Bounded-Leakage Differential Privacy

Katrina Ligett
Computer Science Department, Hebrew University of Jerusalem, Israel
katrina@cs.huji.ac.il

Charlotte Peale
Stanford University, Stanford, CA, USA
cpeale@stanford.edu

Omer Reingold
Computer Science Department, Stanford University, Stanford, CA, USA
reingold@stanford.edu

Abstract

We introduce and study a relaxation of differential privacy [3] that accounts for mechanisms that leak some additional, bounded information about the database. We apply this notion to reason about two distinct settings where the notion of differential privacy is of limited use. First, we consider cases, such as in the 2020 US Census [1], in which some information about the database is released exactly or with small noise. Second, we consider the accumulation of privacy harms for an individual across studies that may not even include the data of this individual. The tools that we develop for bounded-leakage differential privacy allow us reason about privacy loss in these settings, and to show that individuals preserve some meaningful protections.

2012 ACM Subject Classification Theory of computation → Theory of database privacy and security

Keywords and phrases differential privacy, applications, privacy, leakage, auxiliary information

Digital Object Identifier 10.4230/LIPIcs.FORC.2020.10

Funding *Katrina Ligett*: This work was supported in part by Israel Science Foundation (ISF) grant #1044/16, United States Air Force and DARPA under contracts FA8750-16-C-0022 and FA8750-19-2-0222, and the Federmann Cyber Security Center in conjunction with the Israel national cyber directorate. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Air Force and DARPA.
Omer Reingold: Supported in part by NSF Award IIS-1908774 and by VMWare fellowship.

Acknowledgements Part of this work was done while the first and third authors were visiting the Simons Institute for the Theory of Computing.

1 Introduction and Related Work

Differential privacy [3], a notion of the stability of computations, has emerged as the gold standard of mathematically rigorous privacy protection for computations on statistical databases, in part because of its appealing interpretations from the perspective of individuals in the database. For example, an economic interpretation of differential privacy holds that individuals' future utilities will be harmed by at most a small constant factor by providing their true data, rather than lying, to the mechanism [5]. Also appealing is an interpretation that shows that an individual's truthful provision of data to a differentially private mechanism will not substantially change a Bayesian observer's posterior beliefs versus the beliefs that would result if that individual provided false data [10].

One clear advantage of differential privacy over other approaches to defining privacy is that it is not necessary to reason about auxiliary information in order to give differential privacy guarantees. Composition attacks [7] leveraging access to such "outside information"

are a common Achilles' Heel of ad hoc privacy notions. Differential privacy, as it is a property of the *mechanism*, rather than the mechanism's output, is a guarantee that holds regardless of the presence of auxiliary information. However, as observed by Dwork and Naor [4] and Kifer and Machanavajjhala [11, 12], a differentially private release of data against a background of prior data releases can still result in substantial privacy harms. In particular, Dwork and Naor show essentially that innocuous information can always be combined with statistical information to yield a privacy breach. These negative results regarding auxiliary information are used by Dwork and Naor as justification for differential privacy's focus on *relative* guarantees – harms relative to the harm you would have experienced had you not participated in the study or lied about your data.

In this work, we return to the question of auxiliary information, which we term "leakage." We develop a formal framework in which to study differential privacy in the presence of leakage and argue about non-trivial properties of such bounded-leakage differential privacy. Our research is motivated by two applications:

Application: 2020 Census

One context in which auxiliary information has recently gained attention is the 2020 US Census. The majority of data releases, including synthetic data, for the 2020 US Census, are slated to be released subject to differential privacy [1, 8]. However, the Census, by agreement with the Department of Justice, will provide "exact counts,"[1] known as *invariants*, for certain statistics [8, 9, 2]. As Garfinkel et al. [9] allude to, "there is no well-developed theory for how differential privacy operates in the presence of such invariants."

One particularly nefarious problem with auxiliary information emerges if the function that produced the auxiliary information might share randomness with the differentially private algorithm. If so, all guarantees of differential privacy might be lost. Our definitions directly apply to this setting and help clarify the impact of such auxiliary information (and in particular, the role of such shared randomness).

Application: Big-World Privacy

One of the benefits of differential privacy is that it gives a way to quantify the privacy losses of individuals whose data were included in a database input into a mechanism, and even allows quantification of how privacy losses "add up" across multiple mechanisms. In its most basic form, differential privacy tells us that for every ϵ-differentially private study an individual participates in, she incurs an ϵ privacy loss. However, it is also relevant to ask whether a study that an individual *does not* participate in could also degrade that individual's privacy. Consider, for example, a study that included everyone in a certain city who had a particular disease. Then, this would mean that the information that a particular individual from that city did not participate in the study also tells us that this individual doesn't have the disease, and could potentially degrade the individual's privacy in unexpected ways.

If we cannot necessarily quantify an individual's overall privacy loss using only the studies they have participated in, what should we do? One possible solution is to treat all individuals in the world as belonging to a single huge database D. A differentially private mechanism M that computes over some subset of the population would be considered instead to be a composition of two mechanisms: first, a selection function for determining which individuals to include in the study, and then the subsequent differentially private mechanism. This concatenated mechanism, however, need not be differentially private if the selection function

[1] "Exact counts" are of course not really an exact population count, but reflect many sources of error, including non-participation and potentially also Census techniques intended to infer missing data.

is not differentially private. In fact, the selection function could adversarially choose a subset of the database in order to cause sensitive data about an individual to be encoded into the likely outputs of the subsequent differentially private mechanism.

Even if the selection of studies that include the data of a particular individual is independent for other individuals, then the inclusion or exclusion of the data of individual i in study j may depend on i's data. It seems that we cannot avoid the conclusion that the only bound that differential privacy can give us about i's privacy loss is the composition of the ϵ privacy losses for every ϵ-differentially private study that has ever occurred, *even for studies in which i did not participate*. Our framework allows us to separate the privacy loss into that which is incurred from exclusion from particular studies and the loss that is incurred from participation in ϵ-differentially private studies. Once we acknowledge the leakage of an upper-bound on the number of studies i participate in, the privacy loss is incurred as a function of this upper bound rather than as a function of all studies.

Our Contributions

In this paper, we present (Section 2) a definition for *bounded-leakage differential privacy*, a relaxed variant of differential privacy that quantifies the privacy that is maintained by a mechanism despite bounded, additional leakage of information by some "leakage function." We investigate (Section 3) the connections between standard differential privacy and this new variant, and give conditions for when the bounded-leakage privacy of a mechanism can imply something about its differential privacy, and vice versa.

Differentially private mechanisms are known to satisfy some appealing, simple properties such as privacy conservation under post-processing and quantifiable privacy loss for groups. We show that bounded-leakage privacy satisfies the same post-processing results as standard differential privacy (Theorem 11), and prove an analogous result for the group privacy of bounded-leakage differentially private pairs of mechanisms and leakage functions (Theorem 22). Additionally, we show that explicitly "leaking" the value of the leakage function does not affect the bounded-leakage differential privacy of a mechanism/function pair (Corollary 14).

There are numerous results about the composition of differentially private mechanisms, both in a non-adaptive setting where the mechanisms and databases queried are chosen before execution, and additionally in an adaptive setting where intermediate outputs may affect the choice of future queries. We define adaptive composition for bounded-leakage privacy and, using a new reduction technique, we show that any general adaptive or non-adaptive composition bound that holds for differentially private mechanisms must also hold for the class of bounded-leakage private mechanism/function pairs, with the same privacy parameters (Section 4.4).

It is conceivable that leaking additional information about a mechanism could also affect the utility of its outputs. The exponential mechanism for differentially private mechanisms presents a way to construct a differentially private mechanism with a high utility guarantee on its output. We define an analogous mechanism for the bounded-leakage privacy setting, such that given a leakage function and a utility function, we can construct a bounded-leakage private mechanism with high utility guarantees for each possible output of the leakage function (Section E).

Finally, (Section 6) we use the bounded-leakage differential privacy framework to study the Census and Big-World applications, and show that it is possible to formally bound privacy harms in these settings.

Due to space constraints, almost all proofs appear in the appendices.

Additional Related Work

Kasiviswanathan and Smith [10] also provide a formal treatment of auxiliary information, in the context of their Bayesian interpretation of differential privacy. [11] explore a notion of privacy in the context of auxiliary information that is similar to ours, but more limited in the (deterministic) prior releases they consider. The Pufferfish framework [12] is an extremely general privacy framework that allows for reasoning about the conclusions drawn by specific types of attackers. [12] explicitly explores composition of private mechanisms with non-private mechanisms in their Sections 9 and 10, and gives general statements based on the conditional probability distribution of one release given the other. This appears to have analogy in the exploration of independence that we do in Section 3. Given the generality of the Pufferfish framework, it is likely capable of describing our notion of bounded-leakage differential privacy. Our notion is focused on the behavior of differential privacy in the presence of auxiliary information rather than on stronger notions that are resilient to adversaries with auxiliary information. By focusing on a more specific notion, we are able to build up a richer set of properties and consequences of the notion.

2 Model

The standard definition of differential privacy promises that the distribution of results of a randomized mechanism run on any database does not change too much if we change any particular individual's data in that database. Formally, we will represent a database that holds data about n individuals as a tuple from the set $\mathcal{X}^n$ where $\mathcal{X}$ is a data universe of possible characteristics. We call two databases $x, x' \in \mathcal{X}^n$ *neighboring* or *adjacent* if they differ in the data of one individual, and will denote this as $x \sim x'$. Using this notation, the notion of standard differential is formally defined as follows:

▶ **Definition 1** ((ϵ, δ)-Differential Privacy (DP) [3]). *Let $\mathcal{X}$ be some data universe, O an output space, and R a space of random inputs. Given a mechanism $M : \mathcal{X}^n \times R \to O$, we say that M is* (ϵ, δ)-differentially private *if for all neighboring databases $x \sim x' \in \mathcal{X}^n$ and all subsets $S \subseteq O$, we have that*

$$Pr_{r \in R}[M(x, r) \in S] \leq e^{\epsilon} Pr_{r \in R}[M(x', r) \in S] + \delta.$$

Building off of this definition, we introduce a new definition for a variant of differential privacy that we call *bounded-leakage differential privacy*. Intuitively, bounded-leakage differential privacy asserts that, given a database and a mechanism to be run on it, if an additional piece of information about the database were leaked, then the output of the mechanism when run on the database would not leak much more information than what was already leaked.

▶ **Definition 2** ((ϵ, δ)-Bounded-Leakage Differential Privacy (blDP)). *Let $\mathcal{X}$ be some data universe, O_M an output space, O_P a countable output space, and R a space of random inputs. Given a mechanism $M : \mathcal{X}^n \times R \to O_M$, and a* leakage function *$P : \mathcal{X}^n \times R \to O_P$, we say that M is* (ϵ, δ)-bounded-leakage differentially private *with respect to P if for all neighboring databases $x \sim x' \in \mathcal{X}^n$, all $S \subseteq O_M$, and $o \in O_P$, we have that either*

$$Pr_{r \in R}[P(x', r) = o] \cdot Pr_{r \in R}[P(x, r) = o] = 0$$

or

$$Pr_{r \in R}[M(x, r) \in S | P(x, r) = o] \leq e^{\epsilon} Pr_{r \in R}[M(x', r) \in S | P(x', r) = o] + \delta.$$

For the rest of this paper, given any mechanism-function pair (M, P), we will by default denote the output space of M as O_M and the output space of P as O_P.

3 The Relationship Between blDP and DP

3.1 When Does blDP Imply DP?

Bounded-leakage differential privacy is clearly a weaker notion than standard differential privacy due to the fact that the robustness constraint across two neighboring databases is only required to hold conditioning on the value of the leakage function. The following example illustrates one scenario in which we can have perfect bounded-leakage differential privacy, but no meaningful differential privacy.

▶ **Example 3** (perfect blDP does not imply DP for the mechanism)**.** Consider any arbitrary mechanism M. The pair (M, M) satisfies perfect bounded-leakage differential privacy because for any databases $x \sim x'$, $S \subseteq O_M$, and $o \in O_P$ such that $\Pr_r[M(x,r) = o] \cdot \Pr_r[M(x',r) = o] \neq 0$, we will always have

$$\Pr_r[M(x,r) \in S | M(x,r) = o] = \Pr_r[M(x',r) \in S | M(x',r) = o] = \begin{cases} 1 & \text{if } o \in S \\ 0 & \text{otherwise} \end{cases}$$

and so $\Pr_r[M(x,r) \in S | M(x,r) = o] = \Pr_r[M(x',r) \in S | M(x',r) = o]$. Therefore, (M, M) satisfies perfect blDP, but M need not have any sort of differential privacy guarantee.

We should note that this example depends on the fact that blDP is defined such that the privacy mechanism and associated leakage function share the same random input.

However, we *can* say something about the differential privacy of the mechanism in a blDP pair if we additionally know that the leakage function is differentially private.

▶ **Theorem 4** (blDP with a DP leakage function)**.** *Suppose $M : \mathcal{X}^n \times R \to O_M$ is a mechanism satisfying (ϵ_1, δ_1)-blDP with respect to a leakage function $P : \mathcal{X}^n \times R \to O_P$. Additionally, suppose that P satisfies $(\epsilon_2, 0)$-DP. Then M satisfies $(\epsilon_1 + \epsilon_2, \delta_1)$-DP.*

3.2 When Does DP Imply blDP?

In the previous subsection, we noted that intuitively, blDP seems like a weaker notion than DP. We saw that blDP for a privacy mechanism and leakage function pair does not guarantee anything about the DP of the privacy mechanism. Following this intuition further, we might also expect that having a mechanism/function pair where the mechanism is DP should guarantee that the pair satisfies blDP. However, this is actually not necessarily the case. Consider the following example where a perfectly DP mechanism loses all blDP when paired with a particular leakage function:

▶ **Example 5** (perfect DP does not guarantee blDP)**.** Consider the mechanism $M : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$. Then, for $x, r \in \{0,1\}^n$, define M to be $M(x,r) = x \oplus r$. Under this definition, M's output is uniformly distributed over $\{0,1\}^n$, making it perfectly DP.

Now, define the leakage function $P : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$ to be $P(x,r) = r$. Conditioning on P being a particular value means that the randomness used in M will be fixed. For any $x \neq x'$, we will always have $x \oplus r \neq x' \oplus r$ and therefore $M(x,r) \neq M(x',r)$.

Thus, (M, P) does not satisfy blDP for any practical values of ϵ and δ, despite the fact that M is perfectly DP.

Why does our intuition fail here? We can identify three properties of the above example that lead to this unexpected result:

(1) P released information about its random input.
(2) M and P shared the same random input.
(3) The output of M depended on the output of P.

In fact, one can show that if the example above were changed such that *any* of these three properties did not hold, then we would get the opposite result and the differential privacy of M *would* imply bounded-leakage differential privacy for the pair.

Informally, we see that conclusions about the blDP of a pair will depend on how correlated the mechanism and the leakage function are.

▶ **Definition 6.** *We call a mechanism-function pair* (M, P) perfectly independent *if for any database* x *,* $S \subseteq O_M$*, and output* $o \in O_P$ *such that* $\Pr_r[P(x, r) = o] \neq 0$*, we have*

$$\Pr_r[M(x, r) \in S | P(x, r) = o] = \Pr_r[M(x, r) \in S].$$

In other words, the output of M is completely uncorrelated with the value of P when given the same database and random input.

We consider two common examples of perfectly independent leakage functions. One might be where M and/or P are completely deterministic, such as a function that releases an exact summary statistic. A second interesting case is where P uses "fresh randomness," disjoint from M's computation. M and P may share the same random input string, but if the random bits they depend on are completely separate, then their outputs will be perfectly independent.

If a mechanism-function pair is perfectly independent, then a DP mechanism *does* imply blDP for the pair. The following lemma follows immediately from combining the definition of perfect independence with the definitions of DP and blDP, and its proof is omitted:

▶ **Lemma 7.** *If* (M, P) *is a perfectly independent mechanism-function pair and* M *satisfies* (ϵ, δ)*-DP, then* M *must also satisfy* (ϵ, δ)*-blDP with respect to* P*.*

Perfectly independent pairs constitute the class of mechanism-function pairs with completely uncorrelated outputs. On the other end of the spectrum, we have pairs such as the one presented in Example 5 where the output of M is a deterministic function of the output of P. However, there are mechanism-function pairs that lie between these two extremes for which we would also like to derive blDP bounds. Such pairs can be thought of as mechanisms that are only partially dependent on the output of their associated leakage functions (or vice versa).

Where might we see partially independent mechanism-function combinations in practical applications of blDP? One possible scenario is a study that computes its output based on a random sample of individuals chosen from the provided database. For such a study, the leakage function may need to depend on the randomness used to pick the sample. This might be due to space reasons – for example, the study might discard other data after picking the random sample – or for accuracy reasons. In either case, the result of such a leakage function will be somewhat correlated with the randomness used to select the sample in the original study, but may also employ its own randomness as well, giving us a leaked value that partially depends on the randomness used in the original study.

We can quantify this generalized notion of partial independence as follows:

▶ **Definition 8** ((ϵ, δ)-independence)**.** *Consider a mechanism-function pair* (M, P)*. We say that* M *and* P *are* (ϵ, δ)-independent *if for every database* x*, subset* $S \subseteq O_M$*, and output* $o \in O_P$ *such that* $\Pr_r[P(x, r) = o] \neq 0$*, we have that*

$$\Pr_r[M(x, r) \in S | P(x, r) = o] \leq e^\epsilon \Pr_r[M(x, r) \in S] + \delta$$

and

$$\Pr_r[M(x, r) \in S] \leq e^\epsilon \Pr_r[M(x, r) \in S | P(x, r) = o] + \delta.$$

Note that by this definition, (0, 0)-independent pairs are perfectly independent. Using this definition of dependence, we can get a more general version of Theorem 7 which wells us what sort of blDP bounds we can expect from a partially independent pair with a DP mechanism. This is presented in the following theorem, and we note that by substituting in $\epsilon' = \delta' = 0$, we get the result of Lemma 7 as a corollary.

▶ **Theorem 9.** *Suppose we have a mechanism-function pair (M, P) where the outputs of M and P are (ϵ', δ')-independent of one another. If M satisfies (ϵ, δ)-DP, then (M, P) must satisfy $(\epsilon + 2\epsilon', (e^{\epsilon'+\epsilon} + 1)\delta' + e^{\epsilon'}\delta)$-blDP.*

4 Properties

We can show that bounded-leakage differential privacy satisfies many of the properties satisfied by standard differential privacy [3], plus additional desirable properties.

4.1 Post-Processing

We begin by observing that bounded-leakage differential privacy is closed under convex combination.

▶ **Lemma 10.** *A convex combination of (ϵ, δ)-blDP mechanisms with respect to some leakage function $P : \mathcal{X}^n \times R \to O_P$ is also an (ϵ, δ)-blDP mechanism with respect to the same P.*

With this observation in hand, it quickly follows that blDP is preserved under post-processing of the privacy mechanism. The proof is very similar to the proof for post-processing in the DP setting, and is omitted.

▶ **Theorem 11** (Post-processing). *If $M : \mathcal{X}^n \times R \to O_M$ is an (ϵ, δ)-blDP mechanism with respect to $P : \mathcal{X}^n \times R \to O_P$ and $f : O_M \to O'$ is any arbitrary mapping from O_M to an output space O', then $f \circ M$ is also an (ϵ, δ)-blDP mechanism with respect to P.*

4.2 Group Privacy

Group privacy is an important property of differential privacy that quantifies how privacy degrades between databases that differ in the data of more than one individual. In standard differential privacy, group privacy properties hold due to the fact that given any two databases, we can construct a "path" of adjacent databases and apply differential privacy properties to each pair in the path.

Group privacy is a bit more complicated in the case of bounded-leakage differential privacy due to the fact that we cannot always construct a path between two databases such that the leakage function maintains the same value between all pairs of databases along the path. Example 18 shows a situation where bounded-leakage privacy between pairs of databases cannot imply anything about the group privacy of the same mechanism-function pair. However, we can still state properties about group blDP when for every output o of P, we can find a path between the two databases in question such that for every database x on the path, we have $\Pr[P(x, r) = o] > 0$; we see this in Definitions 20 and 21, and Theorem 22.

4.3 What if the Value of P is Leaked?

The definition of blDP we have presented conditions probability on values of P, but the value of P is never explicitly revealed or "leaked." However, intuitively, we would like our definition to satisfy the property that even if the value of of P is explicitly revealed (that is, the output of P is included in the mechanism output), the pair retains its bounded-leakage privacy. The following theorem shows that this property holds.

▶ **Theorem 12** (Value of P leaked). *Let $M : \mathcal{X}^n \times R \to O_M$ be a mechanism that satisfies (ϵ, δ)-blDP with respect to the leakage function $P : \mathcal{X}^n \times R \to O_P$. Consider another mechanism, $M' : \mathcal{X}^n \times R \to O_M \times O_P$ such that M' returns the output of M concatenated with the output of P; that is, $M'(x, r) := M(x, r)||P(x, r)$. Then M' also satisfies (ϵ, δ)-blDP with respect to P.*

We can combine this theorem with the independence results of Lemma 7 or Theorem 9 to get Corollaries 13 and 14, respectively

▶ **Corollary 13.** *Suppose M is an (ϵ, δ)-DP mechanism and P is a leakage function such that the outputs of M and P are perfectly independent. Then the concatenation of M and P, $M'(x, r) = M(x, r)||P(x, r)$, satisfies (ϵ, δ)-blDP with respect to P.*

▶ **Corollary 14.** *Suppose M is an (ϵ, δ)-DP mechanism and P is a leakage function such that the outputs of M and P are (ϵ', δ')-independent. Then the concatenation of M and P, $M'(x, r) = M(x, r)||P(x, r)$, satisfies $(\epsilon + 2\epsilon', (e^{\epsilon'+\epsilon} + 1)\delta' + e^{\epsilon'}\delta)$-blDP with respect to P.*

4.4 Composition

We derive some properties of the composition of multiple blDP mechanism-function pairs. There are two types of composition that we consider: non-adaptive composition, in which the sequence of mechanism-function pairs is fixed in advance; and adaptive composition, in which the choice of future mechanism-function pairs might depend on the results returned by previous mechanisms.

We use a unified reduction technique to obtain results for both settings.

▶ **Definition 15** (DP reduction mechanism). *Given a mechanism $M : \mathcal{X}^n \times R \to O_M$, a leakage function $P : \mathcal{X}^n \times R \to O_P$, some output $o \in O_P$, and two databases $x_0, x_1 \in \mathcal{X}^n$, we define the* DP reduction mechanism *for $(M, P), o, x_0, x_1$ to be the mechanism $M_{P,o}^{x_0,x_1} : \mathcal{X}^n \times (R_{o,x_0} \times R_{o,x_1}) \to O_M \cup \{\text{"null"}\}$, where R_{o,x_b} is defined as the subset of the random input space R such that $R_{o,x_b} = \{r \in R : P(x_b, r) = o\}$. Then, given any $x \in \mathcal{X}^n$ and $(r_0, r_1) \in R_{o,x_0} \times R_{o,x_1}$, $M_{P,o}^{x_0,x_1}(x, (r_0, r_1))$ is defined as*

$$M_{P,o}^{x_0,x_1}(x, (r_0, r_1)) = \begin{cases} \text{"null"} & \text{if } \Pr_r[P(x_0, r) = o] \Pr_r[P(x_1, r) = o] = 0 \\ M(x_0, r_0) & \text{if } x = x_0 \\ M(x_1, r_1) & \text{otherwise} \end{cases}$$

In order to use this mechanism in our reduction proofs, we need it to satisfy two important properties. The first (Proposition 23) is that the distribution of $M_{P,o}^{x_0,x_1}$ on inputs x_0 and x_1 should match the distribution of M conditioned on P outputting o in the blDP setting for those inputs. The second (Proposition 24) states that if (M, P) satisfies bounded-leakage privacy and x_0 and x_1 are neighboring databases, $M_{P,o}^{x_0,x_1}$ must be differentially private.

In Section C, we show how to use this reduction to translate results on non-adaptive composition of differentially private mechanisms to results for blDP. Section D shows the analogous reduction for adaptive composition, yielding the following theorem:

▶ **Theorem 16.** *For all $\epsilon, \delta, \delta' \geq 0$, the class of (ϵ, δ)-blDP mechanisms satisfies $(\epsilon', k\delta + \delta')$-blDP under k-fold adaptive composition for $\epsilon' = \epsilon\sqrt{2k\ln(1/\delta')} + k\epsilon(e^{\epsilon-1})$.*

5 Tools for Achieving blDP

Our Lemma 7 and Theorem 9 give tools for understanding when existing differentially private algorithms can be used to achieve guarantees of blDP. In addition, in Section E, we establish a blDP variant of the exponential mechanism [13]. The exponential mechanism is a foundational differentially private algorithm that performs exponentially weighted sampling from a space of outcomes with weights chosen according to a utility function over databases and outputs. The standard exponential mechanism cannot be directly applied to the blDP scenario because we have no guarantees about how a particular utility function will depend on the additional leaked function P. As an example, if P is the standard deviation of entries in the database and we are seeking to output a result that is close to the average of the database, the value of P will affect the distribution of how "useful" particular outputs are based on their distance from the true mean. To address this, we introduce a notion of a coupled utility function, and show that an analogous mechanism enjoys blDP.

6 Applications

Now that we have presented a definition of bounded-leakage privacy and explored some of its properties, we consider some applications of this definition.

6.1 2020 Census: Releasing Additional Information About a Dataset

In some situations where a differentially private study is run, there may be additional releases of information about the underlying private dataset. This could happen unintentionally via a leak or some sort of adversarial attack, or it could be intentional if those running the study choose to release select pieces of information without noise (or with lower noise levels), such as releasing the number of outliers, or a summary statistic such as the standard deviation or average of the data surveyed.

In this situation, we would like to understand what sort of privacy is maintained after such a leak, and whether the release of such information combined with the results of a differentially private study could degrade the participants' privacy in unexpected ways.

If the leaked information is a deterministic function that depends only on the database or a randomized function that uses independent randomness from that used in the differentially private mechanism, then the differentially private mechanism and this additional function will be perfectly independent. Applying Theorem 7 tells us that releasing this additional information guarantees bounded-leakage privacy with the same bounds as the DP mechanism in the original study, and so other than revealing that the database used in the study was such that it produced the additional statistic in question, the privacy of the results of the original study does not degrade further. This gives a formal language for reasoning about, for example, the privacy properties of the 2020 US Census, where some statistics will be revealed without any noise, and other computations will be subject to differential privacy [9].

6.2 Big World Privacy: Controlling Privacy Degradation due to Absence from Studies

Recall the Big World Privacy problem from the Introduction. Bounded-leakage differential privacy can aid in reasoning about how privacy degrades across many studies, some of which an individual may not have participated in.

Suppose that a sequence of k studies has been run on various subsets of the entire population D. In the Introduction, we discussed modeling each study as the composition of two mechanisms, one being a standard differential-privacy mechanism $M^{(j)}$, and the other a participation function, which we will denote $f_{par}^{(j)}$. Using this definition, the result of the jth study is computed by first getting the output of $f_{par}^{(j)}(D, r_{par}^{(j)}) = D^{(j)}$, which will be a subset of D containing only the data of those chosen to participate. After getting $D^{(j)}$, we compute $M^{(j)}(D^{(j)}, r^{(j)})$ to get the final result of study j. We will assume that each $M^{(j)}$ satisfies ϵ-DP.

When the participation function is arbitrary, then no level of privacy can be maintained. For example, for a particular study of average height, we can either choose a group of NBA players or a group of toddlers. If this decision is based on a sensitive property of individual i, this property will be completely revealed. We will therefore make the simplifying assumption that the participation of an individual i is independent of all other individuals. Our results will hold in more general settings as well, but as the above example demonstrates, some assumption of this sort needs to be made.

With this assumption we are guaranteed that if a mechanism is (ϵ, δ)-DP, then each pair of databases with different i values but the same participation in other individuals is a pair of neighboring databases, and therefore satisfies (ϵ, δ)-DP. The study will be a convex combination of the results of the privacy mechanism on such pairs, and so must also satisfy (ϵ, δ)-DP.

Now, suppose that we would like to reason about the privacy loss of a particular individual i if some of her participation data is leaked. Consider a leakage function $P_i(D, r_{par}(i), r_p)$ that takes in the giant database D, the randomness used to decide i's participation in each study, $r_{par}(i)$, and some additional randomness r_p. We define this function such that it outputs some $t \in \mathbb{Z}$, with $0 \leq t \leq k$, such that t is an upper bound on the total number of studies that our individual i participated in. There are numerous real-world situations where such a bound might be released, such as the observation that "i only participated in studies conducted in the U.S." or that some number of the studies were conducted before i was born.

Let $\overline{M}(D, r_{par}, \overline{r})$ denote the concatenation of all k studies. Conditioning on this leakage function, we can show that the following result holds:

▶ **Theorem 17.** *For any $t \leq k$, subset $S \subseteq O_{\overline{M}}$, and database D_i that differs from D only in i's data, we have that*

$$\Pr_{r_{par}, \overline{r}}[\overline{M}(D, r_{par}, \overline{r}) \in S | P_i(D, r_{par}(i), r_p) = t]$$

$$\leq e^{2t\epsilon} \Pr_{r_{par}, \overline{r}}[\overline{M}(D_i, r_{par}, \overline{r}) \in S | P_i(D_i, r_{par}(i), r_p) = t] + 2t\delta.$$

This bound arises from the standard additive bound (Theorem 32) for the composition of $2t$ (ϵ, δ)-DP mechanisms, but any existing composition bound for the same setting could be substituted in to get an analogous result. It should be noted that this result is a bit different from our standard definition of blDP due to the fact that the resulting privacy bound ($2t\epsilon$) depends on the value of the leakage function (t). We include the proof of this result in Appendix F. As previously noted, standard differential privacy only tells us that i incurs an ϵ privacy loss for every single study in $\overline{M}$. However, considering the same question in the bounded-leakage differential privacy setting allows us to conclude that the privacy loss of i is bounded by the number of studies she may have participated in.

7 Future Directions

We hope that the notion of bounded-leakage differential privacy will aid in the rigorous analysis of privacy guarantees in settings where differential privacy does not hold. This initial exploration suggests many additional avenues to pursue. It would be interesting to develop additional mechanisms that enjoy blDP, and to apply the notion in new domains. Additionally, one might consider variations on the definition of blDP, for example a variant that allows weakened privacy if the probability of the function P attaining a particular set of values is very small.

References

1 John M Abowd. The us census bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2867–2867, 2018.

2 Aref N Dajani, Amy D Lauger, Phyllis E Singer, Daniel Kifer, Jerome P Reiter, Ashwin Machanavajjhala, Simson L Garfinkel, Scot A Dahl, Matthew Graham, Vishesh Karwa, et al. The modernization of statistical disclosure limitation at the us census bureau. In *Washington, DC: US Census Bureau. Available at:* `https://www2.census.gov/cac/sac/meetings/2017-09/statistical-disclosure-limitation.pdf`, 2017.

3 Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

4 Cynthia Dwork and Moni Naor. On the difficulties of disclosure prevention in statistical databases or the case for differential privacy. *Journal of Privacy and Confidentiality*, 2(1), 2010.

5 Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

6 Cynthia Dwork, Guy N Rothblum, and Salil Vadhan. Boosting and differential privacy. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 51–60. IEEE, 2010.

7 Srivatsava Ranjit Ganta, Shiva Prasad Kasiviswanathan, and Adam Smith. Composition attacks and auxiliary information in data privacy. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 265–273, 2008.

8 Simson L Garfinkel. Modernizing disclosure avoidance: Report on the 2020 disclosure avoidance system as implemented for the 2018 end-to-end test, 2018. URL: `https://www.census.gov/about/cac/sac/meetings/2017-09-meeting.html`.

9 Simson L Garfinkel, John M Abowd, and Sarah Powazek. Issues encountered deploying differential privacy. In *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, pages 133–137, 2018.

10 Shiva P Kasiviswanathan and Adam Smith. On the'semantics' of differential privacy: A bayesian formulation. *Journal of Privacy and Confidentiality*, 6(1), 2014.

11 Daniel Kifer and Ashwin Machanavajjhala. No free lunch in data privacy. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, pages 193–204, 2011.

12 Daniel Kifer and Ashwin Machanavajjhala. Pufferfish: A framework for mathematical privacy definitions. *ACM Transactions on Database Systems (TODS)*, 39(1):1–36, 2014.

13 Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *FOCS*, pages 94–103, 2007.

A Missing Proofs from Section 3

Proof of Theorem 4. Consider any neighboring databases $x \sim x'$ and a subset $S \subseteq O_M$. Let $O' = \{o \in O_P : \Pr_r[P(x,r) = o] \Pr_r[P(x',r) = o] \neq 0\}$.

By the definition of O' and the fact that P satisfies $(\epsilon_2, 0)$-DP, we have that

$$\Pr_r[M(x,r) \in S, P(x,r) \in O_P \setminus O'] = 0$$

and therefore

$$\begin{aligned}
\Pr_r[M(x,r) \in S] &= \sum_{o \in O'} \Pr_r[M(x,r) \in S | P(x,r) = o] \Pr_r[P(x,r) = o] \\
&\leq e^{\epsilon_1} \left(\sum_{o \in O'} \Pr_r[M(x',r) \in S | P(x',r) = o] \Pr_r[P(x,r) = o] \right) + \delta_1 \\
&\leq e^{\epsilon_1 + \epsilon_2} \Pr_r[M(x',r) \in S] + \delta_1.
\end{aligned}$$

So M satisfies $(\epsilon_1 + \epsilon_2, \delta_1)$-DP. ◀

▶ Remark. One can extend this theorem to account for a non-zero δ_2 value, at a cost of an additional $(1 + |O'|)\delta_2$ in the δ.

Proof of Theorem 9. Consider any subset $S \subseteq O_M$, neighboring databases $x \sim x'$, and output $o \in O_P$ such that $\Pr_r[P(x,r) = o] \Pr_r[P(x',r) = o] \neq 0$. Combining the definitions of (ϵ', δ')-independence and (ϵ, δ)-DP gives us

$$\begin{aligned}
\Pr_r[M(x,r) \in S | P(x,r) = o] &\leq e^{\epsilon'} \Pr_r[M(x,r) \in S] + \delta' \\
&\leq e^{\epsilon + \epsilon'} \Pr_r[M(x',r) \in S] + e^{\epsilon'} \delta + \delta' \\
&\leq e^{\epsilon + 2\epsilon'} \Pr_r[M(x',r) \in S | P(x',r) = o] + (e^{\epsilon' + \epsilon} + 1)\delta' + e^{\epsilon'} \delta
\end{aligned}$$

and therefore (M, P) satisfies $(\epsilon + 2\epsilon', (e^{\epsilon' + \epsilon} + 1)\delta' + e^{\epsilon'} \delta)$-blDP. ◀

B Some Missing Proofs from Section 4

Proof of Lemma 10. Suppose we have some mechanism $M : \mathcal{X}^n \times R \to O_M$ that is a convex combination of the mechanisms $M_1, ..., M_k : \mathcal{X}^n \times R \to O_M$ such that for each $1 \leq i \leq k$, we have $\Pr_r[M = M_i] = a_i$ for some $a_1, ..., a_k \geq 0$ such that $\sum_{i=1}^k a_i = 1$. Suppose that each M_i satisfies (ϵ, δ)-blDP with respect to a function $P : \mathcal{X}^n \times R \to O_P$.

Now, consider any neighboring databases $x \sim x'$, $S \subseteq O$, and $o \in O_P$ such that $\Pr_r[P(x,r) = o] \Pr_r[P(x',r) = o] \neq 0$. Then we have that

$$\begin{aligned}
\Pr_r[M(x,r) \in S | P(x,r) = o] &= \sum_{i=1}^k \Pr[M = M_i] \Pr_r[M_i(x,r) \in S | P(x,r) = o] \\
&\leq \sum_{i=1}^k a_i (e^{\epsilon} \Pr_r[M_i(x',r) \in S | P(x',r) = o] + \delta) \\
&= e^{\epsilon} \left(\sum_{i=1}^k a_i \Pr_r[M_i(x',r) \in S | P(x',r) = o] \right) + \delta \left(\sum_{i=1}^k a_i \right) \\
&= e^{\epsilon} \Pr_r[M(x',r) \in S | P(x',r) = o] + \delta,
\end{aligned}$$

and so M satisfies (ϵ, δ)-blDP with respect to P. ◀

▶ **Example 18** (A blDP mechanism that fails group privacy). The parity function paired with any arbitrary privacy mechanism acts as a simple example of how a mechanism/function pair can fail to satisfy any sort of group privacy guarantee. If we think of the database as being represented as a vector in $\{0,1\}^n$, and choose the leakage function to be the parity of that vector, then any neighboring databases will have different parity. So, any arbitrary mechanism trivially satisfies perfect blDP with respect to the parity function. However, we can make no guarantees about non-neighboring databases that may share the same parity, and so can make no guarantees about group blDP for the same mechanism/function pair.

▶ Remark 19. The above is also an example of how certain leakage functions can cause blDP to be trivially satisfied when the probability that neighboring databases will leak the same output is zero. If the leakage function has even a small amount of noise, guaranteeing that neighboring databases always have non-zero probability to agree on the leakage (such as in an ϵ- DP with large ϵ), then blDP is guaranteed to be non-trivial. We also note that an alternative definition of blDP could restrict the behavior of the mechanism across *any* two databases inducing the same result under the leakage function, while scaling the corresponding constraint on probabilities according to the Hamming distance between the databases. This strictly stronger definition is worthy of further investigation.

▶ **Definition 20.** *Given a database universe $\mathcal{X}^n$, a* path *of length n between databases x and x' is a sequence of $n+1$ databases from $\mathcal{X}^n$, $x = x_0, x_1, ..., x_n = x'$, such that for all $i \in \{0, ..., n-1\}$, the databases x_i and x_{i+1} are adjacent.*

▶ **Definition 21.** *Given a function $P : \mathcal{X}^n \times R \to O_P$, a path $P = (x_0, ..., x_n)$ is* non-zero *for an output $o \in O_P$ if for all $i \in \{0, ..., n\}$, we have $\Pr_r[P(x_i, r) = o] > 0$.*

▶ **Theorem 22** (Group privacy). *Consider a mechanism M that satisfies (ϵ, δ)-blDP with respect to a function P, and two databases x and x'. If for a particular output $o \in O_P$ there exists a non-zero path of length k between x and x', then for any $S \subseteq O_M$, we have*

$$\Pr_r[M(x,r) \in S | P(x,r) = o] \leq e^{k\epsilon} \Pr_r[M(x',r) \in S | P(x',r) = o] + \delta \left(\frac{e^{k\epsilon} - 1}{e^{\epsilon} - 1} \right).$$

The proof of this theorem follows the same approach as the proof of group privacy in the standard DP setting, and is omitted here.

Proof of Theorem 12. Consider two neighboring databases $x \sin x'$, a subset $S \subseteq O_M \times O_P$, and an output $o \in O_P$ such that $\Pr_r[P(x,r) = o] \Pr_r[P(x',r) = o] \neq 0$. Then,

$$\Pr_r[M'(x,r) \in S | P(x,r) = o] = \Pr_r[M(x,r) \in S_o | P(x,r) = o]$$

where $S_o = \{y \in O_M : (y, o) \in S\}$, with the same holding true when x is replaced with x'.

Therefore, combining the blDP properties of (M, P) and the above equalities gives

$$\Pr_r[M'(x,r) \in S | P(x,r) = o] \leq e^{\epsilon} \Pr_r[M'(x',r) \in S | P(x',r) = o] + \delta$$

and thus M' satisfies (ϵ, δ)-blDP with respect to P. ◀

▶ **Proposition 23.** *Consider any mechanism M, leakage function P, output o of P, and databases x_0 and x_1 such that $\Pr_r[P(x_0,r) = o] \cdot \Pr[P(x_1,r) = o] \neq 0$. Then for any subset $S \subseteq O_M$ and $b \in \{0,1\}$, we have that*

$$\Pr_{r \in (R_{o,x_0} \times R_{o,x_1})}[M_{P,o}^{x_0,x_1}(x_b, r) \in S] = \Pr_{r \in R}[M(x_b, r) \in S | P(x_b, r) = o].$$

▶ **Proposition 24.** *Suppose that a mechanism M satisfies (ϵ, δ)-blDP with respect to a leakage function P. Then, for any $o \in O_P$ and neighboring databases $x_0 \sim x_1$, the DP reduction mechanism $M_{P,o}^{x_0,x_1}$ satisfies (ϵ, δ)-DP.*

The proofs of Propositions 23 and 24 are omitted here for space reasons, but are easily verified.

▶ Remark 25. Another simpler construction can be used to prove a reduction in the reverse direction, reducing DP composition to the blDP setting.

C Details on Non-Adaptive Composition

The following theorem shows how to translate a non-adaptive composition theorem for differential privacy into one for blDP.

▶ **Theorem 26.** *For some $k \geq 1$, suppose that the following implication were to hold: if for any choice of k mechanisms $L_1, ..., L_k$ such that each L_i satisfies (ϵ_i, δ_i)-DP, then the composition of these mechanisms,*

$$L(x, (r_1, ..., r_k)) := L_1(x, r_1)||L_2(x, r_2)||...||L_k(x, r_k),$$

(where each r_i is chosen independently at random) would satisfy (ϵ', δ')-DP. Then, for any choice of k mechanism function pairs $(M_1, P_1), ..., (M_k, P_k)$ such that each M_i satisfies (ϵ_i, δ_i)-blDP with respect to P_i, if we define the composed functions

$$M(x, (r_1, ..., r_k)) := M_1(x, r_1)||...||M(x, r_k) \text{ and } P(x, (r_1, ..., r_k)) := P_1(x, r_1)||...||P_k(x, r_k),$$

then M must also satisfy (ϵ', δ')-blDP with respect to P.

Proof. Consider any neighboring databases $x_0 \sim x_1$, some subset $S \subseteq O_M$, and some $o = (o_1, ..., o_k) \in O_P$ such that $\Pr_r[P(x_0, r) = o] \Pr_r[P(x_1, r) = o] \neq 0$. We note that this requirement implies that $\Pr_r[P_i(x_0, r) = o_i] \Pr_r[P_i(x_1, r) = o_i] \neq 0$ for all i. Then, consider the k DP reduction functions, $(M_1)_{P_1,o_1}^{x_0,x_1}, ..., (M_k)_{P_k,o_k}^{x_0,x_1}$, defined in terms of each (M_i, P_i). By Proposition 24, each $(M_i)_{P_i,o_i}^{x_0,x_1}$ must satisfy (ϵ_i, δ_i)-DP.

Therefore, by our assumption, the composition

$$M'(x, (r_1, ..., r_k)) = (M_1)_{P_1,o_1}^{x_0,x_1}(x, r_1)||...||(M_k)_{P_k,o_k}^{x_0,x_1}(x, r_k)$$

must satisfy (ϵ', δ')-DP.

We can express any subset S as the sum of disjoint rectangles, so it suffices to assume that S is a rectangle. So, $S = S_1 \times S_2 \times ... \times S_k$ where each $S_i \subseteq O_{M_i}$. Then because each r_i is chosen independently, we know that for any $b \in \{0, 1\}$,

$$\Pr[M'(x_b, (r_1, ..., r_k)) \in S] = \prod_{i=1}^{k} \Pr_r[M_i(x_b, r) \in S_i | P_i(x_b, r) = o_i],$$

where because each P_i uses independent randomness as well,

$$\prod_{i=1}^{k} \Pr_r[M_i(x_b, r) \in S_i | P_i(x_b, r) = o_i] = \Pr_r[M(x_b, r) \in S | P(x_b, r) = o].$$

Combining the DP guarantee for M' and the above equality gives us

$$\Pr_r[M(x_0, r) \in S | P(x_0, r) = o] \leq e^{\epsilon'} \Pr_r[M(x_1, r) \in S | P(x_1, r) = o] + \delta'$$

Therefore the composed mechanism M must satisfy (ϵ', δ')-blDP with respect to P. ◀

This result tells us that any nonadaptive composition bounds for the DP setting can be extended to the blDP setting. In particular, we present a corollary below that is reached by applying this statement to a well-known composition theorem for DP.

We first recall the following theorem:

▶ **Theorem 27** ([5]). *Suppose $M_1, ..., M_k$ are mechanisms such that M_i satisfies (ϵ_i, δ_i)-DP. Then the composition of these mechanisms, $M(x, (r_1, ..., r_k)) := M_1(x, r_1)||...||M_k(x, r_k)$, satisfies $(\sum_{i=1}^k \epsilon_i, \sum_{i=1}^k \delta_i)$-DP.*

The following corollary is a direct result of combining this theorem with Theorem 26:

▶ **Corollary 28.** *Suppose $(M_1, P_1), ..., (M_k, P_k)$ are mechanism-function pairs such that each M_i satisfies (ϵ_i, δ_i)-blDP with respect to P_i. Then the composition of these mechanisms,*

$$M(x, (r_1, ..., r_k)) := M_1(x, r_1)||...||M_k(x, r_k)$$

satisfies $(\sum_{i=1}^k \epsilon_i, \sum_{i=1}^k \delta_i)$-blDP with respect to the composition of the P_is,

$$P(x, (r_1, ..., r_k)) := P_1(x, r_1)||...||P_k(x, r_k).$$

Therefore, we can conclude that, like for differential privacy, the rate of bounded-leakage privacy loss as we increase the number of queries to the database, is at most linear.

D Adaptive Composition

It is also important to consider how bounded-leakage-privacy can be affected if the composed mechanisms are chosen adaptively based on the outputs of the previously chosen mechanisms. We analyze how this form of composition affects privacy via an experiment/adversary model. We define two experiments: Experiment 0 and Experiment 1, as follows:

Algorithm 1 Experiment b: blDP of Adaptive k-Fold Composition.

Input: a family of mechanism-function pairs $\mathcal{F} = \{(M_1, P_1), (M_2, P_2), ...\}$, and a probabilistic adversary $\mathcal{A}$.

Repeat k **times:**

$\mathcal{A}$ outputs some query $((x_0, x_1), (M_i, P_i), o_i)$ where (x_0, x_1) is a pair of adjacent databases, $(M_i, P_i) \in \mathcal{F}$, and o_i is some member of the output space of P_i.

if $\Pr[P_i(x_0, r) = o_i] \cdot \Pr[P_i(x_1, r) = o_i] = 0$ **then** $\mathcal{A}$ receives *"null"*.

else $\mathcal{A}$ receives $M_i(x_b, r)$ for some random $r \in R$ such that $P_i(x_b, r) = o_i$.

Intuitively, the definition of bounded-leakage privacy states that if a mechanism-function pair has good bounded-leakage privacy, it should be hard to differentiate the outputs of the mechanism on two adjacent databases for some fixed function output. Extending this to the composition setting, it should still be difficult to distinguish which database was used even if we are able to get more information with multiple queries. Connecting this to the experiments, the adversary should have difficulty distinguishing between the outputs of Experiments 0 and 1 if our family of mechanisms and functions has good bounded-leakage privacy.

To formalize this, we define the "view" of the adversary in a particular experiment to be everything that the adversary sees or knows after the experiment, i.e. the contents of all the adversary's k queries and the responses to those queries. It should be noted that this will not include the random coin flips used to generate the responses to each query, nor whether $b = 0$

or 1. An adversary's view can be denoted by the values of all the queries and their responses, i.e., a "transcript" of the experiment, or just the values of the random coin flips that the adversary used and all of the responses. These are equivalent representations because an adversary's queries can always be reconstructed from the randomness that the adversary used and the responses that it received, and so we will use these two representations of the view interchangeably throughout.

Now that we have formalized this notion of an adversary's view, we can use our model to define bounded-leakage privacy under adaptive composition as follows:

▶ **Definition 29.** *We say that a family $\mathcal{F}$ of mechanism-function pairs satisfies* (ϵ, δ)-blDP under k-fold adaptive composition *if for every adversary $\mathcal{A}$, random variables V^b corresponding to the view of $\mathcal{A}$ in Experiment b, and subset of possible views V, we have that*

$$\Pr[V_0 \in V] \leq e^{\epsilon} \Pr[V_1 \in V] + \delta$$

This definition is designed to parallel the definition for adaptive composition of DP mechanisms given by Dwork and Roth [5]. We include the DP version here so that the two can be easily compared:

Algorithm 2 Experiment b: DP of Adaptive k-Fold Composition [5].

Input: A family $\mathcal{F}$ of mechanisms and a probabilistic adversary $\mathcal{A}$.
Repeat k **times:**
  $\mathcal{A}$ outputs the query $((x_0, x_1), M_i)$ where (x_0, x_1) is a pair of adjacent databases and $M_i \in \mathcal{F}$.
  $\mathcal{A}$ recieves $M_i(x_b, r)$ for some random $r \in R$.

▶ **Definition 30** ([5])**.** *We say that the family of mechanisms $\mathcal{F}$ satisfies (ϵ, δ)-DP under k-fold adaptive composition if for every adversary $\mathcal{A}$, we have $D_\infty^\delta(V_0||V_1) \leq \epsilon$, where V_b denotes the view of A in the DP composition experiment, and $D_\infty^\delta(V_0||V_1)$ is the δ-approximate max divergence between V_0 and V_1, defined as*

$$D_\infty^\delta(V_0||V_1) = \max_{S \subseteq Supp(V_0): \Pr[V_0 \in V] \geq \delta} \left[\ln \frac{\Pr[V_0 \in S] - \delta}{\Pr[V_1 \in S]} \right].$$

We note that this max-divergence definition is equivalent to requiring that for all adversaries and subsets of views of the DP experiment, V, we have $\Pr[V_0 \in V] \leq e^{\epsilon} \Pr[V_1 \in V] + \delta$, which puts the definition in a more familiar form.

We will now connect this model to the standard DP setting. The following theorem states that any bounds for adaptive composition that can be shown to hold in the DP setting must also hold in the blDP setting:

▶ **Theorem 31.** *If a class of (ϵ, δ)-DP mechanism-function pairs satisfies (ϵ', δ')-DP under k-fold adaptive composition, then that class of (ϵ, δ)-blDP mechanisms satisfies (ϵ', δ')-blDP under k-fold adaptive composition.*

Similar to our nonadaptive composition result, the proof of this theorem uses the strategy of reducing the blDP setting to the DP setting by converting any arbitrary blDP adversary to an adversary in the DP setting with the same distribution of views. We can then argue that therefore the original blDP adversary must be constrained by the same bounds as the DP adversary.

Proof of Theorem 31. For the purposes of this proof, we will consider the views of the adversary in both the DP and blDP adaptive composition experiments to contain only the value of the adversary's random bits and the responses it receives for each query so that we can easily compare the views in the DP and blDP settings.

First, assume that the class of (ϵ, δ)-DP mechanisms satisfies (ϵ', δ')-DP under k-fold adaptive composition. Now, consider some adversary $\mathcal{A}^{blDP}$ for the blDP composition experiments for the class of (ϵ, δ)-blDP mechanism-function pairs.

Using $\mathcal{A}^{blDP}$, we construct an adversary for the DP composition experiment on (ϵ, δ)-DP mechanisms as follows: whenever $\mathcal{A}^{blDP}$ would output the query $((x_0, x_1), (M, P), o)$ given the current view of the experiment, $\mathcal{A}^{DP}$ outputs $((x_0, x_1), M_{P,o}^{x_0,x_1})$, where $M_{P,o}^{x_0,x_1}$ is the DP reduction mechanism for $(x_0, x_1), o$, and P.

By Proposition 24, $M_{P,o}^{x_0,x_1})$ satisfies (ϵ, δ)-DP and therefore $\mathcal{A}^{DP}$ is a valid adversary for the class of (ϵ, δ)-DP mechanisms.

Now, we want to show that given this definition of $\mathcal{A}^{DP}$, if V_b^{DP} is a random variable for the view of $\mathcal{A}^{DP}$ in the DP composition Experiment b and V_b^{blDP} is a random variable for the view of $\mathcal{A}^{blDP}$ in the blDP composition Experiment b, then we have $dist(V_b^{DP}) = dist(V_b^{blDP})$.

We split both views into their component random variables corresponding to the randomness of the adversaries and the responses in the experiment such that

$$V_b^{DP} = (R^{DP}, S_1^{DP}, ..., S_k^{DP}) \quad \text{and} \quad V_b^{blDP} = (R^{blDP}, S_1^{blDP}, ..., S_k^{blDP}),$$

where R^{DP} and R^{blDP} are random variables corresponding to the random bits of the adversaries, and each S_i is the response received for the ith query in the experiment.

We will use induction on the number of outputs to show that these distributions must be equal. First, because $\mathcal{A}^{DP}$ uses no additional randomness apart from the randomness of $\mathcal{A}^{blDP}$, we clearly have $dist(R^{DP}) = dist(R^{blDP})$. This forms our base case.

Now, suppose that for some i with $1 \leq i \leq k$, we have

$$dist((R^{DP}, S_1^{DP}, ..., S_{i-1}^{DP})) = dist((R^{blDP}, S_1^{blDP}, ..., S_{i-1}^{blDP})).$$

Then, for any partial view $(r, s_1, ..., s_i)$, we can rewrite $\Pr[(R^{DP}, ..., S_i^{DP}) = (r, ..., s_i)]$ as

$$\Pr[(R^{DP}, ..., S_{i-1}^{DP}) = (r, ..., s_{i-1})] \Pr[S_i^{DP} = s_i | (R^{DP}, ..., S_{i-1}^{DP}) = (r, ..., s_{i-1})],$$

where by fixing $(r, s_1, ..., s_{i-1})$, the ith query from $\mathcal{A}^{DP}$ is deterministically fixed to be some $((x_0, x_1), (M_i, P_i), o_i)$, and therefore the ith query of $\mathcal{A}^{DP}$ is fixed to be $((x_0, x_1), (M_i)_{P_i,o_i}^{x_0,x_1})$. By Proposition 23, if $\Pr_r[P_i(x_0, r) = o_i] \Pr_r[P_i(x_1, r) = o_i] = 0$, then $(M_i)_{P_i,o_i}^{x_0,x_1}(x_b, r)$ will output "null" with probability one. Otherwise, we have that

$$dist_r((M_i)_{P_i,o_i}^{x_0,x_1}(x_b, r)) = dist_{r:P(x_b,r)=o_i}(M_i(x_b, r))$$

and therefore in either case,

$$\begin{aligned} dist(S_i^{DP} | (R^{DP}, ..., S_{i-1}^{DP}) = (r, ..., s_{i-1})) &= dist_r((M_i)_{P_i,o_i}^{x_0,x_1}(x_b, r)) \\ &= dist(S_i^{blDP} | (R^{blDP}, ..., S_{i-1}^{blDP}) = (r, ..., s_{i-1})) \end{aligned}$$

By our inductive assumption, $dist((R^{blDP}, ..., S_{i-1}^{blDP})) = dist((R^{DP}, ..., S_{i-1}^{DP}))$. Putting these together, we must have $dist((R^{blDP}, ..., S_i^{blDP})) = dist((R^{DP}, ..., S_i^{DP}))$. This completes our inductive step, and therefore it follows by induction that

$$dist(V_b^{blDP}) = dist((R^{blDP}, ..., S_k^{blDP})) = dist((R^{DP}, ..., S_k^{DP})) = dist(V_b^{DP})$$

$$dist(V_b^{blDP}) = dist(V_b^{DP}).$$

Because by our initial assumption the class of (ϵ, δ)-DP mechanisms satisfies (ϵ', δ')-DP under k-fold adaptive composition, we must also have that for any subset of views V, we have

$$\Pr[V_0^{blDP} \in V] = \Pr[V_0^{DP} \in V] \leq e^{\epsilon'} \Pr[V_1^{DP} \in V] + \delta' = e^{\epsilon'} \Pr[V_1^{blDP} \in V] + \delta'$$

$$\Pr[V_0^{blDP} \in V] \leq e^{\epsilon'} \Pr[V_1^{blDP} \in V] + \delta'$$

Therefore the class of (ϵ, δ)-blDP mechanisms must also satisfy (ϵ', δ')-blDP under k-fold adaptive composition. ◀

Theorem 31 allows us to apply existing bounds for the adaptive composition of DP mechanisms to the blDP context. Recall the following composition bound for DP mechanisms:

▶ **Theorem 32** ([6]). *For all $\epsilon, \delta, \delta' \geq 0$, the class of (ϵ, δ)-DP mechanisms satisfies $(\epsilon', k\delta+\delta')$-DP under k-fold adaptive composition for:*

$$\epsilon' = \epsilon\sqrt{2k\ln(1/\delta')} + k\epsilon(e^{\epsilon-1}).$$

Combining the results of Theorem 31 and Theorem 32 gives us Theorem 16 as a corollary.

E The Exponential Mechanism

▶ **Definition 33.** *Given a set of outputs O_M and a set of outputs O_P, a* coupled utility function for O_M and O_P *is some function $u_{M,P} : \mathcal{X}^n \times O_M \times O_P \to \mathbb{R}$ that maps triples containing a database, an element of O_M, and an element of O_P to a real-valued score.*

We define this coupled utility function with the intent of defining O_M to be the output space of a particular mechanism, and O_P to be the output space of some associated function. This definition would allow us to define utility functions in the bounded-leakage setting that are inherently stronger than just considering a standard utility function conditioned on a particular output of the leaked function, because here we can have the utility function depend on the output of the associated function even if it is randomized.

We also want to define an analogous notion of utility sensitivity for coupled utilities.

▶ **Definition 34.** *Given a coupled utility function $u_{M,P} : \mathcal{X}^n \times O_M \times O_P \times \mathbb{R}$, we define a function corresponding to the sensitivity of $u_{M,P}$ conditioned on a particular element of O_P, $\Delta u_{M,P} : O_P \to \mathbb{R}$, such that for any $o \in O_P$,*

$$\Delta u_{M,P}(o) = \max_{y \in O_M} \max_{x \sim x'} |u_{M,P}(x, y, o) - u_{M,P}(x', y, o)|$$

This quantifies the sensitivity for our coupled utility given a particular value for the output of P.

Using this new concept of a coupled utility function, we can define a version of the exponential mechanism for blDP as follows:

▶ **Definition 35** (The Exponential Mechanism for blDP). *Given a set of outputs, O_M, a function $P : \mathcal{X}^n \times \mathcal{R} \to O_P$, and a coupled utility function $u_{M,P} : \mathcal{X}^n \times O_M \times O_P \to \mathbb{R}$, the* bounded-leakage exponential mechanism *$M_{E,P}(x, u_{M,P}, O_M, r)$ is defined such that if $P(x, r) = o$, then for any $y \in O_M$, the probability that the mechanism outputs y is proportional to*

$$\exp\left(\frac{\epsilon u_{M,P}(x, y, o)}{2\Delta u_{M,P}(o)}\right).$$

We can now show that in the same way that the standard exponential mechanism guarantees $(\epsilon, 0)$-DP, this version of the mechanism guarantees $(\epsilon, 0)$-blDP.

▶ **Theorem 36.** *The exponential mechanism for blDP satisfies $(\epsilon, 0)$-blDP with respect to its associated function P.*

Proof. Consider any two neighboring databases $x \sim x'$, some output o of P such that $\Pr_r[P(x,r) = o] \cdot \Pr_r[P(x',r) = o] \neq 0$, and some output $s \in O_M$. Applying the definition of the mechanism and utility sensitivity, we have that

$$\frac{\Pr_r[M_{E,P}(x, u_{M,P}, O_M, r) = s | P(x,r) = o]}{\Pr_r[M_{E,P}(x', u_{M,P}, O_M, r) = s | P(x',r) = o]}$$

is at most

$$\exp\left(\frac{\epsilon \Delta u_{M,P}(o}{2\Delta u_{M,P}(o)}\right) \frac{\sum_{y \in O_M} \exp\left(\frac{\epsilon(u_{M,P}(x,y,o) + \Delta u_{M,P}(o))}{2\Delta u_{M,P}(o)}\right)}{\sum_{y \in O_M} \exp\left(\frac{\epsilon u_{M,P}(x,y,o)}{2\Delta u_{M,P}(o)}\right)} = \exp(\epsilon)$$

Therefore, for any subset $S \subseteq O_M$, we have

$$\begin{aligned}\Pr_r[M_{E,P}(x, u_{M,P}, O_M, r) \in S | P(x,r) = o] \\ &\leq \sum_{s \in S} e^{\epsilon} \Pr_r[M_{E,P}(x', u_{M,P}, O_M, r) = s | P(x',r) = o] \\ &= e^{\epsilon} \Pr_r[M_{E,P}(x', u_{M,P}, O_M, r) \in S | P(x',r) = o],\end{aligned}$$

and so $M_{E,P}$ satisfies $(\epsilon, 0)$-blDP with respect to P. ◀

As in the case of the standard exponential mechanism, we also want to show that our mechanism for the bounded-leakage case can give us some guarantee of "good" utility. In the case of the standard exponential mechanism, we recall the following theorem:

▶ **Theorem 37** ([13])**.** *Let M_E be the standard exponential mechanism for a set of outputs S and utility function u. For any database x, let $OPT_u(x) = \max_{y \in S} u(x,y)$, and $S_{OPT} = \{y \in O_M : u(x,y) = OPT_u(x)$. Then, for any $t \in \mathbb{R}$, we have that*

$$\Pr_r[u(M_E(x,u,S,r)) \leq OPT_u(x) - \frac{2\Delta u}{\epsilon}\left(\ln\left(\frac{|S|}{|S_{OPT}|}\right) + t\right)] \leq e^{-t}$$

By the properties of our exponential mechanism for bounded-leakage privacy, we can conclude the following analogous theorem in the blDP setting:

▶ **Theorem 38.** *Let $M_{E,P}$ be the standard bounded-leakage exponential mechanism for a set of outputs S, function P, and utility function $u_{S,P}$. For any database x and output o of P, let $OPT_{u_{S,P}}(x,o) = \max_{y \in S} u_{S,P}(x,y,o)$, and $S_{OPT} = \{y \in O_M : u_{S,P}(x,y,o) = OPT_u(x)$. Then, for any $t \in \mathbb{R}$, database x, and output o of P, we have that*

$$\Pr_{r:P(x,r)=o}[u_{S,P}(M_{E,P}(x, u_{S,P}, S, r)) \leq OPT_{u_{S,P}}(x,o) - \frac{2\Delta u_{S,P}(o)}{\epsilon}\left(\ln\left(\frac{|S|}{|S_{OPT}|}\right) + t\right)] \leq e^{-t}.$$

Proof. This result follows immediately from combining Theorem 37 and the observation that once the output of P is set, $M_{E,P}$ behaves like the standard exponential mechanism for output space S and utility $u(x,y) = u_{S,P}(x,y,o)$. ◀

F Additional Details on the BigWorld Application of blDP

In this section, we provide a proof for the result stated in Theorem 17. To begin, we have the following lemma:

▶ **Lemma 39.** *Let V_i be a leakage function that releases the exact number of studies that i participated in. Then, for any $S \subseteq O_M$, $t \leq k$, and $D \sim D_i$ such that D_i differs from D only in i's data, we have that*

$$\Pr_{\overline{r},r_{par}}[\overline{M}(D,r_{par},\overline{r}) \in S | V_i(D,r_{par}) = t] \leq e^{2t\epsilon} \Pr_{\overline{r},r_{par}}[\overline{M}(D_i,r_{par},\overline{r}) \in S | V_i(D_i,r_{par}) = t] + 2t\delta$$

Proof. We first note that because of our assumption that the participation of i is independent of the participation of all other individuals in D, we can consider the distribution of a particular study $M^{(j)}$'s output to be a convex combination of neighboring databases differing only in i's data.

If i does not participate in study j in either case, then neither mechanism can depend on i, so the distributions of possible results conditioned on v and v' will be equal. Otherwise, we apply the (ϵ,δ)-DP to conclude that

$$\Pr_{r,r_{par}}[M^{(j)}(f^{(j)}_{par}(D,r_{par}),r) \in S_j | V_i(D,r_{par}) = t]$$

$$\leq e^{\epsilon} \Pr_{r,r_{par}}[M^{(j)}(f^{(j)}_{par}(D_i,r_{par}),r) \in S_j | V_i(D_i,r_{par}) = t] + \delta$$

By the definition of our leakage function, the maximum number of j such that i participates in at least one of the two versions of each study is $2t$. So, $\overline{M}$ can be viewed as the composition of at most $2t$ (ϵ,δ)-blDP mechanisms, all with respect to V_i. Meanwhile all other mechanisms are perfectly blDP. Therefore, using the standard composition bound for blDP mechanisms (Corollary 28) we get the desired inequality. ◀

Proof of Theorem 17. With this result in hand, we can now consider our original leakage function P_i, which leaks an upper bound for the magnitude of the participation vector.

For any particular t, we will have that $\Pr_{\overline{r},r_{par}}[\overline{M}(D,r_{par},\overline{r}) \in S | P_i(D,r_{par}) = t]$ is a convex combination of the set of probabilities

$$\left\{ \Pr_{\overline{r},r_{par}}[\overline{M}(D,r_{par},\overline{r}) \in S | V_i(D,r_{par}) = j] \right\}_{0 \leq j \leq t}.$$

By Lemma 39, each of these satisfies $(2t\epsilon, 2t\delta)$-blDP, and so applying Lemma 10 gives the desired inequality. ◀

We should note that we applied the simplest composition bound for (ϵ,δ)-DP or blDP mechanisms in this case, but any bound could be substituted for an analogous result.