# The Privacy Blanket of the Shuffle Model

Borja Balle, James Bell[1]*, Adrià Gascón[1,2] *, and Kobbi Nissim[3]**

[1] The Alan Turing Institute
[2] University of Warwick
[3] Georgetown University

**Abstract.** This work studies differential privacy in the context of the recently proposed *shuffle model*. Unlike in the local model, where the server collecting privatized data from users can track back an input to a specific user, in the shuffle model users submit their privatized inputs to a server anonymously. This setup yields a trust model which sits in between the classical curator and local models for differential privacy. The shuffle model is the core idea in the Encode, Shuffle, Analyze (ESA) model introduced by Bittau et al. (SOPS 2017). Recent work by Cheu et al. (EUROCRYPT 2019) analyzes the differential privacy properties of the shuffle model and shows that in some cases shuffled protocols provide strictly better accuracy than local protocols. Additionally, Erlingsson et al. (SODA 2019) provide a privacy amplification bound quantifying the level of curator differential privacy achieved by the shuffle model in terms of the local differential privacy of the randomizer used by each user.

In this context, we make three contributions. First, we provide an optimal single message protocol for summation of real numbers in the shuffle model. Our protocol is very simple and has better accuracy and communication than the protocols for this same problem proposed by Cheu et al. Optimality of this protocol follows from our second contribution, a new lower bound for the accuracy of private protocols for summation of real numbers in the shuffle model. The third contribution is a new amplification bound for analyzing the privacy of protocols in the shuffle model in terms of the privacy provided by the corresponding local randomizer. Our amplification bound generalizes the results by Erlingsson et al. to a wider range of parameters, and provides a whole family of methods to analyze privacy amplification in the shuffle model.

**Keywords:** Differential Privacy · Privacy Amplification · Secure Shuffling.

* Work supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1, and the UK Government's Defence & Security Programme in support of the Alan Turing Institute.

** Work supported by NSF grant no. 1565387, TWC: Large: Collaborative: Computing Over Distributed Sensitive Data. Work partly done while K. N. was visiting the Alan Turing Institute.

# 1 Introduction

Most of the research in differential privacy focuses on one of two extreme models of distribution. In the curator model, a *trusted* data collector assembles users' sensitive personal information and analyses it while injecting random noise strategically designed to provide both differential privacy and data utility. In the local model, each user $i$ with input $x_i$ applies a local randomizer $\mathcal{R}$ on her data to obtain a message $y_i$, which is then submitted to an *untrusted* analyzer. Crucially, the randomizer $\mathcal{R}$ guarantees differential privacy independently of the analyzer and the other users, even if they collude. Separation results between the local and curator models are well-known since the early research in differential privacy: certain learning tasks that can be performed in the curator model cannot be performed in the local model [23] and, furthermore, for those tasks that can be performed in the local model there are provable large gaps in accuracy when compared with the curator model. An important example is the summation of binary or (bounded) real-valued inputs among $n$ users, which can be performed with $O(1)$ noise in the curator model [14] whereas in the local model the noise level is $\Omega(\sqrt{n})$ [7,11]. Nevertheless, the local model has been the model of choice for recent implementations of differentially private protocols by Google [16], Apple [25], and Microsoft [13]. Not surprisingly, these implementations require a huge user base to overcome the high error level.

The high level of noise required in the local model has motivated a recent search for alternative models. For example, the Encode, Shuffle, Analyze (ESA) model introduces a trusted shuffler that receives user messages and permutes them before they are handled to an untrusted analyzer [9]. A recent work by Cheu et al. [12] provides a formal analytical model for studying the shuffle model and protocols for summation of binary and real-valued inputs, essentially recovering the accuracy of the trusted curator model. The protocol for real-valued inputs requires users to send multiple messages, with a total of $O(\sqrt{n})$ single bit messages sent by each user. Also of relevance is the work of Ishai et al. [18] showing how to combine secret sharing with secure shuffling to implement distributed summation, as it allows to simulate the Laplace mechanism of the curator model. Instead we focus on the single-message shuffle model.

Another recent work by Erlingsson et al. [15] shows that the shuffling primitive provides privacy amplification, as introducing random shuffling in local model protocols reduces $\varepsilon$ to $\varepsilon/\sqrt{n}$.

A word of caution is in place with respect to the shuffle model, as it differs significantly from the local model in terms of the assumed trust. In particular, the privacy guarantee provided by protocols in the shuffle model degrades with the fraction of users who deviate from the protocol. This is because, besides relying on a trusted shuffling step, the shuffle model requires users to provide messages carefully crafted to protect each other's privacy. This is in contrast with the curator model, where this responsibility is entirely held by the trusted curator. Nevertheless, we believe that this model is of interest both for theoretical and practical reasons. On the one hand it allows to explore the space in between the local and curator model, and on the other hand it leads to mechanisms that are easy to explain, verify, and implement; with limited accuracy loss with respect to the curator model.

In this work we do not assume any particular implementation of the shuffling step. Naturally, alternative implementations will lead to different computational trade-offs and trust assumptions. The shuffle model allows to disentangle these aspects from the precise computation at hand, as the result of shuffling the randomized inputs submitted by each user is required to be differentially private, and therefore any subsequent analysis performed by the analyzer will be private due to the postprocessing property of differential privacy.

### 1.1 Overview of Our Results

In this work we focus on single-message shuffle model protocols. In such protocols (i) each user $i$ applies a local randomizer $\mathcal{R}$ on her input $x_i$ to obtain a single message $y_i$; (ii) the messages $(y_1, \dots, y_n)$ are shuffled to obtain $(y_{\sigma(1)}, \dots, y_{\sigma(n)})$ where $\sigma$ is a randomly selected permutation; and (iii) an analyzer post-processes $(y_{\sigma(1)}, \dots, y_{\sigma(n)})$ to produce an outcome. It is required that the mechanism resulting from the combination of the local randomizer $\mathcal{R}$ and the random shuffle should provide differential privacy.

**A protocol for private summation.** Our first contribution is a single-message shuffle model protocol for private summation of (real) numbers $x_i \in [0, 1]$. The resulting estimator is unbiased and has standard deviation $O_{\varepsilon,\delta}(n^{1/6})$.

To reduce the domain size, our protocol uses a fixed-point representation, where users apply randomized rounding to snap their input $x_i$ to a multiple $\bar{x}_i$ of $1/k$ (where $k = O_{\varepsilon,\delta}(n^{1/3})$). We then apply on $\bar{x}_i$ a local randomizer $\mathcal{R}^{PH}$ for computing private histograms over a finite domain of size $k+1$. The randomizer $\mathcal{R}^{PH}$ is simply a randomized response mechanism: with (small) probability $\gamma$ it ignores $\bar{x}_i$ and outputs a uniformly random domain element, otherwise it reports its input $\bar{x}_i$ truthfully. There are hence about $\gamma n$ instances of $\mathcal{R}^{PH}$ whose report is independent to their input, and whose role is to create what we call a *privacy blanket*, which masks the outputs which are reported truthfully. Combining $\mathcal{R}^{PH}$ with a random shuffle, we get the equivalent of a histogram of the sent messages, which, in turn, is the pointwise sum of the histogram of approximately $(1-\gamma)n$ values $\bar{x}_i$ sent truthfully and the privacy blanket, which is a histogram of approximately $\gamma n$ random values.

To see the benefit of creating a privacy blanket, consider the recent shuffle model summation protocol by Cheu et al. [12]. This protocol also applies randomized rounding. However, for privacy reasons, the rounded value needs to be represented in unary across multiple 1-bit messages, which are then fed into a summation protocol for binary values. The resulting error of this protocol is $O(1)$ (as is achieved in the curator model). However, the use of unary representation requires each user to send $O_\varepsilon(\sqrt{n})$ 1-bit messages (whereas in our protocol every user sends a single $O(\log n)$-bit message). We note that Cheu et al. also present a *single message* protocol for real summation with $O(\sqrt{n})$ error.

**A lower bound for private summation.** We also provide a matching lower bound showing that any single-message shuffled protocol for summation must exhibit mean squared error of order $\Omega(n^{1/3})$. In our lower bound argument we consider i.i.d. input distributions, for which we show that without loss of generality the local randomizer's image is the interval $[0, 1]$, and the analyzer is a simple summation of messages. With this view, we can contrast the privacy and accuracy of the protocol. On the one hand, the randomizer may need to output $y \in [0, 1]$ on input $x \in [0, 1]$ such that $|x - y|$ is small, to promote accuracy. However, this interferes with privacy as it may enable distinguishing between the input $x$ and a potential input $x'$ for which $|x' - y|$ is large.

Together with our upper bound, this result shows that the single-message shuffle model sits strictly between the curator and the local models of differential privacy. This had been shown by Cheu et al. [12] in a less direct way by showing that (i) the private selection problem can be solved more accurately in the curator model than the shuffle model, and (ii) the private summation problem can be solved more accurately in the shuffle model than in the local model. For (i) they rely on a generic translation from the shuffle to the local model and known lower bounds for private selection in the local model, while our lower bound operates directly in the shuffle model. For (ii) they propose a single-message protocol that is less accurate than ours.

**Privacy amplification by shuffling.** Lastly, we prove a new privacy amplification result for shuffled mechanisms. We show that shuffling $n$ copies of an $\varepsilon_0$-LDP local randomizer with $\varepsilon_0 = O(\log(n/\log(1/\delta)))$ yields an $(\varepsilon, \delta)$-DP mechanism with $\varepsilon = O((\varepsilon_0 \wedge 1)e^{\varepsilon_0}\sqrt{\log(1/\delta)/n})$, where $a \wedge b = \min\{a, b\}$. The proof formalizes the notion of a *privacy blanket* that we use informally in the privacy analysis of our summation protocol. In particular, we show that the output distribution of local randomizers (for any local differentially private protocol) can be decomposed as a convex combination of an *input-independent* blanket distribution and an *input-dependent* distribution.

Privacy amplification plays a major role in the design of differentially private mechanisms. These include amplification by subsampling [23] and by iteration [17], and the recent seminal work on amplification via shuffling by Erlingsson et al. [15]. In particular, Erlingsson et al. considered a setting more general than ours which allows for interactive protocols in the shuffle model by first generating a random permutation of the users' inputs and then sequentially applying a (possibly different) local randomizer to each element in the permuted vector. Moreover, each local randomizer is chosen depending on the output of previous local randomizers. To distinguish this setting from ours, we shall call the setting of Erlingsson et al. *shuffle-then-randomize* and ours *randomize-then-shuffle*. We also note that both settings are equivalent when there is a single local randomizer that will be applied to all the inputs. Throughout this paper, unless we explicitly say otherwise, the term *shuffle model* refers to the randomize-then-shuffle setting.

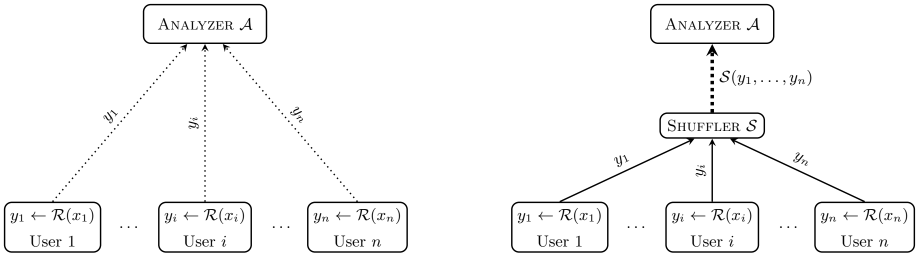
ANALYZER $\mathcal{A}$
$y_1$
$y_i$
$y_n$
$y_1 \leftarrow \mathcal{R}(x_1)$ User 1
$y_i \leftarrow \mathcal{R}(x_i)$ User $i$
$y_n \leftarrow \mathcal{R}(x_n)$ User $n$
$\dots$
ANALYZER $\mathcal{A}$
$\mathcal{S}(y_1, \dots, y_n)$
SHUFFLER $\mathcal{S}$
$y_1$
$y_i$
$y_n$
$y_1 \leftarrow \mathcal{R}(x_1)$ User 1
$y_i \leftarrow \mathcal{R}(x_i)$ User $i$
$y_n \leftarrow \mathcal{R}(x_n)$ User $n$
$\dots$

**Fig. 1.** The local (left) and shuffle (right) models of Differential Privacy. Dotted lines indicate differentially private values with respect to the dataset $\vec{x} = (x_1, \dots, x_n)$, where user $i$ holds $x_i$.

In the shuffle-then-randomize setting, Erlingsson et al. provide an amplification bound with $\varepsilon = O(\varepsilon_0 \sqrt{\log(1/\delta)/n})$ for $\varepsilon_0 = O(1)$. Our result in the randomize-then-shuffle setting recovers this bound for the case of one randomizer, and extends it to $\varepsilon_0$ which is logarithmic in $n$. For example, using the new bound, it is possible to shuffle a local randomizer with $\varepsilon_0 = O(\log(\varepsilon^2 n/\log(1/\delta)))$ to obtain a $(\varepsilon, \delta)$-DP mechanism with $\varepsilon = \Theta(1)$ . Cheu et al. [12] also proved that a level of LDP $\varepsilon_0 = O(\log(\varepsilon^2 n/\log(1/\delta)))$ suffices to achieve $(\varepsilon, \delta)$-DP mechanisms through shuffling, though only for binary randomized response in the randomize-then-shuffle setting. Our amplification bound captures the regimes from both [15] and [12], thus providing a unified analysis of privacy amplification by shuffling for arbitrary local randomizers in the randomize-then-shuffle setting. Our proofs are also conceptually simpler than those in [12, 15] since we do not rely on privacy amplification by subsampling to obtain our results.

## 2 Preliminaries

Our notation is standard. We denote domains as $\mathbb{X}$, $\mathbb{Y}$, $\mathbb{Z}$ and randomized mechanism as $\mathcal{M}$, $\mathcal{P}$, $\mathcal{R}$, $\mathcal{S}$. For denoting sets and multisets we will use uppercase letters $A$, $B$, etc., and denote their elements as $a$, $b$, etc., while we will denote tuples as $\vec{x}$, $\vec{y}$, etc. Random variables, tuples and sets are denoted by $\mathsf{X}$, $\vec{\mathsf{X}}$ and $\boldsymbol{\mathsf{X}}$ respectively. We also use greek letters $\mu$, $\nu$, $\omega$ for distributions. Finally, we write $[k] = \{1, \dots, k\}$, $a \wedge b = \min\{a, b\}$, $[u]_+ = \max\{u, 0\}$ and $\mathbb{N}$ for the natural numbers.

### 2.1 The Curator and Local Models of Differential Privacy

Differential privacy is a formal approach to privacy-preserving data disclosure that prevents attemps to learn private information about specific to individuals in a data release [14]. The definition of differential privacy requires that the contribution $x_i$ of an individual to a dataset $\vec{x} = (x_1, \dots, x_n)$ has not much effect on what the adversary sees. This is formalized by considering a dataset $\vec{x}'$ that differs from $\vec{x}$ only in one element, denoted $\vec{x} \simeq \vec{x}'$, and requiring that the views of a potential adversary when running a mechanism on inputs $\vec{x}$ and $\vec{x}'$ are "indistinguishable". Let $\varepsilon \geq 0$ and $\delta \in [0,1]$. We say that a randomized mechanism $\mathcal{M} : \mathbb{X}^n \to \mathbb{Y}$ is $(\varepsilon, \delta)$-DP if

$$\forall \vec{x} \simeq \vec{x}', \forall E \subseteq \mathbb{Y} : \ \mathbb{P}[\mathcal{M}(\vec{x}) \in E] \leq e^{\varepsilon} \mathbb{P}[\mathcal{M}(\vec{x}') \in E] + \delta \ .$$

As mentioned above, different models of differential privacy arise depending on whether one can assume the availability of a trusted party (a curator) that has access to the information from all users in a centralized location. This setup is the one considered in the definition above. The other extreme scenario is when each user privatizes their data locally and submits the private values to a (potentially untrusted) server for aggregation. This is the domain of *local* differential privacy[4] (see Figure 1, left), where a user owns a data record $x \in \mathbb{X}$ and uses a *local randomizer* $\mathcal{R} : \mathbb{X} \to \mathbb{Y}$ to submit the privatized value $\mathcal{R}(x)$. In this case we say that the local randomizer is $(\varepsilon, \delta)$-LDP if

$$\forall x, x', \forall E \subseteq \mathbb{Y} : \ \mathbb{P}[\mathcal{R}(x) \in E] \leq e^{\varepsilon} \mathbb{P}[\mathcal{R}(x') \in E] + \delta \ .$$

The key difference is that in this case we must protect each user's data, and therefore the definition considers changing a user's value $x$ to another arbitrary value $x'$.

Moving from curator DP to local DP can be seen as effectively redefining the view that an adversary has on the data during the execution of a mechanism. In particular, if $\mathcal{R}$ is an $(\varepsilon, \delta)$-LDP local randomizer, then the mechanism $\mathcal{M} : \mathbb{X}^n \to \mathbb{Y}^n$ given by $\mathcal{M}(x_1, \ldots, x_n) = (\mathcal{R}(x_1), \ldots, \mathcal{R}(x_n))$ is $(\varepsilon, \delta)$-DP in the curator sense. The single-message shuffle model sits in between these two settings.

### 2.2 The Single-Message Shuffle Model

The *single-message shuffle model* of differential privacy considers a data collector that receives one message $y_i$ from each of the $n$ users as in the local model of differential privacy. The crucial difference with the local model is that the shuffle model assumes that a mechanism is in place to provide anonymity to each of the messages, i.e. the data collector is unable to associate messages to users. This is equivalent to assuming that, in the view of the adversary, these messages have been shuffled by a random permutation unknown to the adversary (see Figure 1, right).

Following the notation in [12], we define a single-message protocol $\mathcal{P}$ in the shuffle model to be a pair of algorithms $\mathcal{P} = (\mathcal{R}, \mathcal{A})$, where $\mathcal{R} : \mathbb{X} \to \mathbb{Y}$, and $\mathcal{A} : \mathbb{Y}^n \to \mathbb{Z}$. We call $\mathcal{R}$ the *local randomizer*, $\mathbb{Y}$ the *message space* of the protocol, $\mathcal{A}$ the *analyzer* of $\mathcal{P}$, and $\mathbb{Z}$ the *output space*. The overall protocol implements a mechanism $\mathcal{P} : \mathbb{X}^n \to \mathbb{Z}$ as follows. Each user $i$ holds a data record $x_i$, to which she applies the local randomizer to obtain a message $y_i = \mathcal{R}(x_i)$. The messages $y_i$ are then shuffled and submitted to the analyzer. We write $\mathcal{S}(y_1, \ldots, y_n)$ to denote the random shuffling step, where $\mathcal{S} : \mathbb{Y}^n \to \mathbb{Y}^n$ is a *shuffler* that applies

[4] Of which, in this paper, we only consider the non-interactive version for simplicity.

a random permutation to its inputs. In summary, the output of $\mathcal{P}(x_1, \ldots, x_n)$ is given by $\mathcal{A} \circ \mathcal{S} \circ \mathcal{R}^n(\vec{x}) = \mathcal{A}(\mathcal{S}(\mathcal{R}(x_1), \ldots, \mathcal{R}(x_n)))$.

From a privacy point of view, our threat model assumes that the analyzer $\mathcal{A}$ is applied to the shuffled messages by an untrusted data collector. Therefore, when analyzing the privacy of a protocol in the shuffle model we are interested in the indistinguishability between the shuffles $\mathcal{S} \circ \mathcal{R}^n(\vec{x})$ and $\mathcal{S} \circ \mathcal{R}^n(\vec{x}')$ for datasets $\vec{x} \simeq \vec{x}'$. In this sense, the analyzer's role is to provide utility for the output of the protocol $\mathcal{P}$, whose privacy guarantees follow from those of the *shuffled mechanism* $\mathcal{M} = \mathcal{S} \circ \mathcal{R}^n : \mathbb{X}^n \to \mathbb{Y}^n$ by the post-processing property of differential privacy. That is, the protocol $\mathcal{P}$ is $(\varepsilon, \delta)$-DP whenever the shuffled mechanism $\mathcal{M}$ is $(\varepsilon, \delta)$-DP.

When analyzing the privacy of a shuffled mechanism we assume the shuffler $\mathcal{S}$ is a perfectly secure primitive. This implies that a data collector observing the shuffled messages $\mathcal{S}(y_1, \ldots, y_n)$ obtains no information about which user generated each of the messages. An equivalent way to state this fact, which will sometimes be useful in our analysis of shuffled mechanisms, is to say that the output of the shuffler is a multiset instead of a tuple. Formally, this means that we can also think of the shuffler as a deterministic map $\mathcal{S} : \mathbb{Y}^n \to \mathbb{N}_n^{\mathbb{Y}}$ which takes a tuple $\vec{y} = (y_1, \ldots, y_n)$ with $n$ elements from $\mathbb{Y}$ and returns the multiset $Y = \{y_1, \ldots, y_n\}$ of its coordinates, where $\mathbb{N}_n^{\mathbb{Y}}$ denotes the collection of all multisets over $\mathbb{Y}$ with cardinality $n$. Sometimes we will refer to such multisets $Y \in \mathbb{N}_n^{\mathbb{Y}}$ as *histograms* to emphasize the fact that they can be regarded functions $Y : \mathbb{Y} \to \mathbb{N}$ counting the number of occurrences of each element of $\mathbb{Y}$ in $Y$.

### 2.3 Mean Square Error

When analyzing the utility of shuffled protocols for real summation we will use the *mean square error* (MSE) as accuracy measure. The mean squared error of a randomized protocol $\mathcal{P}(\vec{x})$ for approximating a deterministic quantity $f(\vec{x})$ is given by $\mathrm{MSE}(\mathcal{P}, \vec{x}) = \mathbb{E}[(\mathcal{P}(\vec{x}) - f(\vec{x}))^2]$, where the expectation is taken over the randomness of $\mathcal{P}$. Note that when the protocol is unbiased the MSE is equivalent to the variance, since in this case we have $\mathbb{E}[\mathcal{P}(\vec{x})] = f(\vec{x})$ and therefore

$$\mathrm{MSE}(\mathcal{P}, \vec{x}) = \mathbb{E}[(\mathcal{P}(\vec{x}) - \mathbb{E}[\mathcal{P}(\vec{x})])^2] = \mathbb{V}[\mathcal{P}(\vec{x})] \ .$$

In addition to the MSE for a fixed input, we also consider the *worst-case MSE* over all possible inputs $\mathrm{MSE}(\mathcal{P})$, and the *expected MSE* on a distribution over inputs $\mathrm{MSE}(\mathcal{P}, \vec{\mathsf{X}})$. These quantities are defined as follows:

$$\mathrm{MSE}(\mathcal{P}) = \sup_{\vec{x}} \mathrm{MSE}(\mathcal{P}, \vec{x}) \ ,$$
$$\mathrm{MSE}(\mathcal{P}, \vec{\mathsf{X}}) = \mathbb{E}_{\vec{x} \sim \vec{\mathsf{X}}}[\mathrm{MSE}(\mathcal{P}, \vec{x})] \ .$$

## 3 The Privacy of Shuffled Randomized Response

In this section we show a protocol for $n$ parties to compute a private histogram over the domain $[k]$ in the single-message shuffle model. The local randomizer of

```
Algorithm 1: Private Histogram: Local Randomizer $\mathcal{R}^{PH}_{\gamma,k,n}$
Public Parameters: $\gamma \in [0,1]$, domain size $k$, and number of parties $n$
Input: $x \in [k]$
Output: $y \in [k]$
Sample $b \leftarrow \mathtt{Ber}(\gamma)$
if $b = 0$ then
    Let $y \leftarrow x$
else
    Sample $y \leftarrow \mathtt{Unif}([k])$
return $y$
```

our protocol is shown in Algorithm 1, and the analyzer simply builds a histogram of the received messages. The randomizer is parameterized by a probability $\gamma$, and consists of a $k$-ary randomized response mechanism that returns the true value $x$ with probability $1-\gamma$, and a uniformly random value with probability $\gamma$. This randomizer has been studied and used (in the local model) in several previous works [8, 21, 22]. We discuss how to set $\gamma$ to satisfy differential privacy next.

### 3.1 The *Blanket* Intuition

In each execution of Algorithm 1 a subset $B$ of approximately $\gamma n$ parties will submit a random value, while the remaining parties will submit their true value. The values sent by parties in $B$ form a histogram $Y_1$ of uniformly random values and the values sent by the parties not in $B$ correspond to the true histogram $Y_2$ of their data. An important observation is that in the shuffle model the information obtained by the server is equivalent to the histogram $Y_1 \cup Y_2$. This observation is a simple generalization of the observation made by Cheu et al. [12] that shuffling of binary data corresponds to secure addition. When $k > 2$, shuffling of categorical data corresponds to a secure histogram computation, and in particular secure addition of histograms. In summary, the information collected by the server in an execution corresponds to a histogram $Y$ with approximately $\gamma n$ random entries and $(1-\gamma)n$ truthful entries, which as mentioned above we decompose as $Y = Y_1 \cup Y_2$.

To achieve differential privacy we need to set the value $\gamma$ of Algorithm 1 so that $Y$ changes by an appropriately bounded amount when computed on neighboring datasets where only a certain party's data (say party $n$) changes. Our privacy argument does not rely on the anonymity of the set $B$ and thus we can assume, for the privacy analysis, that the server knows $B$. We further assume in the analysis that the server knows the inputs from all parties except the $n$th one, which gives her the ability to remove from $Y$ the values submitted by any party who responded truthfully among the first $n-1$.

Now consider two datasets of size $n$ that differ on the input from the $n$th party. In an execution where party $n$ is in $B$ we trivially get privacy since the value submitted by this party is independent of its input. Otherwise, party $n$ will be submitting their true value $x_n$, in which case the server can determine $Y_2$ up to the value $x_n$ using that she knows $(x_1, \ldots, x_{n-1})$. Hence, a server trying to break the privacy of party $n$ observes $Y_1 \cup \{x_n\}$, the union of a random histogram with the input of this party. Intuitively, the privacy of the protocol boils down to setting $\gamma$ so that $Y_1$, which we call the random *blanket* of the local randomizer $\mathcal{R}^{PH}_{\gamma,k,n}$, appropriately "hides" $x_n$.

As we will see in Section 5, the intuitive notion of the blanket of a local randomizer can be formally defined for arbitrary local randomizers using a generalization of the notion of total variation distance from pairs to sets of distributions. This will allow us to represent the output distribution of any local randomizer $\mathcal{R}(x)$ as a mixture of the form $(1-\gamma)\nu_x + \gamma\omega$, for some $0 < \gamma < 1$ and probability distributions $\nu_x$ and $\omega$, of which we call $\omega$ the *privacy blanket* of the local randomizer $\mathcal{R}$.

### 3.2 Privacy Analysis of Algorithm 1

Let us now formalize the above intuition, and prove privacy for our protocol for an appropriate choice of $\gamma$. In particular, we prove the following theorem, where the assumption $\varepsilon \leq 1$ is only for technical convenience. A more general approach to obtain privacy guarantees for shuffled mechanisms is provided in Section 5.

**Theorem 1.** *The shuffled mechanism $\mathcal{M} = \mathcal{S} \circ \mathcal{R}^{PH}_{\gamma,k,n}$ is $(\varepsilon,\delta)$-DP for any $k, n \in \mathbb{N}$, $\varepsilon \leq 1$ and $\delta \in (0,1]$ such that $\gamma = \max\{\frac{14k\log(2/\delta)}{(n-1)\varepsilon^2}, \frac{27k}{(n-1)\varepsilon}\} < 1$.*

*Proof.* Let $\vec{x}, \vec{x}' \in [k]^n$ be neighboring databases of the form $\vec{x} = (x_1, x_2, \ldots, x_n)$ and $\vec{x}' = (x_1, x_2, \ldots, x'_n)$. We assume that the server knows the set $B$ of users who submit random values, which is equivalent to revealing to the server a vector $\vec{b} = (b_1, \ldots, b_n)$ of the bits $b$ sampled in the execution of each of the local randomizers. We also assume the server knows the inputs from the first $n-1$ parties.

Hence, we define the view $\mathrm{View}_{\mathcal{M}}$ of the server on a realization of the protocol as the tuple $\mathrm{View}_{\mathcal{M}}(\vec{x}) = (Y, \vec{x}_{\cap}, \vec{b})$ containing:

1. A multiset $Y = \mathcal{M}(\vec{x}) = \{y_1, \ldots, y_n\}$ with the outputs $y_i$ of each local randomizer.
2. A tuple $\vec{x}_{\cap} = (x_1, \ldots, x_{n-1})$ with the inputs from the first $n-1$ users.
3. The tuple $\vec{b} = (b_1, \ldots, b_n)$ of binary values indicating which users submitted their true values.

Proving that the protocol is $(\varepsilon,\delta)$-DP when the server has access to all this information will imply the same level of privacy for the shuffled mechanism $\mathcal{S} \circ \mathcal{R}^{PH}_{\gamma,k,n}$ by the post-processing property of differential privacy.

To show that $\mathrm{View}_{\mathcal{M}}$ satisfies $(\varepsilon,\delta)$-DP it is enough to prove

$$\mathbb{P}_{\mathsf{V}\sim\mathrm{View}_{\mathcal{M}}(\vec{x})}\left[\frac{\mathbb{P}[\mathrm{View}_{\mathcal{M}}(\vec{x}) = \mathsf{V}]}{\mathbb{P}[\mathrm{View}_{\mathcal{M}}(\vec{x}') = \mathsf{V}]} \geq e^{\varepsilon}\right] \leq \delta \enspace .$$

We start by fixing a value $V$ in the range of $\mathrm{View}_{\mathcal{M}}$ and computing the probability ratio above conditioned on $\mathsf{V} = V$.

Consider first the case where $V$ is such that $b_n = 1$, i.e. party $n$ submits a random value independent of her input. In this case privacy holds trivially since $\mathbb{P}[\mathrm{View}_{\mathcal{M}}(\vec{x}) = V] = \mathbb{P}[\mathrm{View}_{\mathcal{M}}(\vec{x}') = V]$. Hence, we focus on the case where party $n$ submits her true value ($b_n = 0$). For $j \in [k]$, let $n_j$ be the number of messages received by the server with value $j$ after removing from $Y$ any truthful answers submitted by the first $n-1$ users. With our notation above, we have $n_j = Y_1(j) + \mathbb{I}[x_n = j]$ and $\sum_{j=1}^{k} n_j = |B| + 1$ for the execution with input $\vec{x}$. Now assume, without loss of generality, that $x_n = 1$ and $x'_n = 2$. As $x_n = 1$, we have that

$$\mathbb{P}[\mathrm{View}_{\mathcal{M}}(\vec{x}) = V] = \binom{|B|}{n_1 - 1, n_2, ..., n_k} \frac{\gamma^{|B|}(1-\gamma)^{n-|B|}}{k^{|B|}} \quad ,$$

corresponding to the probability of a particular pattern $\vec{b}$ of users sampling from the blanket times the probability of obtaining a particular histogram $Y_1$ when sampling $|B|$ elements uniformly at random from $[k]$. Similarly, using that $x'_n = 2$ we have

$$\mathbb{P}[\mathrm{View}_{\mathcal{M}}(\vec{x}') = V] = \binom{|B|}{n_1, n_2 - 1, ..., n_k} \frac{\gamma^{|B|}(1-\gamma)^{n-|B|}}{k^{|B|}} \quad .$$

Therefore, taking the ratio between the last two probabilities we find that, in the case $b_n = 0$,

$$\frac{\mathbb{P}[\mathrm{View}_{\mathcal{M}}(\vec{x}) = V]}{\mathbb{P}[\mathrm{View}_{\mathcal{M}}(\vec{x}') = V]} = \frac{n_1}{n_2} \quad .$$

Now note that for $\mathsf{V} \sim \mathrm{View}_{\mathcal{M}}(\vec{x})$ the count $n_2 = n_2(\mathsf{V})$ follows a binomial distribution $\mathsf{N}_2$ with $n-1$ trials and success probability $\gamma/k$, and $n_1(\mathsf{V}) - 1 = \mathsf{N}_1 - 1$ follows the same distribution. Thus, we have

$$\mathbb{P}_{\mathsf{V} \sim \mathrm{View}_{\mathcal{M}}(\vec{x})}\left[\frac{\mathbb{P}[\mathrm{View}_{\mathcal{M}}(\vec{x}) = \mathsf{V}]}{\mathbb{P}[\mathrm{View}_{\mathcal{M}}(\vec{x}') = \mathsf{V}]} \geq e^{\varepsilon}\right] = \mathbb{P}\left[\frac{\mathsf{N}_1}{\mathsf{N}_2} \geq e^{\varepsilon}\right] \quad ,$$

where $\mathsf{N}_1 \sim \mathtt{Bin}\left(n-1, \frac{\gamma}{k}\right) + 1$ and $\mathsf{N}_2 \sim \mathtt{Bin}\left(n-1, \frac{\gamma}{k}\right)$.

We now bound the probability above using a union bound and the multiplicative Chernoff bound. Let $c = \mathbb{E}[\mathsf{N}_2] = \frac{\gamma(n-1)}{k}$. Since $\mathsf{N}_1/\mathsf{N}_2 \geq e^{\varepsilon}$ implies that either $\mathsf{N}_1 \geq ce^{\varepsilon/2}$ or $\mathsf{N}_2 \leq ce^{-\varepsilon/2}$, we have

$$\begin{aligned}
\mathbb{P}\left[\frac{\mathsf{N}_1}{\mathsf{N}_2} \geq e^{\varepsilon}\right] &\leq \mathbb{P}\left[\mathsf{N}_1 \geq ce^{\varepsilon/2}\right] + \mathbb{P}\left[\mathsf{N}_2 \leq ce^{-\varepsilon/2}\right] \\
&= \mathbb{P}\left[\mathsf{N}_2 \geq ce^{\varepsilon/2} - 1\right] + \mathbb{P}\left[\mathsf{N}_2 \leq ce^{-\varepsilon/2}\right] \\
&= \mathbb{P}\left[\mathsf{N}_2 - \mathbb{E}[\mathsf{N}_1] \geq c\left(e^{\varepsilon/2} - 1 - \frac{1}{c}\right)\right] \\
&\quad + \mathbb{P}\left[\mathsf{N}_2 - \mathbb{E}[\mathsf{N}_2] \leq c(e^{-\varepsilon/2} - 1)\right] \quad .
\end{aligned}$$

Applying the multiplicative Chernoff bound to each of these probabilities then gives that

$$\mathbb{P}\left[\frac{\mathsf{N}_1}{\mathsf{N}_2} \geq e^{\varepsilon}\right] \leq \exp\left(-\frac{c}{3}\left(e^{\varepsilon/2} - 1 - \frac{1}{c}\right)^2\right) + \exp\left(-\frac{c}{2}(1 - e^{-\varepsilon/2})^2\right) \quad .$$

Assuming $\varepsilon \leq 1$, both of the right hand summands are less than or equal to $\frac{\delta}{2}$ if

$$c = \frac{\gamma(n-1)}{k} \geq \max\left\{\frac{14\log\left(\frac{2}{\delta}\right)}{\varepsilon^2}, \frac{27}{\varepsilon}\right\} \quad .$$

Indeed, for the second term this follows from $1 - e^{-\varepsilon/2} \geq (1 - e^{-1/2})\varepsilon \geq \varepsilon/\sqrt{7}$ for $\varepsilon \leq 1$. For the first term we use that $c \geq \frac{27}{\varepsilon}$ implies $e^{\varepsilon/2} - 1 - \frac{1}{c} \geq \frac{25}{54}\varepsilon$ and $14 \geq \frac{3 \cdot 54^2}{25^2}$. □

Two remarks about this result are in order. First, we should emphasize that the assumption of $\varepsilon \leq 1$ is only required for simplicity when using Chernoff's inequality to bound the probability that the privacy loss random variable is large. Without any restriction on $\varepsilon$, a similar result can be achieved by replacing Chernoff's inequality with Bennett's inequality [10, Theorem 2.9] to account for the variance of the privacy loss random variable in the tail bound. Here we decide not to pursue this route because the ad-hoc privacy analysis of Theorem 1 is superseded by the results in Section 5 anyway. The second observation about this result is that, with the choice of $\gamma$ made above, the local randomizer $\mathcal{R}^{PH}_{\gamma,k,n}$ satisfies $\varepsilon_0$-LDP with

$$\varepsilon_0 = O\left(\log\left(\frac{n\varepsilon^2}{\log(1/\delta)} - k\right)\right) = O\left(\log\left(\frac{n\varepsilon^2}{\log(1/\delta)}\left(1 - \frac{\gamma}{14}\right)\right)\right) \quad .$$

This is obtained according to the formula provided by Lemma 6 in Section 5.1. Thus, we see that Theorem 1 can be regarded as a privacy amplification statement showing that shuffling $n$ copies of an $\varepsilon_0$-LDP local randomized with $\varepsilon_0 = O_\delta(\log(n\varepsilon^2))$ yields a mechanism satisfying $(\varepsilon, \delta)$-DP. In Section 5.1 we will show that this is not coincidence, but rather an instance of a general privacy amplification result.

## 4 Optimal Summation in the Shuffle Model

### 4.1 Upper Bound

In this section we present a protocol for the problem of computing the sum of real values $x_i \in [0, 1]$ in the single-message shuffle model. Our protocol is parameterized by values $c, k$, and the number of parties $n$, and its local randomizer and analyzer are shown in Algorithms 2 and 3, respectively.

The protocol uses the protocol depicted in Algorithm 1 in a black-box manner. To compute a differentially private approximation of $\sum_i x_i$, we fix a value $k$.

```
Algorithm 2: Local Randomizer $\mathcal{R}_{c,k,n}$
  Public Parameters: $c$, $k$, and number of parties $n$
  Input: $x \in [0,1]$
  Output: $y \in \{0,1,\ldots,k\}$
  Let $\bar{x} \leftarrow \lfloor xk \rfloor + \mathtt{Ber}(xk - \lfloor xk \rfloor)$  ▷ $\bar{x}$ is the encoding of $x$ with precision $k$
  Sample $b \leftarrow \mathtt{Ber}\left(\frac{c(k+1)}{n}\right)$
  if $b = 0$ then
      Let $y \leftarrow \bar{x}$
  else
      Sample $y \leftarrow \mathtt{Unif}(\{0,1,\ldots,k\})$
  return $y$
```

```
Algorithm 3: Analyzer $\mathcal{A}_{c,k,n}$
  Public Parameters: $c, k$, and number of parties $n$
  Input: Multiset $\{y_i\}_{i\in[n]}$, with $y_i \in \{0,1,\ldots,k\}$
  Output: $z \in [0,1]$
  Let $\hat{z} \leftarrow \frac{1}{k}\sum_{i=1}^{n} y_i$
  Let $z \leftarrow \mathtt{DeBias}(\hat{z})$, where $\mathtt{DeBias}(w) = \left(w - \frac{c(k+1)}{2}\right) / \left(1 - \frac{c(k+1)}{n}\right)$
  return $z$
```

Then we operate on the fixed-point encoding of each input $x_i$, which is an integer $\bar{x}_i \in \{0,\ldots,k\}$. That is, we replace $x_i$ with its fixed-point approximation $\bar{x}_i/k$. The protocol then applies the randomized response mechanism in Algorithm 1 to each $\bar{x}_i$ to submit a value $y_i$ to compute a differentially private histogram of the $(y_1,\ldots,y_n)$ as in the previous section. From these values the server can approximate $\sum_i x_i$ by post processing, which includes a debiasing standard step. The privacy of the protocol described in Algorithms 2 and 3 follows directly from the privacy analysis of Algorithm 1 given in Section 3.

Regarding accuracy, a crucial point in this reduction is that the encoding $\bar{x}_i$ of $x_i$ is via randomized rounding and hence unbiased. In more detail, as shown in Algorithm 2, the value $x$ is encoded as $\bar{x} = \lfloor xk \rfloor + \mathtt{Ber}(xk - \lfloor xk \rfloor)$. This ensures that $\mathbb{E}[\bar{x}/k] = \mathbb{E}[x]$ and that the mean squared error due to rounding (which equals the variance) is at most $\frac{1}{4k^2}$. The local randomizer either sends this fixed-point encoding or a random value in $\{0,1,\ldots,k\}$ with probabilities $1-\gamma$ and $\gamma$, respectively, where (following the analysis in the previous section) we set $\gamma = \frac{k+1}{n}c$. Note that the mean squared error when the local randomizer submits a random value is at most $\frac{1}{2}$. This observations lead to the following accuracy bound.

**Theorem 2.** *For any $\varepsilon \leq 1$, $\delta \in (0,1]$ and $n \in \mathbb{N}$, there exist parameters $c, k$ such that $\mathcal{P}_{c,k,n}$ is $(\varepsilon,\delta)$-DP and*

$$\mathrm{MSE}(\mathcal{P}_{c,k,n}) = O\left(n^{1/3} \cdot \frac{\log^{2/3}(1/\delta)}{\varepsilon^{4/3}}\right) .$$

*Proof.* The following bound on $\mathrm{MSE}(\mathcal{P}_{c,k,n})$ follows from the observations above: unbiasedness of the estimator computed by the analyzer and randomized rounding, and the bounds on the variance of our randomized response.

$$\begin{aligned}
\mathrm{MSE}(\mathcal{P}_{c,k,n}) &= \sup_{\vec{x}} \mathbb{E}[(\mathtt{DeBias}(\hat{z}) - \sum_i x_i)^2] \\
&= \sup_{\vec{x}} \mathbb{E}\left[\left(\sum_i (\mathtt{DeBias}(y_i/k) - x_i)\right)^2\right] \\
&= \sup_{\vec{x}} \sum_i \mathbb{E}\left[(\mathtt{DeBias}(y_i/k) - x_i)^2\right] \\
&= \sup_{\vec{x}} \sum_i \mathbb{V}\left[\mathtt{DeBias}(y_i/k)\right] \\
&= \frac{n}{(1-\gamma)^2} \sup_{x_1} \mathbb{V}[y_1/k] \\
&\leq \frac{n}{(1-\gamma)^2} \left(\frac{1-\gamma}{4k^2} + \frac{\gamma}{2}\right) \\
&\leq \frac{n}{(1-\gamma)^2} \left(\frac{1}{4k^2} + \frac{c(k+1)}{2n}\right) .
\end{aligned}$$

Choosing the parameter $k = (n/c)^{1/3}$ minimizes the sum in the above expression and provides a bound on the MSE of the form $O(c^{2/3}n^{1/3})$. Plugging in $c = \gamma\frac{n}{k+1} = O\left(\frac{\log(1/\delta)}{\varepsilon^2}\right)$ from our analysis in the previous section (Theorem 1) yields the bound in the statement of the theorem. □

Note that as our protocol corresponds to an unbiased estimator, the MSE is equal to the variance in this case. Using this observation we immediately obtain the following corollary for estimation of statistical queries in the single-message shuffle model.

**Corollary 1.** *For every statistical query $q : \mathcal{X} \mapsto [0,1]$, $\varepsilon \leq 1, \delta \in (0,1]$ and $n \in \mathbb{N}$, there is an $(\varepsilon,\delta)$-DP $n$-party unbiased protocol for estimating $\frac{1}{n}\sum_i q(x_i)$ in the single-message shuffle model with standard deviation $O\left(\frac{\log^{1/3}(1/\delta)}{n^{5/6}\varepsilon^{2/3}}\right)$.*

### 4.2 Lower Bound

In this section we show that any differentially private protocol $\mathcal{P}$ for the problem of estimating $\sum_i x_i$ in the single-message shuffle model must have $\mathrm{MSE}(\mathcal{P}) =$

$\Omega(n^{1/3})$ This shows that our protocol from the previous section is optimal, and gives a separation result for the single-message shuffle model, showing that its accuracy lies between the curator and local models of differential privacy.

**Reduction in the i.i.d. setting.** We first show that when the inputs to the protocol $\mathcal{P}$ are sampled i.i.d. one can assume, for the purpose of showing a lower bound, that the protocol $\mathcal{P}$ for estimating $\sum_i x_i$ is of a simplified form. Namely, we show that the local randomizer can be taken to have output values in $[0,1]$, and its analyzer simply adds up all received messages.

**Lemma 1.** *Let $\mathcal{P} = (\mathcal{R}, \mathcal{A})$ be an $n$-party protocol for real summation in the single-message shuffle model. Let $\mathsf{X}$ be a random variable on $[0,1]$ and suppose that users sample their inputs from the distribution $\vec{\mathsf{X}} = (\mathsf{X}_1, \ldots, \mathsf{X}_n)$, where each $\mathsf{X}_i$ is an independent copy of $\mathsf{X}$. Then, there exists a protocol $\mathcal{P}' = (\mathcal{R}', \mathcal{A}')$ such that:*

1. *$\mathcal{A}'(y_1, \ldots, y_n) = \sum_{i=1}^n y_i$ and*[5] *$\mathsf{Im}(\mathcal{R}') \subseteq [0,1]$.*
2. *$MSE(\mathcal{P}', \vec{\mathsf{X}}) \leq MSE(\mathcal{P}, \vec{\mathsf{X}})$.*
3. *If the shuffled mechanism $\mathcal{S} \circ \mathcal{R}^n$ is $(\varepsilon, \delta)$-DP, then $\mathcal{S} \circ \mathcal{R}'^n$ is also $(\varepsilon, \delta)$-DP.*

*Proof.* Consider the post-processed local randomizer $\mathcal{R}' = f \circ \mathcal{R}$ where $f(y) = \mathbb{E}[\mathsf{X} | \mathcal{R}(\mathsf{X}) = y]$. In Bayesian estimation, $f$ is called the posterior mean estimator, and is known to be a minimum MSE estimator [19]. Since $\mathsf{Im}(\mathcal{R}') \subseteq [0,1]$, we have a protocol $\mathcal{P}'$ satisfying claim 1.

Next we show that $\mathrm{MSE}(\mathcal{P}', \vec{\mathsf{X}}) \leq \mathrm{MSE}(\mathcal{P}, \vec{\mathsf{X}})$. Note that the analyzer $\mathcal{A}$ in protocol $\mathcal{P}$ can be seen as an estimator of $\mathsf{Z} = \sum_i \mathsf{X}_i$ given observations from $\vec{\mathsf{Y}} = (\mathsf{Y}_1, \ldots, \mathsf{Y}_n)$, where $\mathsf{Y}_i = \mathcal{R}(\mathsf{X}_i)$. Now consider an arbitrary estimator $h$ of $\mathsf{Z}$ given the observation $\vec{\mathsf{Y}} = \vec{y}$. We have

$$\begin{aligned}\mathrm{MSE}(h, \vec{y}) &= \mathbb{E}[(h(\vec{y}) - \mathsf{Z})^2 | \vec{\mathsf{Y}} = \vec{y}] \\ &= \mathbb{E}[\mathsf{Z}^2 | \vec{\mathsf{Y}} = \vec{y}] - 2h(\vec{y})\mathbb{E}[\mathsf{Z} | \vec{\mathsf{Y}} = \vec{y}] + h(\vec{y})^2 \ .\end{aligned}$$

It follows from minimizing $\mathrm{MSE}(h, \vec{y})$ with respect to $h$ that the minimum MSE estimator of $\mathsf{Z}$ given $\vec{\mathsf{Y}}$ is $h(\vec{y}) = \mathbb{E}[\mathsf{Z} | \vec{\mathsf{Y}} = \vec{y}]$. Hence, by linearity of expectation, and the fact that the $\mathsf{Y}_i$ are independent,

$$\mathbb{E}[\mathsf{Z} | \vec{\mathsf{Y}} = \vec{y}] = \sum_{i=1}^n \mathbb{E}[\mathsf{X}_i | \vec{\mathsf{Y}} = \vec{y}] = \sum_{i=1}^n \mathbb{E}[\mathsf{X}_i | \mathsf{Y}_i = y_i] = \sum_{i=1}^n f(y_i) \ .$$

Therefore, we have shown that $\mathcal{P}' = (\mathcal{R}', \mathcal{A}')$ implements a minimum MSE estimator for $\mathsf{Z}$ given $(\mathcal{R}(\mathsf{X}_1), \ldots, \mathcal{R}(\mathsf{X}_n))$, and in particular $\mathrm{MSE}(\mathcal{P}', \vec{\mathsf{X}}) \leq \mathrm{MSE}(\mathcal{P}, \vec{\mathsf{X}})$.

Part 3 of the lemma follows from the standard post-processing property of differential privacy by observing that the output of $\mathcal{S} \circ \mathcal{R}'^n(\vec{x})$ can be obtained by applying $f$ to each element in the output of $\mathcal{S} \circ \mathcal{R}^n(\vec{x})$. □

[5] Here we use $\mathsf{Im}(\mathcal{R}')$ to denote the image of the local randomizer $\mathcal{R}'$.

**Proof of the lower bound.** It remains to show that, for any protocol $\mathcal{P} = (\mathcal{R}, \mathcal{A})$ satisfying the conditions of Lemma 1, we can find a tuple of i.i.d. random variables $\vec{\mathsf{X}}$ such that $\mathrm{MSE}(\mathcal{P}, \vec{\mathsf{X}}) = \Omega(n^{1/3})$. Recall that by virtue of Lemma 1 we can assume, without loss of generality, that $\mathcal{R}$ is a mapping from $[0,1]$ into itself, $\mathcal{A}$ sums its inputs, and $\vec{\mathsf{X}} = (\mathsf{X}_1, \ldots, \mathsf{X}_n)$ where the $\mathsf{X}_i$ are i.i.d. copies of some random variable $\mathsf{X}$. We first show that under these assumptions we can reduce the search for a lower bound on $\mathrm{MSE}(\mathcal{P}, \vec{\mathsf{X}})$ to consider only the expected square error of an individual run of the local randomizer.

**Lemma 2.** *Let $\mathcal{P} = (\mathcal{R}, \mathcal{A})$ be an $n$-party protocol for real summation in the single-message shuffle model such that $\mathcal{R} : [0,1] \to [0,1]$ and $\mathcal{A}$ is summation. Suppose $\vec{\mathsf{X}} = (\mathsf{X}_1, \ldots, \mathsf{X}_n)$, where the $\mathsf{X}_i$ are i.i.d. copies of some random variable $\mathsf{X}$. Then,*

$$\mathrm{MSE}(\mathcal{P}, \vec{\mathsf{X}}) \geq n\mathbb{E}[(\mathcal{R}(\mathsf{X}) - \mathsf{X})^2] \ .$$

*Proof.* The result follows from an elementary calculation:

$$\begin{aligned}
\mathrm{MSE}(\mathcal{P}, \vec{\mathsf{X}}) &= \mathbb{E}\left[\left(\sum_{i \in [n]} \mathcal{R}(\mathsf{X}_i) - \mathsf{X}_i\right)^2\right] \\
&= \sum_i \mathbb{E}[(\mathcal{R}(\mathsf{X}_i) - \mathsf{X}_i)^2] + \sum_{i \neq j} \mathbb{E}[(\mathcal{R}(\mathsf{X}_i) - \mathsf{X}_i)(\mathcal{R}(\mathsf{X}_j) - \mathsf{X}_j)] \\
&= \sum_i \mathbb{E}[(\mathcal{R}(\mathsf{X}_i) - \mathsf{X}_i)^2] + \sum_{i \neq j} \mathbb{E}[\mathcal{R}(\mathsf{X}_i) - \mathsf{X}_i]^2 \\
&\geq n\mathbb{E}[(\mathcal{R}(\mathsf{X}) - \mathsf{X})^2] \ . \qquad \square
\end{aligned}$$

Therefore, to obtain our lower bound it will suffice to find a distribution on $[0,1]$ such that if $\mathcal{R} : [0,1] \to [0,1]$ is a local randomizer for which the protocol $\mathcal{P} = (\mathcal{R}, \mathcal{A})$ is differentially private, then $\mathcal{R}$ has expected square error $\Omega(n^{-2/3})$ under that distribution. We start by constructing such distribution and then show that it satisfies the desired properties.

Consider the partition of the unit interval $[0,1]$ into $k$ disjoint subintervals of size $1/k$, where $k \in \mathbb{N}$ is a parameter to be determined later. We will take inputs from the set $I = \{m/k - 1/2k \mid m \in [k]\}$ of midpoints of these intervals. For any $a \in I$ we denote by $I(x)$ the subinterval of $[0,1]$ containing $a$. Given a local randomizer $\mathcal{R} : [0,1] \to [0,1]$ we define the probability $p_{a,b} = \mathbb{P}[\mathcal{R}(a) \in I(b)]$ that the local randomizer maps an input $a$ to the subinterval centered at $b$ for any $a, b \in I$.

Now let $\mathsf{X} \sim \mathtt{Unif}(I)$ be a random variable sampled uniformly from $I$. The following observations are central to the proof of our lower bound. First observe that $\mathcal{R}$ maps $\mathsf{X}$ to a value outside of its interval with probability $\frac{1}{k}\sum_{b \in I}(1 - p_{b,b})$. If this event occurs, then $\mathcal{R}(\mathsf{X})$ incurs a squared error of at least $1/(2k)^2$, as the absolute error will be at least half the width of an interval. Similarly, when $\mathcal{R}$ maps an input $a$ to a point inside an interval $I(b)$ with $a \neq b$, the squared error incurred is at least $(|b-a|-1/2k)^2$, as the error is at least the distance between the two interval midpoints minus half the width of an interval. The next lemma encapsulates a useful calculation related to this observation.

**Lemma 3.** *For any $b \in I = \{m/k - 1/2k \mid m \in [k]\}$ we have*

$$\frac{1}{k} \sum_{a \in I \setminus \{b\}} \left( |a-b| - \frac{1}{2k} \right)^2 \geq \frac{1}{48} \left( 1 - \frac{1}{k^2} \right) \quad .$$

*Proof.* Let $b = m/k - 1/2k$ for some $m \in [k]$. Then,

$$\begin{aligned} \frac{1}{k} \sum_{a \in I \setminus \{b\}} \left( |a-b| - \frac{1}{2k} \right)^2 &= \frac{1}{k^3} \sum_{i \in [k] \setminus \{m\}} \left( |i-m| - \frac{1}{2} \right)^2 \\ &\geq \frac{1}{4k^3} \sum_{i \in [k] \setminus \{m\}} (i-m)^2 = \frac{1}{4k^3} \sum_{i \in [k]} (i-m)^2 \quad , \end{aligned}$$

where we used $(u-1/2)^2 \geq u^2/4$ for $u \geq 1$. Now let $\mathsf{U} \sim \mathtt{Unif}([k])$ and observe that for any $m \in [k]$ we have

$$\sum_{i \in [k]} (i-m)^2 \geq \sum_{i \in [k]} (i - \mathbb{E}[\mathsf{U}])^2 = k \mathbb{V}[\mathsf{U}] = \frac{k^3-k}{12} \quad . \qquad \square$$

Now we can combine the two observations about the error of $\mathcal{R}$ under $\mathsf{X}$ into a lower bound for its expected square error. Subsequently we will show how the output probabilities occurring in this bound are related under differential privacy.

**Lemma 4.** *Let $\mathcal{R} : [0,1] \to [0,1]$ be a local randomizer and $\mathsf{X} \sim \mathtt{Unif}(I)$ with $I = \{m/k - 1/2k \mid m \in [k]\}$. Then,*

$$\mathbb{E}[(\mathcal{R}(\mathsf{X}) - \mathsf{X})^2] \geq \sum_{b \in I} \min \left\{ \frac{1-p_{b,b}}{4k^3}, \frac{1}{48} \left( 1 - \frac{1}{k^2} \right) \min_{a \in I} p_{a,b} \right\} \quad .$$

*Proof.* The bound in obtained by formalizing the two observations made above to obtain two different lower bounds for $\mathbb{E}[(\mathcal{R}(\mathsf{X}) - \mathsf{X})^2]$ and then taking their minimum. Our first bound follows directly from the discussion above:

$$\begin{aligned} \mathbb{E}[(\mathcal{R}(\mathsf{X}) - \mathsf{X})^2] &= \sum_{b \in I} \mathbb{E}[(\mathcal{R}(b) - b)^2] \mathbb{P}[\mathsf{X} = b] = \frac{1}{k} \sum_{b \in I} \mathbb{E}[(\mathcal{R}(b) - b)^2] \\ &\geq \frac{1}{k} \sum_{b \in I} (1 - p_{b,b}) \cdot \frac{1}{(2k)^2} = \sum_{b \in I} \frac{1-p_{b,b}}{4k^3} \quad . \end{aligned}$$

Our second bound follows from the fact that the squared error is at least $(|b - a| - \frac{1}{2k})^2$ if $\mathsf{X} = a$ and $\mathcal{R}(a) \in I(b)$, for $a, b \in I$ such that $a \neq b$:

$$\begin{aligned}
\mathbb{E}[(\mathcal{R}(\mathsf{X}) - \mathsf{X})^2] &= \frac{1}{k} \sum_{b \in I} \mathbb{E}[(\mathcal{R}(b) - b)^2] \\
&\geq \frac{1}{k} \sum_{b \in I} \sum_{a \in I \setminus \{b\}} p_{a,b} \left( |b - a| - \frac{1}{2k} \right)^2 \\
&\geq \frac{1}{k} \sum_{b \in I} (\min_{a \in I} p_{a,b}) \sum_{a \in I \setminus \{b\}} \left( |b - a| - \frac{1}{2k} \right)^2 \\
&\geq \sum_{b \in I} (\min_{a \in I} p_{a,b}) \frac{1}{48} \left( 1 - \frac{1}{k^2} \right) \ ,
\end{aligned}$$

where the last inequality uses Lemma 3. Finally, we get

$$\begin{aligned}
\mathbb{E}[(\mathcal{R}(\mathsf{X}) - \mathsf{X})^2] &\geq \min \left\{ \sum_{b \in I} \frac{1 - p_{b,b}}{4k^3}, \sum_{b \in I} (\min_{a \in I} p_{a,b}) \frac{1}{48} \left( 1 - \frac{1}{k^2} \right) \right\} \\
&\geq \sum_{b \in I} \min \left\{ \frac{1 - p_{b,b}}{4k^3}, \frac{1}{48} \left( 1 - \frac{1}{k^2} \right) \min_{a \in I} p_{a,b} \right\} \ . \qquad \square
\end{aligned}$$

**Lemma 5.** *Let $\mathcal{R} : [0,1] \to [0,1]$ be a local randomizer such that the shuffled protocol $\mathcal{M} = \mathcal{S} \circ \mathcal{R}^n$ is $(\varepsilon, \delta)$-DP with $\delta < 1/2$. Then, for any $a, b \in I$, $a \neq b$, either $p_{b,b} < 1 - e^{-\varepsilon}/2$ or $p_{a,b} \geq (1/2 - \delta)/n$.*

*Proof.* If $p_{b,b} < 1 - e^{-\varepsilon}/2$ then the proof is done. Otherwise, consider the neighboring datasets $\vec{x} = (a, \ldots, a)$ and $\vec{x}' = (b, a, \ldots, a)$. Recall that the output of $\mathcal{M}(\vec{x})$ is the multiset obtained from the coordinates of $(\mathcal{R}(x_1), \ldots, \mathcal{R}(x_n))$. By considering the event that this multiset contains no elements from $I(b)$, the definition of differential privacy gives

$$\mathbb{P}[\mathcal{M}(\vec{x}) \cap I(b) = \emptyset] \leq e^{\varepsilon} \mathbb{P}[\mathcal{M}(\vec{x}') \cap I(b) = \emptyset] + \delta \ . \tag{1}$$

As $\mathbb{P}[\mathcal{M}(\vec{x}) \cap I(b) = \emptyset] = (1 - p_{a,b})^n$ and $\mathbb{P}[\mathcal{M}(\vec{x}') \cap I(b) = \emptyset] = (1 - p_{b,b})(1 - p_{a,b})^{n-1} \leq (1 - p_{b,b})$, we get from (1) that

$$(1 - p_{a,b})^n \leq (1 - p_{b,b}) e^{\varepsilon} + \delta \ .$$

As $p_{b,b} \geq 1 - e^{-\varepsilon}/2$ we get that $p_{a,b} \geq 1 - (1/2 + \delta)^{1/n}$ holds. Finally, $p_{a,b} \geq (1/2 - \delta)/n$ follows from the fact that

$$\begin{aligned}
\left( 1 - \frac{1}{n} \left( \frac{1}{2} - \delta \right) \right)^n &= 1 - \left( \frac{1}{2} - \delta \right) + \frac{n-1}{2n} \left( \frac{1}{2} - \delta \right)^2 - \cdots \\
&\geq 1 - \left( \frac{1}{2} - \delta \right) = \frac{1}{2} + \delta \ ,
\end{aligned}$$

which uses that the terms in the binomial expansion are alternating in sign and decreasing in magnitude. $\square$

We can now choose $k = \lceil n^{1/3} \rceil$ and combine Lemmas 2, 4 and 5 to obtain our lower bound.

**Theorem 3.** *Let $\mathcal{P}$ be an $(\varepsilon, \delta)$-DP $n$-party protocol for real summation on $[0,1]$ in the one-message shuffle model with $\delta < 1/2$. Then,* $\mathrm{MSE}(\mathcal{P}) = \Omega(n^{1/3})$.

*Proof.* By the previous lemmas, taking $\vec{\mathsf{X}} = (\mathsf{X}_1, \ldots, \mathsf{X}_n)$ with independent $\mathsf{X}_i \sim \mathtt{Unif}(I)$ we have

$$\begin{aligned} \mathrm{MSE}(\mathcal{P}, \vec{\mathsf{X}}) &\geq n \sum_{b \in I} \min \left\{ \frac{1 - p_{b,b}}{4k^3}, \frac{1}{48} \left( 1 - \frac{1}{k^2} \right) \min_{a \in I} p_{a,b} \right\} \\ &\geq n \sum_{b \in I} \min \left\{ \frac{e^{-\varepsilon}}{8k^3}, \frac{1}{48n} \left( 1 - \frac{1}{k^2} \right) \left( \frac{1}{2} - \delta \right) \right\} \\ &= nk \min \left\{ \frac{e^{-\varepsilon}}{8k^3}, \frac{1}{48n} \left( 1 - \frac{1}{k^2} \right) \left( \frac{1}{2} - \delta \right) \right\} \quad . \end{aligned}$$

Therefore, taking $k = \lceil n^{1/3} \rceil$ yields $\mathrm{MSE}(\mathcal{P}, \vec{\mathsf{X}}) = \Omega(n^{1/3})$. Finally, the result follows from observing that a lower bound for the expected MSE implies a lower bound for worst-case MSE:

$$\mathrm{MSE}(\mathcal{P}) = \sup_{\vec{x} \in [0,1]^n} \mathrm{MSE}(\mathcal{P}, \vec{x}) \geq \sup_{\vec{x} \in I^n} \mathrm{MSE}(\mathcal{P}, \vec{x}) \geq \mathrm{MSE}(\mathcal{P}, \vec{\mathsf{X}}) = \Omega(n^{1/3}) \quad . \quad \square$$

## 5 Privacy Amplification by Shuffling

In this section we prove a new privacy amplification result for shuffled mechanisms. In particular, we will show that shuffling $n$ copies of an $\varepsilon_0$-LDP local randomizer with $\varepsilon_0 = O(\log(n/\log(1/\delta)))$ yields an $(\varepsilon, \delta)$-DP mechanism with $\varepsilon = O((\varepsilon_0 \wedge 1) e^{\varepsilon_0} \sqrt{\log(1/\delta)/n})$, where $a \wedge b = \min\{a, b\}$. For this same problem, the following privacy amplification bound was obtained by Erlingsson et al. in [15], which we state here for the randomize-then-shuffle setting (cf. Section 1.1).

**Theorem 4 ( [15]).** *If $\mathcal{R}$ is a $\varepsilon_0$-LDP local randomizer with $\varepsilon_0 < 1/2$, then the shuffled protocol $\mathcal{S} \circ \mathcal{R}^n$ is $(\varepsilon, \delta)$-DP with*

$$\varepsilon = 12 \varepsilon_0 \sqrt{\frac{\log(1/\delta)}{n}}$$

*for any $n \geq 1000$ and $\delta < 1/100$.*

Note that our result recovers the same dependencies on $\varepsilon_0$, $\delta$ and $n$ in the regime $\varepsilon_0 = O(1)$. However, our bound also shows that privacy amplification can be extended to a wider range of parameters. In particular, this allows us to show that in order to design a shuffled $(\varepsilon, \delta)$-DP mechanism with $\varepsilon = \Theta(1)$ it suffices to take any $\varepsilon_0$-LDP local randomizer with $\varepsilon_0 = O(\log(\varepsilon^2 n/\log(1/\delta)))$. For shuffled binary randomized response, a dependence of the type $\varepsilon_0 = O(\log(\varepsilon^2 n/\log(1/\delta)))$ between the local and central privacy parameters was obtained in [12] using an ad-hoc privacy analysis. Our results show that this amplification phenomenon is not intrinsic to binary randomized response, and in fact holds for any pure LDP local randomizer. Thus, our bound captures the privacy amplification regimes from both [15] and [12], thus providing a unified analysis of privacy amplification by shuffling.

To prove our bound, we first generalize the key idea behind the analysis of shuffled randomized response given in Section 3. This idea was to ignore any users who respond truthfully, and then show that the responses of users who respond randomly provide privacy for the response submitted by a target individual. To generalize this approach beyond randomized response we introduce the notions of *total variation similarity* $\gamma_{\mathcal{R}}$ and *blanket distribution* $\omega_{\mathcal{R}}$ of a local randomizer $\mathcal{R}$. The similarity $\gamma_{\mathcal{R}}$ measures the probability that the local randomizer will produce an output that is independent of the input data. When this happens, the mechanism submits a sample from the blanket probability distribution $\omega_{\mathcal{R}}$. In the case of Algorithm 1 in Section 3, the parameter $\gamma_{\mathcal{R}^{PH}}$ is the probability $\gamma$ of ignoring the input and submitting a sample from $\omega_{\mathcal{R}^{PH}} = \mathtt{Unif}([k])$, the uniform distribution on $[k]$. We define these objects formally in Section 5.1, then give further examples and also study the relation between $\gamma_{\mathcal{R}}$ and the privacy guarantees of $\mathcal{R}$.

The second step of the proof is to extend the argument that allows us to ignore the users who submit truthful responses in the privacy analysis of randomized response. In the general case, with probability $1 - \gamma_{\mathcal{R}}$ the local randomizer's outcome depends on the data but is not necessarily deterministic. Analyzing this step in full generality – where the randomizer is arbitrary and the domain might be uncountable – is technically challenging. We address this challenge by leveraging a characterization of differential privacy in terms of hockey-stick divergences that originated in the formal methods community to address the verification for differentially private programs [4–6] and has also been used to prove tight results on privacy amplification by subsampling [1]. As a result of this step we obtain a privacy amplification bound in terms of the expectation of a function of a sum of i.i.d. random variables. Our final bound is obtained by using a concentration inequality to bound this expectation.

The bound we obtain with this method provides a relation of the form $F(\varepsilon, \varepsilon_0, \gamma, n) \leq \delta$, where $F$ is a complicated non-linear function. By simplifying this function $F$ further we obtain the asymptotic amplification bounds sketched above, where a bound for $\gamma$ in terms of $\varepsilon_0$ is used. One can also obtain better mechanism-dependent bounds by computing the exact $\gamma$ for a given mechanism. In addition, fixing all but one of the parameters of the problem we can numerically solve the inequality $F(\varepsilon, \varepsilon_0, \gamma, n) \leq \delta$ to obtain exact relations between the parameters without having to provide appropriate constants for the asymptotic bounds in closed-form. We experimentally showcase the advantages of this approach to privacy calibration in Section 6.

Due to space constraints, mathematical proofs from this section are omitted from the present version of the paper. All missing proofs can be found in the extended technical report [2].

### 5.1 Blanket Decomposition

The goal of this section is to provide a canonical way of decomposing any local randomizer $\mathcal{R} : \mathbb{X} \to \mathbb{Y}$ as a mixture between an input-dependent and an input-independent mechanism. More specifically, let $\mu_x$ denote the output distribution of $\mathcal{R}(x)$. Given a collection of distributions $\{\mu_x\}_{x \in \mathbb{X}}$ we will show how to find a probability $\gamma$, a distribution $\omega$ and a collection of distribution $\{\nu_x\}_{x \in \mathbb{X}}$ such that for every $x \in \mathbb{X}$ we have the mixture decomposition $\mu_x = (1-\gamma)\nu_x + \gamma\omega$. Since the component $\omega$ does not depend on $x$, this decomposition shows that $\mathcal{R}(x)$ is input oblivious with probability $\gamma$. Furthermore, our construction provides the largest possible $\gamma$ for which this decomposition can be attained.

To motivate the construction sketched above it will be useful to recall a well-known property of the *total variation distance*. Given probability distributions $\mu, \mu'$ over $\mathbb{Y}$, this distance is defined as

$$\mathfrak{T}(\mu \| \mu') = \sup_{E \subseteq \mathbb{Y}} (\mu(E) - \mu'(E)) = \frac{1}{2} \int |\mu(y) - \mu'(y)| dy \ .$$

Note how here we use the notation $\mu(y)$ to denote the "probability" of an individual outcome, which formally is only valid when the space $\mathbb{Y}$ is discrete so that every singleton is an atom. Thus, in the case where $\mathbb{Y}$ is a continuous space we take $\mu(y)$ to denote the density of $\mu$ at $y$, where the density is computed with respect to some base measure on $\mathbb{Y}$. We note that this abuse of notation is introduced for convenience and does not restrict the generality of our results.

The total variation distance admits a number of alternative characterizations. The following one is particularly useful:

$$\mathfrak{T}(\mu \| \mu') = 1 - \int \min\{\mu(y), \mu'(y)\} dy \ . \qquad (2)$$

This shows that $\mathfrak{T}(\mu \| \mu')$ can be computed in terms of the total probability mass that is simultaneously under $\mu$ and $\mu'$. Equation 2 can be derived from the interpretation of the total variation distance in terms of couplings [24]. Using this characterization it is easy to construct mixture decompositions of the form $\mu = (1-\gamma)\nu + \gamma\omega$, $\mu' = (1-\gamma)\nu' + \gamma\omega$, where $\gamma = 1 - \mathfrak{T}(\mu \| \mu')$ and $\omega(y) = \min\{\mu(y), \mu'(y)\}/\gamma$. These decompositions are optimal in the sense that $\gamma$ is maximal and $\nu$ and $\nu'$ have disjoint support.

Extending the ideas above to the case with more than two distributions will provide the desired decomposition for any local randomizer. In particular, we define the *total variation similarity* of a set of distributions $\Lambda = \{\mu_x\}_{x \in \mathbb{X}}$ over $\mathbb{Y}$ as

$$\gamma_\Lambda = \int \inf_x \mu_x(y) dy \ .$$

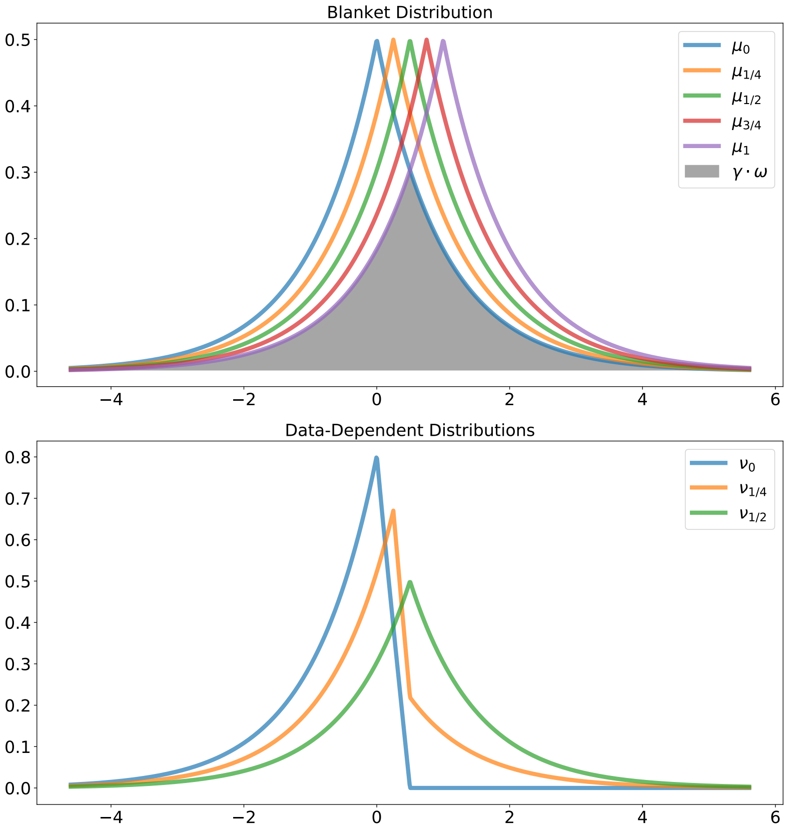

**Fig. 2.** Illustration of the blanket distribution $\omega$ and the data-dependent distributions $\nu_x$ corresponding to a 1-LDP Laplace mechanism with inputs on $[0, 1]$.

We also define the *blanket distribution* of $\Lambda$ as the distribution given by $\omega_\Lambda(y) = \inf_x \mu_x(y)/\gamma_\Lambda$. In this way, given a set of distributions $\Lambda = \{\mu_x\}_{x \in \mathbb{X}}$ with total variation similarity $\gamma$ and blanket distribution $\omega$, we obtain a mixture decomposition $\mu_x = (1 - \gamma)\nu_x + \gamma\omega$ for each distribution in $\Lambda$, where it is immediate to check that $\nu_x = (\mu_x - \gamma\omega)/(1 - \gamma)$ is indeed a probability distribution. It follows from this construction that $\gamma$ is maximal since one can show that, by the definition of $\omega$, for each $y$ there exists an $x$ such that $\nu_x(y) = 0$. Thus, it is not possible to increase $\gamma$ while ensuring that $\nu_x$ are probability distributions.

Accordingly, we can identify a local randomizer $\mathcal{R}$ with the set of distributions $\{\mathcal{R}(x)\}_{x \in \mathbb{X}}$ and define the total variation similarity $\gamma_\mathcal{R}$ and the blanket distributions $\omega_\mathcal{R}$ of the mechanism. As usual, we shall just write $\gamma$ and $\omega$ when the randomizer is clear from the context. Figure 2 plots the blanket distribution and the data-dependent distributions corresponding to the local randomizer obtained by the Laplace mechanism with inputs on $[0, 1]$.

The next result provides expressions for the total variation similarity of three important randomizers: $k$-ary randomized response, the Laplace mechanism on $[0,1]$ and the Gaussian mechanism on $[0,1]$. Note that two of these randomizers offer pure LDP while the third one only offers approximate LDP, showing that the notion of total variation similarity and blanket distribution are widely applicable.

**Lemma 6.** *The following hold:*

1. $\gamma = k/(e^{\varepsilon_0} + k - 1)$ *for* $\varepsilon_0$*-LDP randomized response on* $[k]$*,*
2. $\gamma = e^{-\varepsilon_0/2}$ *for* $\varepsilon_0$*-LDP Laplace on* $[0,1]$*,*
3. $\gamma = 2\mathbb{P}[\mathcal{N}(0,\sigma^2) \leq -1/2]$ *for a Gaussian mechanism with variance* $\sigma^2$ *on* $[0,1]$*.*

This lemma illustrates how the privacy parameters of a local randomizer and its total variation similarity are related in concrete instances. As expected, the probability of sampling from the input-independent blanket grows as the mechanisms become more private. For arbitrary $\varepsilon_0$-LDP local randomizers we are able to show that the probability $\gamma$ of ignoring the input is at least $e^{-\varepsilon_0}$.

**Lemma 7.** *The total variation similarity of any* $\varepsilon_0$*-LDP local randomizer satisfies* $\gamma \geq e^{-\varepsilon_0}$*.*

### 5.2 Privacy Amplification Bounds

Now we proceed to prove the amplification bound stated at the beginning of Section 5. The key ingredient in this proof is to reduce the analysis of the privacy of a shuffled mechanism to the problem of bounding a function of i.i.d. random variables. This reduction is obtained by leveraging the characterization of differential privacy in terms of hockey-stick divergences.

Let $\mu, \mu'$ be distributions over $\mathbb{Y}$. The *hockey-stick divergence* of order $e^\varepsilon$ between $\mu$ and $\mu'$ is defined as

$$\mathfrak{D}_{e^\varepsilon}(\mu \| \mu') = \int [\mu(y) - e^\varepsilon \mu'(y)]_+ dy \ ,$$

where $[u]_+ = \max\{0, u\}$. Using these divergences one obtains the following useful characterization of differential privacy.

**Theorem 5 ( [6]).** *A mechanism* $\mathcal{M} : \mathbb{X}^n \to \mathbb{Y}$ *is* $(\varepsilon,\delta)$*-DP if and only if* $\mathfrak{D}_{e^\varepsilon}(\mathcal{M}(\vec{x}) \| \mathcal{M}(\vec{x}')) \leq \delta$ *for any* $\vec{x} \simeq \vec{x}'$*.*

This result is straightforward once one observes the identity

$$\int [\mu(y) - e^\varepsilon \mu'(y)]_+ dy = \sup_{E \subseteq \mathbb{Y}} (\mu(E) - e^\varepsilon \mu'(E)) \ .$$

An important advantage of the integral formulation is that enables one to reason over individual outputs as opposed to sets of outputs for the case of $(\varepsilon,\delta)$-DP. This is also the case for the usual sufficient condition for $(\varepsilon,\delta)$-DP in terms of a high probability bound for the privacy loss random variable. However, this sufficient condition is not tight for small values of $\varepsilon$ [3], so here we prefer to work with the divergence-based characterization.

The first step in our proof of privacy amplification by shuffling is to provide a bound for the divergence $\mathfrak{D}_{e^\varepsilon}(\mathcal{M}(\vec{x}) \| \mathcal{M}(\vec{x}'))$ for a shuffled mechanism $\mathcal{M} = \mathcal{S} \circ \mathcal{R}^n$ in terms of a random variable that depends on the blanket of the local randomizer. Let $\mathcal{R} : \mathbb{X} \to \mathbb{Y}$ be a local randomizer with blanket $\omega$. Suppose $\mathsf{W} \sim \omega$ is a $\mathbb{Y}$-valued random variable sampled from the blanket. For any $\varepsilon \geq 0$ and $x, x' \in \mathbb{X}$ we define the *privacy amplification random variable* as

$$\mathsf{L}_\varepsilon^{x,x'} = \frac{\mu_x(\mathsf{W}) - e^\varepsilon \mu_{x'}(\mathsf{W})}{\omega(\mathsf{W})} \quad ,$$

where $\mu_x$ (resp. $\mu_{x'}$) is the output distribution of $\mathcal{R}(x)$ (resp. $\mathcal{R}(x')$). This definition allows us to obtain the following result.

**Lemma 8.** *Let $\mathcal{R} : \mathbb{X} \to \mathbb{Y}$ be a local randomizer and let $\mathcal{M} = \mathcal{S} \circ \mathcal{R}^n$ be the shuffling of $\mathcal{R}$. Fix $\varepsilon \geq 0$ and inputs $\vec{x} \simeq \vec{x}'$ with $x_n \neq x'_n$. Suppose $\mathsf{L}_1, \mathsf{L}_2, \ldots$ are i.i.d. copies of $\mathsf{L}_\varepsilon^{x,x'}$ and $\gamma$ is the total variation similarity of $\mathcal{R}$. Then we have the following:*

$$\mathfrak{D}_{e^\varepsilon}(\mathcal{M}(\vec{x}) \| \mathcal{M}(\vec{x}')) \leq \frac{1}{\gamma n} \sum_{m=1}^{n} \binom{n}{m} \gamma^m (1-\gamma)^{n-m} \mathbb{E}\left[\sum_{i=1}^{m} \mathsf{L}_i\right]_+ \quad . \tag{3}$$

The bound above can also be given a more probabilistic formulation as follows. Let $\mathsf{M} \sim \mathtt{Bin}(n, \gamma)$ be the random variable counting the number of users who sample from the blanket of $\mathcal{R}$. Then we can re-write (3) as

$$\mathfrak{D}_{e^\varepsilon}(\mathcal{M}(\vec{x}) \| \mathcal{M}(\vec{x}')) \leq \frac{1}{\gamma n} \mathbb{E}\left[\sum_{i=1}^{\mathsf{M}} \mathsf{L}_i\right]_+ \quad ,$$

where we use the convention $\sum_{i=1}^{m} \mathsf{L}_i = 0$ when $m = 0$.

Leveraging this bound to analyze the privacy of a shuffled mechanism requires some information about the privacy amplification random variables of an arbitrary local randomizer. The main observation here is that $\mathsf{L}_\varepsilon^{x,x'}$ has negative expectation. This means we can expect $\mathbb{E}[\sum_{i=1}^{m} \mathsf{L}_i]_+$ to decrease with $m$ since adding more variables will shift the expectation of $\sum_{i=1}^{m} \mathsf{L}_i$ towards $-\infty$, thus making it less likely to be above 0. Since $m$ represents the number of users who sample from the blanket, this reinforces the intuition that having more users sample from the blanket makes it easier for the data of the $n$th user to be hidden among these samples. The following lemma will help us make this precise by providing the expectation of $\mathsf{L}_\varepsilon^{x,x'}$ as well as its range and second moment.

**Lemma 9.** *Let $\mathcal{R} : \mathbb{X} \to \mathbb{Y}$ be an $\varepsilon_0$-LDP local randomizer with total variation similarity $\gamma$. For any $\varepsilon \geq 0$ and $x, x' \in \mathbb{X}$ the privacy amplification random variable $\mathsf{L} = \mathsf{L}_\varepsilon^{x,x'}$ satisfies:*

1. $\mathbb{E}\mathsf{L} = 1 - e^{\varepsilon}$,
2. $\gamma e^{-\varepsilon_0}(1 - e^{\varepsilon+2\varepsilon_0}) \leq \mathsf{L} \leq \gamma e^{\varepsilon_0}(1 - e^{\varepsilon-2\varepsilon_0})$,
3. $\mathbb{E}\mathsf{L}^2 \leq \gamma e^{\varepsilon_0}(e^{2\varepsilon} + 1) - 2\gamma^2 e^{\varepsilon-2\varepsilon_0}$.

Now we can use the information about the privacy amplification random variables of an $\varepsilon_0$-LDP local randomizer provided by the previous lemma to give upper bounds for $\mathbb{E}[\sum_{i=1}^m \mathsf{L}_i]_+$. This can be achieved by using concentration inequalities to bound the tails of $\sum_{i=1}^m \mathsf{L}_i$. Based on the information provided by Lemma 9 there are multiple ways to achieve this. In this section we unfold a simple strategy based on Hoeffding's inequality that only uses points (1) and (2) above. In Section 5.3 we discuss how to improve these bounds. For now, the following result will suffice to obtain a privacy amplification bound for generic $\varepsilon_0$-LDP local randomizers.

**Lemma 10.** *Let $\mathsf{L}_1, \ldots, \mathsf{L}_m$ be i.i.d. bounded random variables with $\mathbb{E}\mathsf{L}_i = -a \leq 0$. Suppose $b_- \leq \mathsf{L}_i \leq b_+$ and let $b = b_+ - b_-$. Then the following holds:*

$$\mathbb{E}\left[\sum_{i=1}^m \mathsf{L}_i\right]_+ \leq \frac{b^2}{4a} e^{-\frac{2ma^2}{b^2}} \quad .$$

By combining Lemmas 8, 9 and 10 we immediately obtain the main theorem of this section.

**Theorem 6.** *Let $\mathcal{R} : \mathbb{X} \to \mathbb{Y}$ be an $\varepsilon_0$-LDP local randomizer and let $\mathcal{M} = \mathcal{S} \circ \mathcal{R}^n$ be the corresponding shuffled mechanism. Then $\mathcal{M}$ is $(\varepsilon, \delta)$-DP for any $\varepsilon$ and $\delta$ satisfying*

$$\frac{(e^{\varepsilon}+1)^2(e^{\varepsilon_0} - e^{-\varepsilon_0})^2}{4n(e^{\varepsilon}-1)} e^{-Cn\left(\frac{1}{e^{\varepsilon_0}} \wedge \frac{(e^{\varepsilon}-1)^2}{(e^{\varepsilon}+1)^2(e^{\varepsilon_0}-e^{-\varepsilon_0})^2}\right)} \leq \delta \quad , \tag{4}$$

*where $C = 1 - e^{-2} \approx 0.86$.*

While it is easy to numerically test or solve (4), extracting manageable asymptotics from this bound is less straightforward. The following corollary massages this expression to distill insights about privacy amplification by shuffling for generic $\varepsilon_0$-LDP local randomizers.

**Corollary 2.** *Let $\mathcal{R} : \mathbb{X} \to \mathbb{Y}$ be an $\varepsilon_0$-LDP local randomizer and let $\mathcal{M} = \mathcal{S} \circ \mathcal{R}^n$ be the corresponding shuffled mechanism. If $\varepsilon_0 \leq \log(n/\log(1/\delta))/2$, then $\mathcal{M}$ is $(\varepsilon, \delta)$-DP with $\varepsilon = O((1 \wedge \varepsilon_0)e^{\varepsilon_0}\sqrt{\log(1/\delta)/n})$.*

### 5.3 Improved Amplification Bounds

There are at least two ways in which we can improve upon the privacy amplification bound in Theorem 6. One is to leverage the moment information about the privacy amplification random variables provided by point (3) in Lemma 9. The other is to compute more precise information about the privacy amplification random variables for specific mechanisms instead of using the generic bounds provided by Lemma 9. In this section we give the necessary tools to obtain these improvements, which we then evaluate numerically in Section 6.

Hoeffding's inequality provides concentration for sums of bounded random variables. As such, it is easy to apply because it requires little information on the behavior of the individual random variables. On the other hand, this simplicity can sometimes provide sub-optimal results, especially when the random variables being added have standard deviation which is smaller than their range. In this case one can obtain better results by applying one of the many concentration inequalities that take the variance of the summands into account. The following lemma takes this approach by applying Bennett's inequality to bound the quantity $\mathbb{E}[\sum_{i=1}^{m} \mathsf{L}_i]_+$.

**Lemma 11.** *Let $\mathsf{L}_1, \ldots, \mathsf{L}_m$ be i.i.d. bounded random variables with $\mathbb{E}\mathsf{L}_i = -a \leq 0$. Suppose $\mathsf{L}_i \leq b_+$ and $\mathbb{E}\mathsf{L}_i^2 \leq c$. Then the following holds:*

$$\mathbb{E}\left[\sum_{i=1}^{m} \mathsf{L}_i\right]_+ \leq \frac{b_+}{am\log\left(1+\frac{ab_+}{c}\right)} e^{-\frac{mc}{b_+^2}\phi\left(\frac{ab_+}{c}\right)} ,$$

*where $\phi(u) = (1+u)\log(1+u) - u$.*

This results can be combined with Lemmas 7, 8 and 9 to obtain an alternative privacy amplification bound for generic $\varepsilon_0$-LDP local randomizers to the one provided in Theorem 6. However, the resulting bound is cumbersome and does not have a nice closed-form like the one in Theorem 6. Thus, instead of stating the bound explicitly we will evaluate it numerically in the following section.

The other way in which we can provide better privacy bounds is by making them mechanism specific. Lemma 6 already gives exact expression for the total variation similarity $\gamma$ of three local randomizers. To be able to apply Hoeffding's (Lemma 10) and Bennett's (Lemma 11) inequalities to these local randomizers we need information about the range and the second moment of the corresponding privacy amplification random variables. The following results provide this type of information for randomized response and the Laplace mechanism.

**Lemma 12.** *Let $\mathcal{R} : [k] \to [k]$ be the $k$-ary $\varepsilon_0$-LDP randomized response mechanism. Let $\gamma = k/(e^{\varepsilon_0}+k-1)$ be the total variation similarity of $\mathcal{R}$ (cf. Lemma 6). For any $\varepsilon \geq 0$ and $x, x' \in \mathbb{X}$, $x \neq x'$, the privacy amplification random variable $\mathsf{L} = \mathsf{L}_\varepsilon^{x,x'}$ satisfies:*

1. $-(1-\gamma)ke^{\varepsilon} \leq \mathsf{L} - \gamma(1-e^{\varepsilon}) \leq (1-\gamma)k$,
2. $\mathbb{E}\mathsf{L}^2 = \gamma(2-\gamma)(1-e^{\varepsilon})^2 + (1-\gamma)^2 k(1+e^{2\varepsilon})$.

**Lemma 13.** *Let $\mathcal{R} : [0,1] \to \mathbb{R}$ be the $\varepsilon_0$-LDP Laplace mechanism $\mathcal{R}(x) = x + \mathit{Lap}(1/\varepsilon_0)$. For any $\varepsilon \geq 0$ and $x, x' \in \mathbb{X}$ the privacy amplification random variable $\mathsf{L} = \mathsf{L}_\varepsilon^{x,x'}$ satisfies:*

1. $e^{-\varepsilon_0/2}(1-e^{\varepsilon+\varepsilon_0}) \leq \mathsf{L} \leq e^{\varepsilon_0/2}(1-e^{\varepsilon-\varepsilon_0})$,

*2.* $\mathbb{E}\mathsf{L}^2 \leq \frac{e^{2\varepsilon}+1}{3}(2e^{\varepsilon_0/2} + e^{-\varepsilon_0}) - 2e^{\varepsilon}(2e^{-\varepsilon_0/2} - e^{-\varepsilon_0})$.

Again, instead of deriving a closed-form expression like (4) specialized to these two mechanisms, we will numerically evaluate the advantage of using mechanism-specific information in the bounds in the next section. Note that we did not provide a version of these results for the Gaussian mechanism for which we showed how to compute $\gamma$ in Section 5.1. The reason for this is that in this case the resulting privacy amplification random variables are not bounded. This precludes us from using the Hoeffding and Bennett bounds to analyze the privacy amplification in this case. Approaches using concentration bounds that do not rely on boundedness will be explored in future work.

## 6 Experimental Evaluation

In this section we provide a numerical evaluation of the privacy amplification bounds derived in Section 5. We also compare the results obtained with our techniques to the privacy amplification bound of Erlingsson et al. [15].

To obtain values of $\varepsilon$ and $\varepsilon_0$ from bounds on $\delta$ of the form given in Theorem 6 we use a numeric procedure. In particular, we implemented the bounds for $\delta$ in Python and then used SciPy's numeric root finding routines to solve for the desired parameter up to a precision of $10^{-12}$. This leads to a simple and efficient implementation which can be employed in practical applications for the calibration of privacy parameters of local randomizers in shuffled protocols. The resulting code is available at `https://github.com/BorjaBalle/amplification-by-shuffling`.

The results of our evaluation are given in Figure 3. The bounds plotted in this figure are obtained as follows:

1. (EFMRTT'19) is the bound in [15] (see Theorem 4).
2. (Hoeffding, Generic) is the bound from Theorem 6.
3. (Bennett, Generic) is obtained by combining Lemmas 7, 8, 9 and 11.
4. (Hoeffding, RR) is obtained by combining Lemmas 6, 8, 12 and 10.
5. (Bennett, RR) is obtained by combining Lemmas 6, 8, 12 and 11.
6. (Hoeffding, Laplace) is obtained by combining Lemmas 6, 8, 13 and 10.
7. (Bennett, Laplace) is obtained by combining Lemmas 6, 8, 13 and 11.

In panel (i) we observe that our two bounds for generic randomizers give significantly smaller values of $\varepsilon$ than the bound from [15] where the constants where not optimized. Additionally, we see that for generic local randomizers, Hoeffding is better for small values of $n$, while Bennet is better for large values of $n$. In panel (ii) we observe the advantage of incorporating information in the Hoeffding bound about the specific local randomizer. Additionally, this plot allows us to see that for the same level of local DP, binary randomized response has better amplification properties than Laplace, which in turn is better the randomizer response over a domain of size $k = 100$. In panel (iii) we compare the amplification bounds obtained for specific randomizers with the Hoeffding

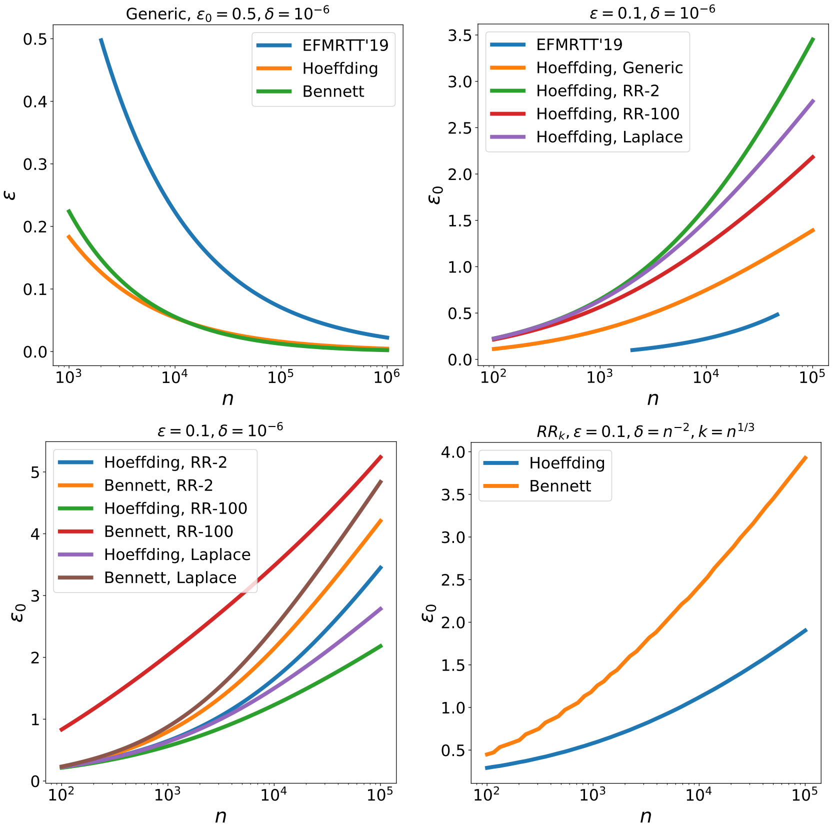

**Fig. 3.** (i) Comparison of $\varepsilon(n)$ for fixed $\varepsilon_0$ and $\delta$ of the bounds obtained for generic $\varepsilon_0$-DP local randomizers using the bound in [15] and our Hoeffding and Bennett bounds. (ii) Comparison of $\varepsilon_0(n)$ for fixed $\varepsilon$ and $\delta$ for generic and specific local randomizers using the Hoeffding bounds. (iii) Comparison of $\varepsilon_0(n)$ for fixed $\varepsilon$ and $\delta$ for specific local randomizers using the Hoeffding and Bennett bound. (iv) Comparison of $\varepsilon_0(n)$ for fixed $\varepsilon$ and $\delta = n^{-2}$ for a randomized response mechanism with domain size $k = n^{1/3}$ using the Hoeffding and Bennett bounds.

and Bennett bounds. We observe that for every mechanism the Bennett bound is better than the Hoeffding bound, especially for large values of $n$. Additionally, the gain of using Bennett instead of Hoeffding is greater for randomized response with $k = 100$ than for other mechanisms. The reason for this is that for fixed $\varepsilon_0$ and large $k$, the total variation similarity of randomized response is close to 1 (cf. Lemma 6). Finally, in panel (iv) we compare the values of $\varepsilon_0$ obtained for a randomized response with domain size growing with the number of users as $k = n^{1/3}$. This is in line with our optimal protocol for real summation in the single-message shuffle model presented in Section 4. We observe that also in this case the Bennett bounds provides a significant advantage over Hoeffding.

To summarize, we showed that our generic bounds outperform the previous amplification bounds developed in [15]. Additionally, we showed that incorporating both information about the variance of the privacy amplification random variable via the use of Bennett's bound, as well as information about the behavior of this random variable for specific mechanisms, leads to significant improvements in the privacy parameters obtained for shuffled protocols. This is important in practice because being able to maximize the $\varepsilon_0$ parameter for the local randomizer – while satisfying a prescribed level of differential privacy in the shuffled protocol – leads to more accurate protocols.

## 7 Conclusion

We have shown a separation result for the single-message shuffle model, showing that it can not achieve the level of accuracy of the curator model of differential privacy, but that it can yield protocols that are significantly more accurate than the ones from the local model. More specifically, we provided a single message protocol for private $n$-party summation of real values in $[0, 1]$ with $O(\log n)$-bit communication and $O(n^{1/6})$ standard deviation. We also showed that our protocol is optimal in terms of accuracy by providing a lower bound for this problem. In previous work, Cheu et al. [12] had shown that the selection problem can be solved more accurately in the central model than in the shuffle model, and that the real summation problem can be solved more accurately in the shuffle model than in the local model. For the former, they rely on lower bounds for selection in the local model by means of a generic reduction from the shuffle to the local model, while our lower bound is directly in the shuffle model, offering additional insight. On the other hand, our single-message protocol for summation is more accurate than theirs.

Moreover, we introduced the notion of the privacy blanket of a local randomizer, and show how it allows us to give a generic treatment to the problem of obtaining privacy amplification bounds in the shuffle model that improves on recent work by Erlingsson et al. [15] and Cheu et al. [12]. Crucially, unlike the proofs in [12, 15], our proof does not rely on privacy amplification by subsampling. We believe that the notion of the privacy blanket is of interest beyond the shuffle model, as it leads to a canonical decomposition of local randomizers that might be useful also in the study of the local model of differential privacy.

For example, Joseph et al. [20] already used a generalization of our blanket decomposition in their study of the role of interactivity in local DP protocols.

## References

1. Balle, B., Barthe, G., Gaboardi, M.: Privacy amplification by subsampling: Tight analyses via couplings and divergences. In: Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada. pp. 6280–6290 (2018)
2. Balle, B., Bell, J., Gascón, A., Nissim, K.: The privacy blanket of the shuffle model. CoRR **abs/1903.02837** (2019), http://arxiv.org/abs/1903.02837
3. Balle, B., Wang, Y.X.: Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In: Proceedings of the 35th International Conference on Machine Learning, ICML (2018)
4. Barthe, G., Gaboardi, M., Grégoire, B., Hsu, J., Strub, P.: Proving differential privacy via probabilistic couplings. In: Symposium on Logic in Computer Science (LICS). pp. 749–758 (2016)
5. Barthe, G., Köpf, B., Olmedo, F., Béguelin, S.Z.: Probabilistic relational reasoning for differential privacy. In: Symposium on Principles of Programming Languages (POPL). pp. 97–110 (2012)
6. Barthe, G., Olmedo, F.: Beyond differential privacy: Composition theorems and relational logic for f-divergences between probabilistic programs. In: International Colloquium on Automata, Languages, and Programming. pp. 49–60. Springer (2013)
7. Beimel, A., Nissim, K., Omri, E.: Distributed private data analysis: Simultaneously solving how and what. In: Wagner, D.A. (ed.) Advances in Cryptology - CRYPTO 2008, 28th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2008. Proceedings. Lecture Notes in Computer Science, vol. 5157, pp. 451–468. Springer (2008). https://doi.org/10.1007/978-3-540-85174-5_25, https://doi.org/10.1007/978-3-540-85174-5_25
8. Bhowmick, A., Duchi, J., Freudiger, J., Kapoor, G., Rogers, R.: Protection Against Reconstruction and Its Applications in Private Federated Learning. arXiv e-prints arXiv:1812.00984 (Dec 2018)
9. Bittau, A., Erlingsson, Ú., Maniatis, P., Mironov, I., Raghunathan, A., Lie, D., Rudominer, M., Kode, U., Tinnés, J., Seefeld, B.: Prochlo: Strong privacy for analytics in the crowd. In: Proceedings of the 26th Symposium on Operating Systems Principles, Shanghai, China, October 28-31, 2017. pp. 441–459. ACM (2017). https://doi.org/10.1145/3132747.3132769, https://doi.org/10.1145/3132747.3132769
10. Boucheron, S., Lugosi, G., Massart, P.: Concentration inequalities: A nonasymptotic theory of independence. Oxford university press (2013)
11. Chan, T.H., Shi, E., Song, D.: Optimal lower bound for differentially private multiparty aggregation. In: Algorithms - ESA 2012 - 20th Annual European Symposium, Ljubljana, Slovenia, September 10-12, 2012. Proceedings. pp. 277–288 (2012)
12. Cheu, A., Smith, A.D., Ullman, J., Zeber, D., Zhilyaev, M.: Distributed differential privacy via shuffling. In: Advances in Cryptology - EUROCRYPT 2019 (2019)
13. Ding, B., Kulkarni, J., Yekhanin, S.: Collecting telemetry data privately. In: Guyon, I., von Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R. (eds.) Advances in Neural Information Processing Systems 30: Annual

Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA. pp. 3574–3583 (2017), `http://papers.nips.cc/paper/6948-collecting-telemetry-data-privately`

14. Dwork, C., McSherry, F., Nissim, K., Smith, A.D.: Calibrating noise to sensitivity in private data analysis. In: Halevi, S., Rabin, T. (eds.) Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings. Lecture Notes in Computer Science, vol. 3876, pp. 265–284. Springer (2006). https://doi.org/10.1007/11681878_14, `https://doi.org/10.1007/11681878_14`
15. Erlingsson, Ú., Feldman, V., Mironov, I., Raghunathan, A., Talwar, K., Thakurta, A.: Amplification by shuffling: From local to central differential privacy via anonymity. In: Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 2468–2479. SIAM (2019)
16. Erlingsson, Ú., Pihur, V., Korolova, A.: RAPPOR: randomized aggregatable privacy-preserving ordinal response. In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014. pp. 1054–1067 (2014)
17. Feldman, V., Mironov, I., Talwar, K., Thakurta, A.: Privacy amplification by iteration. In: 59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018. pp. 521–532 (2018)
18. Ishai, Y., Kushilevitz, E., Ostrovsky, R., Sahai, A.: Cryptography from anonymity. In: FOCS. pp. 239–248. IEEE Computer Society (2006)
19. Jaynes, E.T.: Probability theory: The logic of science. Cambridge university press (2003)
20. Joseph, M., Mao, J., Neel, S., Roth, A.: The role of interactivity in local differential privacy. CoRR **abs/1904.03564** (2019), `http://arxiv.org/abs/1904.03564`
21. Kairouz, P., Bonawitz, K., Ramage, D.: Discrete distribution estimation under local privacy. In: ICML. JMLR Workshop and Conference Proceedings, vol. 48, pp. 2436–2444. JMLR.org (2016)
22. Kairouz, P., Oh, S., Viswanath, P.: Extremal mechanisms for local differential privacy. Journal of Machine Learning Research **17**, 17:1–17:51 (2016)
23. Kasiviswanathan, S.P., Lee, H.K., Nissim, K., Raskhodnikova, S., Smith, A.D.: What can we learn privately? In: 49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA. pp. 531–540. IEEE Computer Society (2008). https://doi.org/10.1109/FOCS.2008.27, `https://doi.org/10.1109/FOCS.2008.27`
24. Lindvall, T.: Lectures on the coupling method. Courier Corporation (2002)
25. Team, A.D.P.: Learning with privacy at scale. Apple Machine Learning Journal **1(9)** (2017)