2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)
Allerton Park and Retreat Center
Monticello, IL, USA, October 2-5, 2018

PROPERTY TESTING FOR DIFFERENTIAL PRIVACY

ANNA C. GILBERT AND AUDRA MCMILLAN

ABSTRACT. We consider the problem of property testing for differential privacy: with black-box access to a purportedly private algorithm, can we verify its privacy guarantees? In particular, we show that *any privacy guarantee that can be efficiently verified is also efficiently breakable* in the sense that there exist two databases between which we can efficiently distinguish. We give lower bounds on the query complexity of verifying pure differential privacy, approximate differential privacy, random pure differential privacy, and random approximate differential privacy. We also give algorithmic upper bounds. The lower bounds obtained in the work are infeasible for the scale of parameters that are typically considered reasonable in the differential privacy literature, even when we suppose that the verifier has access to an (untrusted) description of the algorithm. A central message of this work is that verifying privacy requires compromise by either the verifier or the algorithm owner. Either the verifier has to be satisfied with a weak privacy guarantee, or the algorithm owner has to compromise on side information or access to the algorithm.

1. INTRODUCTION

Recently, differential privacy (DP) has gained traction outside of theoretical research as several companies (Google, Apple, Microsoft, Census, etc.) have announced deployment of large-scale differentially private mechanisms (19, 3, 1, 12). This use of DP, while exciting, might be construed as a marketing tool used to encourage privacy-aware consumers to release more of their sensitive data to the company. In addition, the software behind the deployment of DP is typically proprietary since it ostensibly provides commercial advantage. This raises the question: with limited access to the software, can we verify the privacy guarantees of purportedly DP algorithms?

Suppose there exists some randomised algorithm $\mathcal{A}$ that is claimed to be Ξ- differentially private and we are given query access to $\mathcal{A}$. That is, the domain of $\mathcal{A}$ is the set of databases and we have the power to choose a database D and obtain a (randomised) response $\mathcal{A}(D)$. How many queries are required to verify the privacy guarantee? We formulate this problem in the property testing framework for pure DP, approximate DP, random pure DP, and random approximate DP.

Definition 1 (Property testing with side information)**.** A property testing algorithm with query complexity q, proximity parameter α, privacy parameters Ξ and side information S, makes q queries to the black-box and:

(1) (Completeness) ACCEPTS with probability at least $2/3$ if $\mathcal{A}$ is Ξ-private and S is accurate.
(2) (Soundness) REJECTS with probability at least $2/3$ if $\mathcal{A}$ is α-far from being Ξ-private.

In this early stage of commercial DP algorithms, approaches to transparency have been varied. For some algorithms, like Google's RAPPOR, a full description of the algorithm has been released (19). On the other hand, while Apple has released a white paper (4) and a patent (28), there are still many questions about their exact implementations. We focus on the two extreme settings: when we are given *no information* about the black-box (except the domain and range), and the *full information* setting where we have an untrusted full description of the algorithm $\mathcal{A}$.

Both settings are subject to fundamental limitations. We first show that *verifying* privacy is at least as difficult as *breaking* privacy, even in the full information setting. That is, suppose r samples are sufficient to verify that an algorithm is Ξ-private. Then Theorem 6 implies that for every algorithm that is not Ξ-private, there exists some pair of neighbouring databases D and D' such that r samples from $\mathcal{A}(D)$ is enough to distinguish between D and D'. Differential privacy is designed so that this latter problem requires a large number of samples. This connection has the unfortunate implication that verifiability and privacy are directly at odds: *if a privacy guarantee is efficiently verifiable, then it mustn't be a strong privacy guarantee*.

For the remainder of this work we restrict to discrete distributions on $[n]$. Our upper and lower bounds in each setting are contained in Table 1. We rule out

978-1-5386-6596-1/18/$31.00 ©2018 IEEE

TABLE 1. Per Database Query complexity bounds for property testing of privacy. Bracketed numbers refer to the Theorem number.

	No Information	Full information
pDP	Unverifiable [10]	$\Omega\left(\frac{1}{\beta\alpha^2}\right)$ [16]
		$O\left(\frac{\ln n}{\alpha^2\beta^2}\right)$ [15]
aDP	$\Omega(\max\{n^{1-o(1)}, \frac{1}{\alpha^2}\})$ [12]	$O\left(\frac{\sqrt{n}}{\alpha^2}\right)$[14]
	$O(\frac{n}{\alpha^2})$ [13]	

sublinear verification of privacy in every case except verifying approximate differential privacy in the full information setting. That is, for all other definitions of privacy, the query complexity for property testing of privacy is $\Omega(n)$.

Each privacy notion we consider is a relaxation of pure differential privacy. Generally, the privacy is relaxed in one of two ways: either privacy loss is allowed to occur on unlikely *outputs*, or privacy loss is allowed to occur on unlikely *inputs*. The results in Theorem 8 and the lower bounds in Table 1 imply that for efficient verification, we need to relax in *both* directions. That is, random approximate DP is the only efficiently verifiable privacy notion in the no information setting. Even then, we need about $1/\delta^2$ queries *per database* to verify (ϵ, δ)-approximate differential privacy. Theorem 13 shows that random approximate DP can be verified in (roughly) $O(\frac{4n(1+e^{2\epsilon})\log(1/\gamma)}{\gamma\delta^2})$ samples, where (roughly) δ and γ are the probabilities of choosing a disclosive output or input, respectively. This means verification is efficient if δ and γ are small but not too small. This may seem insufficient to those familiar with DP, where common wisdom decrees that δ and γ should be small enough that this query complexity is infeasibly large.

There have been several other relaxations of pure differential privacy proposed in the literature, chief among them Rényi DP (24) and concentrated DP (18). Many of the results for pure differential privacy in this work can be easily extended to Rényi and concentrated DP. We leave these out of our discussion for brevity.

One might hope to obtain significantly lower query complexity if the property tester algorithm is given side information, even if the side information is untrusted. We find that this is true for both approximate DP and pure DP, if we allow the query complexity to depend on the side information. A randomised algorithm $\mathcal{A}$ can be abstracted as a set of distributions $\{P_D\}$ where $\mathcal{A}(D) \sim P_D$. We obtain a sublinear verifier for approximate DP. For pure DP, we find the quantity that controls the query complexity is

$$\beta = \inf_D \min_i P_D(i).$$

If β is large then efficient verification is possible: verifying that the pure differential privacy parameter is less than $\epsilon + \alpha$ requires $O(\frac{\ln n}{\alpha^2\beta^2})$ queries of each database (Theorem 15). Note that this is not sublinear since $\beta \leq 1/n$ and if $\beta = 0$ then we have no improvement on the no information setting. However, for reasonable β, this is a considerable improvement on the no information lower bounds and may be efficient for reasonable n.

A central theme of this work is that verifying the privacy guarantees that corporations (or any entity entrusted with private data) claim requires compromise by either the verifier or algorithm owner. If the verifier is satisfied with only a weak privacy guarantee (random approximate DP with δ and γ small but not extremely small), then they can achieve this with no side information from the algorithm owner. If the company is willing to compromise by providing information about the algorithm up-front, then much stronger privacy guarantees can be verified. Given this level of transparency, one might be tempted to suggest that the company provide source code instead. While verifying privacy given source code is an important and active area of research, there are many scenarios where the source code itself is proprietary. We have already seen instances where companies have been willing to provide detailed descriptions of their algorithms. In the full information case, we obtain our lowest sample complexity algorithms, including a sublinear algorithm for verifying approximate differential privacy.

This paper proceeds as follows: we start by defining property testing for privacy in Section 2. We then proceed to the main contributions of this work:

- Verifying privacy is as hard as breaking privacy (Section 3).
- In the no information setting, verifying pure differential privacy is impossible while there is a finite query complexity property tester for approximate differential privacy (Section 5).

- If $\beta > 0$, then finite query complexity property testers exist for pure differential privacy in the full information setting (Section 6).
- A sublinear property tester exists for approximate differential privacy in the full information setting.

The main lower bounds and algorithmic upper bounds in this paper are summarized in Table 1. An extended version of this work can be found at (21).

2. Background and Problem Formulation

A database is a vector D in $\mathbb{Z}^{|\Omega|}$ for some data universe Ω. That is, if $\omega \in \Omega$, D_ω is the number of copies of ω in the database. We call two databases D, D' *neighbouring* if they differ on a single data point, that is $\|D - D'\|_1 = 1$. For a randomised algorithm $\mathcal{A}$ and database D, we use $\mathcal{A}(D)$ to denote the output and P_D to denote the distribution of $\mathcal{A}(D)$. We will often prefer to view an algorithm as simply a collection of distributions $\{P_D \mid D \in \mathbb{Z}^{|\Omega|}\}$. We will only consider discrete distributions in this paper, so P_D is a discrete distribution on $[n] = \{1, \cdots, n\}$. For a distribution P, P^r represents r independent copies of P.

For much of this paper we will consider algorithms that accept only two databases as input. We use the notation $\mathcal{A} = (P_0, P_1)$ to denote such an algorithm that accepts only two databases 0 and 1 as input, and $\mathcal{A}(0) \sim P_0$ and $\mathcal{A}(1) \sim P_1$. The databases 0 and 1 are assumed to be neighbouring.

The privacy notions we discuss will all center around the idea that P_D and $P_{D'}$ should be close for neighbouring databases D and D'. As such, we will deal with many measures of closeness between distributions. We collect these definitions for ease of reference.

Definition 2. Let P and Q be two distributions.

- *(Max divergence)* $D_\infty(P, Q) = \sup_E \ln \frac{P(E)}{Q(E)}$.
- *(δ-approximate max divergence)* $D^\delta_\infty(P, Q) = \sup_{E \text{ s.t. } P(E) \geq \delta} \ln \frac{P(E)-\delta}{Q(E)}$.
- *(KL divergence)* $D_{KL}(P\|Q) = \int_R P(x) \ln \frac{P(x)}{Q(x)} dx$.
- *(Total Variance (TV) distance)* $\|P - Q\|_{\text{TV}} = \sup_E |P(E) - Q(E)|$.

where the $\sup_E$ is the supremum over all events E in the outcome space.

2.1. Privacy Definitions.

Pure differential privacy is the gold standard for privacy-preserving data analysis. However, it is a very strong definition and as a result, many relaxations of it have gained traction as the work on differential privacy evolves. These relaxations find various ways to sacrifice privacy, with a view towards allowing a strictly broader class of algorithms to be implemented. Since these definitions are becoming standard, we give only a cursory introduction in this section. An introduction can be found in (17) and more in depth surveys can be found in (30, 15, 22).

The idea is simple; suppose the adversary has narrowed the list of possible databases down to neighbouring databases D and D'. Any output the adversary sees is almost equally as likely to have arisen from P_D or $P_{D'}$. Thus, the adversary gains almost no information that helps them distinguish between D and D'.

Definition 3 (Data Distribution Independent Privacy Definitions)**.** A randomised algorithm $\mathcal{A}$ is

- *ϵ-pure differentially private (pDP)* if $\sup_{D,D'} D_\infty(P_D, P_{D'}) \leq \epsilon$.
- *(ϵ, δ)-approximate differentially private (aDP)* if $\sup_{D,D'} D^\delta_\infty(P_D, P_{D'}) \leq \epsilon$.

where the supremums are over all pairs of neigbouring databases D and D'.

Note that ϵ-pDP is exactly $(\epsilon, 0)$-aDP. The parameter δ can be thought of as our probability of failing to preserve privacy. To see this, suppose the distributions P_D output 0 with probability $1-\delta$, and a unique identifier for the database D with probability δ. Then this algorithm is $(0, \delta)$-DP. Thus, we typically want δ to be small enough that we can almost guarantee that we will not observe this difference in the distributions. In contrast, while it is desirable to have ϵ small, a larger ϵ still gives meaningful guarantees ((16)). Typically one should think of δ as *extremely* small, $\delta \approx 10^{-8}$, and ϵ as *quite* small, $\epsilon \approx 0.1$.

Let $\mathcal{D}$ be a distribution on the data universe Ω. For a database D and datapoint z, let $[D_{-1}, z]$ denote the neighbouring database where the first datapoint of D is replaced by z.

Definition 4 (Data Distribution Dependent Privacy Definitons)**.** An algorithm $\mathcal{A}$ is

- *(ϵ, γ)-Random pure differentially private (RpDP)* if $\mathbb{P}\left(D_\infty(P_D, P_{[D_{-1},z]}) \leq \epsilon\right) \geq 1 - \gamma$.
- *$(\epsilon, \delta, \gamma)$-Random approximate differentially private (RADP)* if $\mathbb{P}\left(D^\delta_\infty(P_D, P_{[D_{-1},z]}) \leq \epsilon\right) \geq 1 - \gamma$.

where the probabilities in RpDP and RaDP are over $D \sim \mathcal{D}^n, z \sim \mathcal{D}$.

TABLE 2. Privacy notions, parameters and metrics.

Privacy Notion	Ξ	$\|\Xi - \Xi'\|$
pDP	ϵ	$\|\epsilon - \epsilon'\|$
aDP	(ϵ, δ)	$\|\delta - \delta'\|$
RpDP	(ϵ, γ)	$\min\{\|\epsilon - \epsilon'\|, \lambda\|\gamma - \gamma'\|\}$
RaDP	$(\epsilon, \delta, \gamma)$	$\min\{\|\delta_\epsilon - \delta'_\epsilon\|, \lambda\|\gamma - \gamma'\|\}$

Similar to δ, γ represents the probability of catastrophic failure in privacy. Therefore, we require that γ is small enough that this event is extremely unlikely to occur.

2.2. **Problem Formulation.** Our goal is to answer the question *given these privacy parameters, is the algorithm $\mathcal{A}$ at least Ξ-private?* where Ξ is an appropriate privacy parameter. A property testing algorithm, which outputs ACCEPT or REJECT, answers this question if it ACCEPTS whenever $\mathcal{A}$ is Ξ-private, and *only* ACCEPTS if the algorithm $\mathcal{A}$ is close to being Ξ-private. A tester with side information may also REJECT simply because the side information is inaccurate.

We say that $\mathcal{A}$ is α-far from being Ξ-private if $\min_{\Xi'} \|\Xi' - \Xi\| > \alpha$, where the minimum is over all Ξ' such that $\mathcal{A}$ is Ξ'-private. The metrics used for each form of privacy are contained in Table 2. We introduce the scalar λ to penalise deviation in one parameter more than deviation in another parameter. For example, it is much worse to mistake a $(0, 0.1)$-RpDP algorithm for $(0, 0)$-RpDP than it is to mistake a $(0.1, 0)$-RpDP algorithm for $(0, 0)$-RpDP. We leave the question of *how much worse* as a parameter of the problem. However, we give the general guideline that if we want an α error to be tolerable in both ϵ and γ then $\lambda \approx \frac{\epsilon}{\gamma}$, which may be large, is an appropriate choice.

The formal definition of a property tester with side information was given in Definition 1. A *no information* property tester is the special case when $S = \emptyset$. A *full information* property tester is the special case when $S = \{Q_D\}$ contains a distribution Q_D for each database D. We use Q_D to denote the distribution on outputs presented in the side information and P_D to denote the true distribution on outputs of the algorithm being tested. For $\alpha > 0$ and privacy parameter Ξ, a full information (FI) property tester for this problem satisfies:

(1) (Completeness) Accepts with probability at least $2/3$ if the algorithm is Ξ-private and $P_D = Q_D$ for all D.
(2) (Soundness) Rejects with probability at least $2/3$ if the algorithm is α-far from being Ξ-private.

We only force the property tester to ACCEPT if the side information is *exactly* accurate ($P_D = Q_D$). It is an interesting question to consider a property tester that is forced to ACCEPT if the side-information is *close* to accurate, for example in TV-distance. We do not consider this in this work as being close in TV-distance does not imply closeness of privacy parameters.

For a database D, we will refer to the process of obtaining a sample from P_D as *querying the black-box*. It will usually be necessary to input each database into the black-box multiple times. We will use m to denote the number of *unique* databases that are queries to the black-box and r to denote the number of times each database is input. We will only consider algorithms where the number of samples from P_D for each input database is r, so our query complexity is mr for each algorithm. Our aim is verify the privacy parameters using as few queries as possible.

2.3. **Related Work.** This work connects to two main bodies of literature. There are several works on verifying privacy with different access models that share the same motivation as this work. In terms of techniques, our work is most closely linked to recent work on property testing of distributions. Several algorithms and tools have been proposed for formal verification of the DP guarantee of an algorithm (6, 27, 26, 20, 29, 6, 27, 26, 20, 29, 7, 5, 23). Much of this work focuses on verifying privacy given access to a description of the algorithm. These tools are aimed at automatic (or simplified) verification of privacy of source code.

Given sample access to two distributions P and Q and a distance measure $d(\cdot, \cdot)$, the question of distinguishing between $d(P, Q) \leq a$ and $d(P, Q) \geq b$ is called *tolerant property testing*. This question is closely related to the question of whether $\mathcal{A} = (P, Q)$ is private. There is a large body of work exploring lower bounds and algorithmic upper bounds for tolerant testing using standard distances (TV, KL, χ^2, etc.) with both $a = 0$ and $a > 0$ (11, 25, 8, 2, 31). In our work, we draw most directly from the techniques of Valiant et. al (32).

A relevant paper in the intersection of privacy and property testing is Dixit et. al (14). The access model in this paper is different to ours but their goal is similar. Recent work by Ding et. al (13) studies privacy verification from a hypothesis testing perspective. They design a privacy verification algorithm which aims to find violations of the privacy guarantee. Their algorithm provides promising experimental results in non-adversarial settings (when the privacy guarantee is frequently violated), although they provide no theoretical guarantees.

3. Lower Bounds via Distinguishability

We now turn to examining the fundamental limitations of property testing for privacy. We find that even in the full information setting, the query complexity to verifying privacy is lower bounded by the number of queries required to distinguish between two possible inputs. We expect the latter to increase with the strength of the privacy guarantee.

Definition 5. Databases D and D' are *r-distinguishable under $\mathcal{A}$* if there exists a testing algorithm such that given a description of $\mathcal{A}$ and $x \sim P^r_{D''}$ where $D'' \in \{D, D'\}$, it accepts with probability at least $2/3$ if $D'' = D$ and rejects with probability at least 2/3 if $D'' = D'$.

The following theorem says that the per database query complexity of a privacy property testing algorithm is lower bounded by the minimal r such that two neighbouring databases are r-distinguishable under $\mathcal{A}$. Recall that we use the notation $\mathcal{A} = (P_0, P_1)$ to denote an algorithm that accepts only two databases 0 and 1 as input, and $\mathcal{A}(0) \sim P_0$ and $\mathcal{A}(1) \sim P_1$. The databases 0 and 1 are assumed to be neighbouring.

Theorem 6. *Consider any privacy definition, privacy parameter Ξ, and let $\alpha > 0$. Suppose there exists a Ξ-privacy property tester with proximity parameter α and (per database) query complexity r. Let $\mathcal{A}$ be an algorithm that is α-far from Ξ-private. If the privacy notion is*

- *pDP or aDP then there exists a pair of neighbouring databases that are r-distinguishable under $\mathcal{A}$.*
- *RpDP or RaDP and $\Xi = (\epsilon, \delta, \gamma)$, then a randomly sampled pair of neighbouring databases has probability at least $\gamma + \frac{\alpha}{\lambda}$ of being r-distinguishable.*

Typically we want the privacy parameters to start small enough that k has to be quite large before any pair of neighbouring databases can be distinguished between using the output from k iterations of the algorithm. In many of our proofs in the following sections proceed by finding two distribution P and Q such that (P, Q) has high privacy parameters (not very private) but it is still difficult to distinguish between P and Q (for example because they only differ on a set with small measure). The proof of Theorem 6 can be found in the (21). It proceeds by relating distinguishing between $D'' = D$ or $D'' = D'$ to determining whether $(P_{D''}, P_{D'})$ is 0-DP.

4. Restriction to Two Distribution Setting

Differential privacy is an inherently local property. That is, verifying that $\mathcal{A}$ is Ξ-private means verifying that $(P_D, P_{D'})$ is Ξ-private, either always or with high probability, for pairs of neighbouring databases D and D'. We refer to the problem of determining whether a pair of distributions (P_0, P_1) satisfies Ξ-privacy as the *two database setting*. We argue in this section that the hard part of privacy property testing is the two database setting. For this reason, from Section 5 onwards, we only consider the two database setting.

An algorithm is non-adaptive if it chooses m pairs of distributions and queries the blackbox with each database r times. It does not choose its queries adaptively.

Theorem 7 (Conversion to random privacy tester)**.** *If there exists a Ξ-privacy property tester for the two database setting with query complexity r per database and proximity parameter α, then there exists a privacy property tester for (Ξ, γ)-random privacy with proximity parameter 2α and query complexity*

$$O\left(r \log\left(\frac{2\lambda}{\alpha}\right) \frac{(\alpha/\lambda + \gamma)^2 + \alpha/\lambda}{(\alpha/\lambda)^2}\right).$$

Theorem 7 is proved by way of an algorithm that converts a two distribution property tester to a random DP property tester. The algorithm (Algorithm **??**), along with soundness and completeness proofs can be found in the (21). Notice that if $\gamma \approx \frac{\alpha}{\lambda}$ then the query complexity is approximately $r \log(\frac{\lambda}{\alpha}) \frac{\lambda}{\alpha} \approx \frac{r \log(\frac{1}{\gamma})}{\gamma}$. One shortcoming of the conversion algorithm in Theorem 7 is that we need to know the data distribution $\mathcal{D}$. We can relax to an approximation $\mathcal{D}'$ that is close in TV-distance, but it is not difficult to see that $\|\mathcal{D} - \mathcal{D}'\|_1 \leq \frac{\alpha}{\lambda}$ is necessary.

Theorem 8 (Lower bound)**.** *Let $\gamma, \alpha > 0$. Let r be a lower bound on the query complexity in the two distribution settting. If $\gamma + \frac{\alpha}{\lambda}$ is sufficiently small then any non-adaptive (Ξ, γ)-random privacy property tester with proximity parameter α has query complexity $\Omega(\max\{r, \frac{\lambda}{\alpha}\})$.*

We conjecture that the lower bound is actually $\Omega(r\frac{\lambda}{\alpha})$. If this is true then Theorem 7 gives an almost optimal conversion from the two database setting to the random setting. The proof of Theorem 8 is found in the (21).

5. No Information Setting

We first show that no privacy property tester with finite query complexity exists for pDP. We then analyse a finite query complexity privacy property tester for aDP, as well query complexity lower bounds. For the remainder of this work we consider the two databases setting, where each algorithm $\mathcal{A} = (P_0, P_1)$ accepts only two databases, 0 and 1, as input and $\mathcal{A}(0) \sim P_0$ and $\mathcal{A}(1) = P_1$. The databases 0 and 1 are assumed to be neighbouring.

5.1. **Unverifiability.** The impossibility of testing pDP arises from the fact that very low probability events can cause the privacy parameters to increase arbitrarily. In each case we can design distributions P and Q that are close in TV-distance but for which the algorithm (P, Q) has arbitrarily large privacy parameters. This intuition allows us to use a corollary of Le Cam's inequality (Corollary 9) to prove our impossibility results.

Lemma 9. *For any privacy definition, let α be the proximity parameter and Ξ be the privacy parameters. Suppose $\mathcal{A} = (P_0, P_1)$ and $\mathcal{B} = (Q_0, Q_1)$ are algorithms such that $\mathcal{A}$ is Ξ-DP and $\mathcal{B}$ is α-far from being Ξ-DP. Then, any privacy property testing algorithm with QC $2r$ must satisfy*

$$\|(P_0^r \times P_1^r) - (Q_0^r \times Q_1^r)\|_{TV} \geq \frac{1}{3}$$

Theorem 10 (pDP lower bound)**.** *Let $\alpha > 0$ and $\epsilon > 0$. No ϵ-pDP property tester with proximity parameter α has finite query complexity.*

The proof of Theorem 10 can be found in the (21). We design two distributions that are equal on a large probability set but for which the ratio $\frac{P_0(x)}{Q_0(x)}$ blows-up on a set with small probability. In Section 6 we will see that testing pure DP becomes possible if we make assumptions on the algorithm $\mathcal{A}$. The assumption we need will ensure that $\frac{P_0(x)}{Q_0(x)}$ is upper bounded.

5.2. **Property Testing for aDP in the No Information Setting.** Fortunately, the situation is less dire for verifying aDP. Finite query complexity property testers do exist for aDP, although their query complexity can be very large. In the previous section, we relied on the fact that two distributions P and Q can be close in TV-distance while (P, Q) has unbounded privacy parameters. In this section, we first show this is not true for aDP. Define

$$(1) \quad \delta_\epsilon^{\mathcal{A}} \geq \max_{D,D' \text{neighbours}} \max_E P_D(E) - e^\epsilon P_{D'}(E).$$

An algorithm is (ϵ, δ)-aDP if and only if $\delta > \delta_\epsilon^{\mathcal{A}}$. The following lemma shows the relationship between the aDP parameters and TV-distance.

Lemma 11. *Let $\mathcal{A} = (P_0, P_1)$ and suppose $\mathcal{A}$ is (ϵ, δ)-aDP and $\alpha > 0$. If $\mathcal{B} = (Q_0, Q_1)$ and*

(1) $\|P_0 - Q_0\|_{TV} \leq \alpha$
(2) $\|P_1 - Q_1\|_{TV} \leq \alpha$,

then $\mathcal{B}$ is $(\epsilon, \delta + (1 + e^\epsilon)\alpha) - aDP$. Furthermore, if $\alpha \leq \frac{1-\delta}{1+e^\epsilon}$ then this bound is tight. That is, if $\delta_\epsilon^{\mathcal{A}} > 0$, then there exists an algorithm $\mathcal{B} = (Q_0, Q_1)$ such that conditions (1) and (2) hold but $\mathcal{B}$ is α-far from $(\epsilon, \delta_\epsilon^{\mathcal{A}})$.

The proof of Lemma 11 can be found in the (21). The first direction, that closeness in TV-distance implies privacy, follows almost by definition.

Theorem 12 (Lower bound)**.** *Let $\alpha, \epsilon, \delta > 0$ and suppose $e^\epsilon/2 + \delta + \alpha < 1$. Any (ϵ, δ)-aDP property tester with proximity parameter α has query complexity*

$$r \geq \max\left\{n^{1-o(1)}, \frac{1}{\alpha^2}\right\}.$$

The proof of the $1/\alpha^2$ component of the lower bound relies on Lemma 9 in a similar way to Theorem 10. The proof of the $n^{1-o(1)}$ lower bound borrows a technique from (32). The lemma uses the fact that if two distributions only differ on elements of $[n]$ with low probability, then many samples are needed to distinguish between them.

At first glance, Theorem 12 doesn't look too bad. We should expect the sample complexity to scale like $1/\alpha^2$ since we need to have enough samples to detect the bad events. Our concern is the size of α. If we would like α to be the same order as δ, then our query complexity must scale as $\frac{1}{\delta^2}$. As we typically require

Algorithm 1 aDP Property Tester

Input: Universe size n, $\epsilon, \delta, \alpha > 0$
$\lambda = \max\{\frac{4n(1+e^{2\epsilon})}{\alpha^2}, \frac{12(1+e^{2\epsilon})}{\alpha^2}\}$
Sample $r \sim$Poi(λ)
Sample $D_0 \sim P^r$, $D_1 \sim Q^r$
for $i \in [m]$ **do**
  $x_i =$ number of i's in D_0
  $y_i =$ number of i's in D_1
  $z_i = \frac{1}{r}(x_i - e^{\epsilon} y_i)$
end for
$z = \sum_{i=1}^{r} \max\{0, z_i\}$
if $z < \delta + \alpha$ **then**
  Output: ACCEPT
else
  Output: REJECT
end if

δ to be extremely small (i.e. $\delta \approx 10^{-8}$), $\frac{1}{\delta}$ may be infeasibly large. If we are willing to accept somewhat larger δ, then $\frac{1}{\delta^2}$ may be reasonable.

Theorem 13 (Upper bound)**.** *Let $\epsilon, \delta, \alpha > 0$. Algorithm 1 is a (ϵ, δ)-aDP property tester with proximity parameter 2α and sample complexity $O(\frac{4n(1+e^{2\epsilon})}{\alpha^2})$.*

We now turn our attention to Algorithm 1, a simple algorithm for testing aDP with query complexity $O(\frac{4n(1+e^{2\epsilon})}{\alpha^2})$. Its sample complexity matches the lower bound in Theorem 12 in α when n is held constant and in n when α is held constant. We are going to use a trick called *Poissonisation* to simplify the proof of soundness and completeness, as in Batu et. al (8). Suppose that, rather than taking r samples from P, the algorithm first samples r_1 from a Poisson distribution with parameter $\lambda = r$ and then takes r_1 samples from P. Let X_i be the random variable corresponding to the number of times the element $i \in [n]$ appears in the sample from P. Then X_i is distributed identically to the Poisson distribution with parameter $\lambda = p_i r$ and all the X_i's are mutually independent. Similarly, we sample r_2 from a Poisson distribution with parameter r and then take r_2 samples from Q. Let Y_i be the the number of times i appears in the sample from Q, so Y_i is Poisson with $\lambda = q_i r$ and the Y_i are independent.

The proof of Theorem 13 can be found in the (21). Let $Z_i = \frac{1}{r}(X_i - e^{\epsilon} Y_i)$. The proof proceeds by controlling the random variable $\sum_{i=1}^{n} \max\{0, \mathbb{E}[Z_i]\}$.

6. Full Information (FI) Setting

The situation is substantially rosier if we have side-information. Although there are some realistic scenarios where one may have *trusted* side-information, we will focus on *untrusted* side-information. In particular, we allow our property tester to REJECT simply because the provided side-information is inaccurate. We will see that the untrusted side-information can still be useful since verifying information is often easier than estimating it.

The usefulness of side-information in property testing is informally lower bounded by how easy it is to generate the same information, and how much the information tells us about the property. For example, in the extended version (21) we show that the means of the distributions are not very helpful side-information since they are efficient to estimate, but do not tell us very much about whether or not the privacy guarantee is satisfied.

For the remainder of the paper we focus on what we call the *full information* setting: we are given sample access to $\mathcal{A}$ and a distribution Q_D for each database D. This case may seem optimistic, however we will find that the lower bounds obtained are still very large. The lower bounds in this optimistic setting also hold for more realistic setting when the verifier has less information. In contrast to the mean, this side information is very informative about the privacy of the algorithm. It is also difficult to generate based on samples. We can *estimate* it using $\Theta(\frac{n}{\alpha^2})$ (9) queries of each database, where α is the accuracy in TV-distance. However, we already know that the only privacy notion for which an estimate is sufficient is aDP.

In Algorithm 2 we use an identity tester rather than density estimation to obtain a lower sample complexity. An identity tester is a property tester T that takes as input a description of the discrete distribution P and m samples from a distribution Q. If $P = Q$ then the tester ACCEPTS with probability at least 2/3 and if $\|P - Q\|_{\text{TV}} \geq \alpha$ then the tester REJECTS with probability at least 2/3. It is also known that testing identity to a known distribution requires asymptotically less samples than estimating an unknown distribution.

Proposition 14. *There exists a identity tester T such that Algorithm 2 is a (ϵ, δ)-aDP FI property tester with query complexity $O\left(\frac{\sqrt{n}}{\alpha^2}\right)$ and proximity parameter α.*

Algorithm 2 aDP FI Property Tester

Input: Universe size n, $\epsilon, \delta, \alpha > 0$, (Q_0, Q_1) and identify tester T with sample complexity r
if (Q_0, Q_1) is not (ϵ, δ)-aDP **then**
  Output: REJECT
else
  if $T(Q_0, x \sim P_0^r) =$ REJECT or $T(Q_1, x \sim P_1^r) =$ REJECT **then**
    Output: REJECT
  else
    Output: ACCEPT
  end if
end if

This is our first, and only, sublinear query complexity property tester for privacy. Since closeness in TV-distance implies closeness in aDP, we only need to check that the true distributions are close to (Q_0, Q_1) and that (Q_0, Q_1) is (ϵ, δ)-aDP. The difficult part is testing closeness of the distributions, for which we borrow from On et. al (10). The full proof can be found in the (21).

Next, we show that for pDP, the side information allows us to obtain a finite query complexity property tester. The side-information gives us an easy way to switch from a worst-case analysis to input specific upper bounds. We argue that

$$\beta = \min_E \min_D P_D(E),$$

where the first $\min$ is over events E and the second is over databases D, is the crucial quantity in understanding verifiability in the full information setting.

The lower bound proofs in the previous section all proceeded by finding two algorithms $\mathcal{A}$ and $\mathcal{B}$ that were close in TV-distance but had very different privacy parameters. The algorithms we chose all had one feature in common: the distributions P_D contained very low probability events. This property allowed us to drive the denominator of $\frac{Q_D(E)}{P_D(E)}$ to 0, and hence the privacy loss to ∞, while remaining close in TV-distance. This method works equally well in the full-information setting *if* low probability events exist in the distributions Q_D.

If the distribution Q_D does not have low probability events, then any distributions close to P_D must have bounded privacy parameters. To see this, suppose $(Q_0, Q_1) = (U, U)$ where U is the uniform distribution U on $\{\psi, \omega\}$. We can establish in approximately $\frac{1}{\alpha^2}$ samples whether or not P_0 and P_1 are both within TV-distance α of uniform. If not, then we REJECT. If so, then the worst case for privacy is $P_0 = (1/2 - \alpha)\chi_\psi + (1/2 + \alpha)\chi_\omega$ and $P_1 = (1/2 + \alpha)\chi_\psi + (1/2 - \alpha)\chi_\omega$. However, the increase in the pDP parameter from (Q_0, Q_1) to (P_0, P_1) is bounded by $\ln \frac{1/2+\alpha}{1/2-\alpha} \approx \alpha$.

Algorithm 3 pDP FI Property Tester

Input: Universe size n, $\epsilon, \alpha > 0$, (Q_0, Q_1)
$\lambda = \frac{\ln n}{\alpha^2 \beta^2}$
Sample $r \sim$Poi(λ)
Sample $D_0 \sim P_0^r$, $D_1 \sim P_1^r$
for $i \in [m]$ **do**
  $x_i =$ number of i's in D_0
  $y_i =$ number of i's in D_1
end for
$\hat{\epsilon} = \sup_i \max\{\ln \frac{x_i}{y_i}, \ln \frac{y_i}{x_i}\}$
if $\hat{\epsilon} > \epsilon + 2\alpha$ **then**
  Output: REJECT
else
  if $\forall i \quad e^{-\alpha} \leq \frac{x_i}{(Q_0)_i} \leq e^{\alpha}$ and $e^{-\alpha} \leq \frac{y_i}{(Q_1)_i} \leq e^{\alpha}$ **then**
    Output: ACCEPT
  else
    Output: REJECT
  end if
end if

Theorem 15 (pDP upper bound)**.** *Let $\epsilon > 0$ and $\alpha > 0$. Algorithm 3 is an ϵ-aDP FI property tester with proximity parameter 10α and query complexity $O\left(\frac{\ln n}{\alpha^2 \beta^2}\right)$.*

Algorithm 3 is a full information property tester for pDP. Note that this algorithm is not sublinear in n since $\beta < \frac{1}{n}$. The proof is a generalisation of the argument made in the previous paragraph.

We now turn to lower bounding the query complexity of aDP testing in the FI setting. The sample complexity is tight in α but deviates by a factor of β.

Theorem 16 (pDP lower bound)**.** *Let $\alpha > 0$ and $\ln 2 > \epsilon > 0$. Given side information (Q_0, Q_1), any ϵ-pDP property tester with proximity parameter α has query complexity $\Omega\left(\frac{1}{\beta \alpha^2}\right)$.*

The proof of Theorem 16 is in (21). Like many of our lower bounds, it proceeds by invoking Theorem 10.

REFERENCES

[1] John M. Abowd and Ian M. Schmutte. Revisiting the economics of privacy: population statistics and confidentiality protection as public goods. https://www2.census.gov/ces/wp/2017/CES-WP-17-37.pdf, 2017.

[2] Jayadev Acharya, Constantinos Daskalakis, and Gautam Kamath. Optimal testing for properties of distributions. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 3591–3599. Curran Associates, Inc., 2015.

[3] Apple. What's new in ios: ios 10.0. *Apple Developer Guide*, 2017.

[4] The Apple DP Team. Learning with privacy at scale. https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html, 2017.

[5] Gilles Barthe, George Danezis, Benjamin Gregoire, Cesar Kunz, and Santiago Zanella-Beguelin. Verified computational differential privacy with applications to smart metering. In *Proceedings of the 2013 IEEE 26th Computer Security Foundations Symposium*, CSF '13, pages 287–301, 2013.

[6] Gilles Barthe, Marco Gaboardi, Emilio Jesús Gallego Arias, Justin Hsu, César Kunz, and Pierre-Yves Strub. Proving differential privacy in hoare logic. In *Proceedings of the 2014 IEEE 27th Computer Security Foundations Symposium*, CSF '14, pages 411–424, 2014.

[7] Gilles Barthe, Boris Köpf, Federico Olmedo, and Santiago Zanella Béguelin. Probabilistic relational reasoning for differential privacy. *SIGPLAN Not.*, 47(1):97–110, January 2012.

[8] Tuğkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D. Smith, and Patrick White. Testing closeness of discrete distributions. *J. ACM*, 60(1):4:1–4:25, February 2013.

[9] Siu-On Chan, Ilias Diakonikolas, Rocco A. Servedio, and Xiaorui Sun. Near-optimal density estimation in near-linear time using variable-width histograms. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'14, pages 1844–1852, Cambridge, MA, USA, 2014. MIT Press.

[10] Siu-On Chan, Ilias Diakonikolas, Gregory Valiant, and Paul Valiant. Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of the Twenty-fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '14, pages 1193–1203, Philadelphia, PA, USA, 2014. Society for Industrial and Applied Mathematics.

[11] Constantinos Daskalakis, Gautam Kamath, and John Wright. *Which Distribution Distances are Sublinearly Testable?*, pages 2747–2764. 2018.

[12] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. In *Advances in Neural Information Processing Systems 30*, December 2017.

[13] Ding Ding, Yuxin Wang, Guanhong Wang, Danfeng Zhang, and Daniel Kifer. Toward detecting violations of differential privacy. arXiv:1805.10277v1, 2018.

[14] Kashyap Dixit, Madhav Jha, Sofya Raskhodnikova, and Abhradeep Thakurta. Testing the lipschitz property over product distributions with applications to data privacy. In Amit Sahai, editor, *Theory of Cryptography*, pages 418–436, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[15] Cynthia Dwork. Differential privacy: A survey of results. In *Theory and Applications of Models of Computation—TAMC*. Springer Verlag, April 2008.

[16] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Differential privacy - a primer for the perplexed. In *Conf. of European Statisticians*, Joint UNECE/Eurostat work session on statistical data confidentiality, 2011.

[17] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9:211–407, August 2014.

[18] Cynthia Dwork and Guy N. Rothblum. Concentrated differential privacy. arXiv:1603.01887v2, 2016.

[19] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, pages 1054–1067, New York, NY, USA, 2014. ACM.

[20] Marco Gaboardi, Andreas Haeberlen, Justin Hsu, Arjun Narayan, and Benjamin C. Pierce. Linear dependent types for differential privacy. *SIGPLAN Not.*, 48(1):357–370, January 2013.

[21] Anna C. Gilbert and Audra McMillan. Property testing for differential privacy. arXiv:1806.06427, 2018.

[22] Zhanglong Ji, Zachary C. Lipton, and Charles Elkan. Differential privacy and machine learning: a survey and review. arXiv:1412.7584, 2014.

[23] Frank D. McSherry. Privacy integrated queries: An extensible platform for privacy-preserving data analysis. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*, SIGMOD '09, pages 19–30, 2009.

[24] I. Mironov. Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 263–275, Aug 2017.

[25] L. Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Trans. Inf. Theor.*, 54(10):4750–4755, October 2008.

[26] Jason Reed and Benjamin C. Pierce. Distance makes the types grow stronger: A calculus for differential privacy. In *Proceedings of the 15th ACM SIGPLAN International Conference on Functional Programming*, ICFP '10, pages 157–168, 2010.

[27] Indrajit Roy, Srinath T. V. Setty, Ann Kilzer, Vitaly Shmatikov, and Emmett Witchel. Airavat: Security and privacy for mapreduce. In *Proceedings of the 7th USENIX Conference on Networked Systems Design and Implementation*, NSDI'10, pages 20–20, 2010.

[28] Abhradeep Thakurta, Andrew Vyrros, Umesh Vaishampayan, Gaurav Kapoor, Julien Freidiger, Vivek Sridhar, and Doug Davidson. Learning New Words. U.S. Patent 9,594,741 B1, March 14 2017.

[29] Michael Carl Tschantz, Dilsun Kaynar, and Anupam Datta. Formal verification of differential privacy for interactive systems (extended abstract). *Electron. Notes Theor. Comput. Sci.*, 276:61–79, September 2011.

[30] Salil Vadhan. The complexity of differential privacy. Harvard University Privacy Tools Project, 2016.

[31] G. Valiant and P. Valiant. An automatic inequality prover and instance optimal identity testing. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 51–60, Oct 2014.

[32] Paul Valiant. Testing symmetric properties of distributions. *SIAM J. Comput.*, 40(6):1927–1968, December 2011.