Privacy-Preserving and Accountable Multi-agent Learning

Extended Abstract

Anudit Nagar
Bennett University
Greater Noida, India
anudit@bennett.edu.in

Cuong Tran
Syracuse University
Syracuse, NY
cutran@syr.edu

Ferdinando Fioretto
Syracuse University
Syracuse, NY
ffiorett@syr.edu

ABSTRACT

Distributed multi-agent learning enables agents to cooperatively train a model without requiring to share their datasets. While this setting ensures some level of privacy, it has been shown that, even when data is not directly shared, the training process is vulnerable to privacy attacks including data reconstruction and model inversion attacks. Additionally, malicious agents that train on inverted labels or random data, may arbitrarily weaken the accuracy of the global model. This paper addresses these challenges and presents Privacy-preserving and Accountable Distributed Learning (PA-DL), a fully decentralized framework that relies on Differential Privacy to guarantee strong privacy protection of the agents data, and Ethereum smart contracts to ensure accountability.

ACM Reference Format:
Anudit Nagar, Cuong Tran, and Ferdinando Fioretto. 2021. Privacy-Preserving and Accountable Multi-agent Learning: Extended Abstract. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021*, IFAAMAS, 2 pages.

1 INTRODUCTION

Distributed multi-agent learning enables agents to cooperatively train a learning model without requiring them to share their dataset. Typical multi-agent learning frameworks, including federated learning [8], allow individual agents to train their local models on their own datasets and share model parameters with a centralized agent that aggregates the received parameters and sends them back to the agents. This simple, yet effective procedure, repeats for several iterations and allows the participating agents to learn a global model without accessing data of other agents.

This paper uses a decentralized computational environment that enables to train a differentially private multi-agent learning model, guaranteeing both privacy and trust. The resulting framework, called *Privacy-preserving and Accountable Distributed Learning (PA-DL)* relies on the Ethereum blockchain, that combines an immutable data storage with a Turing-complete computational environment [10] and guarantees the correctness of the programs executed over the blockchain.

The Ethereum protocol ensures that smart contracts are executed correctly, thus, agents can trust that any data sent to the blockchain will not be corrupted and that smart contracts logic will be executed as intended. While this environment guarantees that the data stored on the blockchain is immutable, it does not guarantee data privacy. The privacy requirement is enforced by ensuring that the learned model is *differentially private* [3]. PA-DL uses a clipping approach on the model parameters and the privacy analysis relies on composition methods [4] and the moment accountant for the Sampled Gaussian Mechanism [9]. Trustworthiness is achieved by running the computation on the immutable blockchain combined with a decentralized procedure that validates the genuineness of the agent contributions to the model.

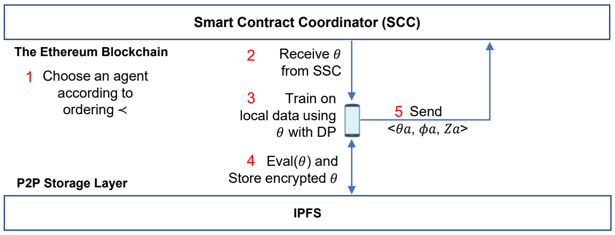
Smart Contract Coordinator (SCC)
The Ethereum Blockchain
1 Choose an agent according to ordering $\prec$
2 Receive θ from SSC
3 Train on local data using θ with DP
5 Send
<θa, ϕa, Za>
4 Eval(θ) and Store encrypted θ
P2P Storage Layer
IPFS

Figure 1: Flow diagram of the PA-DL Framework

Problem Settings and Goals The paper considers a collection of K agents, each holding a dataset D_a ($a \in [K]$) consisting of n_a individuals' data points (X_i, Y_i), with $i \in [n_a]$ drawn from an unknown distribution. Therein, $X_i \in \mathcal{X}$ is a feature vector and $Y_i \in \mathcal{Y}$ is a label. The goal is to learn a *global* model $\mathcal{M}_\theta : \mathcal{X} \to \mathcal{Y}$, where θ is a real valued vector describing the model parameters. The model quality is measured in terms of a *loss function* $\mathcal{L} : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_+$, and the problem is that of minimizing the empirical risk function:

$$\min_\theta J(\mathcal{M}_\theta, D) = \frac{1}{K} \sum_{a \in [K]} J_a(\mathcal{M}^a_\theta, D_a), \tag{1}$$

where $\mathcal{M}^a_\theta$ is a local classifier associated to agent $a \in [K]$ and J_a is its empirical risk function, defined as $J_a(\mathcal{M}^a_\theta, D_a) = \frac{1}{n_a} \sum_{(X_i,Y_i) \in D_a} \mathcal{L}\left(\mathcal{M}^a_\theta(X_i), Y_i\right)$. The paper focuses on problems (1) with non-convex components J_a, as in deep learning tasks.

2 THE PA-DL FRAMEWORK

Private and Accountable Distributed Learning (PA-DL) is a fully distributed learning framework that ensures privacy and accountability while keeping the network bandwidth low. The framework is schematically shown in Figure 1 and its main components are:

- A *Smart Contract Coordinator (SCC)*: It is a program executed on the blockchain that orchestrates the interaction among PA-DL agents to ensure the correct data exchange aimed at training a global model. The SCC operates in *rounds*. At each round a set of agents is invoked according to a predefined ordering and their responses are aggregated.

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.

Agents	10	100	1000
PA-DL	0.102	1.020	10.20
FedAvg	0.600	6.000	60.00

(a)

Method	10^3 Tasks	10^6 Tasks
createTask	$2.76 * 10^{-3}$ USD	2.76 USD
startNextRound	$7.23 * 10^{-4}$ USD	0.723 USD

(b)

Table 1: Network bandwidth in GB (left) and Cost in USD (right) for execution of a method in SCC for N parallelly executing tasks, required to complete 1 round on MNIST data.

- A provably private *PA-DL agent* training procedure: At each round, the invoked agents use the parameters obtained by the SCC to train a model over their dataset. Each training step is ensured to guarantee (ϵ, δ)-differential privacy.
- *Accountability*: Prior being able to submit a model update, a PA-DL agent is required to invoke a verification step that ensures its trustworthiness.

3 EXPERIMENTAL ANALYSIS

Datasets, Models, and Metrics. This section studies the behavior of the proposed PA-DL architectures on three classification tasks: (1) *MNIST* and (2) *Fashion MNIST* comprises of 0 to 9 handwritten digits and articles images, respectively. The task is to correctly classify the class associated with an image. (3) *COVID-19 Chest X-Ray* is the first public COVID-19 CXR image data collection [2] combined with Healthy Chest X-Ray images [5]. The task is to correctly classify from an XRAY image whether a person is COVID-19 positive.

The experiments compare the proposed model against a classic, centralized, Stochastic Gradient Descent (SGD) method, and FedAvg [6], a standard Federated Learning algorithm.

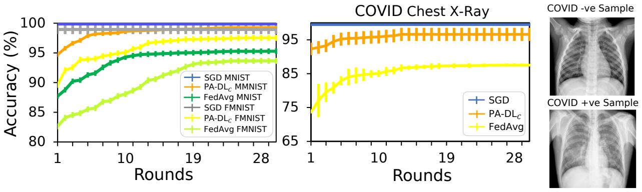

Figure 2: Algorithms accuracy per round on MNIST & FMNIST (left), COVID-19 X-Ray(Right) for K = 1000 agents.

Accuracy and Scalability. Figure 2 reports a comparison of the accuracy of all models. Observe that the PA-DL outperforms FedAvg in all cases. Next, Table 1 (left) compares the network bandwidth consumption analyzed at varying of the number of agents. The table illustrates a clear trend: The proposed PA-DL framework is able to compute models more efficiently than FedAvg as it requires less network bandwidth. Table 1 (right) shows the cost incurred by running the SCC on an Ethereum based Layer-2 Plasma [1] side chain.

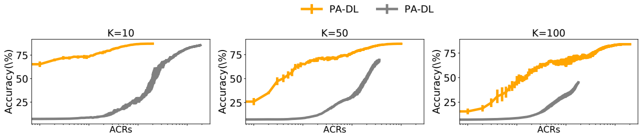

Figure 3: Accuracy on MNIST for K = 10 (top), 50 (middle), and 100 (bottom) agents. The final privacy losses for the model with K = 10, 50, and 100, respectively, are 0.5, 1.1 and 1.6.

Privacy/Accuracy trade-off. Finally, the analysis focuses on the accuracy of the distributed models under the differential privacy constraints. It follows similar settings as those described above. To ensure privacy, this section uses DP-FedAvg [7] (in lieu of FedAvg), under the privacy model adopted in the paper. The privacy settings for all models are: $c = 10$ and $\sigma = 2.0$ and the probability of pure DP violation $\delta = 1e - 3$. The total privacy loss is computed based on the moment accountant method.

Firstly, observe that PA-DL produces private models that are significantly more accurate to those produced by DP-FedAvg, under the same privacy constraints. Our analysis indicates that this is due to the different number of aggregation operations performed by the various algorithms. This crucial behavior was also observed in [7] that noted that for general non-convex objectives, averaging model parameters could produce an arbitrarily bad model. Under a very tight privacy constraint $\epsilon \approx 1$, PA-DL consistently achieves more than 80% accuracy.

The experiment also analyzed biased datasets (not reported due to space constraints) and observed that the results follow the same trends as those outlined above.

These experiments demonstrate the robustness of the proposed models under a non-IID setting, a varying number of agents, and strict privacy constraints. and shows that PA-DL may represent a promising step towards a practical tool for privacy-preserving and accountable multi-agent learning.

Acknowledgments

This research is partially supported by NSF grant 2007164. Its views and conclusions are those of the authors only.

REFERENCES

[1] Vitalik Buterin and Joseph Poon. 2017. Plasma: Scalable Autonomous Smart Contracts. *Plasma.io* (2017). http://plasma.io/plasma.pdf
[2] Joseph Paul Cohen and et al. 2020. COVID-19 Image Data Collection: Prospective Predictions Are the Future. *arXiv 2006.11988* (2020). https://github.com/ieee8023/covid-chestxray-dataset
[3] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*. Springer, 265–284.
[4] Cynthia Dwork, Aaron Roth, et al. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9, 3–4 (2014), 211–407.
[5] Daniel S. Kermany and et al. 2018. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* 172, 5 (2018), 1122 – 1131.e9. https://doi.org/10.1016/j.cell.2018.02.010
[6] Jakub Konečnỳ, Brendan McMahan, and Daniel Ramage. 2015. Federated optimization: Distributed optimization beyond the datacenter. *arXiv preprint arXiv:1511.03575* (2015).
[7] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*. PMLR, 1273–1282.
[8] H. Brendan McMahan, Daniel Ramage, Kunal Talwar, and Li Zhang. 2017. Learning Differentially Private Recurrent Language Models. arXiv:1710.06963 [cs.LG]
[9] Ilya Mironov, Kunal Talwar, and Li Zhang. 2019. Rényi Differential Privacy of the Sampled Gaussian Mechanism. arXiv:1908.10530 [cs.LG]
[10] Gavin Wood et al. 2014. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper* 151, 2014 (2014), 1–32.