PhD Position: "Private and Byzantine-Robust Federated Learning"

1 Research project

1.1 Context

The increasing size of data generated by smartphones and IoT devices motivated the development of Federated Learning (FL) [12], a decentralized learning framework for on-device collaborative training of machine learning models. FL algorithms like FedAvg [15] allow clients to train a common global model without sharing their personal data. FL reduces data collection costs and can help to mitigate data privacy issues, making it possible to train models on large datasets that would otherwise be inaccessible. FL is currently used by many big tech companies (e.g., Google, Apple, Facebook) for learning on their users' data, but the research community envisions also promising applications to learning across large data-silos, like hospitals that cannot share their patients' data [19].

While they mitigate privacy concerns by not exchanging raw data, FL does not in itself offer rigorous privacy guarantees, and FL algorithms can be attacked by malicious participants. For these reasons, in recent years a large body of the literature has focused on the design of decentralized algorithms that are more privacy-preserving, using different variants of differential privacy [16, 6]. On the other hand, another line of research has focused on decentralized learning algorithms that are robust to the presence of malicious individuals in the system (Byzantine agents) [8, 7, 9]. Nevertheless, the design and analysis of algorithms that are both robust and privacy-preserving is far less considered and understood. Recently, it has been shown that in the case where the server is honest-but-curious, the combination of differential privacy and robustness induces an additional error term, making them at odds with each other [2]. Specifically, we face a utility-privacy-robustness trilemma (on top of the conventional privacy-utility and robustness-utility trade-offs). Conversely, in the case of a trusted server, some studies [10] have shown that privacy and robustness can actually be mutually beneficial. A key question then arises: in what contexts are these two notions really good for each other?

1.2 Research objectives

The main goal of this PhD is to answer the previous question on the basis of new theoretical analyses, and to design decentralized algorithms that are both robust and differentially private. Several lines of research could be investigated.

A natural direction to seek better trade-offs between privacy, robustness and utility is to relax the notions of privacy and/or robustness. One may consider the general framework of Pufferfish privacy [13, 17], which allows to relax differential privacy by considering more specific secrets to protect

and by constraining the prior belief that the adversary may have about the data. Similarly, while Byzantine robustness has been shown to be at odds with local differential privacy [2], it is possible to consider weaker threat models, such as the hidden state model [21], the shuffle model [5] or the network model [6]. Regarding robustness, current approaches are designed to ensure protection against Byzantine users that can misbehave arbitrarily [9]. However, such robustness is too stringent and leads to conservative learning performance in practice when no user is fully adversarial. For example, in the case of medical applications it is safe to assume that all the users, usually hospitals, clinics or pharmacies, are honest by intention, but misbehavior could occur due to mistakes like mislabelling. Refining (or designing new) Byzantine-robust schemes to weaker adversaries is crucial to fully realize the benefits of robust decentralized learning in real-world applications.

Another line of investigation is to reconsider the notion of utility. In the majority of the aforementioned work, the key quantity to control (utility) is the optimization error of the empirical risk. However, in (decentralized) machine learning one is often interested in also controlling the generalization error [3, 14], namely the error that will be made on unobserved data points. In this case, it will be interesting to study how robustness and privacy can jointly improve algorithm stability and thus help generalization, e.g., by studying the connections between gradient coherence [4] and robust aggregation [1, 22].

A last direction of research is to consider model update compression or sparsification techniques [20] that have been independently shown to help privacy (see e.g., [11]). Whether the benefits of these scheme hold true when aiming for robustness along with privacy remains unclear. Some technical challenges are as follows. (i) While sparsification (in decentralized learning) improves the overall privacy-utility trade-off, the same need not be true for the privacy-robustness trade-off. (ii) The compression noise can be amplified in the presence of malicious clients in the system [18].

1.3 Timeline

The tentative work-plan for this PhD is as follows:

1. M1-M6: Review the existing literature on decentralized algorithms, differential privacy and byzantine robustness.
2. M4-M16: Quantify the fundamental trade-offs between differential privacy, Byzantine robustness and utility in various settings and threat models. Design algorithms that provably provide a good trade-off in some of these settings, and evaluate them in practice through simulations and experiments.
3. M12-M30: Prove guarantees for the generalization error. Extend the previous results and algorithms to relaxed notions of privacy and/or robustness so as to obtain better trade-offs. Explore the potential role of compression.
4. M24-M32: Show the relevance of the proposed approaches on real-world data from concrete applications. In particular, it will be possible to apply the methods developed in the thesis to medical applications through existing collaborations with hospitals (see below).
5. M30-M36: Write the thesis manuscript and prepare for the defense.

1.4 Expected skills

The applicant is expected to have studied machine learning and/or optimization, and to have good mathematical skills. Some knowledge in distributed algorithms and broad interest for the topic of trustworthy AI is a plus.

2 Research environment

This PhD position is a collaboration between two Inria research teams: PreMeDICaL and Magnet. The position is funded by the IPoP project, a large interdisciplinary project on privacy. The hired PhD student will be mainly based in PreMeDICaL (Montpellier, France) but will have the opportunity to make regular visits to Magnet in Lille.

The PhD student will be jointly supervised by Aurélien Bellet, Nirupam Gupta, Batiste Le Bars and Marc Tommasi. Together, they gather a world-leading expertise in all three key aspects of the topic: federated learning, privacy and robustness.

This project will stimulate existing and emerging collaborations with other research groups on themes at the intersection between machine learning, privacy, robustness and decentralized algorithms. For instance, there will be opportunities to collaborate with other members of the IPoP project, the members of FedMalin (a large Inria project on federated learning), the members of the SSF-ML-DH project (on secure, safe and fairness machine learning for health), and the Distributed Computing Lab at EPFL led by Rachid Guerraoui.

In terms of concrete applications, both PreMeDICaL and Magnet have ongoing collaborations with hospitals and other clinical partners. These collaborations will provide opportunities to apply the approaches developed during the PhD to concrete use-cases, for instance to run multi-centric decentralized medical studies while preserving the confidentiality of the datasets held in each institution and providing robustness guarantees.

References

[1] Y. Allouah, S. Farhadkhani, R. Guerraoui, N. Gupta, R. Pinot, and J. Stephan. Fixing by mixing: A recipe for optimal byzantine ML under heterogeneity. In F. J. R. Ruiz, J. G. Dy, and J. van de Meent, editors, *AISTATS*, 2023.

[2] Y. Allouah, R. Guerraoui, N. Gupta, R. Pinot, and J. Stephan. On the privacy-robustness-utility trilemma in distributed learning. In A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, editors, *ICML*, 2023.

[3] R. Bassily, V. Feldman, C. Guzmán, and K. Talwar. Stability of stochastic gradient descent on nonsmooth convex losses. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *NeurIPS*, 2020.

[4] S. Chatterjee. Coherent gradients: An approach to understanding generalization in gradient descent-based optimization. In *ICLR*, 2020.

[5] A. Cheu, A. D. Smith, J. R. Ullman, D. Zeber, and M. Zhilyaev. Distributed differential privacy via shuffling. In Y. Ishai and V. Rijmen, editors, *EUROCRYPT*, 2019.

[6] E. Cyffers and A. Bellet. Privacy amplification by decentralization. In G. Camps-Valls, F. J. R. Ruiz, and I. Valera, editors, *AISTATS*, 2022.

[7] S. Farhadkhani, R. Guerraoui, N. Gupta, L. Hoang, R. Pinot, and J. Stephan. Robust collaborative learning with linear gradient overhead. In A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, editors, *ICML*, 2023.

[8] S. Farhadkhani, R. Guerraoui, N. Gupta, R. Pinot, and J. Stephan. Byzantine machine learning made easy by resilient averaging of momentums. In K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvári, G. Niu, and S. Sabato, editors, *ICML*, 2022.

[9] R. Guerraoui, N. Gupta, and R. Pinot. Byzantine machine learning: A primer. *ACM Comput. Surv.*, 56(7), 2024.

[10] S. B. Hopkins, G. Kamath, M. Majid, and S. Narayanan. Robustness implies privacy in statistical estimation. In B. Saha and R. A. Servedio, editors, *STOC*, 2023.

[11] R. Hu, Y. Gong, and Y. Guo. Federated learning with sparsified model perturbation: Improving accuracy under client-level differential privacy. *CoRR*, arXiv:2202.07178, 2022.

[12] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, et al. Advances and open problems in federated learning. *Found. Trends Mach. Learn.*, 14(1-2):1–210, 2021.

[13] D. Kifer and A. Machanavajjhala. Pufferfish: A framework for mathematical privacy definitions. *ACM Trans. Database Syst.*, 39(1):3:1–3:36, 2014.

[14] B. Le Bars, A. Bellet, M. Tommasi, K. Scaman, and G. Neglia. Improved stability and generalization guarantees of the decentralized SGD algorithm. In *ICML*, 2024.

[15] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas. Communication-efficient learning of deep networks from decentralized data. In A. Singh and X. J. Zhu, editors, *AISTATS*, 2017.

[16] M. Noble, A. Bellet, and A. Dieuleveut. Differentially private federated learning on heterogeneous data. In G. Camps-Valls, F. J. R. Ruiz, and I. Valera, editors, *AISTATS*, 2022.

[17] C. Pierquin, A. Bellet, M. Tommasi, and atthieu Boussard. Rényi pufferfish privacy: General additive noise mechanisms and privacy amplification by iteration via shift reduction lemmas. In *ICML*, 2024.

[18] A. Rammal, K. Gruntkowska, N. Fedin, E. Gorbunov, and P. Richtárik. Communication compression for byzantine robust learning: New efficient algorithms and improved rates. In *AISTATS*, 2024.

[19] N. Rieke, J. Hancox, W. Li, F. Milletarì, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. H. Maier-Hein, S. Ourselin, M. J. Sheller, R. M. Summers, A. Trask, D. Xu, M. Baust, and M. J. Cardoso. The future of digital health with federated learning. *npj Digit. Medicine*, 3, 2020.

[20] S. U. Stich, J. Cordonnier, and M. Jaggi. Sparsified SGD with memory. In S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *NeurIPS*, 2018.

[21] J. Ye and R. Shokri. Differentially private learning needs hidden state (or much faster convergence). In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *NeurIPS*, 2022.

[22] D. Yin, Y. Chen, K. Ramchandran, and P. L. Bartlett. Byzantine-robust distributed learning: Towards optimal statistical rates. In J. G. Dy and A. Krause, editors, *ICML*, 2018.