ELSA Document Intelligence Challenge: Track 2 - Federated Learning + Privacy-preserving

1 Private Baseline

This document describes the private baseline for the ELSA document intelligence challenge. We first like to introduce the terminology in a way that should be understood both for members of the document intelligence as well as for members of the privacy-preserving community. The term that might be more familiar to members of the privacy-preserving community is in brackets.

Each of the $N = 10$ clients has a dataset that contains documents from different providers (groups). Each document and each provider (group) is only part of one dataset, i.e. the documents of each provider (group) are not distributed among multiple clients. The privacy scenario in this challenge aims at reducing the risk that a malicious competitor (adversary) could identify whether a particular provider (group) is part of the training dataset. Thus, the usage of group privacy is required which is explained below.

Our private baseline FL-GROUP-DP is shown in Alg. 1. First, each client's dataset D_k is partitioned into a set $\mathbb{G}_k$ of disjoint and pre-defined groups, and each client has different groups. Each selected client runs a local instance of federated learning where each group acts as the training data of a "virtual client" within the real client. The model update $(\mathbf{w}'_t - \mathbf{w}'_{t-1})$ built by these "virtual clients" is forwarded to the server, as an update vector, for aggregation.

In particular, a subset $\mathbb{K}$ of all N clients are randomly selected at each round to update the global model, and $C = |\mathbb{K}|/N$ denotes the fraction of selected clients. At round t, suppose that client k receives a snapshot of the common model denoted by $\mathbf{w}_{t-1}$. The client first selects a random subset $\mathbb{M}$ of groups. Then, for every group $G \in \mathbb{M}$, the update $\Delta\mathbf{w}_t^G$ is computed by **AdamW** [1], which is then clipped into $\Delta\hat{\mathbf{w}}_t^G$ to have a bounded L_2-norm S. The sum of $\Delta\hat{\mathbf{w}}_t^G$ over all groups is computed and perturbed with the Gaussian mechanism introduced in Sec. 2. Finally, the virtual clients' common model $\mathbf{w}'_t$ is used to compute the model update of client k as $(\mathbf{w}'_t - \mathbf{w}_{t-1})$, and this update is sent for aggregation. The update $(\mathbf{w}'_t - \mathbf{w}_{t-1})$ is differentially private with respect to all groups in $\cup_k \mathbb{G}_k$.

2 Privacy Analysis

The privacy analysis of our differentially private baseline is discussed in this section. The provided python script to compute the privacy budget ε is derived from the following analysis.

2.1 Definitions

Definition 2.1 (Differential Privacy [2])**.** A randomized mechanism $\mathcal{M}$ with range $\mathcal{R}$ satisfies (ε, δ)-differential privacy, if for any two adjacent datasets E and E', i.e., $E' = E \cup \{x\}$ for some x in the

Algorithm 1: FL-GROUP-DP: Federated Learning with Group Privacy

Server:
  Initialize common model w_0
  for $t = 1$ **to** T_{cl} **do**
    Select set $\mathbb{K}$ of clients randomly
    for *each client* k *in* $\mathbb{K}$ **do**
      $\mathbf{u}_t^k = \mathbf{Client}_k(\mathbf{w}_{t-1})$
    end
    $\mathbf{w}_t = \mathbf{w}_{t-1} + \frac{\sum_k \mathbf{u}_t^k}{|\mathbb{K}|}$
  end
  Output: Global model $\mathbf{w}_t$

Client$_k(\mathbf{w}_{t-1})$**:**
  $\mathbb{G}_k$ is a set of predefined disjoint groups of records in D_k
  Select $\mathbb{M} \subseteq \mathbb{G}_k$ randomly
  for *each group* G *in* $\mathbb{M}$ **do**
    $\mathbf{w}' = \mathbf{w}_{t-1}$
    $\Delta \mathbf{w}_t^G = \mathbf{AdamW}(G, \mathbf{w}', T_{\mathsf{gd}}) - \mathbf{w}_{t-1}$
    $\Delta \hat{\mathbf{w}}_t^G = \Delta \mathbf{w}_t^G / \max\left(1, \frac{||\Delta \mathbf{w}_t^G||_2}{S}\right)$
  end
  $\mathbf{w}'_t = \mathbf{w}_{t-1} + \frac{\sum_G \Delta \hat{\mathbf{w}}_t^G + \mathcal{G}(0, S\mathbf{I}\sigma)}{|\mathbb{M}|}$
  Output: Model update $(\mathbf{w}'_t - \mathbf{w}_{t-1})$

data domain (or vice versa), and for any subset of outputs $O \subseteq \mathcal{R}$, it holds that

$$\Pr[\mathcal{M}(E) \in O] \leq e^{\varepsilon} \Pr[\mathcal{M}(E') \in O] + \delta \tag{1}$$

Intuitively, DP guarantees that an adversary, provided with the output of $\mathcal{M}$, can draw almost the same conclusions (up to ε with probability larger than $1 - \delta$) about any group no matter if it is included in the input of $\mathcal{M}$ or not [2]. This means, for any group owner, a privacy breach is unlikely to be due to its participation in the dataset.

In Federated Learning, the notion of *adjacent (neighboring) datasets* used in DP generally refers to pairs of datasets differing by one client (*client-level* DP), or by one group of one user (*group-level* DP), or by one data point of one user (*record-level* DP). Our challenge focuses on the *group-level* DP [3], where each group refers to a provider.

We use the Gaussian mechanism to upper bound privacy leakage when transmitting information from clients to the server.

Definition 2.2. (Gaussian Mechanism [2]) Let $f : \mathbb{R}^n \to \mathbb{R}^d$ be an arbitrary function that maps n-dimensional input to d logits with sensitivity being:

$$S = \max_{E, E'} \|f(E) - f(E')\|_2 \tag{2}$$

over all adjacent datasets E and $E' \in \mathcal{E}$. The Gaussian Mechanism $\mathcal{M}_\sigma$, parameterized by σ, adds noise into the output, i.e.,

$$\mathcal{M}_\sigma(x) = f(x) + \mathcal{N}(0, \sigma^2 I). \tag{3}$$

$\mathcal{M}_\sigma$ is (ε, δ)-DP for $\sigma \geq \sqrt{2 \ln (1.25/\delta)} S/\varepsilon$.

As in [4, 5], we consider the Sampled Gaussian Mechanism (SGM) —a composition of subsampling and the additive Gaussian noise (defined in 2.5)— for privacy amplification. Moreover, we first compute the SGM's Renyi Differential Privacy as in [5] and then we use conversion Theorem 2.8 from [6] for switching back to Differential Privacy.

Definition 2.3 (Rényi divergence)**.** Let P and Q two distributions on $\mathcal{X}$ defined over the same probability space, and let p and q be their respective densities. The Rényi divergence of a finite order $\alpha \neq 1$ between P and Q is defined as follows:

$$D_\alpha \left(P \parallel Q\right) \triangleq \frac{1}{\alpha - 1} \ln \int_{\mathcal{X}} q(x) \left(\frac{p(x)}{q(x)}\right)^\alpha dx \,.$$

Rényi divergence at orders $\alpha = 1, \infty$ are defined by continuity.

Definition 2.4 (Rényi differential privacy (RDP))**.** A randomized mechanism $\mathcal{M} : \mathcal{E} \to \mathcal{R}$ satisfies (α, ρ)-Rényi differential privacy (RDP) if for any two adjacent inputs E, $E' \in \mathcal{E}$ it holds that

$$D_\alpha \left(\mathcal{M}(E) \parallel \mathcal{M}(E')\right) \leq \rho$$

In this work, we call two datasets E, E' to be adjacent if $E' = E \cup \{x\}$ (or vice versa).

Definition 2.5 (Sampled Gaussian Mechanism (SGM))**.** Let f be an arbitrary function mapping subsets of $\mathcal{E}$ to $\mathbb{R}^d$. We define the Sampled Gaussian mechanism (SGM) parametrized with the sampling rate $0 < q \leq 1$ and the noise $\sigma > 0$ as

$$\mathrm{SG}_{q,\sigma} \triangleq f\left(\{x : x \in E \text{ is sampled with probability } q\}\right) + \mathcal{N}(0, \sigma^2 \mathbb{I}^d),$$

where each element of E is independently and randomly sampled with probability q without replacement.

As for the Gaussian Mechanism, the sampled Gaussian mechanism consists of adding i.i.d Gaussian noise with zero mean and variance σ^2 to each coordinate value of the true output of f. In fact, the sampled Gaussian mechanism draws vector values from a multivariate spherical (or isotropic) Gaussian distribution which is described by random variable $\mathcal{N}(0, \sigma^2 \mathbb{I}^d)$, where d is omitted if it is unambiguous in the given context.

2.2 Analysis

The privacy guarantee of FL-GROUP-DP is quantified using the revisited moment accountant [5] that restates the moments accountant introduced in [4] using the notion of Rényi differential privacy (RDP) defined in [7].

Let μ_0 denote the pdf of $\mathcal{N}(0, \sigma^2)$ and let μ_1 denote the pdf of $\mathcal{N}(1, \sigma^2)$. Let μ be the mixture of two Gaussians $\mu = (1 - q)\mu_0 + a\mu_1$, where q is the sampling probability of a single record in a single round.

Theorem 2.6. *[5]. Let $\mathrm{SG}_{q,\sigma}$ be the Sampled Gaussian mechanism for some function f and under the assumption $\Delta_2 f \leq 1$ for any adjacent $E, E' \in \mathcal{E}$. Then $\mathrm{SG}_{q,\sigma}$ satisfies (α, ρ)-RDP if*

$$\rho \leq \frac{1}{\alpha - 1} \log \max(A_\alpha, B_\alpha) \tag{4}$$

where $A_\alpha \triangleq \mathbb{E}_{z \sim \mu_0}[(\mu(z)/\mu_0(z))^\alpha]$ and $B_\alpha \triangleq \mathbb{E}_{z \sim \mu}[(\mu_0(z)/\mu(z))^\alpha]$

Theorem 2.6 states that applying SGM to a function of sensitivity (Equation 2.2) at most 1 (which also holds for larger values without loss of generality) satisfies (α, ρ)-RDP if $\rho \leq \frac{1}{\alpha-1}\log(\max\{A_\alpha, B_\alpha\})$. Thus, analyzing RDP properties of SGM is equivalent to upper bounding A_α and B_α.

From Corollary 7. in [5], $A_\alpha \geq B_\alpha$ for any $\alpha \geq 1$. Therefore, we can reformulate 4 as

$$\rho \leq \xi_{\mathcal{N}}(\alpha|q) \coloneqq \frac{1}{\alpha-1}\log A_\alpha \tag{5}$$

To compute A_α, we use the numerically stable computation approach proposed in [5] (Sec. 3.3) depending on whether α is expressed as an integer or a real value.

Theorem 2.7 (Composability [7])**.** *Suppose that a mechanism $\mathcal{M}$ consists of a sequence of adaptive mechanisms $\mathcal{M}_1, \ldots, \mathcal{M}_k$ where $\mathcal{M}_i : \prod_{j=1}^{i-1} \mathcal{R}_j \times \mathcal{E} \to \mathcal{R}_i$. If all the mechanisms in the sequence are (α, ρ)-RDP, then the composition of the sequence is $(\alpha, k\rho)$-RDP.*

In particular, Theorem 2.7 holds when the mechanisms themselves are chosen based on the (public) output of the previous mechanisms. By Theorem 2.7, it suffices to compute $\xi_{\mathcal{N}}(\alpha|q)$ at each step and sum them up to bound the overall RDP privacy budget of an iterative mechanism composed of single DP mechanisms at each step.

Theorem 2.8 (Conversion from RDP to DP [6])**.** *If a mechanism $\mathcal{M}$ is (α, ρ)-RDP then it is $((\rho + \log((\alpha-1)/\alpha) - (\log\delta + \log\alpha)/(\alpha-1), \delta)$-DP for any $0 < \delta < 1$.*

Theorem 2.9 (Privacy of FL-GROUP-DP)**.** *For any $0 < \delta < 1$ and $\alpha \geq 1$, FL-GROUP-DP is $(\min_\alpha(T_{\mathsf{cl}} \cdot \xi(\alpha|q) + \log((\alpha-1)/\alpha) - (\log\delta + \log\alpha)/(\alpha-1)), \delta)$-DP, where $\xi_{\mathcal{N}}(\alpha|q)$ is defined in Eq. 5, $q = \frac{C \cdot |\mathbb{M}|}{\min_k |\mathbb{G}_k|}$.*

The proof follows from Theorems 2.6, 2.7,2.8 and the fact that a group (provider) is sampled in every federated round if (1) the corresponding client is sampled, which has a probability of C, and (2) the batch of groups sampled locally at this client contains the group, which has a probability of at most $\frac{|\mathbb{M}|}{\min_k |\mathbb{G}_k|}$. Therefore, a group is sampled with a probability of $q = \frac{C \cdot |\mathbb{M}|}{\min_k |\mathbb{G}_k|}$.

3 Overall ε calculation using the provided script.

As mentioned, the Python script is derived from the previous analysis. The sampled Gaussian mechanism has a variance $\sigma^2 = c^2 \cdot S^2$, where c is the noise multiplier and S is the sensitivity.

To calculate our overall ε budget, we need to set the following hyperparameters inside the provided script: In L.281, we set the noise multiplier c. In L.283, we set the sampling probability q. In L.285 and L.287, we set the number of federated rounds T_{cl} and δ, respectively. Finally, we compute the overall ε budget in L.289.

References

[1] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.

[2] C. Dwork, A. 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] F. Galli, S. Biswas, K. Jung, C. Palamidessi, and T. Cucinotta, "Group privacy for personalized federated learning," *arXiv preprint arXiv:2206.03396*, 2022.

[4] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pp. 308–318, 2016.

[5] I. Mironov, K. Talwar, and L. Zhang, "R\'enyi differential privacy of the sampled gaussian mechanism," *arXiv preprint arXiv:1908.10530*, 2019.

[6] B. Balle, G. Barthe, M. Gaboardi, J. Hsu, and T. Sato, "Hypothesis testing interpretations and renyi differential privacy," in *International Conference on Artificial Intelligence and Statistics*, pp. 2496–2506, PMLR, 2020.

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