Towards Practical Privacy-Preserving Collaborative Machine Learning at a Scale

Rania Talbi

Towards Practical Privacy-Preserving Collaborative Machine Learning at a Scale

Rania Talbi
LIRIS, INSA-Lyon
Lyon, France
rania.talbi@insa-lyon.fr

***Abstract*—Collaborative machine learning allows multiple participants to get a global and valuable insight over their joint data. Nonetheless, in data-sensitive applications, it is crucial to maintain confidentiality across the end-to-end path the data follows from model training phase to the inference phase, preventing any form of information leakage about training data, the learned model, or the inference queries. In this paper, we present our approach to address this problem through PrivML, a framework for end-to-end outsourced privacy-preserving data classification over encrypted data. We provide some preliminary results comparing our proposal with state of the art solutions as well as some insight on our prospective research plan.**

I. Context and Motivation

Nowadays, an increasing number of companies are migrating towards data-driven business models relying on machine learning methods to uncover strategic and insightful information out of data. By following this mainstream, these businesses are entirely capitalizing their production process on the availability of data. In many cases, this critical resource can be just not available, or its amount is not sufficient to capture useful knowledge. Collaborative Machine Learning can be very helpful in similar situations, where multiple parties with common interests can collaborate to get better and more accurate machine learning models using their joint data. Nonetheless, the before-mentioned scheme is not always feasible, due to legal, financial and competitive constraints, particularly when the manipulated data is considered to be sensitive. Such privacy concerns can be observed in health care systems, collaborative fraud detection, recommender systems, etc. Considering the advantages that collaborative learning can bring, many research works have been conducted over the last few years to cope with these concerns. These works fall into two main categories, depending on whether they rely on non-cryptographic techniques such as data randomization [1], or on cryptographic techniques such as homomorphic encryption [2]. Solutions following the first approach usually have better performance but at the expense of lower privacy guarantees and lower utility of ML models. In contrast, solutions following the second approach usually do not affect models' utility and provide higher privacy guarantees, but at the expense of high computational costs. We thus identify the need for privacy-preserving solutions that ensure a proper equilibrium between privacy guarantees, computational efficiency, service utility, and usability. In this paper, we present our approach for resolving this issue using cryptography-based techniques and briefly discuss our preliminary results comparing to state of the art solutions. Finally, we present our future research directions in this scope of work.

II. Research Objectives

Our main objective is to design and implement a privacy-preserving framework that ensures a proper trade-off between the following criteria that we have identified as relevant for (i) *Privacy guarantees:* the provided solution must ensure end–to-end privacy-preservation of training data, machine learning models, users' requests and system responses; (ii) *Computational efficiency:* the computational overhead caused by privacy-preservation mechanisms must be minimal; (iii) *Service utility:* the utility of the provided privacy-preserving ML service must be as close as possible to their original implementations; (iv) *Service usability:* privacy-preservation strategies must not affect the service usability and must provide a proper user experience.

III. Approach

To achieve the goals described above, we propose PrivML a framework for end-to-end outsourced online privacy-preserving data classification over homomorphically encrypted data. In the following, we explain the design principles underlying PrivML and provide some preliminary results.

A. Design Principles.

PrivML delivers a standalone end-to-end practical online privacy-preserving data classification service that guarantees confidentiality throughout the entire classification process. To do that, all of the computations required in both of the training and classification phases are done over data encrypted via a lightweight partially homomorphic cryptosystem called DT-PKC cryptosystem (Distributed Two Trapdoors Public-Key Cryptosystem) [3]. In PrivML, a set of Data Owners (DO) continuously outsource chunks of encrypted training data to a Machine Learning Service Provider ($MLSP$), which builds a classification model over their joint data in a privacy-preserving manner. During the inference phase, a set of clients (C) send encrypted classification queries to the $MLSP$ that responds using the previously learned model without breaching their privacy, nor that of the responses corresponding to them. We initially implemented three classification algorithms in PrivML which are: Decision

Trees (Priv-DT), Naive Bayes classifier (Priv-NB), and Logistic Regression (Priv-LR). We decompose each one of these classification algorithms into a set of elementary operations that we implement by composing a set of two-party computation building blocks based on the homomorphic properties of the DT-PKC cryptosystem and cryptotographic blinding. These building blocks are run by two non-colluding computational sub-units which we call (Master computation unit MU) and (Secondary computation unit SU). These sub-units collaborate to achieve computations such as private logarithm or private product over encrypted data without learning at any time the inputs, intermediate or final results of these computations. A general scheme of the building blocks implemented in PrivML is the following: At first, the first unit MU cryptographically blinds the building block's inputs using randomly generated numbers that only she knows by applying DT-PKC's homomorphic properties. Later on, these inputs are decrypted via a special two-step decryption mechanism provided in DT-PKC and communicated to the second unit SU, which applies the required computations on the blinded inputs, encrypts them then sends them back to the unit MU. Since SU has no knowledge of the random numbers used for blinding the inputs, she cannot recover their exact values. Finally, the MU tries to recover the final result by using the homomorphic properties of the DT-PKC cryptosystem. PrivML absorbs the computational overhead incurred by using homomorphic computations by using building blocks with minimal round complexity instead of the straightforward combination of multiple low granularity ones as proposed for the DT-PKC cryptosystem [3]. Moreover, PrivML uses parallel computing to improve the efficiency and scalability of the proposed ML services. Also, it is worth mentioning that in PrivML, we implement privacy-preserving online classifier learning algorithms which allow updating private data models when new data owners join the collaborative learning process, or when new training data is available, without having to reparse all the data handled so far and start the training from scratch which considerably improves the usability of the system.

B. Preliminary Results.

We compared the classifiers implemented in PrivML with state-of-the-art solutions [2], [4], [5].
PrivML is about 46 times faster than the work proposed in [4], and reduces the bandwidth consumption by a factor of 2241. We also compare the Priv-NB protocol proposed in PrivML with another outsourced solution for privacy-preserving Naive Bayes classification [5] that relies on Gentry's fully homomorphic cryptosystem. We measure a learning time that is 151 times faster and a speedup factor of up to 2647 during the inference phase. As for the Priv-LR protocol, we compare it with the winning solution of iDASH security and privacy competition [6]. In PrivML, we obtain a learning time twice faster than the FHE-LR solution with 150 times lower bandwidth consumption.

IV. Future Directions

In our future work, we consider enhancing privacy-preservation guarantees in a more decentralized collaborative learning framework which is federated learning. In this setting, devices can collaboratively train a model on their joint data without having to send them to an external untrusted service provider as we've done in PrivML. To do that, these devices interactively update an initial model according to their local training data and only exchange these updates with an orchestrator who aggregates them to a global model. Although in this scheme, private data is not shared with external parties, it is still vulnerable to a set of exploratory and causative adversarial attacks [7] that try to infer information about private training data based on communicated model updates (model inversion) or try to make models fail by injecting poisoned data during the learning process (poisoning attacks) or using crafted adversarial examples at inference time (evasion attacks). We are particularly interested in poisoning attacks in our future research since this kind of attacks is much more severe in the context of federated learning and much harder to detect, since the orchestrator cannot run any detection mechanism based on data inspection and since the updates sent by the participants are non-iid distributed. We plan to propose a detection mechanism of model poisoning attacks that is feasible under the assumptions of federated learning. Later on, we want to investigate the possibility of implementing poisoning detection mechanisms that operates over protected model updates via privacy-preservation mechanisms such as homomorphic encryption or secret sharing. Thus allowing the orchestrator to inspect the integrity of model updates without allowing him to infer sensitive information about private training data using these updates.

Acknowledgments

This work is conducted under the supervision of Prof. Sara Bouchenak. It is supported by the French National Research Agency (ANR), through the SIBIL-Lab project.

References

[1] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep Learning with Differential Privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2016, pp. 308–318.

[2] M. Kim, Y. Song, S. Wang, Y. Xia, and X. Jiang, "Secure Logistic Regression Based on Homomorphic Encryption: Design and Evaluation," *JMIR Medical Informatics*, vol. 6, no. 2, p. e19, 2018.

[3] X. Liu, R. H. Deng, K.-K. R. Choo, and J. Weng, "An Efficient Privacy-Preserving Outsourced Calculation Toolkit with Multiple Keys," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 11, pp. 2401–2414, 2016.

[4] R. Bost, R. A. Popa, S. Tu, and S. Goldwasser, "Machine Learning Classification over Encrypted Data." in *NDSS*, vol. 4324, 2015, p. 4325.

[5] S. Kim, M. Omori, T. Hayashi, T. Omori, L. Wang, and S. Ozawa, "Privacy-Preserving Naive Bayes Classification Using Fully Homomorphic Encryption," in *International Conference on Neural Information Processing*. Springer, 2018, pp. 349–358.

[6] X. Wang, H. Tang, S. Wang, X. Jiang, W. Wang, D. Bu, L. Wang, Y. Jiang, and C. Wang, "idash secure genome analysis competition 2017," 2018.

[7] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings *et al.*, "Advances and open problems in federated learning," *arXiv preprint arXiv:1912.04977*, 2019.