Tight Accounting in the Shuffle Model of Differential Privacy

Antti Koskela[1], **Mikko Heikkilä**[2] **and Antti Honkela**[1]

[1] Helsinki Institute for Information Technology HIIT,
Department of Computer Science, University of Helsinki, Finland
[2] Helsinki Institute for Information Technology HIIT,
Department of Mathematics and Statistics, University of Helsinki, Finland

1 Introduction

Our main contribution is to show how the privacy loss distribution (PLD) formalism (Sommer et al., 2019) combined with the Fourier accountant (Koskela et al., 2021) can be used in the shuffle model of DP for several common privacy mechanisms. This provides an efficient method for numerically calculating tight privacy bounds for shuffled mechanisms.

2 Background

Before analysing the shuffled mechanisms we need to introduce some theory and notations. With apologies for conciseness, we start by defining DP and PLD, and finish with the Fourier accountant. For more details, we refer to (Koskela et al., 2021).

2.1 Differential privacy and privacy loss distribution

An input data set containing n data points is denoted as $X = (x_1, \ldots, x_n) \in \mathcal{X}^n$, 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 $X \sim_S X'$).

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*

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

Definition 2. *Let $\varepsilon > 0$ and $\delta \in [0, 1]$. Let $\sim$ define a neighbouring relation. Mechanism $\mathcal{M} : \mathcal{X}^n \to \mathcal{O}$ is $(\varepsilon, \delta, \sim)$-DP if for every $X \sim X'$: $\mathcal{M}(X) \simeq_{(\varepsilon,\delta)} \mathcal{M}(X')$. When the relation is clear from context or irrelevant, we will abbreviate it as (ε, δ)-DP. We call $\mathcal{M}$ tightly $(\varepsilon, \delta, \sim)$-DP, if there does not exist $\delta' < \delta$ such that $\mathcal{M}$ is $(\varepsilon, \delta', \sim)$-DP.*

Definition 3. *Let $\varepsilon > 0$. Mechanism $\mathcal{M} : \mathcal{X} \to \mathcal{O}$ is ε-LDP if for every pair of data points $X, X' \in \mathcal{X}$ and every measurable set $E \subset \mathcal{O}$ we have*

$$\Pr(\mathcal{M}(X) \in E) \leq \mathrm{e}^{\varepsilon} \Pr(\mathcal{M}(X') \in E).$$

We consider discrete-valued mechanisms $\mathcal{M}$ which can be seen as mappings from $\mathcal{X}^n$ to the set of discrete-valued random variables. The generalised probability density functions of $\mathcal{M}(X)$ and $\mathcal{M}(X')$, denoted $f_X(t)$ and $f_{X'}(t)$, respectively, are given by

$$f_X(t) = \sum\nolimits_i a_{X,i} \cdot \delta_{t_{X,i}}(t), \quad f_{X'}(t) = \sum\nolimits_i a_{X',i} \cdot \delta_{t_{X',i}}(t), \tag{2.1}$$

Preprint. Under review.

where $\delta_t(\cdot)$, $t \in \mathbb{R}^d$, denotes the Dirac delta function centred at t, and $t_{X,i}, t_{X',i} \in \mathbb{R}^d$ and $a_{X,i}, a_{X',i} \geq 0$. The privacy loss distribution is defined as follows.

Definition 4. *Let $\mathcal{M} : \mathcal{X}^n \to \mathcal{O}$, $\mathcal{O} \subset \mathbb{R}^d$, be a discrete-valued randomised mechanism and let $f_X(t)$ and $f_{X'}(t)$ be probability density functions as defined by* (2.1)*. We define the generalised privacy loss distribution (PLD) $\omega_{X/X'}$ as*

$$\omega_{X/X'}(s) = \sum\nolimits_{t_{X,i}=t_{X',j}} a_{X,i} \cdot \delta_{s_i}(s), \quad s_i = \log\left(\frac{a_{X,i}}{a_{Y,j}}\right). \tag{2.2}$$

The following theorem (Koskela et al., 2021; Sommer et al., 2019) shows that the tight (ε, δ)-bounds for compositions of non-adaptive mechanisms are obtained using convolutions of PLDs.

Theorem 5. *Consider an n_c-fold non-adaptive composition of a mechanism $\mathcal{M}$. The composition is tightly (ε, δ)-DP for $\delta(\varepsilon)$ given by $\delta(\varepsilon) = \max_{X \sim X'}\{\delta_{X/X'}(\varepsilon), \delta_{X'/X}(\varepsilon)\}$, where*

$$\delta_{X/X'}(\varepsilon) = 1 - \left(1 - \delta_{X/X'}(\infty)\right)^{n_c} + \int_{\varepsilon}^{\infty} (1 - \mathrm{e}^{\varepsilon - s}) \left(\omega_{X/X'} *^{n_c} \omega_{X/X'}\right)(s)\, \mathrm{d}s, \tag{2.3}$$

$$\delta_{X/X'}(\infty) = \sum\nolimits_{\{t_i\,:\,\mathbb{P}(\mathcal{M}(X)=t_i)>0,\,\mathbb{P}(\mathcal{M}(X')=t_i)=0\}} \mathbb{P}(\mathcal{M}(X) = t_i)$$

and $\omega_{X/X'} *^{n_c} \omega_{X/X'}$ denotes the k-fold convolution of the density function $\omega_{X/X'}$ (an analogous expression holds for $\delta_{X'/X}(\varepsilon)$).

In this work, finding the tight (ε, δ)-bounds amounts to finding a pair of random variables P and Q corresponding to neighbouring data sets that determine the PLDs $\omega_{P/Q}$ and $\omega_{Q/P}$. The Fourier accountant algorithm (Koskela et al., 2021) is then used to evaluate (2.3). We remark that the algorithm proposed by Gopi et al. (2021) also gives accurate numerical upper bounds for (2.3) and is more computationally efficient.

3 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 mechanism $\mathcal{M} = \mathcal{S} \circ \mathcal{R}^{PH}_{\gamma,k,n}$.

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, their analysis assumes a fairly strong adversary who knows the full set of the parties who submit random values, the values $x_1, \ldots, x_{n-1}$ and the output $\mathcal{M}(X) = Y = (y_{\pi(1)}, \ldots, y_{\pi(n)})$ after applying the local randomisers and shuffling, where we write π for a uniformly random permutation. That is, writing $\text{View}^A_{\mathcal{M}}$ for the view of adversary A when $\mathcal{M}$ is run on data set X, we define A_s as follows:

Definition 6. *Let $\mathcal{M} = \mathcal{S} \circ \mathcal{R}^{PH}_{\gamma,k,n}$ be the shuffled k-RR mechanism. We define adversary A_s as an adversary with the view*

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

where β is a binary vector identifying which parties answered truthfully.

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, $\frac{\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}}(X)=V)}{\mathbb{P}(\text{View}^{A_s}_{\mathcal{M}}(X')=V)} = \frac{n_1}{n_2}$, 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, it is shown that the corresponding random variables

$$N_1 \sim P_s \quad \text{and} \quad N_2 \sim Q_s,$$

where

$$P_s = \text{Bin}\left(n-1, \frac{\gamma}{k}\right) + 1, \quad \text{and} \quad Q_s = \text{Bin}\left(n-1, \frac{\gamma}{k}\right). \tag{3.1}$$

Thus, $\text{View}^{A_s}_{\mathcal{M}}(X) \simeq_{(\varepsilon,\delta)} \text{View}^{A_s}_{\mathcal{M}}(X')$ if $P_s \simeq_{(\varepsilon,\delta)} Q_s$.

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

3.1 Tight bounds for varying adversaries

Following the reasoning of the proof of Balle et al. (2019, Thm. 3.1), we assume w.l.o.g. that the neighbouring (worst-case) data sets are $X = (x_1, \ldots, x_{n-1}, 1)$ and $X' = (x_1, \ldots, x_{n-1}, 2)$ and define

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

where β is a binary vector identifying which of the first $n-1$ parties answered truthfully. Note that compared to the stronger adversary A_s defined in Section 3, 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.

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 $n_i, i = 1, \ldots, k$, so $\sum_{i=1}^{k} n_i = b + 1$. Using standard techniques and deferring the details to the full version of the paper, writing $N_i|B, i \in \{1,2\}$ for the random variable N_i conditional on B, we show 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 analyse the neighbouring random variables

$$\begin{aligned} P_w &= P_1 + P_2, \quad P_1 \sim (1-\gamma) \cdot N_1|B, \quad P_2 \sim \frac{\gamma}{k} \cdot (B+1), \\ Q_w &= Q_1 + Q_2, \quad Q_1 \sim (1-\gamma) \cdot N_2|B, \quad Q_2 \sim \frac{\gamma}{k} \cdot (B+1). \end{aligned} \tag{3.2}$$

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

$$B \sim \text{Bin}(n-1, \gamma) \quad \text{and} \quad N_i^B|B \sim \text{Bin}(B, 1/k), \quad i = 1, \ldots, k. \tag{3.3}$$

As $V \sim \text{View}_{\mathcal{M}}^{A_w}(X)$, we finally have

$$N_1|B = N_1^B|B + \text{Bern}(1 - \gamma + \gamma/k) \quad \text{and} \quad N_2|B = N_2^B|B + \text{Bern}(\gamma/k). \tag{3.4}$$

The distributions (3.3) and (3.4) determine the neighbouring distributions P_w and Q_w given in (3.2).

3.2 From single message to multi-message protocols

Assuming that at each round of the multi-message protocol the adversary has a view that corresponds to the ones analysed in the single-message case in Section 3, an (ε, δ)-DP upper bound for the multi-message protocol is then obtained from a non-adaptive composition of certain pairs of random variables. This can be seen as follows.

As noted previously in Section 3, for the single message k-RR protocol and adversary A_s, for any possible view $V = \{(x_1, \ldots, x_{n-1}), \beta \in \{0,1\}^n, Y = (y_{\pi(1)}, \ldots, y_{\pi(n)})\}$ of the adversary, $\text{View}_{\mathcal{M}}^{A_s}(X) \simeq_{(\varepsilon,\delta)} \text{View}_{\mathcal{M}}^{A_s}(X')$ if $P_s \simeq_{(\varepsilon,\delta)} Q_s$ where P_s, Q_s are given by (3.1).

In the multi-message setting with n_c communication rounds, where the messages of each party depend on the same fixed data set throughout the protocol run, we may similarly assume that besides the results from the previous rounds, the strong adversary A_s knows

1. The input data $(x_1^i, \ldots, x_{n-1}^i)$ at each round i (assuming that the differing element is x_n^i).
2. The full set $\beta_i \in \{0,1\}^n$ of users who submit random values at each round i.
3. The output $\mathcal{M}_i(X) = Y^i = (y^i_{\pi_i(1)}, \ldots, y^i_{\pi_i(n)})$ after each round i.

Then clearly, when the local randomisers and the shuffler are all independent over the rounds, this multi-message protocol is (ε, δ)-DP if n_c-wise compositions of P_s and Q_s are (ε, δ)-indistinguishable, i.e., if

$$(P_s, \ldots, P_s) \simeq_{(\varepsilon,\delta)} (Q_s, \ldots, Q_s).$$

Figure 1a 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 (3.1), or the weaker adversary A_w, corresponding to P_w, Q_w from (3.2), together with the loose analytic bounds from Balle et al. (2019, Thm. 3.1). As shown in the Figure, the tight bounds are considerably better 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.

4 General analysis via clones of ε_0-LDP local randomisers

Feldman et al. (2020) consider general ε_0-LDP local randomisers combined with a shuffler. The model allows adaptive compositions of user contributions allowing e.g. a novel analysis of the differentially private stochastic gradient descent. The analysis is based on obtaining (ε, δ)-bound for the 2-dimensional distributions (see Remark 3.5 of Feldman et al., 2020)

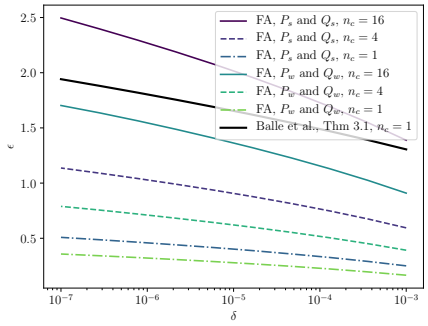
$$P = (A + \Delta, C - A) \quad \text{and} \quad Q = (A, C - A + \Delta), \tag{4.1}$$

where

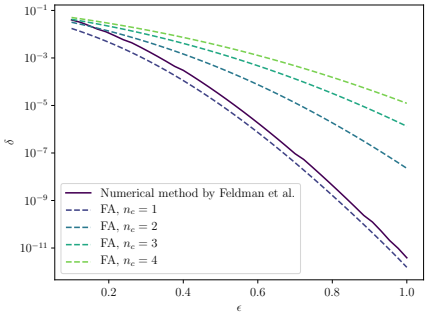
$$C \sim \mathrm{Bin}(n-1, \mathrm{e}^{-\varepsilon_0}), \quad A \sim \mathrm{Bin}(C, \tfrac{1}{2}) \quad \text{and} \quad \Delta \sim \mathrm{Bern}\left(\tfrac{\mathrm{e}^{\varepsilon_0}}{\mathrm{e}^{\varepsilon_0}+1}\right), \tag{4.2}$$

$n \in \mathbb{N}$, $p \in [0, 1]$. Feldman et al. (2020) give also a numerical method for obtaining a tight (ε, δ)-bound (i.e., for evaluating the hockey-stick divergence between P and Q). We apply the Fourier Accountant to the PLD determined by P and Q and this way obtain tight (ε, δ)-bounds also for non-adaptive compositions of the shuffling mechanism, i.e., for the multi-message protocols (see Section 3.2). We leave the details to the full version of the paper.

Figure 1b shows a comparison between the PLD approach and the numerical method proposed by Feldman et al. (2020) (in the implementation we use the parameter value $S = 1$). We see that for a single call of the shuffling mechanism the results given by this method are not far from the results given by the Fourier Accountant. This is expected as the method by Feldman et al. (2020) gives an accurate upper bound for the hockey-stick divergence between P and Q which is exactly what the Fourier Accountant does, however here also for non-adaptive compositions of the mechanism.



(a) Shuffled single and multi-message 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). We apply FA to distributions P_s and Q_s of equation (3.1), and to P_w and Q_w of equation (3.2); both are tight bounds under the assumed adversary. 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 $k = 4$.



(b) Evaluation of $\delta(\varepsilon)$ for a single call of the shuffling mechanism using the numerical method by Feldman et al. (2020), and for n_c number of non-adaptive compositions ($n_c = 1, 2, 3, 4$) using the Fourier Accountant. Notice that the numerical method by Feldman et al. (2020) is only applicable for a single call of the mechanism. Here number of users $n = 10^4$ and the LDP parameter $\varepsilon_0 = 4.0$.

References

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.

Feldman, V., McMillan, A., and Talwar, K. (2020). Hiding among the clones: A simple and nearly optimal analysis of privacy amplification by shuffling. *arXiv preprint arXiv:2012.12803*.

Gopi, S., Lee, Y. T., and Wutschitz, L. (2021). Numerical composition of differential privacy. *arXiv preprint arXiv:2106.02848*.

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.

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.