RÉNYI DIFFERENTIAL PRIVACY IN THE SHUFFLE MODEL: ENHANCED AMPLIFICATION BOUNDS

Name of author

Address - Line 1
Address - Line 2
Address - Line 3

ABSTRACT

The shuffle model of Differential Privacy has gained significant attention in privacy-preserving data analysis due to its remarkable tradeoff between privacy and utility. This can be attributed to the privacy amplification effects, which enable stronger privacy guarantees by incorporating a shuffling procedure prior to transmitting locally randomized data to the analyzer. A key focus in this field is to achieve more precise bounds for privacy amplification. In this work, our primary contribution is enhancing the bounds for Rényi-Differential Privacy in the shuffle model. This is achieved through analyzing the distance between a fixed pair of distributions comprising three values. As a result, we have successfully achieved a nearly optimal bound of $O(\frac{2e^{\epsilon_0}\lambda}{n})$, with only a negligible difference from the lower bound.

Index Terms— Rényi Differential Privacy, Shuffle Model, Privacy Amplification

1. INTRODUCTION

Rényi Differential Privacy (RDP) [1] is a variant of Differential Privacy (DP), which is currently the standard and elegant framework for protecting privacy [2, 3]. RDP provides a flexible framework for quantifying privacy guarantees by introducing the Rényi order parameter, allowing fine-grained control over privacy levels and facilitating the composition of privacy guarantees [4].

The interest in the shuffle model has been driven by its privacy amplification effect [5, 6, 7, 8]. This effect is significant because when adding local noise to protect individual privacy, particularly in sensitive data scenarios, utility is often compromised [9]. Therefore, it is crucial to accurately characterize the amplification effects of a shuffler when applying the shuffle model to various algorithms.

The primary focus in this area is to achieve more precise RDP bounds for privacy amplification [7, 8, 10, 11]. In the shuffle model, individuals' outputs from local randomizers are released through a trusted shuffler (as depicted in Figure 1). While advanced composition theorems for DP [2] can quantify privacy leakage, those bounds may not be sufficiently tight. To address this limitation, the ”moment account” framework was developed by [4], enabling a much tighter composition. This is achieved by providing the composition privacy guarantee in terms of RDP and subsequently mapping it back to the DP guarantee [1]. Therefore, the development of RDP privacy guarantees can lead to stronger composition privacy results. For clarity and convenience, Table 1 outlines previous findings as well as our own regarding privacy amplification through shuffling.

However, as noted in [10], there is still a multiplicative gap of the order e^{ϵ_0} between the lower and upper bounds. To address this gap, we have developed new techniques that enable us to provide an RDP bound that is nearly optimal. Moreover, we have been able to relax constraints on the parameters involved. Our work thus contributes to a better understanding of RDP bounds for the shuffle model.

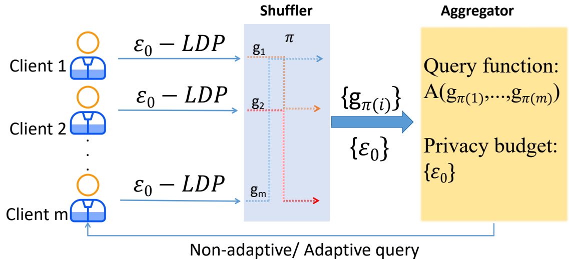

Fig. 1: The shuffle model with differential privacy

2. PRELIMINARIES AND NOTATIONS

This section gives essential terminology, definitions and properties related to differential privacy. Due to space limitations, proofs will be included in the full version.

Definition 1. *((ϵ, δ)-Central Differential Privacy) A randomized algorithm M is called (ϵ, δ)-indistinguishable if for all*

Methods	Results	LB [10]
Erlingsson et al. (SODA, 2019) [7]	$O(\frac{e^{6\epsilon_0}\lambda}{n})$	$O(\frac{e^{\epsilon_0}\lambda}{n})$
Girgis et al. (CCS, 2021) [10]	$O(\frac{e^{2\epsilon_0}\lambda}{n})$	
Feldman et al. (FOCS, 2022)[8]	$O(\frac{64e^{\epsilon_0}\lambda}{n})$	
Feldman et al. (SODA, 2023)[11]		
This work	$O(\frac{2e^{\epsilon_0}\lambda}{n})$	

Table 1: Privacy amplification of Rényi-DP via shuffling. λ is the order of RDP and LB represents the lower bound.

$S \subseteq Range(M)$ and for all neighboring databases D_0, D_1:

$$\mathbb{P}(M(D_0) \in S) \leq e^{\epsilon}\mathbb{P}(M(D_1) \in S) + \delta, \tag{1}$$

where D_0 and D_1 are considered neighboring if they differ by exactly one record, and we denote this relationship as $D_0 \sim D_1$.

Central Differential Privacy requires a trustworthy server, which is difficult in practice. A stronger privacy guarantee for each individual users can be given in the local setting as there is no need to trust a centralized authority [12].

Definition 2. *((ϵ, δ)-Local Differential Privacy) An algorithm $\mathcal{R} : \mathcal{D} \to \mathcal{S}$ is a (ϵ, δ)-DP local randomizer if for all pairs $x, x' \in \mathcal{D}$, $\mathcal{R}(x)$ and $\mathcal{R}(x')$ are (ϵ, δ)-indistinguishable.*

Lemma 1 (Laplace mechanism [2])**.** *For a function $g : D \to \mathbb{R}^d$, let l_1 sensitivity be defined as $\Delta(g) = \max_{D_0 \sim D_1} \|g(D_0) - g(D_1)\|_1$, then for any $\epsilon \in (0, 1)$, the noisy output $h(D) = g(D) + Lap(\Delta(g)/\epsilon)$ satisfies $(\epsilon, 0)$-DP.*

Definition 3. *(Rényi Divergence) For two random variables U and V, the Rényi divergence of U and V of order $\lambda > 1$ is defined as:*

$$D^{\lambda}(U\|V) = \frac{1}{\lambda - 1} log \mathop{\mathbb{E}}_{x \sim V}\left[\left(\frac{U(x)}{V(x)}\right)^{\lambda}\right]. \tag{2}$$

Introduced in [1], Rényi differential privacy (RDP) can be defined based on Rényi divergence.

Definition 4 (Rényi Differential Privacy)**.** *A mechanism M is said to be $(\lambda, \epsilon(\lambda))$-RDP if for all neighbouring pairs X_0, X_1, it holds that*

$$D^{\lambda}(M(X_0)\|M(X_1)) \leq \epsilon. \tag{3}$$

Finally, we establish the framework of privacy protection algorithm under consideration in this paper. The notation $[n]$ represents the set of natural numbers from 1 to n.

Definition 5. *For a domain $\mathcal{D}$, let $\mathcal{R}^{(i)} : \mathcal{S}^{(1)} \times \mathcal{S}^{(2)} \times \cdots \times \mathcal{S}^{(i-1)} \times \mathcal{D} \to \mathcal{S}^{(i)}$ for $i \in [n]$, where $\mathcal{S}^{(i)}$ is the range space of $\mathcal{R}^{(i)}$, be a sequence of algorithms such that $\mathcal{R}^{(i)}(z_{1:i-1}, \cdot)$ is an (ϵ_0, δ_0)-LDP randomizer for all values of auxiliary inputs $z_{1:i-1} \in \mathcal{S}^{(1)} \times \mathcal{S}^{(2)} \times \cdots \mathcal{S}^{(i-1)}$. Let $\mathcal{A}_R : \mathcal{D} \to \mathcal{S}^{(1)} \times \mathcal{S}^{(2)} \times \cdots \times \mathcal{S}^{(n)}$ represent the algorithm applied to the given dataset $x_{1:n} \in \mathcal{D}^n$. The algorithm sequentially computes $z_i = \mathcal{R}^{(i)}(z_{1:i-1}, x_i)$ for $i \in [n]$ and outputs $z_{1:n}$. We refer to $\mathcal{A}_R(\mathcal{D})$ as an (ϵ_0, δ_0)-LDP adaptive process. Alternatively, if we first uniformly sample a permutation $\pi : [n] \to [n]$, and then sequentially compute $z_i = \mathcal{R}^{(i)}(z_{1:i-1}, x_{\pi_i})$ for $i \in [n]$, we say it is a shuffled process and denote it as $\mathcal{A}_{R,S}(\mathcal{D})$. Here, $\pi_i = \pi(i)$ represents the position of i after permutation.*

Remark 1. *Especially, $\mathcal{A}_R(\mathcal{D})$ is an $(\epsilon_0, 0)$-DP adaptive process if $\delta_0 = 0$. For the sake of brevity and convenience of notation, we omit D and use $\mathcal{A}_R$, $\mathcal{A}_{R,S}$ to represent the adaptive process and the shuffled adaptive process, respectively.*

Proposition 1. *(Feldman et al. [8]) For a domain $\mathcal{D}$, let $\mathcal{A}_R$ be the $(\epsilon_0, 0)$-LDP adaptive process and $\mathcal{A}_{R,S}$ be the related shuffled $(\epsilon_0, 0)$-LDP adaptive process. Assume $X_0 = (x_1^0, x_2, \ldots, x_n)$ and $X_1 = (x_1^1, x_2, \ldots, x_n)$ be two neighbouring datasets such that for all $j \neq 1, x_j \notin \{x_1^0, x_1^1\}$. Suppose that there exists a positive value $p \in (0, 1]$ such that for all $i \in [n], x \in \mathcal{D}\backslash\{x_1^0, x_1^1\}$ and $z_{1:i-1} \in \mathcal{S}^{(1)} \times \mathcal{S}^{(2)} \times \cdots \times \mathcal{S}^{(i-1)}$, there exists a distribution $LO^{(i)}(z_{1:i-1}, x)$ such that*

$$\begin{aligned}\mathcal{R}^{(i)}(z_{1:i-1}, x) =& \frac{p}{2}\mathcal{R}^{(i)}(z_{1:i-1}, x_1^0) + \frac{p}{2}\mathcal{R}^{(i)}(z_{1:i-1}, x_1^1) \\ &+ (1-p)LO^{(i)}(z_{1:i-1,x}).\end{aligned} \tag{4}$$

Then there exists a randomized postprocessing algorithm f such that $\mathcal{A}_s(X_0)$ is distributed identically to $f(A + \Delta, C - A + 1 - \Delta)$ and $\mathcal{A}_s(X_1)$ is distributed identically to $f(A + 1 - \Delta, C - A + \Delta)$, where $p = e^{-\epsilon_0}, \Delta \sim Bern(\frac{e^{\epsilon_0}}{e^{\epsilon_0}+1}), C \sim Bin(n-1, p), A \sim Bin(C, 1/2)$.

The starting point of this paper is Proposition 1, which transforms the original problem into a simpler task of analyzing a non-adaptive protocol. It is worth noting that Proposition 1 mentions the joint distribution of A and C, which corresponds to the multinomial distribution $Multinom(n-1; p/2, p/2, 1-p)$. Here, A and $C - A$ represent the number of 0s and 1s, respectively.

3. PRIVACY AMPLIFICATION BY SHUFFLING BASED ON MULTINOMIAL DISTRIBUTION

3.1. The Exact RDP Bound for the Shuffle Model

In order to provide our tighter exact closed-form bound, we employ a combination of maximum likelihood and hypothesis testing methods, leveraging their strengths to derive a more accurate and precise result.

Theorem 1. *Let $P = (A, C - A + \Delta)$ and $Q = (A + \Delta, C - A)$, where $p = e^{-\epsilon_0}, C \sim Bin(n-1, p), A \sim Bin(C, 1/2), \Delta \sim Bern(p)$. Then*

$$D^\lambda(P\|Q) = \frac{1}{\lambda - 1} \log \int_0^1 |h'(x)|^{1-\lambda} dx.$$

Here, $\alpha(t), \beta(t)$ and $h(\alpha)$ can be obtained as follows.

$$\alpha(t) = \sum_{v=1}^{n-1} \mathbb{P}(A > \frac{tv + t}{tv - 1})\mathbb{P}(C = v), \tag{5}$$

$$g(\alpha) = \inf_t \{t : \alpha(t) \leq \alpha\}, \tag{6}$$

$$\beta(t) = h(\alpha) = 1 - \alpha - \sum_{v=1}^{n-1} \mathbb{P}(A = \lceil \frac{g(\alpha)v - 1}{g(\alpha) + 1} \rceil)\mathbb{P}(C = v). \tag{7}$$

Theorem 1 gives exact RDP bound for the shuffled output, however, since RDP is symmetric, we also need to provide result of $D^\lambda(Q\|P)$. Without causing any ambiguity, we use notations P and Q to represent the same value in Theorem 1 in the following context. In fact, we can directly obtain the Rényi divergence between two multinomial distributions by numerical computation [8].

Corollary 1. *The $(\epsilon_0, 0)$-LDP shuffled adaptive process satisfies $(\lambda, \max\{D^\lambda(P\|Q), D^\lambda(Q\|P)\}$-RDP, where P, Q are defined in Theorem 1.*

3.2. The Asymptotic RDP Bound for the Shuffle Model

Although Corollary 1 provides a RDP bound for pure differential privacy, it is excessively complicated and lacks intuitive understanding. In the following context, we provide an asymptotic RDP bound. This not only simplifies computation but also facilitates the extension to other divergence-based privacy definitions.

Lemma 2. *Assume $\xi = (n_0, n_1, n_2)'$ is a random variable which obeys multinomial distribution with parameters $(n - 1; \frac{p}{2}, \frac{p}{2}, 1 - p)$, then ξ approximately follows the multivariate normal distribution $N(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$ as $n \to \infty$, where $\tilde{\boldsymbol{\mu}} = \left(\frac{(n-1)p}{2}, \frac{(n-1)p}{2}, (n-1)(1-p)\right)'$ and covariance matrix of ξ is*

$$\tilde{\boldsymbol{\Sigma}} = (n-1)\begin{pmatrix} \frac{p}{2}(1 - \frac{p}{2}) & -\frac{p^2}{4} & -\frac{p(1-p)}{2} \\ -\frac{p^2}{4} & \frac{p}{2}(1 - \frac{p}{2}) & -\frac{p(1-p)}{2} \\ -\frac{p(1-p)}{2} & -\frac{p(1-p)}{2} & p(1-p) \end{pmatrix}.$$

Lemma 2 provides the asymptotic normality of the multinomial distribution, which is closely related to Gaussian differential privacy (GDP) [13]. The Berry-Esseen type central limit theorem [14] ensures that the convergence rate is $O\left(\frac{1}{\sqrt{n}}\right)$, which is crucial for understanding the rate of convergence, and the numerical calculation confirms a convergence rate of approximately $O\left(\frac{1}{n}\right)$.

Theorem 2. *For a domain $\mathcal{D}$, the shuffled $(\epsilon_0, 0)$-LDP adaptive process approximately satisfies $\frac{2e^{\epsilon_0/2}}{\sqrt{n-1}}$-GDP.*

According to the fact that μ-GDP implies $(\lambda, \frac{1}{2}\mu^2\lambda)$-RDP [13], we have the asymptotic RDP bound of the shuffled output.

Corollary 2. *For a domain $\mathcal{D}$, the shuffled $(\epsilon_0, 0)$-LDP adaptive process approximately satisfies $(\lambda, \frac{2e^{\epsilon_0}\lambda}{n-1})$-RDP for any $\lambda > 1$.*

3.3. Comparison of RDP Bounds under the Shuffle Model

Bounds on RDP for privacy amplification via shuffling were initially introduced by Erlingsson et al. [7]. Girgis et al. [10] improved the bound to $O(\lambda e^{2\epsilon_0}/n)$ and gave a lower bound of $O(\lambda e^{\epsilon_0}/n)$ for all $\epsilon_0 \geq 0$ and all integer $\lambda \geq 2$. Subsequently, Feldman et al. [11] improved on the results for big ϵ_0 when $\lambda < \frac{n}{16\epsilon_0 e^{\epsilon_0}}$.

Girgis et al. [10] gave an upper bound and an lower bound of RDP for $\epsilon_0 \geq 0$ and any integer $\lambda \geq 2$. That is, the RDP of the shuffled $(\epsilon_0, 0)$-LDP adaptive process is upper-bounded by

$$\epsilon(\lambda) \leq \frac{1}{\lambda - 1} \log \left(e^{\lambda^2 \frac{(e^{\epsilon_0} - 1)^2}{\bar{n}}} + e^{\epsilon_0 \lambda - \frac{n-1}{8e^{\epsilon_0}}} \right), \tag{8}$$

where $\bar{n} = \lfloor \frac{n-1}{2e_0^\epsilon} \rfloor + 1$. And the RDP of the shuffled $(\epsilon_0, 0)$-LDP adaptive process is lower-bounded by

$$\epsilon(\lambda) \geq \frac{1}{\lambda - 1} \log \left(1 + \frac{\lambda(\lambda - 1)}{2} \frac{(e^{\epsilon_0} - 1)^2}{ne^{\epsilon_0}} \right). \tag{9}$$

The exponential term $e^{\epsilon_0 \lambda - \frac{n-1}{8e^{\epsilon_0}}}$ in the upper bound comes from the Chernoff bound, it goes to 0 rapidly as n increases. If we omit this term, the upper bound is nearly $\frac{2}{n-1} \frac{\lambda}{\lambda - 1} e^{\epsilon_0}(e^{\epsilon_0} - 1)^2$, which is worse than our simplified bound in Corollary 2 by a multiplicative factor of $\frac{\lambda}{\lambda - 1}(e^{\epsilon_0} - 1)^2$. Although the RDP bound has been improved to $O(\frac{64e^{\epsilon_0}\lambda}{n})$ [11], the RDP bound we provide is significantly better.

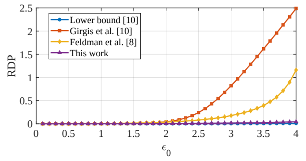

Fig. 2: RDP as a function of ϵ_0 for $\lambda = 4$ and $n = 10^4$

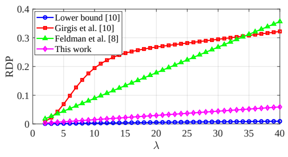

Fig. 3: RDP as a function of λ for $\epsilon_0 = 2$ and $n = 10^4$

Figure 5 illustrates that our RDP bound provides a significantly tighter bound compared to the ones presented in [**?**] for a fixed value of λ. Moreover, the figure demonstrates that for both $\epsilon_0 > 1$ and $\epsilon_0 < 1$, our bound and the lower bound are in close proximity. Similarly, Figure 6 shows that our RDP bound and the lower bound are nearly identical for each $\lambda \geq 1$.

4. NUMERICAL RESULTS

Stochastic gradient descent (SGD) is a crucial algorithm for empirical risk minimization (ERM), which aims to minimize a parameterized function given by $\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell(\boldsymbol{\theta}, x_i)$, where $\boldsymbol{\theta} \in \mathbb{R}^d$. Several studies have focused on a differentially private variant of Stochastic Gradient Descent (SGD) [15, 16]. Additionally, researchers have investigated a deep learning version of SGD tailored for the popular MNIST handwritten digit dataset [4]. Recently, many literature have focused on the performance of SGD under the shuffle model [17, 8].

As a result, our approach achieves an accuracy of 97.07% after approximately 50 epochs. This accuracy result is consistent with the findings of a vanilla neural network trained on the same MNIST dataset [18]. By employing this methodology, we can effectively train a simple classifier that achieves high accuracy in recognizing handwritten digits from the MNIST dataset.

Table 2: Experiment setting for the shuffled SGD on the MNIST dataset

Parameters/Setting	Value	Explanation
Activation function	ReLU	
Output layer	Softmax	
Loss function	Cross-entropy	
Input layer	60 variables	60 PCA components
C	10	Clipping bound
ϵ_0	$[0.1, 2]$	Privacy budget
η	0.05	Step size
m	300	Batch size
n	$60,000$	Sample size
T	50	Epoch count

To the best of our knowledge, the best RDP bound for the shuffled noisy SGD with $(\epsilon_0, 0)$-LDP adaptive process is listed in [11], while Figure 2 and 3 show that the privacy bound in this work is tighter. Furthermore, our technique is based on Laplace mechanism and can be applied to stochastic gradient descent with batch size m.

Proposition 2. *The k-fold composition of μ_i-GDP mechanisms is $\sqrt{\mu_1^2 + \cdots + \mu_k^2}$.*

Theorem 3. *Algorithm 1 approximately satisfies $(\lambda, \frac{2Te^{\epsilon_0}\lambda}{m-1})$-RDP.*

Proof. In each epoch, the algorithm is consisted of two main steps splitting and shuffling, let

$$\begin{aligned}&\mathcal{R}^{(i)}(z_{1:i-1}, D_{\pi(i)})\\ =&\tilde{\boldsymbol{\theta}}_i = \tilde{\boldsymbol{\theta}}_{i-1}(z_{1:i-1}) - \eta_i(\nabla \ell(\tilde{\boldsymbol{\theta}}_{i-1}(z_{1:i-1}), D_{\pi(i)}) + \boldsymbol{b}_i),\end{aligned}$$

then the output of Algorithm 1 can be seen as post processing of the shuffled m blocks. Since l_1 sensitivity of each $\mathcal{R}^{(i)}(z_{1:i-1}, \cdot)$ is $\frac{2C}{m}$, then it is $(\epsilon_0, 0)$-LDP according to Lemma 1. Combined with Theorem 2 and Proposition 2, Algorithm 1 approximately satisfies $\frac{2\sqrt{T}}{\sqrt{m-1}} e^{\frac{\epsilon_0}{2}}$-GDP. Since μ-GDP implies $(\lambda, \frac{1}{2}\mu^2\lambda)$-RDP, the proof is completed. □

Algorithm 1 Shuffled noisy SGD for $(\epsilon_0, 0)$-LDP

```
Require: $X = (x_1, \dots, x_n)$, $\mathcal{L}(\boldsymbol{\theta}, x)$, $\boldsymbol{\theta}_0$, $\eta$, $T$, $\epsilon_0$, $d$, $m$, $C$
Ensure: $\hat{\boldsymbol{\theta}}_{T,m}$
Split $[n]$ into $m$ disjoint subsets $S_1, \cdots, S_m$ with equal size $n/m$
Choose arbitrary initial point $\hat{\boldsymbol{\theta}}_{0,m}$
for each $t \in [T]$ do
    $\tilde{\boldsymbol{\theta}}_0 = \hat{\boldsymbol{\theta}}_{t-1,m}$
    Choose a random permutation $\pi$ of $[m]$
    for each $i \in [m]$ do
        $\boldsymbol{b}_i \sim Lap(0, \frac{2C}{\epsilon m}\boldsymbol{I}_d)$
        for each $j \in S_{\pi(i)}$ do
            $\boldsymbol{g}_i^j = \nabla \ell(\tilde{\boldsymbol{\theta}}_{i-1}, x_j)$
            $\tilde{\boldsymbol{g}}_i^j = \boldsymbol{g}_i^j / \max(1, \|\boldsymbol{g}_i^j\|_1 / C)$
        end for
        $\tilde{\boldsymbol{\theta}}_i = \tilde{\boldsymbol{\theta}}_{i-1} - \eta(\frac{1}{m}\sum_j \tilde{\boldsymbol{g}}_i^j + \boldsymbol{b}_i)$
    end for
    $\hat{\boldsymbol{\theta}}_{t,m} = \tilde{\boldsymbol{\theta}}_m$
end for
return $\hat{\boldsymbol{\theta}}_{T,m}$
```

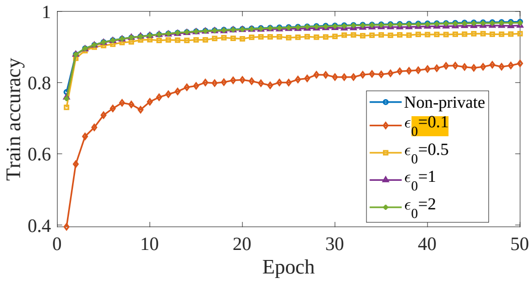

Fig. 4: Comparison of train accuracy with different ϵ_0

5. CONCLUSION

We have achieved a nearly optimal bound of $O(\frac{2e^{\epsilon_0}\lambda}{n})$ with only a negligible difference from the lower bound. This result contributes to a better understanding of privacy amplification effects in the shuffle model. The numerical results indicate that the fitting accuracy approaches the true accuracy when $\epsilon_0 > 1$. In the future, we hope to consider the privacy amplification effects of the shuffle model under more general metrics.

6. REFERENCES

[1] Ilya Mironov, "Rényi differential privacy," in *2017 IEEE 30th computer security foundations symposium (CSF)*. IEEE, 2017, pp. 263–275.

[2] Cynthia Dwork, Aaron Roth, et al., "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[3] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith, "Calibrating noise to sensitivity in private data analysis," *Journal of Privacy and Confidentiality*, vol. 7, no. 3, pp. 17–51, 2016.

[4] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.

[5] Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim, "The privacy blanket of the shuffle model," in *Advances in Cryptology–CRYPTO 2019: 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18–22, 2019, Proceedings, Part II 39*. Springer, 2019, pp. 638–667.

[6] Albert Cheu, Adam Smith, Jonathan Ullman, David Zeber, and Maxim Zhilyaev, "Distributed differential privacy via shuffling," in *Advances in Cryptology–EUROCRYPT 2019: 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19–23, 2019, Proceedings, Part I 38*. Springer, 2019, pp. 375–403.

[7] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta, "Amplification by shuffling: From local to central differential privacy via anonymity," in *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2019, pp. 2468–2479.

[8] Vitaly Feldman, Audra McMillan, and Kunal Talwar, "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 (FOCS)*. IEEE, 2022, pp. 954–964.

[9] Xingxing Xiong, Shubo Liu, Dan Li, Zhaohui Cai, and Xiaoguang Niu, "A comprehensive survey on local differential privacy," *Security and Communication Networks*, vol. 2020, pp. 1–29, 2020.

[10] Antonious M Girgis, Deepesh Data, Suhas Diggavi, Ananda Theertha Suresh, and Peter Kairouz, "On the renyi differential privacy of the shuffle model," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, 2021, pp. 2321–2341.

[11] Vitaly Feldman, Audra McMillan, and Kunal Talwar, "Stronger privacy amplification by shuffling for rényi and approximate differential privacy," in *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 2023, pp. 4966–4981.

[12] Shiva Prasad Kasiviswanathan, Homin K Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith, "What can we learn privately?," *SIAM Journal on Computing*, vol. 40, no. 3, pp. 793–826, 2011.

[13] Jinshuo Dong, Aaron Roth, and Weijie J Su, "Gaussian differential privacy," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 84, no. 1, pp. 3–37, 2022.

[14] Iosif Pinelis and Raymond Molzon, "Optimal-order bounds on the rate of convergence to normality in the multivariate delta method," *Electronic Journal of Statistics*, vol. 10, no. 1, pp. 1001, 2016.

[15] Shuang Song, Kamalika Chaudhuri, and Anand D Sarwate, "Stochastic gradient descent with differentially private updates," in *2013 IEEE global conference on signal and information processing*. IEEE, 2013, pp. 245–248.

[16] Raef Bassily, Adam Smith, and Abhradeep Thakurta, "Private empirical risk minimization: Efficient algorithms and tight error bounds," in *2014 IEEE 55th annual symposium on foundations of computer science*. IEEE, 2014, pp. 464–473.

[17] Antonious Girgis, Deepesh Data, Suhas Diggavi, Peter Kairouz, and Ananda Theertha Suresh, "Shuffled model of differential privacy in federated learning," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2021, pp. 2521–2529.

[18] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

7. APPENDIX

First, we introduce a powerful tool called f-DP [13] based on hypothesis testing.

Let U and V denote the probability distributions of $M(\mathbf{X})$ and $M(\mathbf{X}')$, respectively. We consider a rejection rule $0 \leq \phi \leq 1$, with type I and type II error rates defined as

$$\alpha_\phi = \mathbb{E}_U[\phi], \quad \beta_\phi = 1 - \mathbb{E}_V[\phi]. \tag{10}$$

Definition 6. *(Trade-off function) For any two probability distributions U and V on the same space Ω, the trade-off function $T(U,V) : [0,1] \to [0,1]$ is defined as*

$$T(U,V)(\alpha) = \inf\{\beta_\phi : \alpha_\phi \leq \alpha\},$$

where the infimum is taken over all measurable rejection rules ϕ, and $\alpha_\phi = \mathbb{E}_U(\phi)$ and $\beta_\phi = 1 - \mathbb{E}_V(\phi)$.

Definition 7. *(f-differential privacy, f-DP) Let f be a trade-off function, a mechanism M is said to be f-differentially private if*

$$T(M(X_0), M(X_1))(\alpha) \geq f(\alpha), \tag{11}$$

for all neighboring data sets X_0 and X_1 and $0 \leq \alpha \leq 1$. Under the assumption of no ambiguity, we assume that the trade-off function is a function of α and denote the abbreviated form of equation (11) as $T(M(X_0), M(X_1)) \geq f$.

Definition 8. *(Tensor product) The tensor product of two trade-off functions $f = T(U,V)$ and $g = T(U',V')$ is defined as*

$$f \otimes g := T(U \otimes V, U' \otimes V').$$

Before proving the theorem, we need to list some useful facts about f-DP. Detailed proofs can be found in [**?**] .

Fact 1. *(ϵ_0, δ_0)-DP is equivalent to f_{ϵ_0,δ_0}-DP, where*

$$f_{\epsilon_0,\delta_0}(\alpha) = \max\{0, 1 - \delta_0 - e^{\epsilon_0}\alpha, e^{-\epsilon_0}(1 - \delta_0 - \alpha)\}.$$

Fact 2. *Tensor product of two trade-off functions has commutative and associative properties, and $f_{\epsilon_1,\delta_1} \otimes \cdots \otimes f_{\epsilon_n,\delta_n} = (f_{\epsilon_0,0}, \otimes, \cdots, \otimes f_{\epsilon_n,0}) \otimes (f_{0,\delta_1} \otimes \cdots \otimes f_{0,\delta_n})$, especially, $f_{\epsilon,\delta} = f_{\epsilon,0} \otimes f_{0,\delta}$.*

Fact 3.

$$f \otimes f_{0,\delta} = \begin{cases} (1-\delta) f(\frac{\alpha}{1-\delta}), & 0 \leq \alpha \leq 1 - \delta, \\ 0, & 1 - \delta \leq \alpha \leq 1. \end{cases} \tag{12}$$

Especially, $f_{0,\delta_1} \otimes f_{0,\delta_2} = f_{0,1-(1-\delta_1)(1-\delta_2)}$ and $f_{0,\delta_0}^{\otimes n} = f_{1-(1-\delta_0)^n}$, where $h^{\otimes k}$ denotes the k-fold iterative composition of a function h.

Fact 4. *A f-DP mechanism is called μ-GDP if f can be obtained by $f = T(N(0,1), N(\mu,1)) = \Phi(\Phi^{-1}(1-\alpha) - \mu)$. In other words, A mechanism is μ-GDP if and only if it is $(\epsilon, \delta(\epsilon))$-DP for all $\epsilon \geq 0$, where*

$$\delta(\epsilon) = \Phi(-\frac{\epsilon}{\mu} + \frac{\mu}{2}) - e^\epsilon \Phi(-\frac{\epsilon}{\mu} - \frac{\mu}{2}),$$

where $\Phi(\cdot)$ is cumulative distribution function of standard normal distribution $N(0,1)$.

Fact 5. *The k-fold composition of μ_i-GDP mechanisms is $\sqrt{\mu_1^2 + \cdots + \mu_k^2}$.*