PRIVILEGE: PRIvacy and Homomorphic Encryption for ArtificiaL IntElliGencE

***Abstract*—PRIVILEGE project advances the state of the art of defence technology using Artificial Intelligence (AI) systems, with respect to data security and privacy preservation in a collaborative setting. The PRIVILEGE collaborative solution, based on distributed AI frameworks, such as federated learning and PATE, together with privacy-preservation and security tools such as Differential Privacy (DP), (Fully) Homomorphic Encryption (FHE), Verifiable Computation (VC), or Multi-Party Computation (MPC) will strengthen collaboration among different allies on the secure analysis of sensitive defence and military data.**

***Index Terms*—Federated Learning, PATE, homomorphic encryption, multi-party computation, differential privacy**

I. Introduction

The objective of the PRIVILEGE project is to address the issue of secure collaboration of allied military forces without exchanging sensitive data. PRIVILEGE has designed, delivered specifications and developed "privacy-by-design" machine learning training techniques. More specifically, the goal of PRIVILEGE is to enrich AI with privacy-preserving techniques tools such as Homomorphic Encryption (HE) or Differential Privacy (DP), in order to be able to exploit confidential military and defence data throughout the life-cycle of AI methods with a focus on the learning step. Additionally, the novel privacy-preserving AI algorithms will be suitable for public domain (civilian) AI applications, making them GDPR compliant.

II. Context

AI technologies are nowadays widely applied across various sectors such as medicine, finance and image recognition. In the defence and military domain, AI techniques are now applied in novel tools for surveillance, weapons control, autonomous driving and cybersecurity. All these applications require a large volume of training data, and the sharing of data between allies (e.g. in the context of NATO) may be useful even if it is difficult, and in some cases impossible.

The main technical target of PRIVILEGE is to address the data-privacy issues dealing with the collaborative training of Artificial Neural Network systems by means of secure computation and privacy-preserving techniques. This will allow multiple owners of learning datasets to build better models over the union of their training data without disclosing these datasets to one another nor to some third party.

III. Privilege research and technical solution

A. Science and technology breakthroughs

Public studies about using FHE in the training phase for neural nets are sparse or concentrate on toy examples for collaborative training. As such, the design of a private-by-design learning step for collaborative algorithms such as Federated Learning and PATE requires the adaptation of the existing homomorphic encryption schemes, as well as the design of specifically tailored homomorphic learning methods.

Another breakthrough of the project is the design of an execution integrity layer for the aggregation function through Verifiable Computing. The use of Verifiable Computing for attesting the integrity of an aggregated learning result is an innovative idea that was not previously investigated in the open literature.

Another significant advancement of PRIVILEGE is the use of Differential Privacy methods for collaborative learning in an application context. The implementation, application and validation of DP mechanisms for the three PRIVILEGE use-cases are a significant step forward. Moreover, its combination with homomorphic encryption in the context of PATE will be an important achievement towards secure AI tools, concerning both training data and model confidentiality.

Finally, in contrast with the current state of the art solutions, which are mainly tested with the textbook example of the MNIST dataset (handwritten digit database) or some variants of it, PRIVILEGE demonstrates the practical performances of the privacy-preserving framework for collaborative learning on three real use-cases from the defence field: the classification of radio waves in defence operation, the classification of malicious network logs, and the video processing for unmanned vehicles.

B. Methodological approach and Results

1) Differential Privacy and AI: Differential Privacy (DP) [1] allows to bound the leak of information when using a random algorithm in the sense that it gives a bound of the difference of the algorithm outputs distribution when the algorithm is computed on two adjacent datasets. It is largely used because of its good properties (composition theorem, post-processing property, etc.).

Several mechanisms that achieve DP have been proposed for machine learning model training. For instance, [2] proposes an adaptation of stochastic gradient descent basically used to train deep neural network algorithm by adding a Gaussian mechanism during the update of gradient/weights step and by clipping to avoid large sensibility on the gradients or the weights norm. Moreover, it is paired with a moments accountant algorithm to amplify the privacy loss across weight updates. Private Aggregation Teacher Ensembles (PATE) [3] is another approach used to introduce differential privacy in machine learning. Here, a set of machine learning teachers is trained on disjoint sensitive datasets. Then these teachers' models are used to label a less sensitive dataset with some noise added during the vote aggregation to ensure differential privacy. Finally, a student machine learning model, which is expose to potential attacker, is trained on the less labeled sensitive dataset.

Federated Learning [4] is used when several data owners want train a joint machine-learning model without sharing data. It is possible to add global differential privacy during Federated Learning training [5]. In this case, the central aggregator manages to clip the gradients/weights update send by the participants and add a Gaussian noise to achieve differential privacy.

2) Secure computation: Fully Homomorphic Encryption (FHE), part of the provably-secure cryptography, allows encryption and

general computation directly over encrypted data and has now a well-founded corpus of security properties. PRIVILEGE proposes a beyond-state-of-the-art solution based on Paillier [6] cryptosystem with faster encryption and batching for Federated Learning (FL) in the case of a semi-honest aggregation server. Moreover, for FL training, it also provides alternatives in the form of optimized batched versions of BFV [7] and CKKS [8]. As for PATE, a novel and efficient argmax operator exploiting the properties of TFHE cryptosystem [9] has been conceived. For the case of a malicious aggregation server, the HE solutions have been complemented with authentication mechanisms for the encrypted data such as the ones from [10] and [11].

PRIVILEGE also proposes Multi-Party Computation (MPC) as an alternative to FHE. MPC distributes the encrypted computation across multiple servers, protecting the inputs' privacy unless all of the involved parties collude. In PRIVILEGE, different MPC protocols available in the MP-SPDZ framework [12] were tested and applied for FL and PATE and can be deployed in the cases where there are several aggregation servers.

For more details on the different privacy-preserving and security solutions developed in the context of PRIVILEGE project, we refer the interested reader to [13], [14], [15].

3) Overall PRIVILEGE framework: The PRIVILEGE framework is separated into several components:

- A **Communication Layer**: This component is based mainly on an *Apache Kafka*[1] message broker and is used by PRIVILEGE services communicating with each other. Simpler communications use REST calls over HTTP.
- The **Registration and Orchestration** Service: A python *Flask*[2] service that is responsible for registering the Data Providers and communicate this information to the User Interface component.
- **Data Providers**: These components emulate participants in a Collaborative Learning. They have machine learning models trained on local private data. They register with the Registration and Orchestration service and participate in Collaborative Learning tasks.
- An **End-User**: This component emulates a participant willing to have the results of a Collaborative Learning. It launches computations and collects the final results. It is also responsible for the FHE keys deployment before a Collaborative Learning task using FHE.
- The **User Interface** component: This graphical component displays information about the registered Computing Parties and Data Providers. An end-user can connect to this interface to set up and launch collaborative learning jobs.
- The **Federated Learning Aggregator** component: A third-party node, responsible for implementing Federated Learning operations on the Data Providers' inputs.
- The **FHE Aggregator** component: A third-party node, possibly semi-honest, responsible for implementing PATE aggregation operations on the Data Providers inputs, using the FHE primitive.
- **Computing Parties**: Components embedding MPC capabilities. They are responsible for implementing the PATE aggregation operations on the Data Providers' inputs, using MPC primitives.

Each component is containerized using *Docker*[3], standardizing the deployment of each component. For demonstration, a *docker-compose*[4] configuration file is used in single-node deployments.

[1]https://kafka.apache.org/

[2]https://flask.palletsprojects.com/en/2.2.x/

[3]https://www.docker.com/

[4]https://docs.docker.com/compose/

IV. Conclusion

PRIVILEGE proposes a unique concept of an enhanced collaborative learning technology that includes privacy-preserving techniques. It combines advanced cryptographic tools with collaborative learning frameworks such as Federated Learning and PATE. This approach has been initially validated using three real-world defence use cases but its applicability is much wider.

Acknowledgements

The research leading to these results has received funding from the European Union's Preparatory Action on Defence Research (PADR-FDDT-OPEN-03-2019). This paper reflects only the authors' views and the Commission is not liable for any use that may be made of the information contained therein.

PRIVILEGE project (2020-2023) is coordinated by THALES ThereSIS (Romain Ferrari) and involves four European partners: THALES ThereSIS (France), CEA (France), CESNET (Czechia Republic), Intracom Defense S.A. (Greece). The privacy-preserving collaborative PRIVILEGE framework and the associated innovations have a targeted TRL between 3-5.

References

[1] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, pp. 211–407, 2014.

[2] 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*, Oct. 2016, pp. 308–318.

[3] N. Papernot, S. Song, I. Mironov, A. Raghunathan, K. Talwar, and Erlingsson, "Scalable Private Learning with PATE," Feb. 2018. [Online]. Available: http://arxiv.org/abs/1802.08908

[4] X. Yin, Y. Zhu, and J. Hu, "A comprehensive survey of privacy-preserving federated learning," *ACM Computing Surveys (CSUR)*, vol. 54, pp. 1 – 36, 2021.

[5] 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.

[6] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Advances in Cryptology — EUROCRYPT '99*, J. Stern, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 223–238.

[7] J. Fan and F. Vercauteren, "Somewhat practical fully homomorphic encryption," Cryptology ePrint Archive, Report 2012/144, 2012, https://ia.cr/2012/144.

[8] J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," Cryptology ePrint Archive, Report 2016/421, 2016, https://ia.cr/2016/421.

[9] I. Chillotti, N. Gama, M. Georgieva, and M. Izabachène, "Faster fully homomorphic encryption: Bootstrapping in less than 0.1 seconds," in *ASIACRYPT*, 2016, pp. 3–33.

[10] P. Struck, L. Schabhüser, D. Demirel, and J. Buchmann, "Linearly homomorphic authenticated encryption with provable correctness and public verifiability," in *International Conference on Codes, Cryptology, and Information Security*. Springer, 2017, pp. 142–160.

[11] D. Fiore, R. Gennaro, and V. Pastro, "Efficiently verifiable computation on encrypted data," in *Proceedings of the 2014 ACM SIGSAC*, 2014, pp. 844–855.

[12] M. Keller, "MP-SPDZ: A versatile framework for multi-party computation," Cryptology ePrint Archive, Report 2020/521, 2020, https://eprint.iacr.org/2020/521.

[13] A. Madi, O. Stan, A. Mayoue, A. Grivet-Sébert, C. Gouy-Pailler, and R. Sirdey, "A secure federated learning framework using homomorphic encryption and verifiable computing," in *2021 (RDAAPS)*, 2021, pp. 1–8.

[14] A. G. Sébert, R. Sirdey, O. Stan, and C. Gouy-Pailler, "Protecting data from all parties: Combining fhe and dp in federated learning," 2022. [Online]. Available: https://arxiv.org/abs/2205.04330

[15] O. Stan, V. Thouvenot, A. Boudguiga, K. Kapusta, M. Zuber, and R. Sirdey, "A secure federated learning: Analysis of different cryptographic tools," in *Proceedings of the 19th International Conference on Security and Cryptography, SECRYPT 2022*, S. D. C. di Vimercati and P. Samarati, Eds. SCITEPRESS, 2022, pp. 669–674.