Random Orthogonalization for Private Wireless Federated Learning (Extended Abstract)

Sadaf ul Zuhra †, Mohamed Seif †, Karim Banawan, and H. Vincent Poor

***Abstract*—We consider the problem of private wireless federated learning through a massive MIMO multiple-access channel (MAC). In this problem, a parameter server (PS) having M antennas needs to train a global machine-learning model with the aid of K single-antenna users. Each user trains a local model to update the PS's global model without leaking information about the user's local model. By harnessing the additive nature of the MAC, the PS aggregates the local updates and updates the global model. We show that by adopting the random orthogonalization technique and careful noise injection by the users, maintaining the privacy of local models is possible under local differential privacy metrics without sacrificing the accuracy/convergence rate of the global machine-learning model. We derive the exact achievable privacy level. Our results show that the privacy level is a function of the channel gains. We substantiate our findings by carrying out a standard classification task, which achieves an accuracy of 89% in less than 15 communication rounds while maintaining an acceptable privacy level of the users' local models. Moreover, numerical results show that the privacy leakage is decreasing in the number of users K, while it is increasing in the number of antennas M.**

Background and Objective: Random orthogonalization is an uplink communication mechanism for federated learning (FL) [1] in a massive multiple-input multiple-output (MIMO) setting [2]. Random orthogonalization leverages the channel hardening and favorable propagation properties of massive MIMO to achieve notable performance gains in FL with no channel state information required at the users and a significant reduction in the channel estimation overhead at the receiver. A prominent challenge in FL settings is preserving the privacy of personal dataset, which may be leaked by observing the users' local models [3], [4]. In this work, we explore the potential of random orthogonalization for preserving a certain level of privacy in FL. This is motivated by the fact that the parameter server (PS) makes use of the superposition of the updates received from the users instead of decoding the update from individual users.

System Model: Consider a wireless FL system with K single-antenna users and a parameter server (PS) with M antennas. Users communicate with the PS through a wireless fading multiple access channel (MAC). In each iteration, users train a local model using stochastic gradient descent (SGD) and then communicate the resulting d dimensional update back to the PS over d channel uses. The training process continues until a specified number of iterations T. Denote by $x_{k,i}^{(t)} \in \mathbb{R}$ the update transmitted by user k in the i-th channel use of iteration t. The transmitted signal vector over d channel uses is given by $\mathbf{x}_k^{(t)} = [x_{k,1}^{(t)}\ x_{k,2}^{(t)}\ \cdots\ x_{k,d}^{(t)}]$, and is subject to an average power constraint P, i.e., $\mathbb{E}(\|\mathbf{x}_k^{(t)}\|_2^2) \leq P$. Consequently, for all $i \in \{1, 2, \ldots, d\}$ and $t \in \{1, 2, \ldots, T\}$, the received signal $\mathbf{y}_i^{(t)} \in \mathbb{R}^M$ at PS in channel use i of iteration t is given by,

$$\mathbf{y}_i^{(t)} = \sum_{k=1}^{K} \mathbf{h}_k x_{k,i}^{(t)} + \mathbf{m}_i^{(t)}, \tag{1}$$

where, $\mathbf{h}_k \in \mathbb{R}_+^{M\times 1}$ is the wireless channel vector between user k and the PS.

We assume a block flat-fading channel, where the channel coefficient remains constant within the duration of a communication block. $\mathbf{m}_i^{(t)} \in \mathbb{R}^{M\times 1}$ in (1) is the receiver noise, whose elements are independent and identically distributed (i.i.d.) with zero-mean and variance σ_m^2. We assume that the channel state information is unavailable to the users. The PS carries out channel estimation before the start of the training process to estimate the channel vectors sum $\mathbf{h}_s = \sum_{k=1}^{K} \mathbf{h}_k$. Note that, under the random orthogonalization scheme, the PS does not need to know the individual channels of all the users. This, in turn, results in significant reductions in channel estimation overheads, especially for large values of K and M.

Federated Learning Model: Each user k has a local private dataset $\mathcal{D}_k$ with D_k data points, denoted as $\mathcal{D}_k = \{(\mathbf{u}_j^{(k)}, v_j^{(k)})\}_{j=1}^{D_k}$, where $\mathbf{u}_j^{(k)}$ is the j-th data point and $v_j^{(k)}$ is the corresponding label at user k. For some training loss function $f(\cdot)$, the local loss function at user k is given by,

$$f_k(\mathbf{w}) = \frac{1}{D_k} \sum_{j=1}^{D_k} f(\mathbf{w}; (\mathbf{u}_j^{(k)}, v_j^{(k)})),$$

where $\mathbf{w} \in \mathbb{R}^d$ is the parameter vector to be optimized. The objective of the FL system is to obtain the optimal global model $\mathbf{w}^*$ at the PS by minimizing the global loss function

$$F(\mathbf{w}) \triangleq \frac{1}{\sum_{k=1}^{K} D_k} \sum_{k=1}^{K} D_k f_k(\mathbf{w}).$$

The minimization of $F(\mathbf{w})$ is carried out iteratively through a distributed SGD algorithm. More specifically, in the t-th training iteration, the PS broadcasts the global parameter vector $\mathbf{w}^{(t)}$ to all users. User k then computes its local gradient $\mathbf{g}_k^{(t)}$ using stochastic mini batch $\mathcal{B}_k^{(t)} \subseteq \mathcal{D}_k$, with size n_k, i.e., for all $t \in \{1, 2, \ldots, T\}$, $|\mathcal{B}_k^{(t)}| = n_k$. Therefore, for all $k \in \{1, 2, \ldots, K\}$ and all $t \in \{1, 2, \ldots, T\}$, the local gradient $\mathbf{g}_k^{(t)}$ is given by,

$$\mathbf{g}_k^{(t)} = \frac{1}{n_k} \sum_{j\in\mathcal{B}_k^{(t)}} \nabla f_k(\mathbf{w}^{(t)}; (\mathbf{u}_j^{(k)}, v_j^{(k)})),$$

Subsequently, the local parameter at user k, $\mathbf{w}_k^{(t)} = [w_{k,1}^{(t)}, w_{k,2}^{(t)}, \ldots, w_{k,d}^{(t)}]$ is updated according to the update rule

Sadaf ul Zuhra, Mohamed Seif, and H. Vincent Poor are with the Department of Electrical and Computer Engineering, Princeton University, 08540 Princeton, NJ, USA. ({sadaf.zuhra,mseif,poor}@princeton.edu)

Karim Banawan is with the Electrical Engineering Department, Faculty of Engineering, Alexandria University, 21544 Alexandria Egypt. (kbanawan@alexu.edu.eg)

†These authors contributed equally to this work.

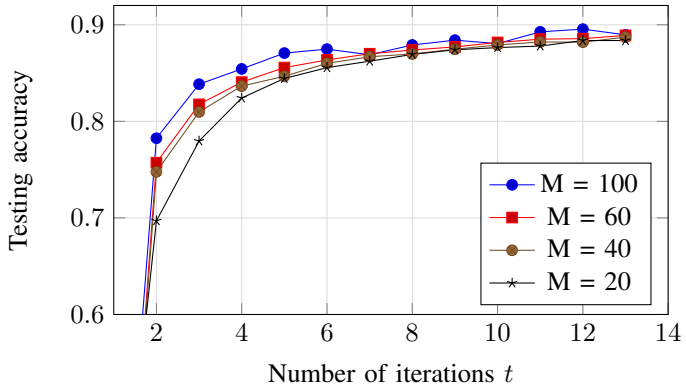

Fig. 1. Testing accuracy of the proposed mechanism as a function of the number of iterations t for different antenna settings.

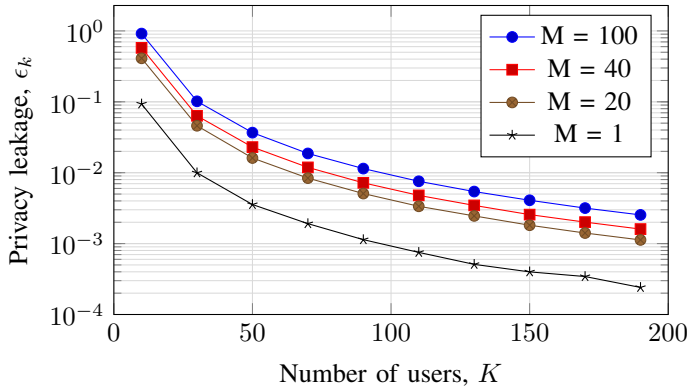

Fig. 2. Privacy leakage per user ϵ_k vs number of users K for different antenna settings, with $\sigma^2 = 0.1$, $\sigma_m^2 = 1$, $\alpha = 3$, $C = 1$ and $\delta = 10^{-5}$.

$\mathbf{w}_k^{(t+1)} = \mathbf{w}_k^{(t)} - \eta_t \mathbf{g}_k^{(t)}$, where η_t is the learning rate of the distributed learning algorithm at iteration t. The PS aggregates the local model updates from the users to obtain a new global model $\mathbf{w}^{(t+1)} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{w}_k^{(t+1)}$.

Privacy: To guarantee an acceptable level of data privacy for the FL users, the SGD algorithm needs to satisfy local differential privacy (LDP) constraints for each user. Specifically, we employ the (ϵ, δ)-LDP notion of [5] to quantify the data privacy level provided by the algorithm. To realize such privacy guarantees, we employ the Gaussian mechanism [5], where each user adds a small amount of artificial Gaussian noise to its model update before transmitting it to the PS.

Proposed Transmission Scheme: Under the proposed privacy enhancing FL scheme, the signal transmitted by the kth user during channel use $i \in \{1, 2, \ldots, d\}$ of iteration $t \in \{1, 2, \ldots, T\}$ is given by,

$$x_{k,i}^{(t)} = \alpha_k \left(w_{k,i}^{(t)} + n_{k,i}^{(t)} \right), \tag{2}$$

where $n_{k,i}^{(t)}$ is the local perturbation noise generated independently at the k-th user such that $n_{k,i}^{(t)} \sim \mathcal{N}(0, \sigma_k^2)$, and α_k is a scaling coefficient to ensure that the average power constraint is met. We further assume that the norm of the model update $\mathbf{w}_k^{(t)}$ is bounded by some constant $C \geq 0$. To ensure this, we normalize the update vector as $\mathbf{w}_k^{(t)} := \min\left(1, C/\|\mathbf{w}_k^{(t)}\|_2\right) \cdot \mathbf{w}_k^{(t)}$. In order to ensure coherent superposition for obtaining unbiased estimates of the model updates, for all $k \in \{1, 2, \ldots, K\}$, set $\alpha_k = \alpha = \sqrt{P}$. Then, the received signal at the PS in channel use i of iteration t is

$$\begin{aligned} \mathbf{y}_i^{(t)} &= \sum_{k=1}^{K} \mathbf{h}_k x_{k,i}^{(t)} + \mathbf{m}_i^{(t)} \\ &= \alpha \sum_{k=1}^{K} \mathbf{h}_k w_{k,i}^{(t)} + \alpha \sum_{k=1}^{K} \mathbf{h}_k n_{k,i}^{(t)} + \mathbf{m}_i^{(t)}, \end{aligned} \tag{3}$$

where $\mathbf{h}_k$ is the channel between the kth user and the PS. The PS performs the following post-processing on the received signal in channel use i to obtain the aggregate update $\tilde{\mathbf{w}}_i^{(t+1)}$:

$$\begin{aligned} \tilde{\mathbf{w}}_i^{(t+1)} &= \frac{1}{\alpha K} \mathbf{h}_s^H \mathbf{y}_i^{(t)} \\ &= \frac{1}{K}\sum_{k=1}^{K} (\mathbf{h}_k^H \mathbf{h}_k) w_{k,i}^{(t)} + \frac{1}{K}\sum_{k=1}^{K}\sum_{j\neq k} (\mathbf{h}_k^H \mathbf{h}_j) w_{j,i}^{(t)} \\ &+ \underbrace{\frac{1}{K}\sum_{k=1}^{K} \mathbf{h}_k^H \mathbf{h}_k n_{k,i}^{(t)} + \frac{1}{K}\sum_{k=1}^{K}\sum_{j\neq k} \mathbf{h}_k^H \mathbf{h}_j n_{j,i}^{(t)} + \frac{1}{\alpha K}\sum_{k=1}^{K} \mathbf{h}_k^H \mathbf{m}_i^{(t)}}_{z_i{}^{(t)}}, \end{aligned} \tag{4}$$

where $z_i{}^{(t)}$ is the effective noise in $\tilde{\mathbf{w}}_i^{(t)}$ with variance

$$\begin{aligned} \sigma_z^2 &= \frac{1}{K^2}\left[\sum_{k=1}^{K} \|\mathbf{h}_k\|_2^4 \sigma_k^2 + \sum_{k=1}^{K}\sum_{j\neq k} |\mathbf{h}_k^H \mathbf{h}_j|^2 \sigma_j^2\right] \\ &\quad + \frac{1}{\alpha^2 K^2}\sum_{k=1}^{K} \|\mathbf{h}_k\|_2^2 \sigma_m^2 \\ &\overset{(a)}{=} \frac{\sigma^2}{K^2}\left[\sum_{k=1}^{K} \|\mathbf{h}_k\|_2^4 + \sum_{k=1}^{K}\sum_{j\neq k} |\mathbf{h}_k^H \mathbf{h}_j|^2\right] + \frac{\sigma_m^2}{\alpha^2 K^2}\sum_{k=1}^{K} \|\mathbf{h}_k\|_2^2, \end{aligned} \tag{5}$$

where, in step (a), the amount of local perturbation across users is chosen to be the same, i.e., $\sigma_k^2 = \sigma^2, \forall k$.

The privacy attained by the proposed scheme is quantified by the following theorem:

Theorem 1. *The proposed transmission scheme achieves (ϵ_k, δ)-LDP per iteration for each user k, where*

$$\epsilon_k = \frac{\|\mathbf{h}_k\|_2^2}{K} \cdot \frac{2C}{\sigma_z}\sqrt{2\log\frac{1.25}{\delta}}. \tag{6}$$

Preliminary numerical analysis: The accuracy and convergence rate of the proposed mechanism are illustrated in Fig. 1 for a handwritten digit classification task. It is observed that the mechanism achieves an accuracy of 89% in less than 15 iterations while maintaining the privacy of the users' local models. Fig. 2 shows the variation of the privacy leakage in (6) as a function of the number of users.

REFERENCES

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of Artificial Intelligence and Statistics (AISTATS)*, 2017, pp. 1273–1282.

[2] X. Wei, C. Shen, J. Yang, and H. V. Poor, "Random orthogonalization for federated learning in massive MIMO systems," in *Proceedings of the IEEE International Conference on Communications (ICC)*, 2022, pp. 3382–3387.

[3] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. S. Quek, and H. Vincent Poor, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454–3469, 2020.

[4] K. Wei, J. Li, C. Ma, M. Ding, C. Chen, S. Jin, Z. Han, and H. V. Poor, "Low-latency federated learning over wireless channels with differential privacy," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 1, pp. 290–307, 2022.

[5] 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.