k)\right)/z(k)$, take k large enough so that $b(k) \cdot c(k) \leq \mathsf{err}(k)$ (this is always possible, since c is negligible while err and b are polynomial). We then have

$$\frac{\mathsf{err}(k) + b(k) \cdot c(k)}{z(k)} \leq 2 \cdot \mathsf{err}(k) \cdot \sqrt[4n]{\mathsf{negl}_1(k)},$$

which is negligible. We conclude that $\{\hat{f}_k\}$ is $(\epsilon, \mathsf{negl})$-DP.

The case of IND-CDP. A result analogous to the above holds also for the case of IND-CDP. This follows fairly easily using the equivalent formulation of IND-CDP in terms of $\mathrm{SIM}_{\forall\exists}$-CDP. The difference between SIM-CDP and $\mathrm{SIM}_{\forall\exists}$-CDP is with respect to the order of quantifiers, but this has no real effect on our proof. Note, in particular, that our construction of $\{\hat{f}_k\}$ does not depend, either explicitly or implicitly, on $\{F_k\}$.

Acknowledgments

We thank the referees for their detailed comments which allowed us to simplify parts of our proof. The second author thanks Adam Smith for suggesting the problem, and for several interesting discussions on the topic of computational differential privacy.

References

1. A. Blum, C. Dwork, F. McSherry, and K. Nissim. Practical privacy: The SuLQ framework. In *24th ACM Symposium on Principles of Database Systems (PODS)*, pages 128–138. ACM Press, 2005.
2. A. Blum, K. Ligett, and A. Roth. A learning theory approach to non-interactive database privacy. In *40th Annual ACM Symposium on Theory of Computing (STOC)*, pages 609–618. ACM Press, 2008.
3. C. Dwork. Differential privacy. In *33rd Intl. Colloquium on Automata, Languages, and Programming (ICALP), Part II*, volume 4052 of *LNCS*, pages 1–12. Springer, 2006.
4. C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology — Eurocrypt 2006*, volume 4004 of *LNCS*, pages 486–503. Springer, 2006.
5. C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *3rd Theory of Cryptography Conference — TCC 2006*, volume 3876 of *LNCS*, pages 265–284. Springer, 2006.
6. C. Dwork and K. Nissim. Privacy-preserving datamining on vertically partitioned databases. In *Advances in Cryptology — Crypto 2004*, volume 3152 of *LNCS*, pages 528–544. Springer, 2004.
7. R. Gennaro, Y. Gertner, J. Katz, and L. Trevisan. Bounds on the efficiency of generic cryptographic constructions. *SIAM Journal on Computing*, 35(1):217–246, 2005.

8. R. Impagliazzo and S. Rudich. Limits on the provable consequences of one-way permutations. In *21st Annual ACM Symposium on Theory of Computing (STOC)*, pages 44–61. ACM Press, 1989.
9. S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately? In *49th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 531–540. IEEE, 2008.
10. A. McGregor, I. Mironov, T. Pitassi, O. Reingold, K. Talwar, and S. P. Vadhan. The limits of two-party differential privacy. In *51st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 81–90. IEEE, 2010.
11. I. Mironov, O. Pandey, O. Reingold, and S. Vadhan. Computational differential privacy. In *Advances in Cryptology — Crypto 2009*, volume 5677 of *LNCS*, pages 126–142. Springer, 2009.
12. K. Nissim, S. Raskhodnikova, and A. Smith. Smooth sensitivity and sampling in private data analysis. In *39th Annual ACM Symposium on Theory of Computing (STOC)*, pages 75–84. ACM Press, 2007.
13. O. Reingold, L. Trevisan, and S. P. Vadhan. Notions of reducibility between cryptographic primitives. In *1st Theory of Cryptography Conference — TCC 2004*, volume 2951 of *LNCS*, pages 1–20. Springer, 2004.

A Proof of Lemma 2

Let Y_k be the set of possible distances between two points in $\mathbb{R}_k^n$; i.e.,

$$Y_k \stackrel{\text{def}}{=} \{y \in \mathbb{R} \mid y = ||\mathbf{x}_1 - \mathbf{x}_2||_p \text{ for some } \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}_k^n\}.$$

Let $p_{y,k} \stackrel{\text{def}}{=} \Pr\left[y - 2^{-k} < ||f_k(D) - q(D)||_p \leq y\right]$. Then, by the assumption of our theorem,

$$\sigma_{p,v}(q, D, f_k) \leq \textstyle\sum_{y \in Y_k} p_{y,k} \cdot y^v \leq \mathsf{err}(k).$$

We can upper-bound $\sigma_{p,v}(q, D, \hat{f}_k)$ by assuming that the noise added by $\hat{f}_k$ moves the output further away from the correct answer $q(D)$. In the worst case (when $p = 1$), this increases the distance between the output and $q(D)$ by at most $c'(k) \stackrel{\text{def}}{=} n \cdot c(k)$. Therefore,

$$\sigma_{p,v}(q, D, \hat{f}_k) \leq \textstyle\sum_{y \in Y_k} p_{y,k} \cdot (y + c'(k))^v.$$

Using Taylor's theorem, $(y + c'(k))^v \leq y^v + v \cdot (y + c'(k))^{v-1} \cdot c'(k)$. Thus, for k sufficiently large it holds that

$$\begin{aligned}
\sigma_{p,v}(q, D, \hat{f}_k) &\leq \textstyle\sum_{y \in Y_k} p_{y,k} \cdot \left(y^v + v \cdot (y + c'(k))^{v-1} \cdot c'(k)\right) \\
&\leq \mathsf{err}(k) + \textstyle\sum_{y \in Y_k} p_{y,k} \cdot \left(v \cdot (y + c'(k))^{v-1} \cdot c'(k)\right) \\
&\leq \mathsf{err}(k) + v \cdot c'(k) \cdot \textstyle\sum_{y \in Y_k} p_{y,k} \cdot (y + n)^{v-1},
\end{aligned}$$

using for the last inequality the fact that $c'(k) \leq n$ for k large enough.

If $y \leq n$ then $(y+n)^{v-1} \leq (2n)^{v-1}$, while if $y \geq n$ then $(y+n)^{v-1} \leq (2y)^{v-1}$. As a result, we can bound the expression above as

$$\begin{aligned}
&\sigma_{p,v}(q, D, \hat{f}_k) \\
&\leq \mathsf{err}(k) + v \cdot c'(k) \cdot \sum_{y \in Y_k} p_{y,k} \cdot 2^{v-1} \cdot (n^{v-1} + y^{v-1}) \\
&\leq \mathsf{err}(k) + v \cdot c'(k) \cdot \left(\sum_{y \in Y_k} p_{y,k} \cdot 2^{v-1} n^{v-1} + \sum_{y \in Y_k} p_{y,k} \cdot 2^{v-1} y^{v-1} \right) \\
&\leq \mathsf{err}(k) + v \cdot c'(k) \cdot \left(2^{v-1} n^{v-1} + 2^{v-1} \sum_{y \in Y_k} p_{y,k} \cdot y^{v-1} \right).
\end{aligned}$$

Since $y > 0$, we have $y^{v-1} \leq y^v + 1$. Then:

$$\begin{aligned}
\sigma_{p,v}(q, D, \hat{f}_k) &\leq \mathsf{err}(k) + v \cdot c'(k) \cdot \left(2^{v-1} n^{v-1} + 2^{v-1} \sum_{y \in Y_k} p_{y,k} \cdot (y^v + 1) \right) \\
&\leq \mathsf{err}(k) + v \cdot c'(k) \cdot \left(2^{v-1} n^{v-1} + 2^{v-1} \cdot (\mathsf{err}(k) + 1) \right) \\
&\leq \mathsf{err}(k) + c(k) \cdot \left(2^{v-1} v \cdot n^v + 2^{v-1} v \cdot n \cdot (\mathsf{err}(k) + 1) \right).
\end{aligned}$$

Since err is polynomial and n, v are constants, this completes the proof.