Computational Differential Privacy

Ilya Mironov[1], Omkant Pandey[2,⋆], Omer Reingold[3,⋆,⋆⋆], and
Salil Vadhan[4,⋆,⋆⋆]

[1] Microsoft Research, Silicon Valley Campus
mironov@microsoft.com
[2] University of California, Los Angeles
omkant@cs.ucla.edu
[3] Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel
omer.reingold@weizmann.ac.il
[4] School of Engineering and Applied Sciences and Center for Research on Computation and Society, Harvard University
salil@eecs.harvard.edu

Abstract. The definition of differential privacy has recently emerged as a leading standard of privacy guarantees for algorithms on statistical databases. We offer several relaxations of the definition which require privacy guarantees to hold only against *efficient*—i.e., computationally-bounded—adversaries. We establish various relationships among these notions, and in doing so, we observe their close connection with the theory of pseudodense sets by Reingold et al. [1]. We extend the dense model theorem of Reingold et al. to demonstrate equivalence between two definitions (indistinguishability- and simulatability-based) of computational differential privacy.

Our computational analogues of differential privacy seem to allow for more accurate constructions than the standard information-theoretic analogues. In particular, in the context of private approximation of the distance between two vectors, we present a differentially-private protocol for computing the approximation, and contrast it with a substantially more accurate protocol that is only *computationally* differentially private.

1 Introduction

A curator of a statistical database may promote valuable social purposes in his operation. At the same time, non-careful procedures for managing access to the database may expose sensitive information (in potentially subtle ways), damaging individual contributors and putting the curator at the risk of legal liability.

The statistics, database, and datamining communities have long understood that there is a complicated space of possible trade-offs between usability of statistical databases and secrecy of individual records. A recent

⋆ Work done in part while visiting Microsoft Research Silicon Valley.
⋆⋆ Supported by US-Israel BSF grant 2006060.

line of research in privacy in statistical databases is focussed on formalizing and quantifying the notions of privacy and usability, and developing privacy-preserving analogues for many types of queries or algorithms one may want to run on a database (surveyed by Dwork [2]).

The cornerstone of the new approach to privacy is the definition of *differential privacy*, which first appeared in [3]. Intuitively, the definition captures the risk of joining the database, where the risk is measured as the adversary's success in predicting whether a single record is present in the database, given the rest of the database. The definition gives unconditional guarantees (including privacy for (small) groups) against a powerful adversary, preserved by sequential composition, and still allows many types of statistical or machine learning analyses, as shown in [4–6]. We note that the adversary's gain in success probability typically tolerated in applications of differential privacy is not zero or even "cryptographically" small (and cannot be so under any reasonable utility guarantees [3]).

The standard definition of differential privacy is very strong in that it provides privacy even against a computationally unbounded adversary. While there has been substantial success in designing mechanisms that achieve this strong definition (e.g., [5, 7, 6, 8]), in this paper we suggest that such information-theoretic privacy may sometimes have a significant price (in utility or complexity). Thus we propose several computational analogues of differential privacy, where we only require privacy against a feasible (i.e., polynomial time) adversary.

Immediate benefits of the relaxation include combining pseudo-random generators with differentially-private mechanisms, or running such mechanisms in a distributed manner with only computational guarantees of security. More importantly, computational differentially-private mechanisms may exist for problems for which standard differentially-private mechanisms are impossible or unknown. We propose the problem of constructing a two-party protocol with two-sided guarantees of privacy for approximating the Hamming distance between two bit-vectors as a candidate for separating the power of computational and information-theoretic definitions of privacy.

Definition of computational differential privacy. There are two natural approaches to defining differential privacy with a computational flavor. The first one, which may be characterized as *indistinguishability*-based, goes back to the definition of differential privacy and replaces an unrestricted adversary with a computationally-bounded one. Doing so, at least in the non-uniform case, does not expand the class of privacy-preserving algorithms as the new definition can be shown equivalent to the old one. If instead, we start with the weaker definition of (ϵ, δ)-differential privacy [9], which allows some negligible additive distinguishing advantage, we do obtain a new class of mechanisms that are private under the new definition, which we call IND-CDP.

The second approach to defining computational differential privacy, which we naturally call *simulation*-based or SIM-CDP, builds upon the definition of differential privacy and its properties. It asserts that the view of the adversary can be simulated given access to a differentially-private function of the database (and thus the simulation is differentially private).

The simulated view must be computationally indistinguishable from the real mechanism's transcript.

Given these two equally compelling definitional approaches it is quite natural to consider their relative power. We show that both definitions are closed under sequential composition and provide privacy for (small) groups of records. We thus switch to study the relationship between these definitions. One can easily demonstrate that SIM-CDP implies IND-CDP. The converse of this statement, however, is an intriguing question that we leave open in this work. Instead, in the main technical contribution of this paper we establish equivalence between a weaker (though still natural) simulation-based definition (called $\text{SIM}_{\forall\exists}$-CDP) and IND-CDP (Section 3). We also generalize our definition to interactive mechanisms, where we uncover one more definition, called SIM^{+}-CDP. A summary of our results relating various definitions of privacy is presented in Figure 1.

$$\begin{tikzcd}
\text{IND-CDP} \arrow[r, shift right, "\text{Theorem 3}"'] & \text{SIM}_{\forall\exists}\text{-CDP} \arrow[l, shift right, "\text{Theorem 1}"'] & \text{SIM-CDP} \arrow[l, "\text{immediate}"'] \arrow[r, shift right, "/" marking, "\text{Theorem 4}"'] & \text{SIM}^{+}\text{-CDP} \arrow[l, shift right, "\text{Theorem 4}"']
\end{tikzcd}$$

Fig. 1. Relations between definitions of computational differential privacy.

Our approach to proving equivalence between IND-CDP and $\text{SIM}_{\forall\exists}$-CDP establishes a surprising connection between computational differential privacy and "pseudodense sets" studied by Reingold et al. [1], who were in turn motivated by the work of Green, Tao, and Ziegler in additive combinatorics [10, 11] (closely related notions were previously studied by [12]). In fact, IND-CDP can be stated in terms of *pseudodensity* of the mechanism's distribution over adjacent datasets, and SIM-CDP is equivalent to existence of *models* that are *dense* for all adjacent pairs, to use the language of [1]. We extend the Dense Model Theorem of [1] to account for the symmetry of the definitions of differential privacy.

As mentioned above, we also construct protocols achieving computational differential privacy that seem to admit significantly better accuracy than any information-theoretically private protocols. Our main example is a private approximation of the Hamming distance between two vectors in a two-party setting. We propose three protocols for this problem: one protocol with information-theoretic differential privacy guarantees and multiplicative approximation error, and two protocols handling the semi-honest and malicious cases achieving computational differential privacy (specifically SIM-CDP) and with error *independent of the size of the input* (it only depends on the privacy and security parameters). Subsequent to our results, this gap in accuracy between information-theoretic and computational differential privacy was shown to be inherent [13].

Secure vs differentially-private computations. The problem of secure Hamming distance computation, together with closely related problems of secure scalar product and secure set-intersection cardinality [14–18], may benefit from casting them in the differential privacy framework. Indeed, the standard cryptographic guarantee of letting the parties compute the output of a function, such as the Hamming distance between two vectors, while hiding everything else about their inputs may be insufficient to argue security of a sequential composition of this protocol when one has to consider the information leaked through the *output* of the function. For example, if Alice varies her input while Bob's vector stays constant, by observing the output of the protocol Alice may learn individual values of Bob's input bits. Differential privacy treatment addresses the orthogonal question of *what* is computed rather than *how*; in particular, it may be used to analyze effects of adaptive sequential or concurrent composition on the adversary's confidence in predicting any particular bit, even in the presence of auxiliary information.

2 Definitions

We describe our definitions in this section. We start by introducing some notation.

Mechanism f. In our definitions, we will be interested in measuring privacy guarantees provided by randomized mechanisms, denoted f. Mechanism f operates on subsets D of a (potentially infinite) universe U, which we associate with databases, and outputs a value in the range $\mathcal{R}$. The size of the input D will be denoted by n. We say that two databases D and D' are *adjacent* if their symmetric difference contains at most one record (i.e., $|D \Delta D'| \leq 1$)[5]. Further, the maximum size of the output of f is m.

As we are dealing with computational notion, we will mostly be concerned with *efficient* adversaries. Unless specified otherwise, throughout the paper, an *efficient* adversary is modeled by a family of polynomial-sized circuits $\{A_\kappa\}_{\kappa \in \mathbb{N}}$, or equivalently, a nonuniform probabilistic polynomial time (PPT) Turing machine.

Parameter κ. A "security" parameter κ controls various quantities in our definitions/constructions as follows. The size of the adversary will be polynomial in κ. The mechanism is parameterized by κ, which lets us consider a family $\{f_\kappa\}_{\kappa \in \mathbb{N}}$, where $f_\kappa \colon \mathcal{D} \to \mathcal{R}_\kappa$. The output size m of f is required to be (at most) polynomial in κ. We say that a function in κ is *negligible* if it approaches zero faster than the reciprocal of any polynomial in κ.

We first recall the standard definition of ϵ-differential privacy [3]:

Definition 1 (ϵ-DP privacy) *A randomized mechanism $f \colon \mathcal{D} \to \mathcal{R}$ provides ϵ-DP if for all adjacent inputs $D, D' \in \mathcal{D}$ (i.e., $|D \Delta D'| \leq 1$) and all subsets $S \subseteq \mathcal{R}$*

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

where the probability space is f's coin tosses.

[5] Δ denotes symmetric difference of two sets and $|\cdot|$ denotes the size when the argument is a set.

A closely related notion of (ϵ, δ)-differential privacy [9] has an additive parameter that allows the probabilities to diverge when they are both relatively small:

Definition 2 ((ϵ, δ)-DP privacy) *A randomized mechanism $f \colon \mathcal{D} \to \mathcal{R}$ provides (ϵ, δ)-DP if for all adjacent inputs D and D' and all subsets $S \subseteq \mathcal{R}$*

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

where the probability space is f's coin tosses.

Our first new definition, ϵ_κ-IND-CDP, is an adaptation of ϵ-differential-privacy to the computational setting. This adaptation is obtained by considering a PPT adversary A, and requiring that f "looks differentially private" to every such A.

Definition 3 (IND-CDP privacy) *An ensemble $\{f_\kappa\}_{\kappa \in \mathbb{N}}$ of* randomized *functions $f_\kappa \colon \mathcal{D} \to \mathcal{R}_\kappa$ provides ϵ_κ-IND-CDP if there exists a negligible function* $\mathrm{negl}(\cdot)$ *such that for every nonuniform* PPT TM *("distinguisher") A, every polynomial $p(\cdot)$, every sufficiently large $\kappa \in \mathbb{N}$, all data sets $D, D' \in \mathcal{D}$ of size at most $p(\kappa)$ such that $|D \Delta D'| \leq 1$, and every advice string z_κ of size at most $p(\kappa)$, it holds that*

$$\Pr\left[A_\kappa(f_\kappa(D)) = 1\right] \leq e^{\epsilon_\kappa} \times \Pr\left[A_\kappa(f_\kappa(D')) = 1\right] + \mathrm{negl}(\kappa),$$

where we write $A_\kappa(x)$ for $A(1^\kappa, z_\kappa, x)$ and the probability is taken over the randomness of mechanism f_κ and adversary A_κ.

Notice that if the adversary A is allowed unbounded computation time, then the definition simply says that for any fixed κ the mechanism f_κ is $(e_\kappa, \delta_\kappa)$-DP for δ_κ being $\mathrm{negl}(\kappa)$. The reason we do not consider the computational analogue of ϵ_κ-DP (with $\delta = 0$) is that it ends up being equivalent to information-theoretic ϵ-DP. Indeed, for any singleton $r \in \mathcal{R}_\kappa$, we can choose A_κ to be the indicator function for $\{r\}$, implying that $\Pr[f_\kappa(D) = r] \leq e^{\epsilon_\kappa} \Pr[f_\kappa(D') = r]$. This immediately implies ϵ_κ-DP by summing both sides over all $r \in S$ for any subset $S \subset \mathcal{R}_\kappa$.
Our second definition, ϵ_κ-SIM-CDP, is described next. This definition interprets "looks differentially private" differently from our first definition: it says that f "looks differentially private" if there exists an ϵ-DP mechanism F (called simulator) such that $F(D)$ and $f(D)$ are computationally indistinguishable for every D.

Definition 4 (SIM-CDP privacy) *An ensemble $\{f_\kappa\}_{\kappa \in \mathbb{N}}$ of randomized functions $f_\kappa \colon \mathcal{D} \to \mathcal{R}_\kappa$ provides ϵ_κ-SIM-CDP if there exists an ensemble $\{F_\kappa\}_{\kappa \in \mathbb{N}}$ of ϵ_κ-*differentially-private mechanisms *$F_\kappa \colon \mathcal{D} \to \mathcal{R}_\kappa$ and a negligible function* $\mathrm{negl}(\cdot)$*, such that for every non-uniform* PPT TM *A, every polynomial $p(\cdot)$, every sufficiently large $\kappa \in \mathbb{N}$, every data* *set $D \in \mathcal{D}$ of size at most $p(\kappa)$, and every advice string z_κ of size at most $p(\kappa)$, it holds that,*

$$|\Pr\left[A_\kappa(f_\kappa(D)) = 1\right] - \Pr\left[A_\kappa(F_\kappa(D)) = 1\right]| \leq \mathrm{negl}(\kappa).$$

That is, $f_\kappa(D)$ and $F_\kappa(D)$ are computationally indistinguishable*.*

Note that the definition does not require F to be computable in probabilistic polynomial time; it only has to exist, and be (information theoretically) differentially private.
The definition of SIM-CDP requires that there *exists* a simulator F that acts in a differentially-private manner on *all* pairs of adjacent inputs. This suggests a weakening of the definition, where the order of quantifiers is reversed, i.e., instead of requiring a global simulator that works for all pairs of databases, we require that for *any* pair of adjacent databases there *exists* a simulator whose distributions on these two inputs satisfy the differential privacy condition.

Definition 5 (SIM$_{\forall\exists}$-CDP privacy) *An ensemble $\{f_\kappa\}_{\kappa\in\mathbb{N}}$ of randomized functions $f_\kappa\colon \mathcal{D} \to \mathcal{R}_\kappa$ provides ϵ_κ-SIM$_{\forall\exists}$-CDP if for all polynomials $p(\cdot)$, all sequences $\{(D_\kappa, D'_\kappa)\}_{\kappa\in\mathbb{N}}$ of pairs of datasets such that $|D_\kappa| \leq p(\kappa), |D'_\kappa| \leq p(\kappa)$ and $|D_\kappa \Delta D'_\kappa| \leq 1$, there exist ensembles $\{F_\kappa(D_\kappa)\}_{\kappa\in\mathbb{N}}$ and $\{F_\kappa(D'_\kappa)\}_{\kappa\in\mathbb{N}}$, such that the following two conditions hold:*

1. *[F_κ is ϵ_κ-DP.] For all subsets $S \subset \mathcal{R}$:*

$$e^{-\epsilon_\kappa} \times \Pr[F_\kappa(D'_\kappa) \in S] \leq \Pr[F_\kappa(D_\kappa) \in S] \leq e^{\epsilon_\kappa} \times \Pr[F_\kappa(D'_\kappa) \in S].$$

2. *[$f_\kappa(D_\kappa), f_\kappa(D'_\kappa)$ are indistinguishable from $F_\kappa(D_\kappa), F_\kappa(D'_\kappa)$ respectively.] For every non-uniform PPT TM A, every polynomial $q(\cdot)$, every sufficiently large $\kappa \in \mathbb{N}$, and every advice string z_κ of size at most $q(\kappa)$:*

$$|\Pr\left[A_\kappa(f_\kappa(D)) = 1\right] - \Pr\left[A_\kappa(F_\kappa(D)) = 1\right]| \leq \mathrm{negl}(\kappa) \text{ for } D \in \{D_\kappa, D'_\kappa\}$$

where we write $A_\kappa(x)$ for $A(1^\kappa, z_\kappa, x)$.

We may also consider an even weaker definition, where the probability $\Pr[F_\kappa(D_\kappa) \in S]$ is only bounded from above by $e^\epsilon \times \Pr[F_\kappa(D'_\kappa) \in S]$ (and a second pair of simulators exist for the ordered (D'_κ, D_κ) pair), but as we shall see in Section 3, it turns out to be equivalent to SIM$_{\forall\exists}$-CDP.
Robustness of our definitions. Protocols satisfying our definitions retain their privacy guarantees under sequential composition, and for all but SIM$_{\forall\exists}$-CDP we directly prove group privacy (for records); in both cases the privacy parameters ϵ and δ deteriorate linearly with the number of compositions or records, respectively. Due to space constraints, a detailed exposition is presented in the full version.Informally, under the definition of group privacy the adversary has to guess whether two or more elements are simultaneously present or absent in the database, given the rest of the database. The definition of group privacy is often applicable when

the differentially-private mechanism is preceded by computations that may amplify (up to a constant) the number of records affected by any individual [19].

Interactive case. For simplicity, current definitions consider only non-interactive mechanisms. An extension to the interactive case will be presented in Section 4. As it turns out, a variation of ϵ_κ-SIM-CDP, called ϵ_κ-SIM$^+$-CDP, can also be defined and proven separate from ϵ_κ-SIM-CDP (see Section 4).

3 Relations among various notions of CDP

In this section we establish reductions between the three definitions of computational differential privacy. The first implication, namely, that SIM-CDP implies SIM$_{\forall\exists}$-CDP, which implies IND-CDP, is the easiest (Section 3.1). The proof that IND-CDP implies SIM$_{\forall\exists}$-CDP (and thus that the two definitions are equivalent) extends the Dense Model Theorem of [1] and is significantly more involved (Section 3.2).

3.1 Simulatability implies indistinguishability

Theorem 1 (SIM-CDP ⇒ SIM$_{\forall\exists}$-CDP ⇒ IND-CDP) *If an ensemble $\{f_\kappa\}_{\kappa\in\mathbb{N}}$ of randomized functions $f_\kappa\colon \mathcal{D} \to \mathcal{R}_\kappa$ provides ϵ_κ-SIM-CDP, then it also provides ϵ_κ-SIM$_{\forall\exists}$-CDP; if it provides ϵ_κ-SIM$_{\forall\exists}$-CDP, it also provides ϵ_κ-IND-CDP.*

Proof. The first implication is by construction, the second follows by a hybrid argument. The full proof appears in the full version. □

In the section that follows, we will prove that ϵ_κ-IND-CDP ⇒ ϵ_κ-SIM$_{\forall\exists}$-CDP by giving an extension of the Dense Model Theorem of [1] (which may be of independent interest).

3.2 Dense sets and IND-CDP ⇒ SIM$_{\forall\exists}$-CDP

First, we define or recall notions of (non-uniform) density, pseudodensity, and indistinguishability for distributions, closely following [1].

Consider two distributions X and Y defined over $\mathcal{R}$, and a collection $\mathcal{A}$ of randomized predicates $A\colon \mathcal{R} \to \{0,1\}$, which may be, for instance, all circuits of size at most $s(\kappa)$, where κ is the security parameter.

We say that X is e^ϵ-*dense* in Y if

$$\forall x \in \mathcal{R} \qquad \Pr[X = x] \le e^\epsilon \cdot \Pr[Y = x].$$

We define X as δ-*indistinguishable* from Y with respect to $\mathcal{A}$ if

$$\forall A \in \mathcal{A} \qquad |\Pr[A(X) = 1] - \Pr[A(Y) = 1]| \le \delta,$$

where here and elsewhere in this section we write $A(X)$ for the distribution on the range of A obtained by applying A to the variable sampled according to X and the probability space is that of X and A's coins.

Finally, a "combination" of the two definitions is rather naturally defined as X being (e^{ϵ}, δ)-*pseudodense* in Y with respect to $\mathcal{A}$ if

$$\forall A \in \mathcal{A} \qquad \Pr[A(X) = 1] \leq e^{\epsilon} \cdot \Pr[A(Y) = 1] + \delta.$$

The connections between notions of differential privacy, SIM-CDP and IND-CDP and the above definitions are immediate:

A randomized mechanism $f \colon \mathcal{D} \to \mathcal{R}$ is ϵ_κ-DP if and only if $f(D)$ is e^{ϵ_κ}-dense in $f(D')$ for all adjacent pairs D and D', where the probability space of the distributions $f(D)$ and $f(D')$ over $\mathcal{R}$ is f's randomness.

An ensemble $\{f_\kappa\}_{\kappa \in \mathbb{N}}$ is ϵ_κ-IND-CDP if and only if there is a super-polynomial function $s(\kappa) = \kappa^{\omega(1)}$ such that for all sufficiently large κ, all adjacent pairs $D, D' \in \mathcal{D}$ of size at most $s(\kappa)$, the distribution $f_\kappa(D)$ is $(e^{\epsilon_\kappa}, \frac{1}{s(\kappa)}))$-pseudodense in $f_\kappa(D')$, with respect to the set $\mathcal{A}_\kappa$ of circuits of size at most $s(\kappa)$.

Similarly, $\{f_\kappa\}_{\kappa \in \mathbb{N}}$ is ϵ_κ-SIM-CDP if there exists an ensemble $\{F_\kappa\}_{\kappa \in \mathbb{N}}$ and a super-polynomial function $s(\kappa) = \kappa^{\omega(1)}$ such that all randomized mechanisms $F_\kappa \colon \mathcal{D} \to \mathcal{R}_\kappa$ are ϵ_κ-DP and for all sufficiently large κ, all $D \in \mathcal{D}$ of size at most $s(\kappa)$, distributions $f_\kappa(D)$ and $F_\kappa(D)$ are $\frac{1}{s(\kappa)}$-indistinguishable for the set $\mathcal{A}_\kappa$ of circuits of size at most $s(\kappa)$.

It is convenient to consider the two-sided notions of *mutually* e^{ϵ}*-dense* and *mutually* (e^{ϵ}, δ)*-pseudodense* sets, where X and Y are e^{ϵ}-dense (resp., (e^{ϵ}, δ)-pseudodense) in each other. Since the definitions of IND-CDP and SIM-CDP are symmetric in terms of the databases D and D', all relationships between distributions of f, f_κ, and F_κ on D and D' in the formulations above are, in fact, mutual.

Reingold et al. [1] showed that pseudodensity is indeed a composition of density and indistinguishability for some classes of distinguishers. One implication is immediate: If there are X, Y, and M over $\mathcal{R}$ such that M is e^{ϵ}-dense in Y and X is δ-indistinguishable from M, then X is (e^{ϵ}, δ)-pseudodense in Y (all–with respect to the same class $\mathcal{A}$ of distinguishers). The first claim of the following theorem establishes the converse (with a caveat that indistinguishability is required to hold with respect to a class of functions of slightly higher complexity, as is common in proofs by reduction). The second claim is new to our work, and is the key to relating IND-CDP and SIM$_{\forall\exists}$-CDP.

Theorem 2 *Let X and Y be distributions over a finite universe $\mathcal{R}$ such that X is (e^{ϵ}, δ)-pseudodense in Y with respect to the family $\mathcal{T}(\mathcal{A})$ defined below.*

Claim I. *There exists a distribution M over $\mathcal{R}$ such that M is e^{ϵ}-dense in Y and X is 4δ-indistinguishable from M with respect to the family $\mathcal{A}$.*

Claim II. *Furthermore, if Y is e^{ϵ}-dense in X, then it can also be guaranteed that Y is e^{ϵ}-dense in M (i.e., Y and M are mutually e^{ϵ}-dense).*

If $\mathcal{A}$ is a family of predicates, we define $\mathcal{T}(\mathcal{A})$ as the collection of functions of the following type:

$$b(x) = \begin{cases} 1 & \text{if } h_1(x) + \cdots + h_k(x) > t; \\ 0 & \text{otherwise,} \end{cases}$$

where $h_i \in \mathcal{A} \cup \bar{\mathcal{A}}$, $t \in \mathbb{N}$, and $k = O(1/\delta^2 \log(e^\epsilon/4\delta)))$. $\bar{\mathcal{A}}$ is the set of negations of $\mathcal{A}$.

Proof. **Claim I.** The proof of Claim I appears in [1], where it is stated for the case when Y is the uniform distribution (but the proof generalizes to arbitrary Y).

Claim II. Assume towards a contradiction that for any M that is mutually e^ϵ-dense in Y there is a function A_M from $\mathcal{A}$ that distinguishes it from X with probability more than $\mu = 4\delta$. Note that the same automatically holds for M that is a convex combination of distributions that are mutually e^ϵ-dense in Y, because the set of such distributions is convex. By the min-max principle of game theory, or equivalently, duality of linear programming, there exists a convex combination $\bar{b}$ of functions from $\mathcal{A} \cup \bar{\mathcal{A}}$ that distinguishes *any* such M from X:

$$\Pr[\bar{b}(X) = 1] > \Pr[\bar{b}(M) = 1] + \mu. \tag{1}$$

The function $\bar{b}$ can be viewed as a distribution over predicates in $\mathcal{A} \cup \bar{\mathcal{A}}$. Arrange elements x of $\mathcal{R}$ in the order of decreasing $\Pr[\bar{b}(x) = 1]$. Choose the set $S \subset \mathcal{R}$ as the initial part of the list so that $\Pr[Y \in S] = 1/(1+e^\epsilon)$. [6]

Define Y_S as follows:

$$\Pr[Y_S = y] = \Pr[Y = y] \cdot \begin{cases} e^\epsilon & \text{if } y \in S; \\ e^{-\epsilon} & \text{otherwise.} \end{cases}$$

It is easy to verify that Y_S is a distribution:

$$\sum_{y \in \mathcal{R}} \Pr[Y_S = y] = e^\epsilon \sum_{y \in S} \Pr[Y = y] + e^{-\epsilon} \sum_{y \notin S} \Pr[Y = s]$$
$$= e^\epsilon \frac{1}{1+e^\epsilon} + e^{-\epsilon} \frac{e^\epsilon}{1+e^\epsilon} = 1.$$

By construction Y_S and Y are e^ϵ-dense in each other, and therefore Y_S can be distinguished from X by $\bar{b}$ with probability at least μ (think of Y_S as the "hardest" distribution for $\bar{b}$ from among those that are mutually e^ϵ-dense in Y).

We make use of the following lemma proved in [20]:

Lemma 1 ([20, Claim 2.3]). *Let $F \colon X \to [0,1]$ be a bounded function, let Z and W be distributions such that $\mathbb{E}[F(Z)] \geq \mathbb{E}[F(W)] + \mu$. Then there is a real number $t \in [\mu/2, 1]$ such that*

$$\Pr[F(Z) \geq t] \geq \Pr[F(W) \geq t - \mu/2] + \mu/2.$$

[6] If exact equality cannot be achieved here, we take S to be the largest initial of the list such that $\Pr[Y \in S] < 1/(1+e^\epsilon)$, and for the next element r of the list, define $\Pr[Y_S = r] \in \left[\Pr[Y = r]e^{-\epsilon}, \Pr[Y = r]e^\epsilon\right]$ in order to make Y_S a distribution.

Applying the lemma to $F(x) = \Pr[\bar{b}(x) = 1]$, X and Y_S, there exists a real t so that a deterministic function b defined as

$$b(x) = \begin{cases} 1 & \text{if } \Pr[\bar{b}(x) = 1] \geq t + \mu/2; \\ 0 & \text{if } \Pr[\bar{b}(x) = 1] \leq t; \\ \perp & \text{otherwise,} \end{cases}$$

is such that $\Pr[b(X) = 1] > \Pr[b(Y_S) \neq 0] + \mu/2$. In other words, classifying $x \in \mathcal{R}$ as "X" when $b(x) = 1$ and "Y_S" when $b(x) = 0$ is a good distinguisher between X and Y_S (Notice that there is some slack left between $b(x) = 1$ and $b(x) = 0$).
We claim that $b(y) = 0$ for all $y \in S$. Assume the opposite. By construction of the set S, $b(y) \neq 0$ for all $y \in S$. Since Y is e^{ϵ}-dense in X (this is the only time we use this condition), for all $y \notin S$ it holds that $\Pr[Y_S = y] = e^{-\epsilon}\Pr[Y = y] \leq e^{-\epsilon}e^{\epsilon}\Pr[X = y]$, i.e., the density of X dominates the density of Y_S outside S, including the set where b is zero. Therefore

$$\Pr[b(Y_S) = 0] = \sum_{y \notin S, b(y)=0} \Pr[Y_S = y] \leq \sum_{y \notin S, b(y)=0} \Pr[X = y] = \Pr[b(X) = 0],$$

which contradicts the fact that $\Pr[b(X) = 1] > \Pr[b(Y_S) \neq 0] + \mu/2$.
Now we know that $b(y) = 0$ outside S and we conclude that

$$\Pr[b(Y) \neq 0] = \Pr[b(Y_S) \neq 0] \cdot e^{-\epsilon} < (\Pr[b(X) = 1] - \mu/2) \cdot e^{-\epsilon}.$$

That is,

$$\Pr[b(X) = 1] > e^{\epsilon} \cdot \Pr[b(Y) \neq 0] + \mu/2. \tag{2}$$

This would contradict the pseudodensity condition except that b is not part of the family of functions $\mathcal{T}(\mathcal{A})$. The following lemma approximates b with a function from $\mathcal{T}(\mathcal{A})$:

Lemma 2 ([20, Claim 2.4]). *Let $F \colon \Omega \to [0;1]$ be a convex combination of bounded functions from a class G, let Z_1, Z_2 be two distributions on Ω, and let $\alpha, \beta > 0$. Then there are functions $f_1, \dots, f_k \in G$ (not necessarily distinct) where $k = O(1/\alpha^2 \cdot \log(1/\beta))$, such that*

$$\Pr\left[\left|F(Z_i) - \frac{1}{k}(f_1(Z_i) + \cdots + f_k(Z_i))\right| > \alpha\right] \leq \beta \textit{ for } i = 1, 2.$$

We apply the lemma with parameters $\alpha = \mu/10$, $\beta = e^{-\epsilon}\mu/10$, and $F = \bar{b}$, we find an approximation to b with a function $\tilde{b}$ from $\mathcal{T}(\mathcal{A})$ with the property that

$$\Pr[\tilde{b}(X) = 1] \geq \Pr[b(X) = 1] - e^{-\epsilon}\mu/10,$$
$$\Pr[\tilde{b}(Y) = 1] \leq \Pr[b(Y) \neq 0] + e^{-\epsilon}\mu/10,$$

Combining these with equation (2) contradicts the pseudodensity of X in Y. Since we only consider predicates (0-1 functions), the threshold value t can be taken as an integer. □

Observe that if $\mathcal{A}_\kappa$ is the set of circuits of size $s(\kappa)$ for some $s(\kappa) = \kappa^{\omega(1)}$ and we take $\delta = 1/s(\kappa), \epsilon_\kappa \leq s(\kappa)$, then $\mathcal{T}(\mathcal{A})$ consists of circuits of size at most $t(\kappa) = s(\kappa)^{O(1)}$.
By applying both claims of Theorem 2, we obtain equivalence between the notions of IND-CDP and SIM$_{\forall\exists}$-CDP.

Theorem 3 *If a family of randomized mechanisms $\{f_\kappa\} \colon \mathcal{D} \to \mathcal{R}_\kappa$ is ϵ_κ-IND-CDP for $\epsilon_\kappa \in O(\log \kappa)$, it is also ϵ_κ-SIM$_{\forall\exists}$-CDP.*

Proof. If $\{f_\kappa\}$ is ϵ_κ-IND-CDP, then there is a super-polynomial function $s(\kappa) = \kappa^{\omega(1)}$ such that for all sufficiently large κ, and $D, D' \in \mathcal{D}$ of size at most $s(\kappa)$ and $|D_\kappa \Delta D'_\kappa| \leq 1$ the distribution $f_\kappa(D_\kappa)$ is $(e^{\epsilon_\kappa}, \frac{1}{s(\kappa)})$)-pseudodense in $f_\kappa(D')$, with respect to the set $\mathcal{A}_\kappa$ of circuits of size at most $s(\kappa)$. Let D, D' be adjacent data sets of size at most $s(\kappa)$. The pairs of distributions $f_\kappa(D)$ and $f_\kappa(D')$, where $f_\kappa(D)$ is $(e^{\epsilon}, 1/t(\kappa))$-pseudodense in $f_\kappa(D')$, are in situation of Claim I of Theorem 2. Therefore there exists a family of distributions $\{F_\kappa(D)\}_{\kappa\in\mathbb{N}}$ such that (a) $F_\kappa(D)$ and $f_\kappa(D)$ are $1/t(\kappa)^{\Omega(1)}$-indistinguishable for circuits of size $t(\kappa)^{\Omega(1)}$, and (b) $F_\kappa(D)$ is e^{ϵ_κ}-dense in $f_\kappa(D')$.
Since, in turn, $f_\kappa(D')$ is $(e^{\epsilon_\kappa}, 1/t(\kappa))$-pseudodense in $f_\kappa(D)$, which is indistinguishable from $F_\kappa(D)$, then $f_\kappa(D')$ is $(e^{\epsilon_\kappa}, 1/t(\kappa)^{\Omega(1)})$-pseudodense in $F_\kappa(D)$Indeed, for circuits $\{A_\kappa\}$ of size $t(\kappa)^{\Omega(1)}$, we have

$$\begin{aligned}\Pr[A_\kappa(f_\kappa(D')) = 1] \leq e^{\epsilon_\kappa} \cdot \Pr[A_\kappa(f_\kappa(D)) = 1] + 1/t(\kappa) \leq \\ e^{\epsilon_\kappa} \cdot \Big(\Pr[A_\kappa(F_\kappa(D)) = 1] + 1/t(\kappa)^{\Omega(1)}\Big) + 1/t(\kappa) \\ = e^{\epsilon_\kappa} \cdot \Pr[A_\kappa(F_\kappa(D)) = 1] + 1/t(\kappa)^{\Omega(1)},\end{aligned}$$

where the last part uses the conditions $\epsilon_\kappa = O(\log \kappa)$ and $t(\kappa) = \kappa^{\omega(1)}$.
Thus, we are in the situation of Claim II of Theorem 2 (two distributions, which are dense and pseudodense in one another). Therefore there exists a family of distributions $\{F_\kappa(D')\}_{\kappa\in\mathbb{N}}$, such that they are mutually e^{ϵ_κ}-dense in $\{F_\kappa(D)\}_{\kappa\in\mathbb{N}}$ and are $1/s(\kappa)$-indistinguishable from $\{f_\kappa(D')\}_{\kappa\in\mathbb{N}}$ by circuits of size $s(\kappa)$ for $s(\kappa) = \Big(t(\kappa)^{\Omega(1)}\Big)^{\Omega(1)} = t(\kappa)^{\Omega(1)}$.
Pictorially the proof of the theorem is represented in Figure 2. □

4 Privacy-preserving two-party computation

We now extend our definitions to the interactive case. We work in the general two-party computation setting. A motivating scenario for a two-party "private" computation involves two hospitals H_1, H_2 (holding patient records D_1, D_2) who would like to compute some statistical function $h(D_1, D_2)$. Both hospitals are concerned about the privacy of patient records, and may not be willing or even legally allowed to share data.
Differentially-private multi-party computation (MPC) was considered by Beimel et al. [21], who mainly studied the efficiency trade-offs of the

$$\begin{tikzcd}
f_\kappa(D) \arrow[r, dashed, leftrightarrow] & f_\kappa(D') & & f_\kappa(D) \arrow[r, dashed, leftrightarrow] & f_\kappa(D') \arrow[ddl] & & f_\kappa(D) \arrow[r, dashed, leftrightarrow] & f_\kappa(D') \arrow[ddl] \\
& & \xrightarrow{\text{Claim I}} & \overset{c}{\wr} & & \xrightarrow{\text{Claim II}} & \overset{c}{\wr} & \overset{c}{\wr} \\
& & & F_\kappa(D) \arrow[uur, dashed] & & & F_\kappa(D) \arrow[uur, dashed] \arrow[r, leftrightarrow] & F_\kappa(D')
\end{tikzcd}$$

Fig. 2. Schematic proof of Theorem 3. $X \dashleftarrow Y$ means X is pseudodense in Y, $X \leftarrow Y$ means X is dense in Y. Claim I of Theorem 2 is applied to the pair $f_\kappa(D)$ and $f_\kappa(D')$; Claim II is applied to the pair $F_\kappa(D)$ and $f_\kappa(D')$.

following natural paradigm for differentially-private computation of a function h: design an ϵ-DP mechanism $\widehat{h}$ that approximates h and then do secure MPC computation to obtain $\widehat{h}(D_1, D_2)$. [21] work only in the semi-honest/honest-majority models as it allows them to use information-theoretic MPC, which fits well with differential privacy.

The case of two-party computation (TWO-PC), however, is somewhat trickier as information-theoretically secure computation is impossible for generic functionalities [22]. Hence, one must resort to the computational security which interferes with the (standard) information-theoretic notion of differential privacy.

Dwork et al. [9] present a multi-party protocol run on top of a verifiable secret sharing scheme. Depending on the availability of secret channels, the protocol may only be secure against a computationally bounded adversary; however, no definition of computational differential privacy is given.

4.1 Definitions

We will now present our definitions for interactive protocols defined using *interactive functions* [23]. The reason for this choice (instead of interactive Turing machines) is that the concept of differential privacy is *orthogonal* to the choice of the computational model. In addition, many useful privacy mechanisms *may not* necessarily be efficiently computable (e.g., noise calibrated to smooth sensitivity [24] or exponential mechanisms [7]). Of course, when considering our computational definitions, we will require that the function corresponding to the *adversary* be implementable using a non-uniform PPT interactive Turing machine.

Notation. For ensembles $\{f_\kappa\}_{\kappa\in\mathbb{N}}$ and $\{g_\kappa\}_{\kappa\in\mathbb{N}}$ of randomized interactive functions f_κ, g_κ respectively, $\{\langle f_\kappa, g_\kappa\rangle\}_{\kappa\in\mathbb{N}}$ will denote the ensemble of interactive protocols defined by them. Further, in an execution $\langle f_\kappa, g_\kappa^*\rangle$ with inputs $x \in \mathcal{D}$ for the honest party, we will denote the view of the adversary (defined by interactive function g_κ^*) by $\mathtt{VIEW}_{\kappa, g_\kappa^*}(x)$.

Informally, a function ensemble $\{g_\kappa\}_{\kappa\in\mathbb{N}}$ is said to be an ensemble of *efficiently computable randomized interactive functions* if every function g_κ in the ensemble can be computed by a (non-uniform) PPT TM (a formal definition can be found in the full version). We now present our definitions. For an efficiently computable randomized interactive function g_κ, let $[g_\kappa]$ denote the binary string representing the interactive (non-uniform) Turing machine (equivalently, circuit) that implements g_κ.

Definition 6 *An ensemble* $\{\langle f_\kappa(\cdot), g_\kappa(\cdot)\rangle\}_{\kappa\in\mathbb{N}}$ *of interactive protocols,* ensures for $\{f_\kappa\}_{\kappa\in\mathbb{N}}$,

- ϵ_κ-DP, *if for every ensemble* $\{g^*_\kappa\}_{\kappa\in\mathbb{N}}$ *of randomized interactive functions, it holds that the ensemble* $\{\mathtt{VIEW}_{\kappa,g^*_\kappa}(x)\}_{\kappa\in\mathbb{N}}$ *provides* ϵ_κ-DP *with respect to* $x \in \mathcal{D}$.
- ϵ_κ-IND-CDP, *if for every ensemble* $\{g^*_\kappa\}_{\kappa\in\mathbb{N}}$ *of* efficiently computable *randomized interactive functions, and all sufficiently large* κ*, it holds that the ensemble* $\{\mathtt{VIEW}_{\kappa,g^*_\kappa}(x)\}_{\kappa\in\mathbb{N}}$ *provides* ϵ_κ-IND-CDP *(as per definition 3) with respect to* $x \in \mathcal{D}$.
- ϵ_κ-SIM-CDP, *if for every ensemble* $\{g^*_\kappa\}_{\kappa\in\mathbb{N}}$ *of* efficiently computable *randomized interactive functions, there exists an ensemble* $\{F_\kappa\}_{\kappa\in\mathbb{N}}$ *of* ϵ_κ-differentially-private mechanisms $F_\kappa(\cdot)$ *such that for every* $x \in \mathcal{D}$*, the probability ensembles* $\{\mathtt{VIEW}_{\kappa,g^*_\kappa}(x)\}_{\kappa\in\mathbb{N}}$ *and* $\{F_\kappa(x)\}_{\kappa\in\mathbb{N}}$ *are computationally indistinguishable.*

All three notions are defined symmetrically for the other ensemble $\{g_\kappa\}_{\kappa\in\mathbb{N}}$.

A protocol should be "useful" in some sense (analogous to correctness property of standard TWO-PC protocols). For example, a TWO-PC protocol (output denoted by $\widehat{h}(x, y)$ for inputs x, y) for computing the Hamming distance $h(x, y)$ is (γ, ξ)-additive-useful [25, 21, 6] if and only if $\Pr[|h(x) - \widehat{h}(x)| > \gamma(\kappa)] \leq \xi(\kappa)$. We define and work with a somewhat more general notion, (s, ξ)-usefulness with respect to a predicate P, which can be found in the full version. We can now define privacy-preserving TWO-PC.

Definition 7 (Privacy-preserving two-party computation) *An interactive protocol ensemble* $\{\langle f_\kappa(\cdot), g_\kappa(\cdot)\rangle\}_{\kappa\in\mathbb{N}}$ *is* (s, ξ) ϵ_κ-type private two-party computation protocol for $h = (h_f, h_g)$ with respect to P *if for* both f_κ, g_κ*, the ensemble ensures* ϵ-type *and provides* (ξ, s)*-usefulness for* f_κ *with respect to predicate* P*, where* $\mathsf{type} \in \{$DP, IND-CDP, SIM-CDP$\}$.

There is rich literature considering notions of security for MPC/TWO-PC simulation-based security [26, 27], super-polynomial simulation [28, 29], input indistinguishable computation [30], etc. Our notions of private TWO-PC (definition 7) can be seen as new notions of "security" where the only concern for the parties is the privacy of their inputs—here the notion of privacy being (computational) differential privacy. As these notions do not demand efficient simulation (note that even in ϵ_κ-IND-CDP, we do not require the "ideal" F_κ is efficiently computable), they may be easier to achieve; though accuracy may now be the difficult dimension of this aspect.

We note that although our presentation is only for TWO-PC, an extension to MPC is straightforward.

"Ideal/Real" Style Definition of Privacy: ϵ-SIM$^+$-CDP. We now present a new definition, ϵ_κ-SIM$^+$-CDP, which is of particular interest in the context of interactive TWO-PC (for the non-interactive case, it reduces to ϵ_κ-SIM-CDP). This definition is inspired from the "ideal/real" paradigm style definitions used for defining secure TWO-PC/MPC (see [31, 32]).
Let P_1, P_2 be two parties, with private inputs a, b respectively, who would like to compute a function $h(a, b)$. What would be the "ideal" situation for the two parties? If there were a trusted third party $\mathcal{T}$ available, P_1, P_2 could first fix a ϵ-*differentially-private* mechanism $\widehat{h}$ that would be "useful" for approximating h according to some metric, and then hand over their inputs a, b to $\mathcal{T}$, who could then compute $\widehat{h}(a, b)$ with uniformly chosen randomness and provide the output to both the parties. Clearly, this informally described "ideal process" (which is literally known as IDEAL world, in secure MPC literature) provides ϵ-DP. Thus, if we had a secure TWO-PC protocol π that emulates this ideal world for all PPT adversaries, intuitively π would "look differentially private" to these adversaries. Moreover, since $\widehat{h}$ is (information-theoretically) differentially private, we can use any π proven "secure" by simulation and privacy is intuitively maintained even if the simulation is not efficient.
We now present the formal definition. In what follows, we assume familiarity with "ideal/real"-paradigm. For a complete and formal description of IDEAL and REAL experiments, we refer the reader to standard texts (e.g., [31, 32]). Our definition differs from the standard definition, solely in the sense that the simulator is not necessarily efficient. Though clear from the definitions, we point out that we are working with the *static* corruption model. To maintain consistency with our notation, our definitions are described via interactive function ensembles.

Definition 8 (SIM$^+$-CDP private two-party computation) *An interactive protocol ensemble* $\{\langle f_\kappa(\cdot), g_\kappa(\cdot)\rangle\}_{\kappa\in\mathbb{N}}$ *is* (s, ξ) ϵ_κ-SIM$^+$-CDP private two-party computation protocol *for* $h = (h_f, h_g)$ *with respect to the predicate* P *if there exists an* ϵ_κ-DP *randomized mechanism* $\widehat{h} = (\widehat{h}_f, \widehat{h}_g)$ *such that*

- *Mechanism* $\widehat{h}$ *provides* (s, ξ)*-usefulness for* h *with respect to the predicate* P*.*
- *The protocol ensemble* $\{\langle f_\kappa(\cdot), g_\kappa(\cdot)\rangle\}_{\kappa\in\mathbb{N}}$ *is a secure two-party computation protocol ensemble for the randomized functionality* $\widehat{h}$ *as per the "ideal/real"-style definition of secure two party computation. (see full version)*

Clearly, the definitions of SIM-CDP and SIM$^+$-CDP are similar in asserting the existence of simulators whose output is computationally indistinguishable from the real world's transcripts. The difference between the definitions is that in the former the simulator is restricted to being differentially private but otherwise has unfettered access to the input, while the latter only has access to a differentially-private output, from which it has to reconstruct the entire view. The proofs of the following two theorems are provided in the full version.

Theorem 4 *1. SIM$^+$-CDP $\Rightarrow$ SIM-CDP* If a protocol ensemble $\{\langle f_\kappa(\cdot), g_\kappa(\cdot)\rangle\}_{\kappa\in\mathbb{N}}$ satisfies definition 8, then it also satisfies definition 7 for ***type***=SIM-CDP.
2. SIM-CDP $\not\Rightarrow$-SIM$^+$-CDP] There exists a protocol ensemble $\{\langle f_\kappa(\cdot), g_\kappa(\cdot)\rangle\}_{\kappa\in\mathbb{N}}$ and a function $h(\cdot,\cdot)$ such that the protocol ensemble $\{\langle f_\kappa(\cdot), g_\kappa(\cdot)\rangle\}_{\kappa\in\mathbb{N}}$ is a $(0,0)$-additive-useful private two-party computation protocol for h (computing h exactly), that provides ϵ_κ-SIM-CDP but not ϵ_κ-SIM$^+$-CDP.

Protocols: Private two-party computation of the Hamming distance. We now demonstrate the usefulness of our definitions by constructing a *simple* and *efficient* protocol which allows two parties to compute the Hamming distance between their respective inputs in just *two* rounds. This protocol will demonstrate the flexibility that comes with our definitions for designing "privacy" protocols. Note that our "privacy" definitions makes the problem quite different from the work on private set intersection protocols (see, for example [33] and the references therein).

For vectors $a, b \in \{0,1\}^n = \mathcal{D}$ define the *Hamming distance*, denoted $h(a,b)$, to be the number of positions in which a, b differ (equivalently, the vectors can be associated with subsets of an n-element universe). Then using *additive homomorphic encryption* and the transformation by [25], we can construct a protocol private TWO-PC for approximating $h(a,b)$ more efficient than generic constructions. Due to space constraints, this protocol π_h (semi-honest and malicious versions) along with the proof of following theorem, is presented in full version.

Theorem 5 *For every $0 < \epsilon < 1$ there exists a $\gamma \in O(1)$ and a constant ξ (depending only on ϵ, γ) such that for every $a, b \in \{0,1\}^n$ it holds that the output $\widehat{h}$ of the protocol π_h satisfies the following*

$$\Pr[|h^* - \widehat{h}| \geq \gamma] \leq \xi,$$

where $h^* = h(a,b)$ and the probability is taken over the randomness of π_h. Further, π_h provides ϵ_κ-SIM-CDP.

Note that NO protocol which satisfies ϵ-DP with above accuracy guarantee on additive error is known so far. For small *multiplicative* error, however, in the full version of this paper, we present a protocol, π_h^*, which ensures ϵ-DP. We also prove the following theorem there:

Theorem 6 *For every $0 < \epsilon, \gamma, \eta < 1$, there exists $\delta \in \Theta(1 - e^{-\gamma/2})$ such that for every $a, b \in \{0,1\}^n$ satisfying $h(a,b) \in \omega\left(\frac{1}{\epsilon^2\delta^2}(\ln(5n/\eta))^4\right)$, it holds that the output $\widehat{h}$ of the protocol π_h^* satisfies the following*

$$\Pr[h^* \leq \widehat{h} \leq (1+\gamma)h^*] \geq 1 - \eta,$$

where $h^* = h(a,b)$ and the probability is taken over the randomness of π_h^*. Further, π_h^* provides ϵ-DP.

The high level idea behind π_h^* is to construct a differentially-private version of the communication-efficient KOR algorithm based on sketches [34]. This is done by properly sanitizing the sketches using a standard randomized response mechanism, which has additive error $\Theta(\sqrt{n})$. The two error sources (intrinsic to KOR and due to randomized response) cannot vanish simultaneously, restricting the minimal value of $h(a,b)$ for which π_h^*'s usefulness guarantee hold (its privacy guarantees are always preserved).

We clarify that although differentially-private protocols compute an approximation to the actual function h, they should not be confused with the *fundamentally different* line of research on "secure approximations" (introduced by Feigenbaum et al. [35]). Due to space constraints, this discussion, along with future research directions is deferred to the full version.

Acknowledgements. We would like to thank Cynthia Dwork, Thomas Holenstein, Frank McSherry, Kunal Talwar, and Adam Smith for inspiring discussions and the anonymous reviewers for helpful comments.

References

1. Reingold, O., Trevisan, L., Tulsiani, M., Vadhan, S.: Dense subsets of pseudorandom sets. In: FOCS 2008
2. Dwork, C.: Differential privacy: A survey of results. In: Theory and Applications of Models of Computation, TAMC 2008. Volume 4978 of Lecture Notes in Computer Science., Springer 1–19
3. Dwork, C.: Differential privacy. Invited talk. In: ICALP (2). (2006)
4. Dwork, C., Nissim, K.: Privacy-preserving datamining on vertically partitioned databases. In: CRYPTO 2004
5. Barak, B., Chaudhuri, K., Dwork, C., Kale, S., McSherry, F., Talwar, K.: Privacy, accuracy, and consistency too: A holistic solution to contingency table release. In: Principles Of Database Systems 2007. 273–282
6. Blum, A., Ligett, K., Roth, A.: A learning theory approach to non-interactive database privacy. In: STOC 2008. 609–618
7. McSherry, F., Talwar, K.: Mechanism design via differential privacy. In: FOCS, IEEE Computer Society (2007) 94–103
8. Feldman, D., Fiat, A., Kaplan, H., Nissim, K.: Private coresets. In: STOC. (2009) To appear.
9. Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., Naor, M.: Our data, ourselves: Privacy via distributed noise generation. In: EUROCRYPT 2006. 486–503
10. Green, B., Tao, T.: The primes contain arbitrarily long arithmetic progressions. pre-print arXiv:math/0404188 [math.NT] (April 2004)
11. Tao, T., Ziegler, T.: The primes contain arbitrarily long polynomial progressions. pre-print arXiv:math/0404188 [math.NT] (October 2006)
12. Barak, B., Shaltiel, R., Wigderson, A.: Computational analogues of entropy. In: RANDOM-APPROX. (2003) 200–215
13. Reingold, O., Vadhan, S. Personal Communication.
14. Agrawal, R., Evfimievski, A.V., Srikant, R.: Information sharing across private databases. In: ACM SIGMOD Conference. (2003) 86–97
15. Wright, R.N., Yang, Z.: Privacy-preserving Bayesian network structure computation on distributed heterogeneous data. In: KDD. (2004) 713–718

16. Goethals, B., Laur, S., Lipmaa, H., Mielikäinen, T.: On private scalar product computation for privacy-preserving data mining. In: ICISC. (2004) 104–120
17. Freedman, M.J., Nissim, K., Pinkas, B.: Efficient private matching and set intersection. In: EUROCRYPT. (2004) 1–19
18. Kissner, L., Song, D.X.: Privacy-preserving set operations. In: CRYPTO. (2005) 241–257
19. McSherry, F.: Privacy integrated queries. In: ACM SIGMOD 2009
20. Reingold, O., Trevisan, L., Tulsiani, M., Vadhan, S.: Dense subsets of pseudorandom sets. Electronic Colloquium on Computational Complexity (ECCC) (045) (2008)
21. Beimel, A., Nissim, K., Omri, E.: Distributed private data analysis: Simultaneously solving how and what. In: CRYPTO 2008. 451–468
22. Chor, B., Kushilevitz, E.: A zero-one law for boolean privacy. SIAM J. Discrete Math. **4**(1) (1991) 36–47
23. Goldwasser, S., Sipser, M.: Private coins versus public coins in interactive proof systems. In: STOC, ACM (1986) 59–68
24. Nissim, K., Raskhodnikova, S., Smith, A.: Smooth sensitivity and sampling in private data analysis. In: STOC. (2007) 75–84
25. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: TCC. (2006) 265–284
26. Yao, A.C.C.: Protocols for secure computations (extended abstract). In: FOCS'82. 160–164
27. Goldreich, O., Micali, S., Wigderson, A.: How to play any mental game or A completeness theorem for protocols with honest majority. In: STOC 1987. 218–229
28. Pass, R.: Simulation in quasi-polynomial time, and its application to protocol composition. In Biham, E., ed.: Advances in Cryptology—EUROCRYPT 2003. Volume 2656 of Lecture Notes in Computer Science., Springer (2003) 160–176
29. Prabhakaran, M., Sahai, A.: New notions of security: achieving universal composability without trusted setup. In: STOC 2004. 242–251
30. Micali, S., Pass, R., Rosen, A.: Input-indistinguishable computation. In: FOCS 2006. 367–378
31. Goldreich, O.: Secure Multiparty Computation. Manuscript, Preliminary Version, 1998 Available from http://www.wisdom.weizmann.ac.il/ oded/pp.html.
32. Canetti, R.: Security and composition of multiparty cryptographic protocols. J. Cryptology **13**(1) (2000) 143–202
33. Camenisch, J., Zaverucha, G.M.: Private intersection of certified sets. In: Financial Cryptography and Data Security. (2009) To appear.
34. Kushilevitz, E., Ostrovsky, R., Rabani, Y.: Efficient search for approximate nearest neighbor in high dimensional spaces. In: STOC'98. 614–623
35. Feigenbaum, J., Ishai, Y., Malkin, T., Nissim, K., Strauss, M., Wright, R.N.: Secure multiparty computation of approximations. In: ICALP