On Multi-Round Privacy in Federated Learning

Abstract

Secure Aggregation is essential in federated learning to ensure that, in any given round, the server learns nothing about the local models of the users beyond their aggregate. Secure aggregation, however, does not protect the privacy of the users over multiple training rounds due to the partial user participation at each round. To quantify such long-term privacy leakage, a new metric termed as multi-round privacy has been recently introduced that requires that the server cannot reconstruct any individual model using the aggregate models from any number of training rounds. In addition, a privacy-preserving structured user selection strategy known as Multi-RoundSecAgg has been developed to ensure multi-round privacy while taking into account the convergence rate and the fairness in the user selection. Multi-RoundSecAgg, however, provides a trade-off between the multi-round privacy guarantee and the convergence rate in the sense that stronger multi-round privacy requires a larger number of training rounds. In this paper, we consider two weaker notions of multi-round privacy, termed as weak multi-round privacy and semi-strong multi-round privacy, which still require that the server cannot get any individual model. We show that considering these weaker notions allow for better convergence rates compared to Multi-RoundSecAgg while still protecting the privacy of the individual users in a weaker sense.

Introduction

Federated Learning (FL) allows users' data to remain private while used to train an AI. This is done by sending a copy of the server's model to each user, and having the users train the model and send it back. However, sharing the local models with the server still reveals information about the local datasets as demonstrated by the model inversion attacks [3, 6, 10, 4]. Secure aggregation protocols address this challenge by ensuring that the server does not learn anything about the local models beyond their aggregate [2, 8, 5, 9, 1].

The Problem

However, secure aggregation may not protect the privacy of the users over multiple training rounds, when different sets of users participate in different rounds [7].

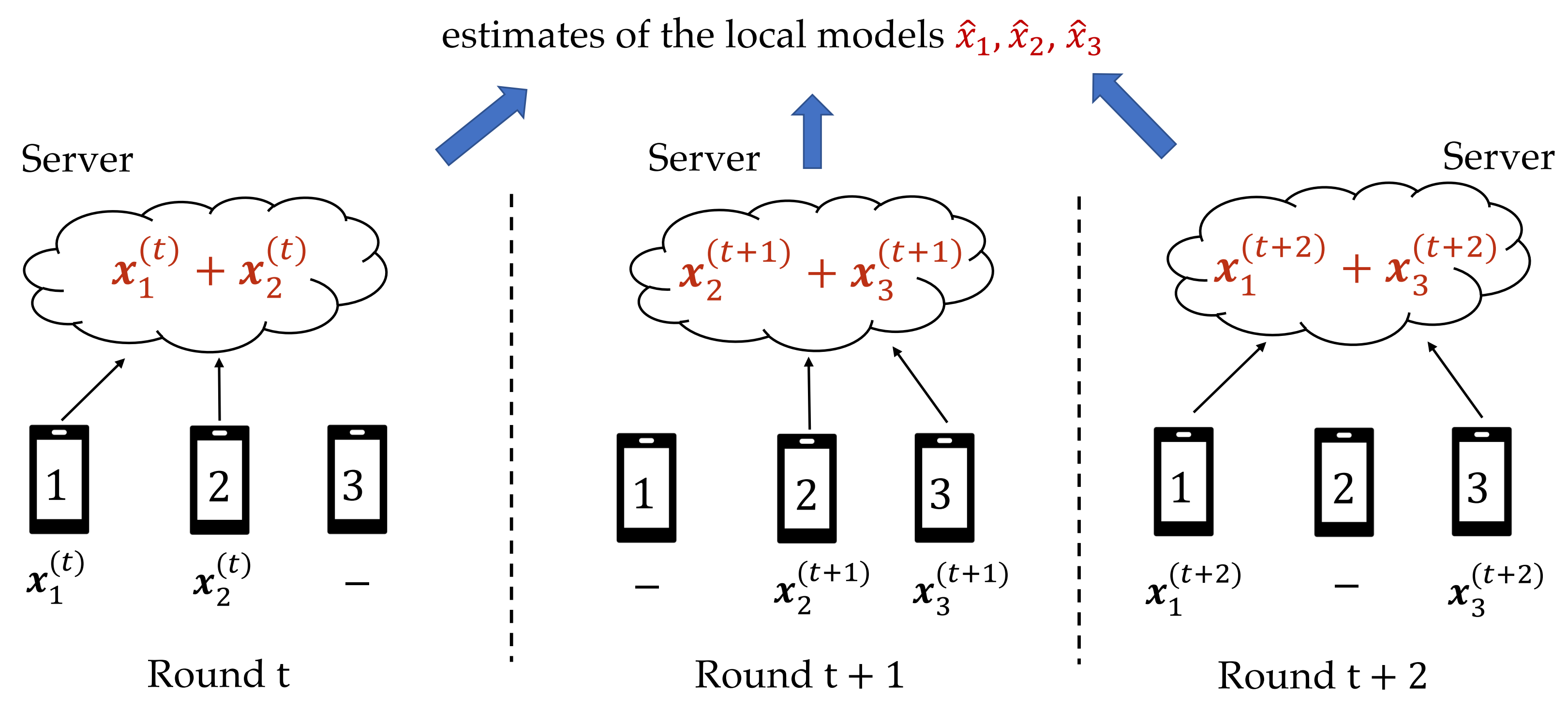

Figure 1. A multi-round secure aggregation example, where the server can reconstruct all local models. The set of participating users at round t is $\mathcal{S}^{(t)} = \{1,2\}$, at round $t+1$ is $\mathcal{S}^{(t+1)} = \{2,3\}$ and at round $t+1$ is $\mathcal{S}^{(t+1)} = \{1,3\}$. If the local models do not change significantly over these three rounds (e.g., the models start to converge), the server can reconstruct all local models from the aggregate models of the three rounds.

The Solution

To solve this, we restrict which users can participate in each round of training, according to some protocol. We can compare these protocols according to two metrics:

1. A *Multi-Round Privacy Guarantee*. A multi-round privacy guarantee T guarantees that the best the server can get is the aggregate of at least T local models.
2. The *Average Aggregation Cardinality*. This measures, on average, how many users participate in each round of training. Because we restrict how we can select which users participate, sometimes it will be impossible to find a valid subset of users, so this takes into account how likely that is to happen.

Definitions

There are N users and we select K of them to participate each round.

The *participation vector* $\mathbf{p}^{(t)} \in \{0,1\}^N$ is the characteristic vector corresponding to the t^{th} round of training whose i-th entry is 1 when user i is selected and 0 otherwise.

The *participation matrix* $mathbf P^{(t)} = \left[\mathbf{p}^{(0)}, \mathbf{p}^{(1)}, \dots, \mathbf{p}^{(t-1)}\right]^\top \in \{0,1\}^{t\times N}$.

Although there are several ways of defining multi-round privacy, we will use the following:

Definition (Semi-Strong Multi-Round Privacy). (Semi-Strong Multi-Round Privacy). The semi-strong multi-round privacy guarantee T requires that, for all $t < T$, if $({\mathbf{P}^{(J)}}^\top z)_{i1}, \dots, ({\mathbf{P}^{(J)}}^\top z)_{it}$ are the t elements of greatest magnitude, the sum of the magnitudes is, at most, $t/(T-t)$ times the sum of the magnitudes of the rest of the elements.

We also need to define a metric for comparing different ways to restrict user selection:

Definition (Average Aggregation Cardinality). The average aggregation cardinality quantifies the expected number of participating users as follows

$$C = \liminf_{J\to\infty} \frac{\mathbb{E}\left[\sum_{t=0}^{J-1} \|\mathbf{p}^{(t)}\|_0\right]}{J}, \tag{1}$$

where the expectation is over the randomness in $\mathcal{A}$ and the user availability.

Batch Partitioning

The originally proposed technique in Multi-RoundSecAgg[7] was Batch Partitioning. This strategy divides the users into N/T batches, and selects K/T batches each round, where every user in each batch is online. The number of user sets in this family is given by

$$R_{\text{BP}} = \binom{N/T}{K/T}$$

For example, if $N = 6$, $K = 4$, and $T = 2$, the batch partition matrix is given by

$$\mathbf{B} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

This Batch Partitioning technique is provably optimal given a previous, stronger Multi-Round Privacy definition.[7]

Half Partitioning

However, given the weaker Semi-Strong Multi-Round Privacy definition, we can define a half partitoning technique:

Half Partitioning. A *Half Partitioning Matrix* with $T = 2$ is a binary matrix created by stacking all rows created with the following process: start with N 0s, choose $K/2$ nonconsecutive terms (the first and last terms count as consecutive), and replace each selected term and the following term with a 1. The number of user sets is given by

$$R_{HP} = \frac{2N}{K}\binom{2N/T - K/T - 1}{K/T - 1}$$

For example, when $N = 6$, $K = 4$, and $T = 2$, the half partition matrix is given by

$$H = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

Comparison

- Batch Partitioning and Half Partitioning have a semi-strong privacy guarantee of T.
- Half Partitioning more than doubles the number of rows.
- The aggregation cardinality is also increased by Half Partitioning.

Conclusion

In this presentation, we have considered a weaker notion of multi-round privacy and developed a user selection strategy that allow for better convergence rates compared to batch partitioning while still protecting the privacy of the individual users in a slightly weaker sense. An interesting future direction is to investigating the robustness of half partitioning compared to batch partitioning against attacks such as model inversion attacks.

References

[1] James Henry Bell, Kallista A Bonawitz, Adrià Gascón, Tancrède Lepoint, and Mariana Raykova. Secure single-server aggregation with (poly) logarithmic overhead. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, pages 1253–1269, 2020.

[2] Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1175–1191, 2017.

[3] Matt Fredrikson, Somesh Jha, and Thomas Ristenpart. Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pages 1322–1333, 2015.

[4] Jonas Geiping, Hartmut Bauermeister, Hannah Dröge, and Michael Moeller. Inverting gradients–how easy is it to break privacy in federated learning? *arXiv preprint arXiv:2003.14053*, 2020.

[5] Swanand Kadhe, Nived Rajaraman, O Ozan Koyluoglu, and Kannan Ramchandran. Fastsecagg: Scalable secure aggregation for privacy-preserving federated learning. *arXiv preprint arXiv:2009.11248*, 2020.

[6] Milad Nasr, Reza Shokri, and Amir Houmansadr. Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In *2019 IEEE symposium on security and privacy (SP)*, pages 739–753. IEEE, 2019.

[7] Jinhyun So, Ramy E Ali, Basak Guler, Jiantao Jiao, and Salman Avestimehr. Securing secure aggregation: Mitigating multi-round privacy leakage in federated learning. *arXiv preprint arXiv:2106.03328*, 2021.

[8] Jinhyun So, Başak Güler, and A Salman Avestimehr. Turbo-aggregate: Breaking the quadratic aggregation barrier in secure federated learning. *IEEE Journal on Selected Areas in Information Theory*, 2(1):479–489, 2021.

[9] Yizhou Zhao and Hua Sun. Information theoretic secure aggregation with user dropouts. *arXiv preprint arXiv:2101.07750*, 2021.

[10] Ligeng Zhu and Song Han. Deep leakage from gradients. In *Federated Learning*, pages 17–31. Springer, 2020.