Tight Accounting in the Shuffle Model of Differential Privacy

Anonymous authors
Paper under double-blind review

Abstract

Shuffle model of differential privacy is a novel distributed privacy model based on a combination of local privacy mechanisms and a secure shuffler. It has been shown that the additional randomisation provided by the shuffler improves privacy bounds compared to the purely local mechanisms. Accounting tight bounds, however, is complicated by the complexity brought by the shuffler. The recently proposed numerical techniques for evaluating (ε, δ)-differential privacy guarantees have been shown to give tighter bounds than commonly used methods for compositions of various complex mechanisms. In this paper, we show how to utilise these numerical accountants for adaptive compositions of general ε-LDP shufflers and for shufflers of k-randomised response mechanisms, including their subsampled variants. This is enabled by an approximation that speeds up the evaluation of the corresponding privacy loss distribution from $\mathcal{O}(n^2)$ to $\mathcal{O}(n)$, where n is the number of users, without noticeable change in the resulting $\delta(\varepsilon)$-upper bounds. We also demonstrate looseness of the existing bounds and methods found in the literature, improving previous composition results for shufflers significantly.

1 Introduction

The shuffle model of differential privacy (DP) is a distributed privacy model which sits between the high trust–high utility centralised DP, and the low trust–low utility local DP (LDP). In the shuffle model, the individual results from local randomisers are only released through a secure shuffler. This additional randomisation leads to "amplification by shuffling", resulting in better privacy bounds against adversaries without access to the unshuffled local results.

We consider computing privacy bounds for both single and composite shuffle protocols, where by composite protocol we mean a protocol, where the subsequent user-wise local randomisers depend on the same local datasets and possibly on the previous output of the shuffler, and at each round the results from the local randomisers are independently shuffled. Moreover, using the analysis by Feldman et al. (2021), we provide bounds in the case the subsequent local randomisers are allowed to depend adaptively on the output of the previous ones.

In this paper we show how numerical accounting (Koskela et al., 2020; 2021; Gopi et al., 2021) can be employed for privacy analysis of both single and composite shuffle DP mechanisms. To our knowledge, ours is the only existing method enabling tight privacy accounting for composite protocols in the shuffle model. We demonstrate that thus obtained bounds can be up to orders of magnitudes tighter than the existing bounds from the literature. By using the tight privacy bounds we can also evaluate how significantly adversaries with varying capabilities differ in terms of the resulting privacy bounds. Due to limited space, we have placed most of the proofs to the Appendix.

1.1 Related work

DP was originally defined in the central model assuming a trusted aggregator by Dwork et al. (2006), while the fully distributed LDP was formally introduced and analysed by Kasiviswanathan et al. (2011). Closely related to the shuffle model of DP, Bittau et al. (2017) proposed the Encode, Shuffle, Analyze framework for distributed learning, which uses the idea of secure shuffler for enhancing privacy. The shuffle model of DP was formally defined by Cheu et al. (2019), who also provided the first separation result showing that the shuffle model is strictly between the central and the local models of DP. Another direction initiated by Cheu et al. (2019) and continued, e.g., by Balle et al. (2020b); Ghazi et al. (2021) has established a separation between single- and multi-message shuffle protocols.

There exists several papers on privacy amplification by shuffling, some of which are central to this paper. Erlingsson et al. (2019) showed that the introduction of a secure shuffler amplifies the privacy guarantees against an adversary, who is not able to access the outputs from the local randomisers but only sees the shuffled output. Balle et al. (2019) improved the amplification results and introduced the idea of privacy blanket, which we also utilise in our analysis of k-randomised response. We compare our bounds with those of Balle et al. (2019) in Section 4.1. Feldman et al. (2021) used a related idea of hiding in the crowd to improve on the previous results, while Girgis et al. (2021) generalised shuffling amplification further to scenarios with composite protocols and parties with more than one local sample under simultaneous communication and privacy restrictions. We use some results of Feldman et al. (2021) in the analysis of general LDP mechanisms, and compare our bounds with theirs in Section 3.3. We also calculate privacy bounds in the setting considered by Girgis et al. (2021), namely in the case a subset of users sending contributions to the shufflers are sampled randomly. This can be seen as a subsampled mechanism and we are able to combine the analysis of Feldman et al. (2021), the privacy loss distribution related subsampling results of Zhu et al. (2022) and FFT accounting to obtain tighter (ε, δ)-bounds than Girgis et al. (2021), as shown in Section 3.4.

2 Background: numerical privacy accounting

Before analysing the shuffled mechanisms we introduce some required theory and notations. In particular, we use the so called privacy loss distribution formalism which is based on finding so called dominating pairs of distributions for the mechanisms at hand. This introduction is dense, for more details we refer to Koskela et al. (2021); Gopi et al. (2021); Zhu et al. (2022).

2.1 Differential privacy and privacy loss distribution

An input data set containing n data points is denoted as $X = (x_1, \dots, x_n) \in \mathcal{X}^n$, where $x_i \in \mathcal{X}$, $1 \leq i \leq n$. We say X and X' are neighbours if we get one by substituting one element in the other (denoted $X \sim X'$).

Definition 1. *Let $\varepsilon > 0$ and $\delta \in [0, 1]$. Let P and Q be two random variables taking values in the same measurable space $\mathcal{O}$. We say that P and Q are (ε, δ)-indistinguishable, denoted $P \simeq_{(\varepsilon,\delta)} Q$, if for every measurable set $E \subset \mathcal{O}$ we have*

$$\Pr(P \in E) \leq \mathrm{e}^{\varepsilon} \Pr(Q \in E) + \delta, \qquad \Pr(Q \in E) \leq \mathrm{e}^{\varepsilon} \Pr(P \in E) + \delta.$$

Definition 2. *Let $\varepsilon > 0$ and $\delta \in [0, 1]$. Mechanism $\mathcal{M} : \mathcal{X}^n \to \mathcal{O}$ is (ε, δ)-DP if for every $X \sim X'$: $\mathcal{M}(X) \simeq_{(\varepsilon,\delta)} \mathcal{M}(X')$. We call $\mathcal{M}$ tightly (ε, δ)-DP, if there does not exist $\delta' < \delta$ such that $\mathcal{M}$ is (ε, δ')-DP. The case when $n = 1$ and $\delta = 0$ is called ε-LDP.*

We rely on the results of Zhu et al. (2022) and characterise (ε, δ)-DP bounds using the hockey-stick divergence which for $\alpha > 0$ is defined as

$$H_\alpha(P||Q) = \int [P(t) - \alpha \cdot Q(t)]_+ \, \mathrm{d}t,$$

where for $x \in \mathbb{R}$, $x_+ = \max\{0, x\}$. Using the hockey-stick divergence, by (Lemma 5, Zhu et al., 2022), tight (ε, δ)-DP bounds can also be characterised as

$$\delta(\varepsilon) = \max_{X \sim X'} H_{\mathrm{e}^\varepsilon}(\mathcal{M}(X)||\mathcal{M}(X')).$$

We can generally find (ε, δ)-bounds by analysing dominating pairs of distributions:

Definition 3 (Zhu et al. 2022)**.** *A pair of distributions (P, Q) is a dominating pair of distributions for mechanism $\mathcal{M}(X)$ if for all $\alpha > 0$,*

$$\max_{X \sim X'} H_\alpha(\mathcal{M}(X) || \mathcal{M}(X')) \leq H_\alpha(P || Q).$$

If the equality holds for all α for some X, X', then (P, Q) is tightly dominating.

Using dominating pairs of distributions, we can obtain $\delta(\varepsilon)$-upper bounds for adaptive compositions:

Theorem 4 (Zhu et al. 2022)**.** *If (P, Q) dominates $\mathcal{M}$ and (P', Q') dominates $\mathcal{M}'$, then $(P \times P', Q \times Q')$ dominates the adaptive composition $\mathcal{M} \circ \mathcal{M}'$.*

Having dominating pairs of distributions for each individual mechanism in a composition, the hockey-stick divergence can be transformed into a more easily computable form by using the privacy loss random variables (PRVs). PRV for a pair of distributions (P, Q) is defined as follows.

Definition 5. *Let $P(t)$ and $Q(t)$ be probability density functions. We define the PRV $\omega_{P/Q}$ as*

$$\omega_{P/Q} = \log \frac{P(t)}{Q(t)}, \quad t \sim P(t),$$

where $t \sim P(t)$ means that t is distributed according to $P(t)$.

With slight abuse of notation, we denote the probability density function of the random variable $\omega_{P/Q}$ by $\omega_{P/Q}(t)$, and call it the privacy loss distribution (PLD).

The $\delta(\varepsilon)$-bounds can be represented using the following representation that involves the PRV.

Theorem 6 (Gopi et al. 2021)**.** *We have:*

$$H_{\mathrm{e}^\varepsilon}(P || Q) = \mathbb{E}_{s \sim \omega_{P/Q}} \left[1 - \mathrm{e}^{\varepsilon - s} \right]_+, \tag{2.1}$$

Moreover, if $\omega_{P/Q}$ is a PRV for the pair of distributions (P, Q) and $\omega_{P'/Q'}$ a PRV for the pair of distributions (P', Q'), then the PRV for the pair of distributions $(P \times P', Q \times Q')$ is given by $\omega_{P/Q} + \omega_{P'/Q'}$

By identifying dominating pairs of distributions for each mechanism in a composition and by formulating the $\delta(\varepsilon)$-bound via hockey-stick divergence as an integral of the form equation 2.1, the numerical PLD accountants (Koskela et al., 2021; Gopi et al., 2021) can be utilised for computing accurate $\delta(\varepsilon)$-bounds.

We will also use the following subsampling amplification result by Zhu et al. (2022) which gives a dominating pair of distributions for the composed mechanism $\mathcal{M} \circ S_{Subset}$, where S_{Subset} denotes a subsampling procedure where, from an input of n entries, a fixed sized subset of $\gamma \cdot n$, $0 < \gamma \leq 1$, entries is sampled.

Lemma 7 (Zhu et al. 2022)**.** *Suppose a pair of distributions (P, Q) is a dominating pair of distributions for a mechanism $\mathcal{M}$ for all datasets of size $\gamma \cdot n$ under the $\sim$-neighbourhood relation (i.e., the substitute relation), where $\gamma > 0$ is the subsampling ratio (size of the subset divided by n). Then, $(\gamma \cdot P + (1 - \gamma) \cdot Q, Q)$ is a dominating pair of distributions for the subsampled mechanism $\mathcal{M} \circ S_{Subset}$.*

When computing tight $\delta(\varepsilon)$-bounds for the shufflers of the k-RR local randomisers and their compositions, instead of equation 2.1, for certain random variables ω_i, $1 \leq i \leq n_c$, where n_c is the number of compositions, we need to evaluate expressions of the form

$$\delta(\varepsilon) = \mathbb{P} \left(\sum\nolimits_{i=1}^{n_c} \omega_i \geq \varepsilon \right). \tag{2.2}$$

The numerical PLD accounting methods are straightforwardly applied to equation 2.2 as well.

2.2 Numerical PLD accounting using FFT

In order to evaluate integrals of the form equation 2.1, we use the Fast Fourier Transform (FFT)-based method by Koskela et al. (2021) called the Fourier Accountant (FA). This means that each PLD is truncated and placed on an equidistant numerical grid over an interval $[-L, L]$, $L > 0$. The distributions for the sums of the PRVs are given by convolutions of the individual PLDs and are evaluated using the FFT algorithm. By a careful error analysis the error incurred by the numerical method can be bounded and an upper $\delta(\varepsilon)$-bound obtained. Expressions of the form equation 2.2 can be similarly evaluated with FFT-convolutions. We note that alternatively, for accurately computing the integrals we could also use the FFT-based method proposed by Gopi et al. (2021).

3 General shuffled ε_0-LDP mechanisms

Feldman et al. (2021) consider general ε_0-LDP local randomisers combined with a shuffler. The analysis allows also sequential adaptive compositions of the user contributions before shuffling. The analysis is based on decomposing individual LDP contributions to mixtures of data dependent part and noise, which leads to finding (ε, δ)-bound for the 2-dimensional distributions (see Thm. 3.2 of Feldman et al., 2021)

$$P = (A + \Delta, C - A + 1 - \Delta), \qquad Q = (A + 1 - \Delta, C - A + \Delta), \tag{3.1}$$

where for $n \in \mathbb{N}$,

$$C \sim \mathrm{Bin}(n-1, \mathrm{e}^{-\varepsilon_0}), \quad A \sim \mathrm{Bin}(C, \tfrac{1}{2}), \quad \Delta \sim \mathrm{Bern}\left(\tfrac{\mathrm{e}^{\varepsilon_0}}{\mathrm{e}^{\varepsilon_0}+1}\right). \tag{3.2}$$

Intuitively, C denotes the number of other users whose mechanism outputs are indistinguishable "clones" of the two differing users, with A denoting random split between these. Feldman et al. (2021) also propose a numerical method to compute the hockey-stick divergence $H_{\mathrm{e}^\varepsilon}(P||Q)$. Using the following lemma, we can use the FFT-based numerical accountants to obtain accurate bounds also for adaptive compositions of general ε_0-LDP shuffling mechanisms:

Lemma 8. *Let X and X' be neighbouring datasets and denote by $\mathcal{A}_s(X)$ and $\mathcal{A}_s(X')$ outputs of the shufflers of adaptive ε_0-LDP local randomisers (for more detailed description, see Thm. 3.2 of Feldman et al., 2021). Then, for all $\alpha > 0$,*

$$H_\alpha(\mathcal{A}_s(X)||\mathcal{A}_s(X')) \leq H_\alpha(P||Q),$$

where P and Q are given as in equation A.1.

Proof. By Thm. 3.2 of Feldman et al. (2021) there exists a post-processing algorithm Φ such $\mathcal{A}_s(X)$ is distributed identically to $\Phi(P)$ and $\mathcal{A}_s(X')$ identically to $\Phi(Q)$. The claim follows then from the data-processing inequality which holds for the hockey-stick divergence (Balle et al., 2020a). ☐

Corollary 9. *The pair of distributions (P, Q) in equation A.1 is a dominating pair of distributions for the shuffling mechanism $\mathcal{A}_s(X)$.*

Furthermore, using Thm. 4, we can bound the $\delta(\varepsilon)$ of n_c-wise adaptive composition of the shuffler $\mathcal{A}_s$ using product distributions of Ps and Qs:

Corollary 10. *Denote $\mathcal{A}_s^{n_c}(X, z_0) = \mathcal{A}_s(X, \mathcal{A}_s(X, ...\mathcal{A}_s(X, z_0)))$ for some initial state z_0. For all neighbouring datasets X and X' and for all $\alpha > 0$,*

$$H_\alpha(\mathcal{A}_s^{n_c}(X)||\mathcal{A}_s^{n_c}(X')) \leq H_\alpha(P \times \ldots \times P||Q \times \ldots \times Q), \tag{3.3}$$

We remark that the case of heterogeneous adaptive compositions (e.g. varying n and ε_0) can be handled analogously using Thm. 4.

Thus, using equation 3.3 for $\alpha = \mathrm{e}^\varepsilon$, we get upper bounds for adaptive compositions of general shuffled ε_0-LDP mechanisms with the Fourier accountant by finding the PLD for the distributions P, Q (given in Eq. equation A.1). Note that even though the resulting (ε, δ)-bound is tight for P's and Q's, it need not be tight for a specific mechanism like the shuffled k-RR. The bound simply gives an upper bound for any shuffled ε_0-LDP mechanisms.

3.1 PLD for shuffled ε_0-LDP mechanisms

As already noted, we can find $\delta(\varepsilon)$-upper bounds for general shuffled ε_0-LDP mechanisms by analysing the pair of distributions (P, Q) of Eq. equation A.1. To analyse the compositions, we need to determine the PLD $\omega_{P/Q}$. Denoting $q = \frac{\mathrm{e}^{\varepsilon_0}}{\mathrm{e}^{\varepsilon_0}+1}$, we see that the distributions in equation A.1 are given by the mixture distributions

$$P = q \cdot P_1 + (1-q) \cdot P_0, \quad Q = (1-q) \cdot P_1 + q \cdot P_0,$$

where

$$P_1 \sim (A+1, C-A), \quad P_0 \sim (A, C-A+1),$$

and A and C are as given in equation 3.2. In the Appendix we give the expressions for the probabilities $\mathbb{P}(P = (a,b))$ and $\mathbb{P}(Q = (a,b))$ needed to determine the discrete-valued PLD

$$\omega_{P/Q}(s) = \sum\nolimits_{a,b} \mathbb{P}(P = (a,b)) \cdot \delta_{s_{a,b}}(s), \quad s_{a,b} = \log\left(\tfrac{\mathbb{P}(P=(a,b))}{\mathbb{P}(Q=(a,b))}\right), \tag{3.4}$$

where $\delta_s(\cdot)$, $s \in \mathbb{R}$, denotes the Dirac delta function centred at s.

3.2 Lowering PLD computational complexity using Hoeffding's inequality

The PLD equation A.2 has $\mathcal{O}(n^2)$ terms which makes its naive evaluation overly expensive for a large number of users n. Using an appropriate tail bound (Hoeffding) for the binomial distribution, we can truncate part of the probability mass and add it directly to δ. More specifically, if each PLD ω_i, $1 \leq i \leq n_c$, in an n_c-composition is approximated by a truncated distribution $\widetilde{\omega}_i$ such that the truncated probability masses are $\tau_i \geq 0$, respectively, then $\delta(\varepsilon) = \widetilde{\delta}(\varepsilon) + \delta(\infty)$, where $\widetilde{\delta}(\varepsilon)$ is the value of the integral of Thm. 4 obtained with the truncated PLDs and $\delta(\infty) = 1 - \prod_i (1-\tau_i) \leq \sum_i \tau_i$, gives an upper bound for the composition without truncations (Koskela et al., 2021). Using the Hoeffding inequality we obtain an accurate approximation of $\omega_{P/Q}$ with only $\mathcal{O}(n)$ terms. We formalise this approximation as follows.

Lemma 11. *Let the PLD $\omega_{P/Q}$ be defined as in equation A.2 and let $\tau > 0$. Consider the set*

$$S_n = [\max\left(0, (p-c_n)(n-1)\right), \min\left(n-1, (p+c_n)(n-1)\right)],$$

where $c_n = \sqrt{\frac{\log(4/\tau)}{2(n-1)}}$ and the set

$$S_i = [\max\big(0, (\tfrac{1}{2} - c_i) \cdot i\big), \min\big(n-1, (\tfrac{1}{2} + c_i) \cdot i\big)],$$

where $c_i = \sqrt{\frac{\log(4/\tau)}{2 \cdot i}}$. Then, the distribution $\widetilde{\omega}_{P/Q}$ defined by

$$\widetilde{\omega}_{P/Q}(s) = \sum_{i \in S_n} \sum_{j \in S_i} \mathbb{P}(P = (j+1, i-j)) \cdot \delta_{s_{j+1,i-j}}(s), \quad s_{a,b} = \log\left(\tfrac{\mathbb{P}(P=(a,b))}{\mathbb{P}(Q=(a,b))}\right) \tag{3.5}$$

has $\mathcal{O}(n \cdot \log(4/\tau))$ terms and differs from $\omega_{P/Q}$ at most mass τ.

Proof. As A is conditioned on C, we first use a tail bound on C and then on A, to reduce the number of terms. Using Hoeffding's inequality for $C \sim \mathrm{Bin}(n-1, p)$ states that for $c > 0$,

$$\mathbb{P}\big(C \leq (p-c)(n-1)\big) \leq \exp\big(-2(n-1)c^2\big),$$
$$\mathbb{P}\big(C \geq (p+c)(n-1)\big) \leq \exp\big(-2(n-1)c^2\big).$$

Requiring that $2 \cdot \exp\left(-2(n-1)c^2\right) \leq \tau/2$ gives the condition $c \geq \sqrt{\frac{\log(4/\tau)}{2(n-1)}}$ and the expressions for c_n and S_n. Similarly, we use Hoeffding's inequality for $A \sim \mathrm{Bin}(C, \frac{1}{2})$ and get expressions for c_i and S_i. The total neglected mass is at most $\tau/2 + \tau/2 = \tau$. For the number of terms, we see that S_n contains at most $2c_n(n-1) = \sqrt{n-1}\sqrt{2 \cdot \log(4/\tau)}$ terms and for each i, and S_i contains at most $2c_i i = \sqrt{i}\sqrt{2 \cdot \log(4/\tau)} \leq \sqrt{n-1}\sqrt{2 \cdot \log(4/\tau)}$ terms. Thus $\widetilde{\omega}_{P/Q}$ has at most $\mathcal{O}(n \cdot \log(4/\tau))$ terms. We get the form equation 3.5 by an appropriate change of variables. □

When evaluating $\widetilde{\omega}_{P/Q}$, we require that the neglected mass is smaller than some prescribed tolerance τ (e.g. $\tau = 10^{-12}$). When computing guarantees for compositions, the cost of FFT for evaluating the convolutions dominates the rest of the computation.

3.3 Experimental comparison to the numerical method of Feldman et al. (2021)

Figure 1 shows a comparison between the PLD approach and the numerical method proposed by Feldman et al. (2021). We see that for a single composition the results given by this method are not far from the results given by the Fourier Accountant (FA). This is expected as their method aims for giving an accurate upper bound for the hockey-stick divergence between P and Q, which for a single composition is equivalent to what FA does. However, the method of Feldman et al. (2021) only works for a single round, whereas FA also gives tight bounds for composite protocols. We emphasise here that FA gives strict upper (ε, δ)-bounds. Moreover, due to the speed-up given by Lemma 11, for a single round protocol, evaluating tight bounds for $n = 10^4$ was approximately 10 times faster than by using the method of Feldman et al. (2021), taking approximately one second on a standard CPU. As the main cost of our approach consists of forming the PLD, the overhead cost of computing guarantees for compositions via FFT is small.

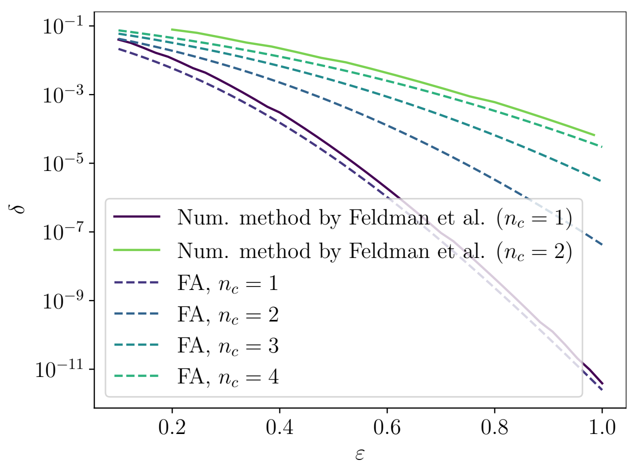

Figure 1: Evaluation of $\delta(\varepsilon)$ for general single and composite shuffle $(\varepsilon_0, 0)$-LDP mechanisms: for single composition protocols the numerical method by Feldman et al. (2021) is close to the tight bounds from FA ($n_c = 1$). Their method is not directly applicable to compositions, for which the Fourier accountant also gives tight bounds. For $n_c = 2$ the bound for the method by Feldman et al. (2021) is computed using the strong composition theorem (Kairouz et al., 2015). Number of users $n = 10^4$ and the LDP parameter $\varepsilon_0 = 4.0$.

3.4 Experimental comparison to the RDP bounds of Girgis et al. (2021)

Girgis et al. (2021) consider a protocol where a randomly sampled, fixed sized subset of users sends contributions to the shuffler on each round, and the local randomisers are assumed to be integer-valued ε_0-LDP mechanisms. This can be seen as a composition of a shuffler and a subsampling mechanism. We can generalise our analysis to this case via Lemma 7, which states that if (P, Q) is a dominating pair of distributions for subset of size $\gamma \cdot n$, , where γ is the subsampling ratio (size of the subset divided by n), then the mixture distribution $(\gamma \cdot P + (1 - \gamma) \cdot Q, Q)$ gives a dominating pair of distributions for the subsampled mechanism. Notice also that by Lemma 8 we obtain such a pair (P, Q) from the expression equation A.1 with n replaced by $\gamma \cdot n$, as Lemma 8 gives a dominating pair (P, Q) for general shuffled ε_0-LDP mechanisms. As we see from Figure 2, the PLD-based approach leads to considerably lower $\varepsilon(\delta)$-bounds. Notice that the tightness of the PLD-based bound is that of the analysis of (Feldman et al., 2021) which gives the dominating pair (P, Q) of equation A.1 and that the RDP-based analysis of (Girgis et al., 2021) is fundamentally different. This explains the fact that the bounds seem to cross as the number of compositions grows large.

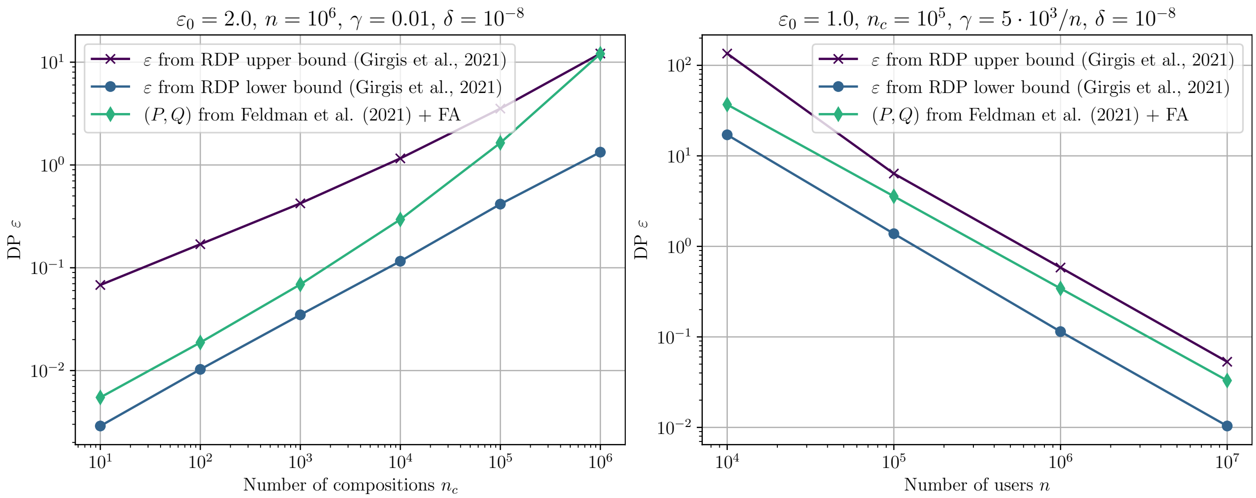

Figure 2: Evaluation of $\varepsilon(\delta)$ for compositions of subsampled shufflers. We compare the bounds obtained using FA and the PLD determined by Lemma 7, and the RDP-bounds given in Thm. 2 of (Girgis et al., 2021) that are mapped to $\varepsilon(\delta)$-bounds using Lemma 1 of (Girgis et al., 2021). Left: bounds for different numbers of users n when number of compositions n_c is fixed. Right: number of compositions n_c varies and n is fixed. Here γ denotes the subsampling ratio.

4 Shuffled k-randomised response

Balle et al. (2019) give a protocol for n parties to compute a private histogram over the domain $[k]$ in the single-message shuffle model. The randomiser is parameterised by a probability γ, and consists of a k-ary randomised response mechanism (k-RR) that returns the true value with probability $1-\gamma$ and a uniformly random value with probability γ. Denote this k-RR randomiser by $\mathcal{R}^{PH}_{\gamma,k,n}$ and the shuffling operation by $\mathcal{S}$. Thus, we are studying the privacy of the shuffled randomiser $\mathcal{M} = \mathcal{S} \circ \mathcal{R}^{PH}_{\gamma,k,n}$.

Consider first the proof of Balle et al. (2019, Thm. 3.1). Assuming without loss of generality that the differing data element between X and X', $X, X' \in [k]^n$, is x_n, the (strong) adversary A_s used by Balle et al. (2019, Thm. 3.1) is defined as follows.

Definition 12. *Let $\mathcal{M} = \mathcal{S} \circ \mathcal{R}^{PH}_{\gamma,k,n}$ be the shuffled k-RR mechanism, and w.l.o.g. let the differing element be x_n. We define adversary A_s as an adversary with the view*

$$\mathit{View}^{A_s}_{\mathcal{M}}(X) = \big((x_1, \ldots, x_{n-1}), \quad \beta \in \{0,1\}^n, \quad (y_{\pi(1)}, \ldots, y_{\pi(n)})\big),$$

where y are the outputs from the shuffler, β is a binary vector identifying which parties answered randomly, and π is a uniformly random permutation applied by the shuffler.

Assuming w.l.o.g. that the differing element $x_n = 1$ and $x'_n = 2$, the proof then shows that for any possible view V of the adversary A_s, $\frac{\mathbb{P}(\mathrm{View}^{A_s}_{\mathcal{M}}(X)=V)}{\mathbb{P}(\mathrm{View}^{A_s}_{\mathcal{M}}(X')=V)} = \frac{n_1}{n_2}$, where n_i denotes the number of messages received by the server with value i after removing from the output Y any truthful answers submitted by the first $n-1$ users. Moreover, Balle et al. (2019) show that for all neighbouring X and X',

$$\mathrm{View}^{A_s}_{\mathcal{M}}(X) \simeq_{(\varepsilon,\delta)} \mathrm{View}^{A_s}_{\mathcal{M}}(X') \tag{4.1}$$

for

$$\delta(\varepsilon) = \mathbb{P}\left(\frac{N_1}{N_2} \geq \mathrm{e}^{\varepsilon}\right), \tag{4.2}$$

where

$$N_1 \sim \mathrm{Bin}\left(n-1, \frac{\gamma}{k}\right) + 1, \quad N_2 \sim \mathrm{Bin}\left(n-1, \frac{\gamma}{k}\right). \tag{4.3}$$

From the proof of Balle et al. (2019, Thm. 3.1) we directly get the following representation for tight $\delta(\varepsilon)$ for non-adaptive compositions of the k-RR shuffler. With certain techniques used by Zhu et al. (2022) to show that adaptive compositions can be analysed using dominating pairs of distributions, we can show that the result holds also for adaptive compositions.

Theorem 13. *Consider n_c adaptive compositions of the k-RR shuffler mechanism $\mathcal{M}$ and an adversary A_s as described in Def. 12 above. Then, the tight (ε, δ)-bound is given by*

$$\delta(\varepsilon) = \mathbb{P}\left(\sum_{i=1}^{n_c} Z_i \geq \varepsilon\right),$$

where Z_i's are independent and for all $1 \leq i \leq n_c$, $Z_i \sim \log\left(\frac{N_1}{N_2}\right)$, where N_1 and N_2 are distributed as in equation C.7.

Balle et al. (2019) showed that for adversary A_s, the shuffled mechanism $\mathcal{M} = \mathcal{S} \circ \mathcal{R}^{PH}_{\gamma,k,n}$ is (ε, δ)-DP for any $k, n \in \mathbb{N}$, $\varepsilon \leq 1$ and $\delta \in (0, 1]$ such that $\gamma = \max\left\{\frac{14 \cdot k \cdot \log(2/\delta)}{(n-1)\cdot\varepsilon^2}, \frac{27 \cdot k}{(n-1)\cdot\varepsilon}\right\}$. Comparison to this bound is shown in Figure 3.

4.1 Tight bounds for varying adversaries using the Fourier accountant

Following the reasoning of the proof of Balle et al. (2019, Thm. 3.1), for adversary A_s (see Def. 12), we can compute tight $\delta(\varepsilon)$-bounds using Thm. 13. Having tight bounds also enables us to evaluate exactly how much different assumptions on the adversary cost us in terms of privacy. For example, instead of the adversary A_s we can analyse a weaker adversary A_w, who has extra information only on the first $n-1$ parties. We formalise this as follows.

Definition 14. *Let $\mathcal{M} = \mathcal{S} \circ \mathcal{R}^{PH}_{\gamma,k,n}$ be the shuffled k-RR mechanism, and w.l.o.g. let the differing element be x_n. Adversary A_w is an adversary with the view*

$$View^{A_w}_{\mathcal{M}}(X) = \big((x_1, \dots, x_{n-1}), \quad \beta \in \{0,1\}^{n-1}, \quad (y_{\pi(1)}, \dots, y_{\pi(n)})\big),$$

where y are the outputs from the shuffler, β is a binary vector identifying which of the first $n-1$ parties answered randomly, and π is a uniformly random permutation applied by the shuffler.

Note that compared to the stronger adversary A_s formalised in Def. 12, the difference is only in the vector β. We write $b = \sum_i \beta_i$, and B for the corresponding random variable in the following. The next theorem gives the random variables we need to calculate privacy bounds for adversary A_w.

Theorem 15. *Assume w.l.o.g. differing elements $x_n = 1, x'_n = 2$, and adversary A_w as given in Def. 14. To find a tight DP bound for $\mathcal{M} = \mathcal{S} \circ \mathcal{R}^{PH}_{\gamma,k,n}$ we can equivalently analyse the random variables P_w, Q_w defined as*

$$P_w = P_1 + P_2, \quad Q_w = Q_1 + Q_2, \tag{4.4}$$

where

$$P_1 \sim (1-\gamma)\cdot N_1|B, \quad P_2 \sim \tfrac{\gamma}{k}\cdot(B+1), \quad Q_1 \sim (1-\gamma)\cdot N_2|B, \quad Q_2 \sim \tfrac{\gamma}{k}\cdot(B+1),$$

$$N_1|B = \mathrm{Bin}(B, 1/k) + \mathrm{Bern}(1-\gamma+\gamma/k) \quad N_2|B = \mathrm{Bin}(B, 1/k) + \mathrm{Bern}(\gamma/k),$$

$$B \sim \mathrm{Bin}(n-1, \gamma).$$

As a direct corollary to this theorem, and analogously to Thm. 13, we have the following result which allows computing tight $\delta(\varepsilon)$-bounds against the adversary A_w for adaptive compositions.

Theorem 16. *Consider n_c adaptive compositions of the k-RR shuffler mechanism $\mathcal{M}$ and an adversary A_w as described in Def. 14 above. Then, the tight (ε, δ)-bound is given by*

$$\delta(\varepsilon) = \mathbb{P}\left(\sum\nolimits_{i=1}^{n_c} Z_i \geq \varepsilon\right),$$

where Z_i's are independent and for all $1 \leq i \leq m$,

$$Z_i \sim \log\left(\frac{N_1}{N_2}\right), \quad N_1 \sim P_w, \quad N_2 \sim Q_w,$$

where P_w and Q_w are given in equation C.1.

Figure 3 shows an empirical comparison of the tight bounds obtained with Fourier accountant assuming the stronger adversary A_s, which leads to the neighbouring random variables P_s, Q_s from equation C.7, or the weaker adversary A_w, corresponding to P_w, Q_w from Thm. 15, together with the analytic bounds from Balle et al. (2019, Thm. 3.1). As shown in the Figure, the numerical bounds are considerably tighter than the analytic one. There is also a clear difference in the tight bounds resulting from assuming either the strong adversary A_s or the weaker A_w. We remark that the evaluation of the distributions for Z_i's in theorems 13 and 16 can be carried out in high accuracy in $\mathcal{O}(n)$-time using Hoeffding's inequality similarly as in Lemma 3.2.

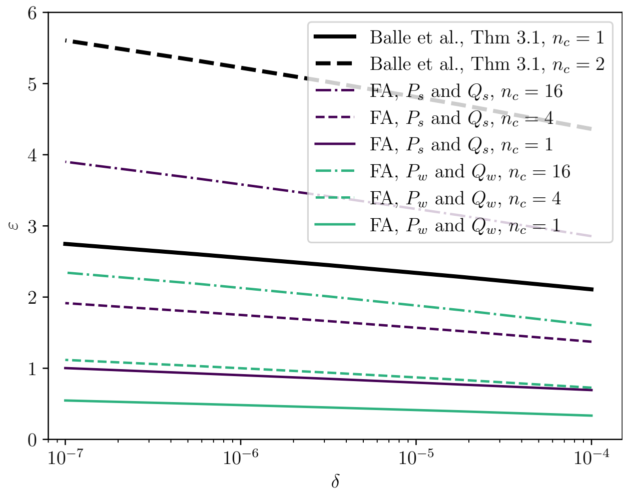

Figure 3: Shuffled k-randomised response: tight bounds are significantly better than the existing analytic one. Tight (ε, δ)-DP bounds obtained using the Fourier accountant (FA) for different number of compositions n_c, and the analytical bound from Balle et al. (2019, Thm. 3.1) for a single composition. For $n_c = 2$ we also compute a bound using the bound Balle et al. (2019, Thm. 3.1) and the strong composition theorem (Kairouz et al., 2015). We apply FA to the $\delta(\varepsilon)$-expression of Thm. 13 (P_s and Q_s), and to the $\delta(\varepsilon)$-expression of Thm. 16 (P_w and Q_w). Both are tight bounds under the assumed adversary (stronger and weaker). FA with P_s, Q_s and $n_c = 1$ is the tight bound with the same assumptions as used in the loose analytic bound. Total number of users $n = 1000$, probability of randomising for each user $\gamma = 0.25$, and $k = 4$.

5 On the difficulty of obtaining bounds in the general case

We have provided means to compute accurate (ε, δ)-bounds for the general ε_0-LDP shuffler using the results by Feldman et al. (2021) and tight bounds for the case of k-randomised response. Using the following example, we illustrate the computational difficulty of obtaining tight bounds for arbitrary local randomisers. Consider neighbouring datasets $X, X' \in \mathbb{R}^n$, where all elements of X are equal, and X' contains one element differing by 1. Without loss of generality (due to shifting and scaling invariance of DP), we may consider the case where X consists of zeros and X' has 1 at some element. Considering a mechanism $\mathcal{M}$ that consists of adding Gaussian noise with variance σ^2 to each element and then shuffling, we see that the adversary sees the output of $\mathcal{M}(X)$ distributed as $\mathcal{M}(X) \sim \mathcal{N}(0, \sigma^2 I_n)$, and the output $\mathcal{M}(X')$ as the mixture distribution $\mathcal{M}(X') \sim \frac{1}{n} \cdot \mathcal{N}(e_1, \sigma^2 I_n) + \ldots + \frac{1}{n} \cdot \mathcal{N}(e_n, \sigma^2 I_n)$, where e_i denotes the ith unit vector. Determining the hockey-stick divergence $H_{e^\varepsilon}(\mathcal{M}(X') || \mathcal{M}(X))$ cannot be projected to a lower-dimensional problem, unlike in the case of the (subsampled) Gaussian mechanism, for example, which is equivalent to a one-dimensional problem (Koskela and Honkela, 2021). This means that in order to obtain tight (ε, δ)-bounds, we need to numerically evaluate the n-dimensional hockey-stick integral $H_{e^\varepsilon}(\mathcal{M}(X')||\mathcal{M}(X))$. Using a numerical grid as in FFT-based accountants is unthinkable due to the curse of the dimensionality. However, we may use the fact that for any data set X, the density function $f_X(t)$ of $\mathcal{M}(X)$ is a permutation-invariant function, meaning that for any $t \in \mathbb{R}^n$ and for any permutation $\sigma \in \pi_n$, $f_X\big(\sigma(t)\big) = f_X(t)$. This allows reducing the number of required points on a regular grid for the hockey stick integral from $O(m^n)$ to $O(m^n/n!)$, where m is the number of discretisation points in each dimension. Recent research on numerical integration of permutation-invariant functions (e.g. Nuyens et al., 2016) suggests it may be possible to significantly reduce or even eliminate the dependence on n using more advanced integration techniques. In the Appendix we give results on experiments where we have computed $H_{e^\varepsilon}(\mathcal{M}(X')||\mathcal{M}(X))$ using Monte Carlo integration on a hypercube $[-L, L]^n$ which requires $\approx 5 \cdot 10^7$ samples for getting two correct significant figures already for $n = 7$.

6 Discussion

We have shown how numerical privacy accounting can be used to calculate accurate upper bounds for compositions of various (ε, δ)-DP mechanisms and different adversaries in the shuffle model. An alternative approach would be to use the Rényi differential privacy (Mironov, 2017). However, as illustrated by the comparison against the results of Girgis et al. (2021) in Fig. 2, numerical PLD accounting leads to considerably tighter bounds. For shuffled mechanisms, the difference appears even more significant than for regular DP-SGD (Koskela et al., 2020; 2021), showing up to an order of magnitude reduction in ε.

Numerical and analytical privacy bounds are in many cases complementary and serve different purposes. Numerical accountants allow finding the tightest possible bounds for production and enable more unbiased comparison of algorithms when accuracy of accounting is not a factor. Analytical bounds enable theoretical research and understanding of scaling properties of algorithms, but the inaccuracy of the bounds raises the risk of misleading conclusions about privacy claims.

While our results provide significant improvements over previous state-of-the-art, they only provide optimal accounting for k-randomised response. Developing optimal accounting for more general mechanisms as well as extending the results to $(\varepsilon_0, \delta_0)$-LDP base mechanisms are important topics for future research.

References

Balle, B., Barthe, G., and Gaboardi, M. (2020a). Privacy profiles and amplification by subsampling. *Journal of Privacy and Confidentiality*, 10(1).

Balle, B., Bell, J., Gascón, A., and Nissim, K. (2019). The privacy blanket of the shuffle model. In *Annual International Cryptology Conference*, pages 638–667. Springer.

Balle, B., Bell, J., Gascón, A., and Nissim, K. (2020b). Private summation in the multi-message shuffle model. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, pages 657–676.

Bittau, A., Erlingsson, Ú., Maniatis, P., Mironov, I., Raghunathan, A., Lie, D., Rudominer, M., Kode, U., Tinnes, J., and Seefeld, B. (2017). Prochlo: Strong privacy for analytics in the crowd. In *Proceedings of the 26th Symposium on Operating Systems Principles*, pages 441–459.

Cheu, A., Smith, A., Ullman, J., Zeber, D., and Zhilyaev, M. (2019). Distributed differential privacy via shuffling. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 375–403. Springer.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Proc. TCC 2006*, pages 265–284.

Erlingsson, Ú., Feldman, V., Mironov, I., Raghunathan, A., Talwar, K., and Thakurta, A. (2019). Amplification by shuffling: From local to central differential privacy via anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2468–2479. SIAM.

Feldman, V., McMillan, A., and Talwar, K. (2021). Hiding among the clones: A simple and nearly optimal analysis of privacy amplification by shuffling. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science*. IEEE.

Ghazi, B., Golowich, N., Kumar, R., Pagh, R., and Velingker, A. (2021). On the power of multiple anonymous messages: Frequency estimation and selection in the shuffle model of differential privacy. In Canteaut, A. and Standaert, F.-X., editors, *Advances in Cryptology – EUROCRYPT 2021*, pages 463–488, Cham. Springer International Publishing.

Girgis, A., Data, D., and Diggavi, S. (2021). Renyi differential privacy of the subsampled shuffle model in distributed learning. *Advances in Neural Information Processing Systems*, 34.

Gopi, S., Lee, Y. T., and Wutschitz, L. (2021). Numerical composition of differential privacy. In *Advances in Neural Information Processing Systems*.

Kairouz, P., Oh, S., and Viswanath, P. (2015). The composition theorem for differential privacy. In *International conference on machine learning*, pages 1376–1385. PMLR.

Kasiviswanathan, S. P., Lee, H. K., Nissim, K., Raskhodnikova, S., and Smith, A. (2011). What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826.

Koskela, A. and Honkela, A. (2021). Computing differential privacy guarantees for heterogeneous compositions using fft. *arXiv preprint arXiv:2102.12412*.

Koskela, A., Jälkö, J., and Honkela, A. (2020). Computing tight differential privacy guarantees using FFT. In *International Conference on Artificial Intelligence and Statistics*, pages 2560–2569. PMLR.

Koskela, A., Jälkö, J., Prediger, L., and Honkela, A. (2021). Tight differential privacy for discrete-valued mechanisms and for the subsampled gaussian mechanism using FFT. In *International Conference on Artificial Intelligence and Statistics*, pages 3358–3366. PMLR.

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

Nuyens, D., Suryanarayana, G., and Weimar, M. (2016). Rank-1 lattice rules for multivariate integration in spaces of permutation-invariant functions - error bounds and tractability. *Adv. Comput. Math.*, 42(1):55–84.

Sommer, D. M., Meiser, S., and Mohammadi, E. (2019). Privacy loss classes: The central limit theorem in differential privacy. *Proceedings on Privacy Enhancing Technologies*, 2019(2):245–269.

Zhu, Y., Dong, J., and Wang, Y.-X. (2022). Optimal accounting of differential privacy via characteristic function. In *International Conference on Artificial Intelligence and Statistics*, pages 4782–4817. PMLR.

A Auxiliary results for determining the PLD in Section 3

The analysis of general ε_0-LDP shufflers is carried out by finding (ε, δ)-bound for the 2-dimensional distributions (see Thm. 3.2 of Feldman et al., 2021)

$$P = (A + \Delta, C - A + 1 - \Delta), \qquad Q = (A + 1 - \Delta, C - A + \Delta), \tag{A.1}$$

In this section we give the needed expressions to determine the PLD

$$\omega_{P/Q}(s) = \sum\nolimits_{a,b} \mathbb{P}(P = (a,b)) \cdot \delta_{s_{a,b}}(s), \tag{A.2}$$

where

$$s_{a,b} = \log \left(\frac{\mathbb{P}(P = (a,b))}{\mathbb{P}(Q = (a,b))} \right).$$

Recall: denoting $q = \frac{\mathrm{e}^{\varepsilon_0}}{\mathrm{e}^{\varepsilon_0}+1}$, the distributions in equation A.1 are given by the mixture distributions

$$\begin{aligned} P &= q \cdot P_1 + (1-q) \cdot P_0, \\ Q &= (1-q) \cdot P_1 + q \cdot P_0, \end{aligned} \tag{A.3}$$

where

$$\begin{gathered} P_1 = (A+1, C-A), \quad P_0 = (A, C-A+1), \\ C \sim \mathrm{Bin}(n-1, \mathrm{e}^{-\varepsilon_0}), \quad A \sim \mathrm{Bin}(C, \tfrac{1}{2}). \end{gathered}$$

A.1 Determining the log ratios $s_{a,b}$

To determine $s_{a,b}$'s, we need the following auxiliary results.

Lemma A.1. *When $b > 0$ and $a > 0$, we have:*

$$\mathbb{P}(P_1 = (a,b)) = \frac{a}{b} \cdot \mathbb{P}(P_0 = (a,b)).$$

Proof. We see that $P_1 = (a,b)$ if and only if $A = a-1$ and $C = a+b-1$. Since

$$\begin{aligned} \mathbb{P}(A = a-1 \,|\, C = a+b-1) &= \binom{a+b-1}{a-1} \frac{1}{2^{a+b-1}} \\ &= \frac{a}{b} \cdot \binom{a+b-1}{a} \frac{1}{2^{a+b-1}} \\ &= \frac{a}{b} \cdot \mathbb{P}(A = a \,|\, C = a+b-1), \end{aligned}$$

we see that

$$\begin{aligned} \mathbb{P}(P_1 = (a,b)) &= \mathbb{P}(C = a+b-1) \cdot \mathbb{P}(A = a-1 \,|\, C = a+b-1) \\ &= \mathbb{P}(C = a+b-1) \cdot \frac{a}{b} \cdot \mathbb{P}(A = a \,|\, C = a+b-1) \\ &= \frac{a}{b} \cdot \mathbb{P}(P_0 = (a,b)), \end{aligned}$$

since $P_0 = (a,b)$ if and only if $A = a$ and $C = a+b-1$. ◻

Using these expressions, and the fact that $\mathbb{P}(P_0 = (a,0)) = 0$ for all a and $\mathbb{P}(P_1 = (0,b)) = 0$ for all b, we get the following expressions needed for $s_{a,b}$'s.

Lemma A.2. *When $b > 0$ and $a \geq 0$,*

$$\frac{\mathbb{P}(P = (a,b))}{\mathbb{P}(Q = (a,b))} = \frac{q \cdot \frac{a}{b} + (1-q)}{q + (1-q)\frac{a}{b}}.$$

When $0 < a \leq n$,

$$\frac{\mathbb{P}(P = (a,0))}{\mathbb{P}(Q = (a,0))} = \frac{q}{1-q}.$$

A.2 Probabilities $\mathbb{P}(P = (a,b))$

To determine $\omega_{P/Q}$, we still need to determine $\mathbb{P}(P = (a,b))$'s. These are given by the following expressions.

Lemma A.3. *When $a > 0$,*

$$\mathbb{P}(P_1 = (a,b)) = \binom{n-1}{i}\binom{i}{j} p^i (1-p)^{n-1-i} \frac{1}{2^i},$$

where $(a,b) = (j+1, i-j)$ (i.e., $C = i$ and $A = j$), and

$$\mathbb{P}(P_0 = (a,b)) = \frac{\mathrm{e}^{-\varepsilon_0}}{1-\mathrm{e}^{-\varepsilon_0}} \frac{n-a-b}{2a} \mathbb{P}(P_1 = (a,b)).$$

For $0 < b \leq n$*,*

$$\mathbb{P}(P_1 = (0,b)) = 0$$

and

$$\mathbb{P}(P_0 = (0,b)) = \binom{n-1}{b-1} \left(\frac{\mathrm{e}^{-\varepsilon_0}}{2}\right)^{b-1} (1-\mathrm{e}^{-\varepsilon_0})^{n-b}.$$

Proof. The expressions follow directly from the definitions of P_0, P_1, A and C. □

B Proof of the Lemma: Lowering PLD computational complexity using Hoeffding's inequality

The PLD equation A.2 has $\mathcal{O}(n^2)$ terms which makes its naive evaluation overly expensive for a large number of users n. Using an appropriate tail bound (Hoeffding) for the binomial distribution, we can truncate part of the probability mass and add it directly to δ. More specifically, if each PLD ω_i, $1 \leq i \leq n_c$, in an n_c-composition is approximated by a truncated distribution $\widetilde{\omega}_i$ such that the truncated probability masses are $\tau_i \geq 0$, respectively, then

$$\delta(\varepsilon) = \widetilde{\delta}(\varepsilon) + \delta(\infty),$$

where $\widetilde{\delta}(\varepsilon)$ is the value of the hockey-stick divergence obtained with the truncated PLDs $\widetilde{\omega}_i$, $1 \leq i \leq n_c$, and where

$$\delta(\infty) = 1 - \prod_i (1-\tau_i) \leq \sum_i \tau_i,$$

gives an upper bound for the composition without truncations, see e.g. Thm 1 in Sommer et al. (2019) or Lemma 4 in Koskela et al. (2021). Using the Hoeffding inequality we obtain an accurate approximation of $\omega_{P/Q}$ with only $\mathcal{O}(n)$ terms. We formalise this approximation as follows.

Lemma 11. *Let* $\tau > 0$*. Consider the set*

$$S_n = [\max\left(0, (p-c_n)(n-1)\right), \min\left(n-1, (p+c_n)(n-1)\right)],$$

where $c_n = \sqrt{\frac{\log(4/\tau)}{2(n-1)}}$ *and the set*

$$S_i = [\max\left(0, (\tfrac{1}{2}-c_i)\cdot i\right), \min\left(n-1, (\tfrac{1}{2}+c_i)\cdot i\right)],$$

where $c_i = \sqrt{\frac{\log(4/\tau)}{2\cdot i}}$*. Then, the distribution* $\widetilde{\omega}_{P/Q}$ *defined by*

$$\widetilde{\omega}_{P/Q}(s) = \sum_{i \in S_n} \sum_{j \in S_i} \mathbb{P}(P = (j+1, i-j)) \cdot \delta_{s_{j+1,i-j}}(s), \quad s_{a,b} = \log\left(\frac{\mathbb{P}(P=(a,b))}{\mathbb{P}(Q=(a,b))}\right) \tag{B.1}$$

has $\mathcal{O}(n \cdot \log(4/\tau))$ *terms and differs from* $\omega_{P/Q}$ *at most mass* τ*.*

Proof. Using Hoeffding's inequality for $C \sim \mathrm{Bin}(n-1,p)$ states that for $c > 0$,

$$\mathbb{P}\big(C \leq (p-c)(n-1)\big) \leq \exp\big(-2(n-1)c^2\big),$$
$$\mathbb{P}\big(C \geq (p+c)(n-1)\big) \leq \exp\big(-2(n-1)c^2\big).$$

Requiring that $2 \cdot \exp\left(-2(n-1)c^2\right) \leq \tau/2$ gives the condition $c \geq \sqrt{\frac{\log(4/\tau)}{2(n-1)}}$ and the expressions for c_n and S_n. Similarly, we use Hoeffding's inequality for $A \sim \mathrm{Bin}(C, \frac{1}{2})$ and get expressions for c_i and S_i. The total neglegted mass is at most $\tau/2 + \tau/2 = \tau$. For the number of terms, we see that S_n contains at most $2c_n(n-1) = \sqrt{n-1}\sqrt{2\cdot\log(4/\tau)}$ terms and for each i, S_i contains at most $2c_i i = \sqrt{i}\sqrt{2\cdot\log(4/\tau)} \leq \sqrt{n-1}\sqrt{2\cdot\log(4/\tau)}$ terms. Thus $\widetilde{\omega}_{P/Q}$ has at most $\mathcal{O}(n\cdot\log(4/\tau))$ terms. We get the expression equation B.1 by the change of variables $a = i+1$ ($A = i$) and $b = i-j$ ($C = j$). □

C Auxiliary results for Section 4

C.1 Proof of Theorem 15

We restate the theorem for ease of reading and then give the proof.

Theorem 15. *Assume w.l.o.g. differing elements $x_n = 1, x'_n = 2$, and adversary A_w as given in Def. 14. To find a tight DP bound for $\mathcal{M} = \mathcal{S} \circ \mathcal{R}^{PH}_{\gamma,k,n}$ we can equivalently analyse the random variables P_w, Q_w defined as*

$$P_w = P_1 + P_2, \quad Q_w = Q_1 + Q_2, \tag{C.1}$$

where

$$\begin{aligned}
&P_1 \sim (1-\gamma)\cdot N_1|B, \quad P_2 \sim \frac{\gamma}{k}\cdot(B+1),\\
&Q_1 \sim (1-\gamma)\cdot N_2|B, \quad Q_2 \sim \frac{\gamma}{k}\cdot(B+1),\\
&B \sim \mathrm{Bin}(n-1,\gamma),\\
&N_i^B|B \sim \mathrm{Bin}(B, 1/k), \quad i = 1,2,\\
&N_1|B = N_1^B|B + \mathrm{Bern}(1-\gamma+\gamma/k),\\
&N_2|B = N_2^B|B + \mathrm{Bern}(\gamma/k).
\end{aligned}$$

Proof. Notice that for k-RR, seeing the shuffler output is equivalent to seeing the total counts for each class resulting from applying the local randomisers to X or X'. The adversary A_w can remove all truthfully reported values by client j, $j \in [n-1]$. Denote the observed counts after this removal by $n_i, i = 1,\dots,k$, so $\sum_{i=1}^k n_i = b+1$.

We now have

$$\begin{aligned}
&\mathbb{P}(\mathrm{View}^{A_w}_{\mathcal{M}}(\mathbf{x}) = V)\\
&= \sum_{i=1}^k \mathbb{P}(N_1 = n_1, \dots, N_i = n_i - 1, N_{i+1} = n_{i+1}, \dots N_k = n_k|B)\cdot\\
&\mathbb{P}(\mathcal{R}(x_n) = i)\cdot\mathbb{P}(B = b)\\
&= \binom{b}{n_1 - 1, n_2, \dots, n_k}\left(\frac{1}{k}\right)^b \cdot \left(1-\gamma+\frac{\gamma}{k}\right)\cdot\gamma^b(1-\gamma)^{n-1-b}\\
&\quad + \sum_{i=2}^k \binom{b}{n_1,\dots,n_i - 1, n_{i+1},\dots,n_k}\left(\frac{1}{k}\right)^b\cdot\frac{\gamma}{k}\cdot\gamma^b(1-\gamma)^{n-1-b}\\
&= \binom{b}{n_1,n_2,\dots,n_k}\frac{\gamma^b(1-\gamma)^{n-1-b}}{k^b}\left[n_1(1-\gamma+\frac{\gamma}{k}) + \sum_{i=2}^k n_i\frac{\gamma}{k}\right]\\
&= \binom{b}{n_1,n_2,\dots,n_k}\frac{\gamma^b(1-\gamma)^{n-1-b}}{k^b}\cdot\\
&\left[n_1(1-\gamma+\frac{\gamma}{k}) + (b+1-n_1)\frac{\gamma}{k}\right]\\
&= \binom{b}{n_1,n_2,\dots,n_k}\frac{\gamma^b(1-\gamma)^{n-1-b}}{k^b}\left[n_1(1-\gamma) + \frac{\gamma}{k}(b+1)\right].
\end{aligned} \tag{C.2}$$

Noting then that $\mathbb{P}(\mathcal{R}^{PH}_{\gamma,k,n}(x'_n) = i) = (1-\gamma+\frac{\gamma}{k})$ when $i = 2$ and $\frac{\gamma}{k}$ otherwise, repeating essentially the same steps gives

$$\mathbb{P}(\mathrm{View}^{A_w}_{\mathcal{M}}(X') = V) = \binom{b}{n_1,n_2,\dots,n_k}\frac{\gamma^b(1-\gamma)^{n-1-b}}{k^b}\left[n_2(1-\gamma) + \frac{\gamma}{k}(b+1)\right]. \tag{C.3}$$

Looking at ratio of the two final probabilities given in Eqs. equation C.2 and equation C.3 we have

$$\mathbb{P}_{V \sim \mathrm{View}^{A_w}_{\mathcal{M}}(X)} \left[\frac{\mathbb{P}(\mathrm{View}^{A_w}_{\mathcal{M}}(X) = V)}{\mathbb{P}(\mathrm{View}^{A_w}_{\mathcal{M}}(X') = V)} \geq e^{\varepsilon} \right] = \mathbb{P} \left[\frac{N_1|B \cdot (1-\gamma) + \frac{\gamma}{k}(B+1)}{N_2|B \cdot (1-\gamma) + \frac{\gamma}{k}(B+1)} \geq e^{\varepsilon} \right],$$

where we write $N_i|B, i \in \{1,2\}$ for the random variable N_i conditional on B. This shows that for DP bounds, the adversaries' full view is equivalent to only considering the joint distribution of $N_i, B, i \in \{1,2\}$, and we can therefore look at the neighbouring random variables

$$P_w = P_1 + P_2, \quad Q_w = Q_1 + Q_2, \tag{C.4}$$

where

$$P_1 \sim (1-\gamma) \cdot N_1|B, \quad P_2 \sim \frac{\gamma}{k} \cdot (B+1),$$
$$Q_1 \sim (1-\gamma) \cdot N_2|B, \quad Q_2 \sim \frac{\gamma}{k} \cdot (B+1).$$

Writing n_i^B for the count in class i resulting from the noise sent by the $n-1$ parties, from k-RR definition we also have

$$B \sim \mathrm{Bin}(n-1, \gamma) \quad \text{and} \quad N_i^B|B \sim \mathrm{Bin}(B, 1/k), \tag{C.5}$$

$i = 1, \ldots, k$. As $V \sim \mathrm{View}^{A_w}_{\mathcal{M}}(X)$, we finally have

$$N_1|B = N_1^B|B + \mathrm{Bern}(1 - \gamma + \gamma/k), \quad N_2|B = N_2^B|B + \mathrm{Bern}(\gamma/k). \tag{C.6}$$

The distributions equation C.5 and equation C.6 determine the neighbouring distributions P_w and Q_w given in equation C.4 which completes the proof. □

C.2 Composition Result for Shuffled k-randomised response

Balle et al. (2019, Thm. 3.1) shows that $\delta(\varepsilon)$-bounds against the strong adversary are obtained by a certain analysis of the random variables:

$$N_1 \sim \mathrm{Bin}\left(n-1, \frac{\gamma}{k}\right) + 1, \quad N_2 \sim \mathrm{Bin}\left(n-1, \frac{\gamma}{k}\right). \tag{C.7}$$

From the proof of Balle et al. (2019, Thm. 3.1) we directly get the following result for adaptive compositions of the k-RR shuffler.

Theorem 13. *Consider n_c adaptive compositions of the k-RR shuffler mechanism $\mathcal{M}$ and an adversary A_s. Then, the tight (ε, δ)-bound is given by*

$$\delta(\varepsilon) = \mathbb{P}\left(\sum_{i=1}^{n_c} Z_i \geq \varepsilon \right),$$

where Z_i's are independent and for all $1 \leq i \leq n_c$, $Z_i \sim \log\left(\frac{N_1}{N_2}\right)$, where N_1 and N_2 are distributed as in equation C.7.

Proof. We first remark that for any $\alpha \geq 0$, i.e., for any neighbouring X and X', when $\alpha \geq 0$,

$$H_\alpha(\mathrm{View}^{A_s}_{\mathcal{M}}(X) || \mathrm{View}^{A_s}_{\mathcal{M}}(X')) = \mathbb{P}\left(\frac{N_1}{N_2} \geq \alpha \right), \tag{C.8}$$

where $N_1 \sim \mathrm{Bin}(n-1, \frac{\gamma}{k}) + 1$, $N_2 \sim \mathrm{Bin}(n-1, \frac{\gamma}{k})$. This can be seen directly from the arguments of the proof of Balle et al. (2019, Thm. 3.1). Next, we may use a similar argument as in the proof of (Zhu et al., 2022, Thm. 10). By using equation C.8 repeatedly, we see that for an adaptive composition of two mechanisms

$\mathcal{M}_1$ and $\mathcal{M}_2$:

$$\begin{aligned}
\delta(\varepsilon) &= \mathbb{P}_{V\sim \text{View}^{A_s}_{\mathcal{M}_1}(X), V'\sim \text{View}^{A_s}_{\mathcal{M}_2}(X,V)}\left[\frac{\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}_1}(X)=V)\cdot\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}_2}(X,V)=V')}{\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}_1}(X')=V)\cdot\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}_2}(X',V)=V')} \geq \mathrm{e}^{\varepsilon}\right] \\
&= \mathbb{P}_{V\sim \text{View}^{A_s}_{\mathcal{M}_1}(X)}\left[\mathbb{P}_{V'\sim \text{View}^{A_s}_{\mathcal{M}_2}(X,V)}\left[\frac{\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}_1}(X)=V)\cdot\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}_2}(X,V)=V')}{\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}_1}(X')=V)\cdot\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}_2}(X',V)=V')} \geq \mathrm{e}^{\varepsilon}\right]\right] \\
&= \mathbb{P}_{V\sim \text{View}^{A_s}_{\mathcal{M}_1}(X)}\left[\mathbb{P}_{V'\sim \text{View}^{A_s}_{\mathcal{M}_2}(X,V)}\left[\frac{\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}_2}(X,V)=V')}{\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}_2}(X',V)=V')} \geq \mathrm{e}^{\varepsilon - \log\frac{\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}_1}(X)=V)}{\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}_1}(X')=V)}}\right]\right] \\
&= \mathbb{P}_{V\sim \text{View}^{A_s}_{\mathcal{M}_1}(X)}\left[\frac{N_1^2}{N_2^2} \geq \mathrm{e}^{\varepsilon - \log\frac{\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}_1}(X)=V)}{\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}_1}(X')=V)}}\right] \\
&= \mathbb{P}_{V\sim \text{View}^{A_s}_{\mathcal{M}_1}(X)}\left[\frac{\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}_1}(X)=V)}{\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}_1}(X')=V)} \geq \mathrm{e}^{\varepsilon - \log\frac{N_1^2}{N_2^2}}\right] \\
&= \mathbb{P}\left[\frac{N_1^1}{N_2^1} \geq \mathrm{e}^{\varepsilon - \log\frac{N_1^2}{N_2^2}}\right] \\
&= \mathbb{P}\left[\frac{N_1^1\cdot N_1^2}{N_2^1\cdot N_2^2} \geq \mathrm{e}^{\varepsilon}\right] \\
&= \mathbb{P}\left[\log\left(\frac{N_1^1}{N_2^1}\right) + \log\left(\frac{N_1^2}{N_2^2}\right) \geq \varepsilon\right],
\end{aligned}$$

where $N_1^1, N_1^2 \sim \text{Bin}(n-1, \frac{\gamma}{k}) + 1$, $N_2^1, N_2^2 \sim \text{Bin}(n-1, \frac{\gamma}{k})$. The proof for $n_c > 2$ goes analogously. □

The proof of the following result which allows computing tight bounds against the adversary A_w, goes analogously to the proof of Thm. 13.

Theorem 16. *Consider m compositions of the k-RR shuffler mechanism $\mathcal{M}$ and an adversary A_w. Then, the tight (ε, δ)-bound is given by*

$$\delta(\varepsilon) = \mathbb{P}\left(\sum_{i=1}^{m} Z_i \geq \varepsilon\right),$$

where Z_i's are independent and for all $1 \leq i \leq m$,

$$Z_i \sim \log\left(\frac{N_1}{N_2}\right), \quad N_1 \sim P_w, \quad N_2 \sim Q_w,$$

where P_w and Q_w are given in equation C.1.

C.3 Experiment for Section 5

Consider neighbouring datasets $X, X' \in \mathbb{R}^n$, where all elements of X are equal, and X' contains one element differing by 1. Without loss of generality (due to shifting and scaling invariance of DP), we may consider the case where X consists of zeros and X' has 1 at some element. Considering a mechanism $\mathcal{M}$ that consists of adding Gaussian noise with variance σ^2 to each element and then shuffling, we see that the adversary sees the output of $\mathcal{M}(X)$ distributed as $\mathcal{M}(X) \sim \mathcal{N}(0, \sigma^2 I_n)$, and the output $\mathcal{M}(X')$ as the mixture distribution $\mathcal{M}(X') \sim \frac{1}{n}\cdot\mathcal{N}(e_1, \sigma^2 I_n) + \ldots + \frac{1}{n}\cdot\mathcal{N}(e_n, \sigma^2 I_n)$, where e_i denotes the ith unit vector. In order to obtain tight (ε, δ)-bounds, we need to numerically evaluate the n-dimensional hockey-stick integral $H_{\mathrm{e}^{\varepsilon}}(\mathcal{M}(X')||\mathcal{M}(X))$.

In Figure 4 we have computed $H_{\mathrm{e}^{\varepsilon}}(\mathcal{M}(X')||\mathcal{M}(X))$ up to $n = 7$ using Monte Carlo integration on a hypercube $[-L, L]^n$ which requires $\approx 5 \cdot 10^7$ samples for getting two correct significant figures for $n = 7$.

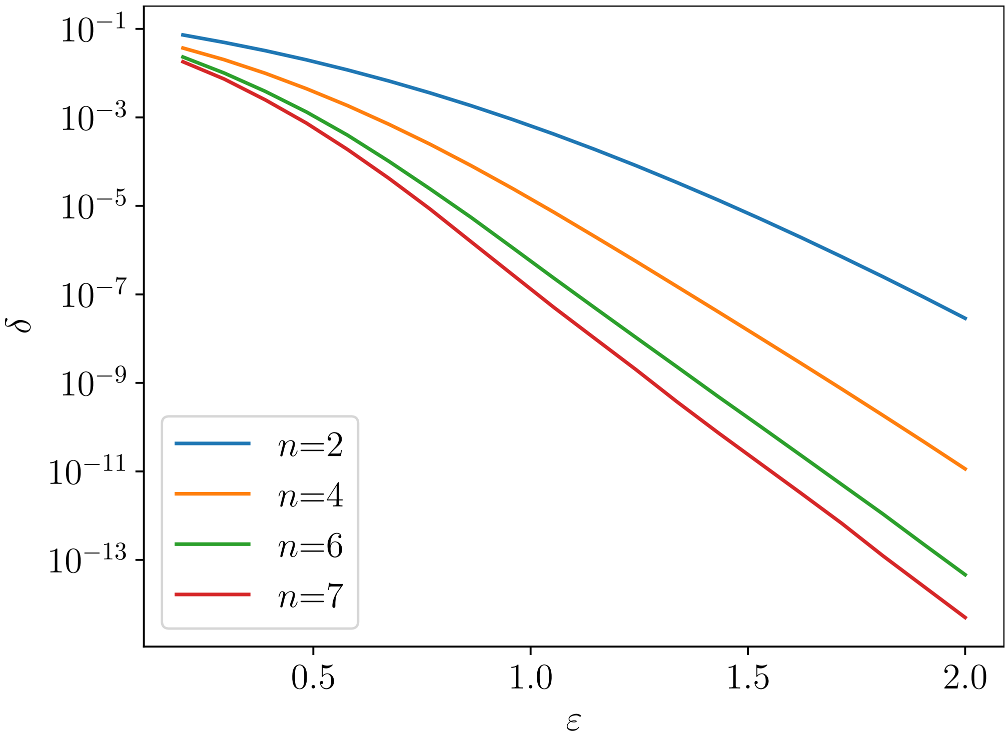

Figure 4: Approximation of tight $\delta(\varepsilon)$ for shuffled outputs of Gaussian mechanisms ($\sigma = 2.0$) by Monte Carlo integration of the hockey-stick divergence $H_{e^\varepsilon}(\mathcal{M}(X')||\mathcal{M}(X))$, using $5 \cdot 10^7$ samples (two correct significant figures).

D Experimental comparison between specialized analysis of k-RR (Balle et al., 2019) and general Clones - analysis (Feldman et al., 2021)

Here we compare the specialized k-RR PLD obtained from (Balle et al., 2019) and the PLD obtained from the 'Clones'-analysis of Feldman et al. (2021) combined with FA. Interestingly, we find that the numerical bounds are very close to each other (Figure 5). We compare using the k-RR parameter values of our Figure 3 in the main text, i.e. $\gamma = 0.25$, and $k = 4$. We increase the parameters of FA so that the discretisation error is negligible. The fact that the lines cross is not a contradiction: the underlying analysis of k-RR is has stronger assumptions about the adversary than the analysis by Feldman et al. (2021), and on the other hand the analysis by Feldman et al. (2021) is not tight as it involves analytical approximations. For the weaker adversary (P_w and Q_w) we already obtain much stronger guarantees than by using the analysis by Feldman et al. (2021).

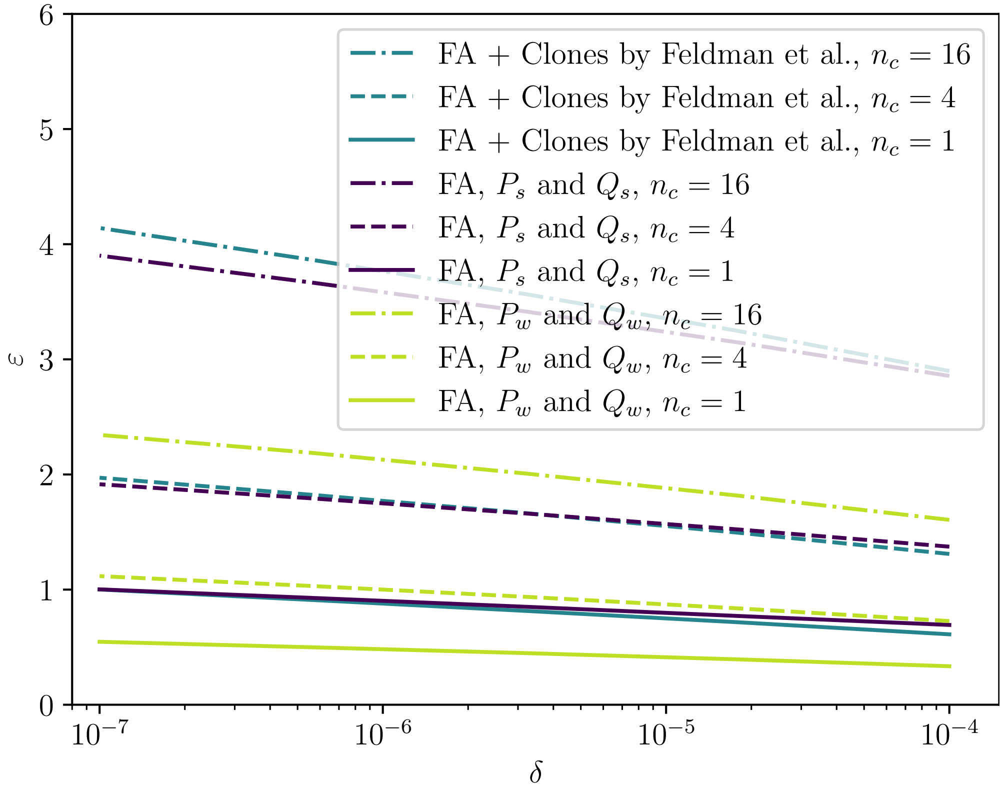

Figure 5: k-RR with the strong adversary (P_s and Q_s from [BBG19]) and our weak adversary (P_w and Q_w) and tight (ε, δ)-DP bounds obtained using FA for different number of compositions n_c. Total number of users $n = 1000$, probability of randomising for each user $\gamma = 0.25$, and $k = 4$. Also shown are the bounds computed using results of [FMT22] combined with FA that hold for general ε_0-LDP local randomizers. In this case (k-RR), $\varepsilon_0 = \log(((1-\gamma)\cdot k + \gamma)/\gamma)$.