Federated Learning With Differential Privacy: Algorithms and Performance Analysis

IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 15, 2020

Kang Wei, *Graduate Student Member, IEEE*, Jun Li , *Senior Member, IEEE*,
Ming Ding , *Senior Member, IEEE*, Chuan Ma , Howard H. Yang , *Member, IEEE*,
Farhad Farokhi , *Senior Member, IEEE*, Shi Jin , *Senior Member, IEEE*,
Tony Q. S. Quek , *Fellow, IEEE*, and H. Vincent Poor , *Life Fellow, IEEE*

Differential privacy

- 差分攻击：通过查询汉明距离为1的两个数据集，获得某单个样本的隐私信息

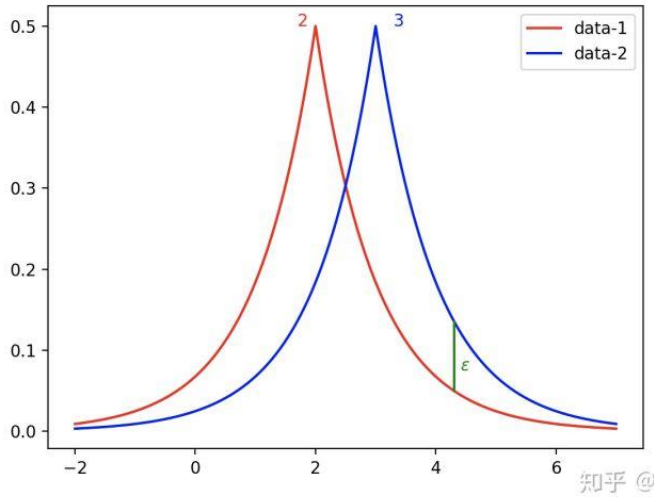
- 例子：
  - 查询某一公司某一人是否结婚。单独的查询个人婚姻情况是高度敏感的，但查询公司总共有多少人结婚是被允许的。
  - 当A登记婚姻情况前查询已婚人数返回3，A登记后再次查询返回人数为4，则可以得知A已婚。

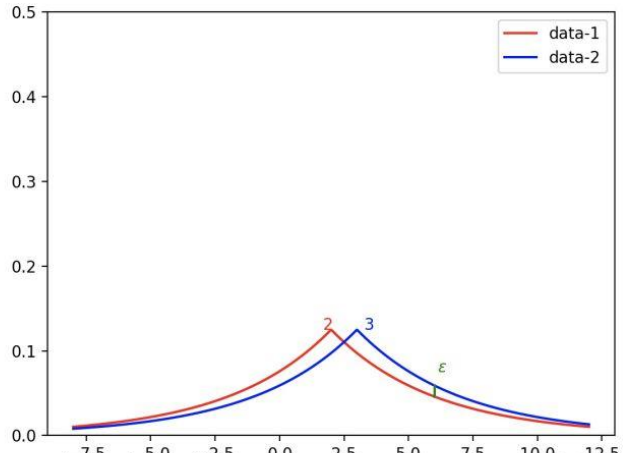
Differential privacy

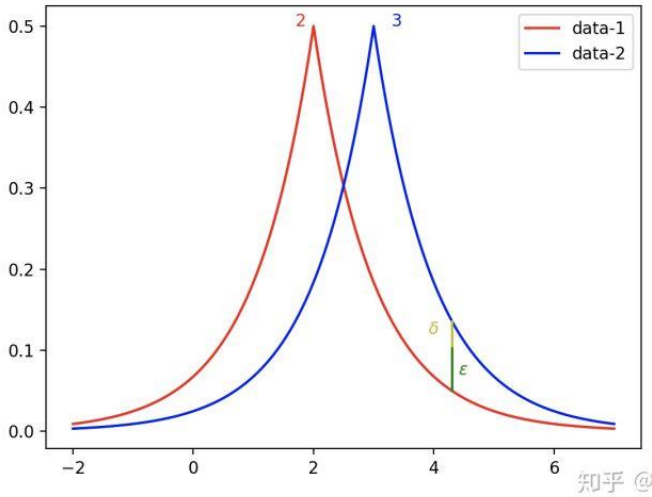
- DP: $\Pr[\mathcal{M}(x) \in S] \leq e^{\varepsilon} Pr[\mathcal{M}(x') \in S]$
- KL-Divergence: for two distributions

$$D_{\infty}(Y \| Z) = \max_{S \subset Supp(Y)} [\ln \frac{\Pr[Y \in S]}{\Pr[Z \in S]}] = \max_{y \in Y} [\ln \frac{\Pr[Y = y]}{\Pr[Z = y]}] \leq \varepsilon$$

$$\Pr[\mathcal{M}(x) \in S] \leq e^{\varepsilon} Pr[\mathcal{M}(x') \in S] + \delta$$







Algorithm

$$
\begin{aligned}
s_{\mathrm{U}}^{\mathcal{D}_i} &\triangleq \mathbf{w}_i = \arg\min_{\mathbf{w}} F_i(\mathbf{w}, \mathcal{D}_i) \\
&= \frac{1}{|\mathcal{D}_i|} \sum_{j=1}^{|\mathcal{D}_i|} \arg\min_{\mathbf{w}} F_i(\mathbf{w}, \mathcal{D}_{i,j}),
\end{aligned}
$$

$$
\begin{aligned}
\Delta s_{\mathrm{U}}^{\mathcal{D}_i} &= \max_{\mathcal{D}_i, \mathcal{D}_i'} \| s_{\mathrm{U}}^{\mathcal{D}_i} - s_{\mathrm{U}}^{\mathcal{D}_i'} \| \\
&= \max_{\mathcal{D}_i, \mathcal{D}_i'} \left\| \frac{1}{|\mathcal{D}_i|} \sum_{j=1}^{|\mathcal{D}_i|} \arg\min_{\mathbf{w}} F_i(\mathbf{w}, \mathcal{D}_{i,j}) \right. \\
&\quad \left. - \frac{1}{|\mathcal{D}_i'|} \sum_{j=1}^{|\mathcal{D}_i'|} \arg\min_{\mathbf{w}} F_i(\mathbf{w}, \mathcal{D}_{i,j}') \right\| = \frac{2C}{|\mathcal{D}_i|},
\end{aligned}
$$

Algorithm 1 Noising Before Aggregation FL

Data: T, $\mathbf{w}^{(0)}$, μ, ϵ and δ

Initialization: $t = 1$ and $\mathbf{w}_i^{(0)} = \mathbf{w}^{(0)}$, $\forall i$

while $t \leq T$ **do**

  Local training process:

  while $\mathcal{C}_i \in \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_N\}$ **do**

    Update the local parameters $\mathbf{w}_i^{(t)}$ as

    $\mathbf{w}_i^{(t)} = \arg\min_{\mathbf{w}_i} \left(F_i(\mathbf{w}_i) + \frac{\mu}{2} \|\mathbf{w}_i - \mathbf{w}^{(t-1)}\|^2 \right)$

    Clip the local parameters

    $\mathbf{w}_i^{(t)} = \mathbf{w}_i^{(t)} / \max\left(1, \frac{\|\mathbf{w}_i^{(t)}\|}{C}\right)$

    Add noise and upload parameters $\widetilde{\mathbf{w}}_i^{(t)} = \mathbf{w}_i^{(t)} + \mathbf{n}_i^{(t)}$

  Model aggregating process:

  Update the global parameters $\mathbf{w}^{(t)}$ as

  $\mathbf{w}^{(t)} = \sum_{i=1}^{N} p_i \widetilde{\mathbf{w}}_i^{(t)}$

  The server broadcasts global noised parameters

  $\widetilde{\mathbf{w}}^{(t)} = \mathbf{w}^{(t)} + \mathbf{n}_{\mathrm{D}}^{(t)}$

  Local testing process:

  while $\mathcal{C}_i \in \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_N\}$ **do**

    Test the aggregating parameters $\widetilde{\mathbf{w}}^{(t)}$ using local dataset

  $t \leftarrow t + 1$

Result: $\widetilde{\mathbf{w}}^{(T)}$

Noise terms

- User, upload link: $\sigma_{\mathrm{U}} = cL\Delta s_{\mathrm{U}}/\epsilon$
- Overall: $\sigma_{\mathrm{A}} = cT\Delta s_{\mathrm{D}}/\epsilon$
- Added to the sever:

$$\sigma_{\mathrm{D}} = \sqrt{\sigma_{\mathrm{A}}^2 - \frac{\sigma_{\mathrm{U}}^2}{N}} = \begin{cases} \dfrac{2cC\sqrt{T^2 - L^2N}}{mN\epsilon} & T > L\sqrt{N}, \\ 0 & T \leq L\sqrt{N}. \end{cases}$$

Result

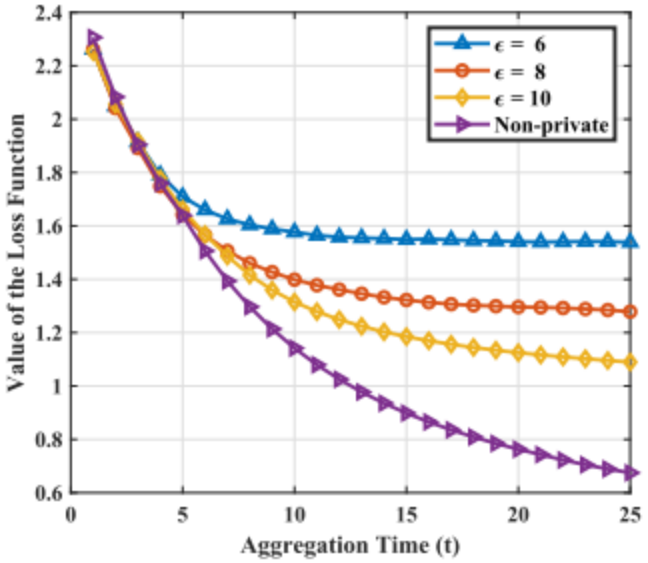


Fig. 3. The comparison of training loss with various protection level for 50 clients using $\epsilon = 6$, $\epsilon = 8$ and $\epsilon = 10$, respectively.

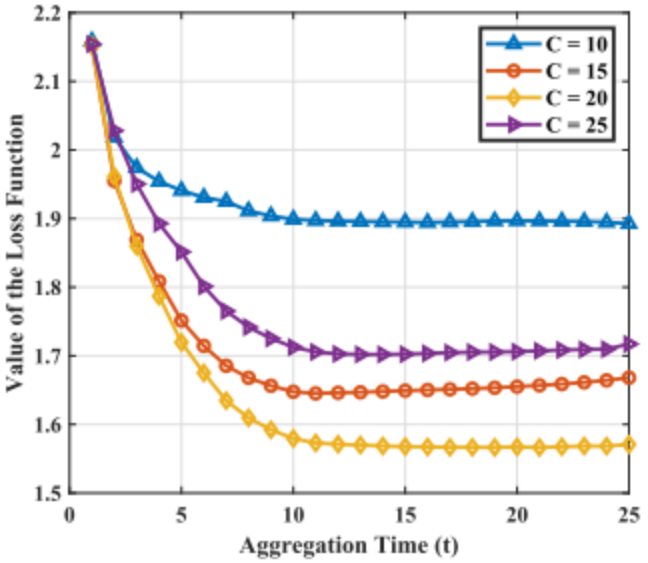


Fig. 5. The comparison of training loss with various clipping thresholds for 50 clients using $\epsilon = 60$.

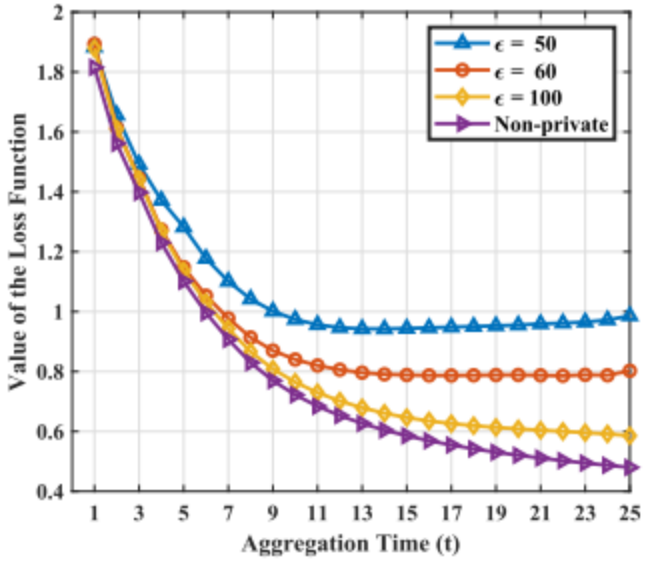


Fig. 4. The comparison of training loss with various privacy levels for 50 clients using $\epsilon = 50$, $\epsilon = 60$ and $\epsilon = 100$, respectively.

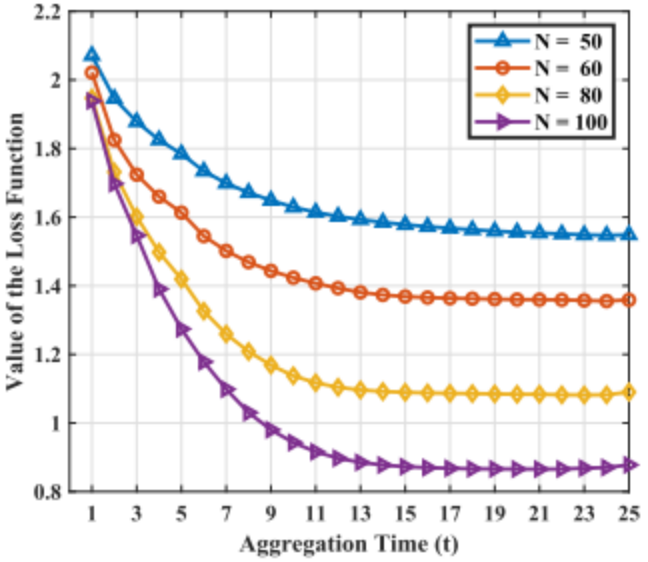


Fig. 6. The value of the loss function with various numbers of clients under $\epsilon = 60$ under NbAFL Algorithm with 50 clients.

Result

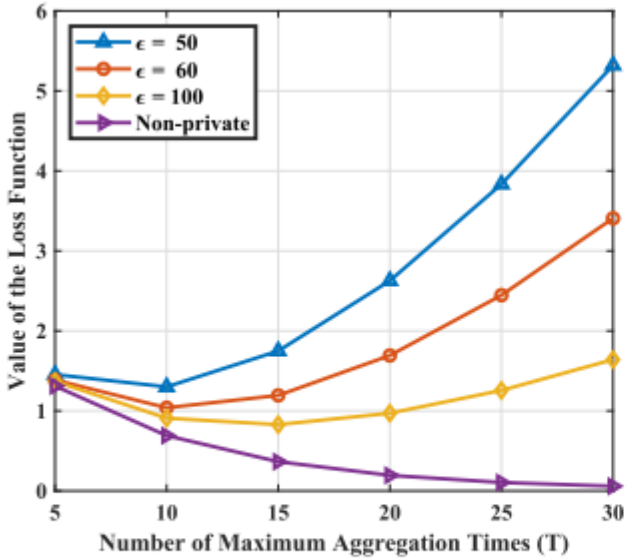


Fig. 7. The convergence upper bounds with various privacy levels $\epsilon = 50$, 60 and 100 under 50-clients' NbAFL algorithm.

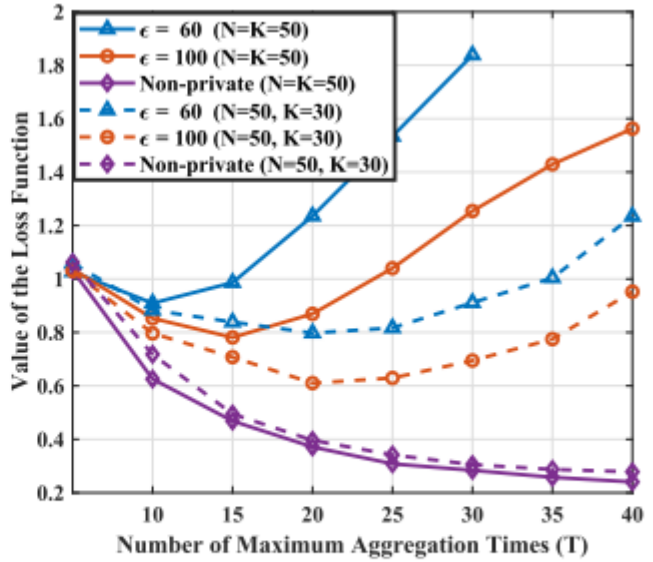


Fig. 9. The value of the loss function with various privacy levels $\epsilon = 60$ and $\epsilon = 100$ under NbAFL Algorithm with 50 clients.

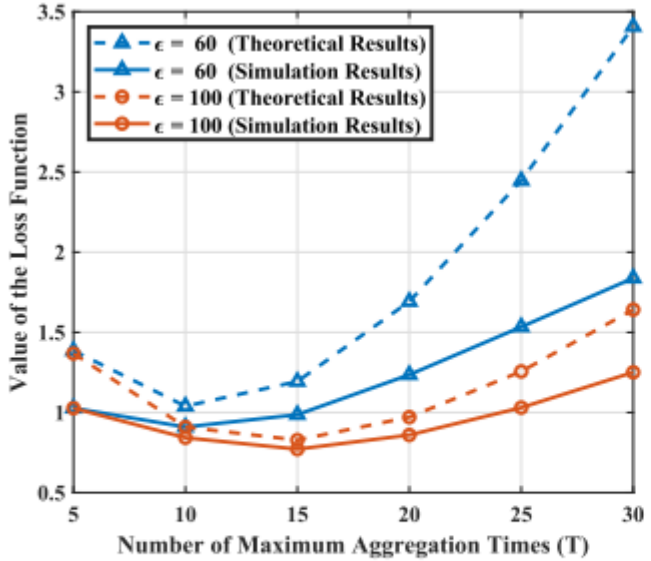


Fig. 8. The comparison of the loss function between experimental and theoretical results with the various aggregation times under NbAFL Algorithm with 50 clients.

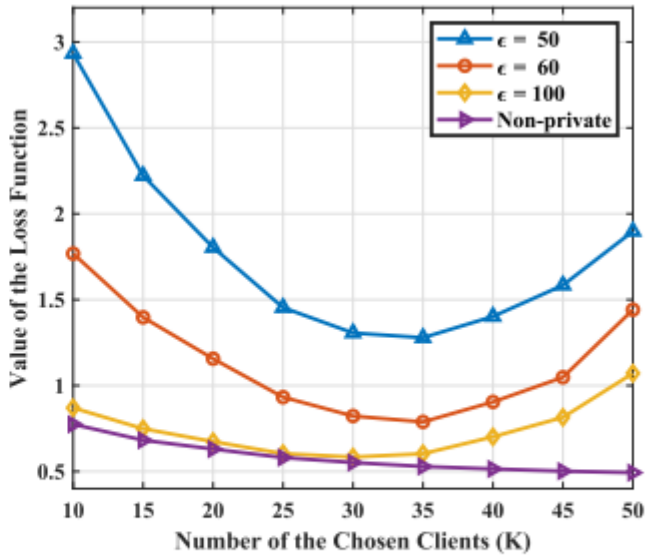


Fig. 10. The value of the loss function with various numbers of chosen clients under $\epsilon = 50$, 60, 100 under NbAFL Algorithm and non-private approach with 50 clients.