DE LA RECHERCHE À L'INDUSTRIE

Machine learning without jeopardising the training data

July, 25th, 2022

Arnaud Grivet Sébert, Renaud Sirdey, Cédric Gouy-Pailler

Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France

Several techniques to achieve privacy

I. Context

II. SPEED : Secure, PrivatE, and Efficient Deep learning

III. Private federated learning

IV. SecTL: Secure and Verifiable Transfer Learning-based inference

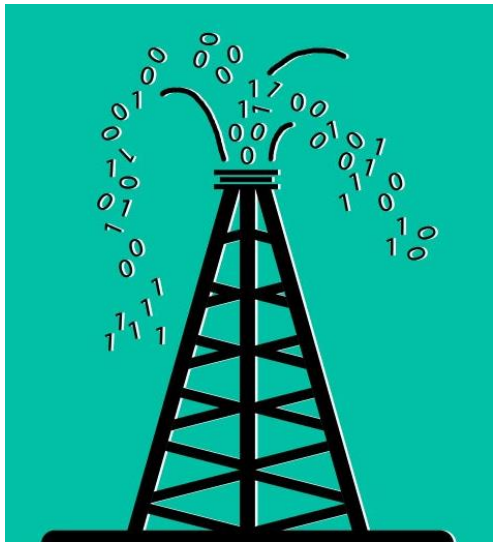
V. Conclusion and perspectives

I. Context

Two opposite objectives

“Data is the new oil”

- Importance of the data for machine learning and even more for deep learning



VS

Data privacy

- Increasing concern for privacy (GDPR in EU, HIPAA in the USA), especially for military, medical data

Several techniques to achieve privacy

- Anonymization: removing information that may enable to identify the person
  - not sufficient (cf. Netflix prize and **Narayanan & Shmatikov, 2008**)

- Secure aggregation (masking with noise)
  - requires communication between the clients before each learning round

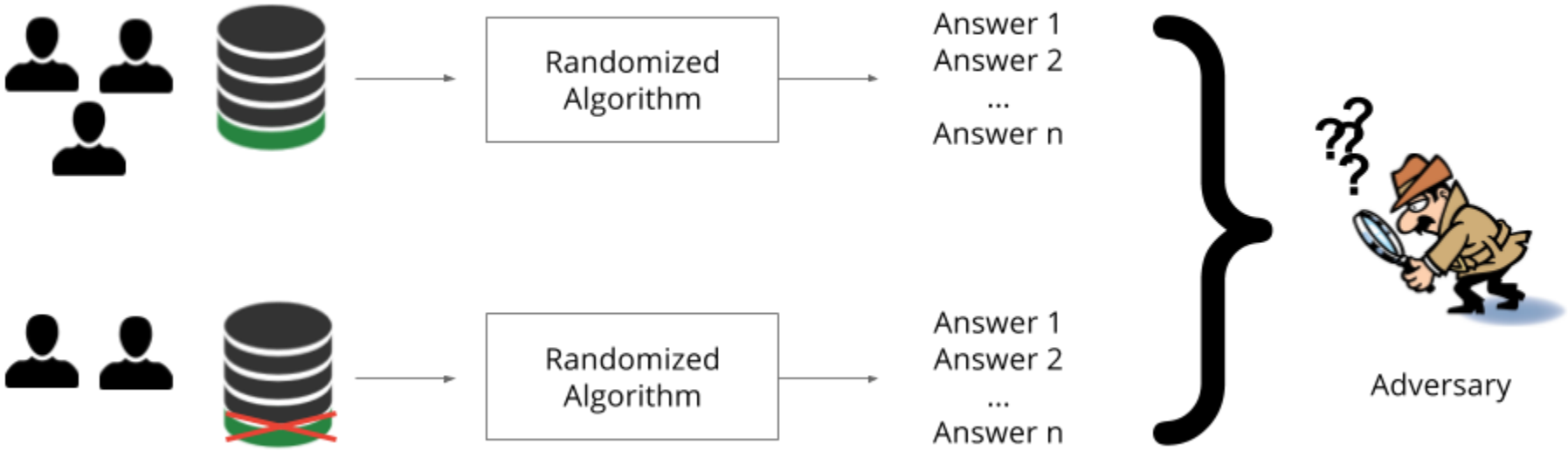
- Homomorphic encryption (HE)

$$\mathsf{Enc}(m_1) \oplus \mathsf{Enc}(m_2) = \mathsf{Enc}(m_1 + m_2) \in \Omega.$$
$$\mathsf{Enc}(m_1) \otimes \mathsf{Enc}(m_2) = \mathsf{Enc}(m_1 m_2) \in \Omega.$$

- Differential privacy (DP)

Differential privacy

- Numerical measure of the **indistinguishability** of two adjacent databases → **theoretical guarantees**
- (ε, δ)-DP : $$\mathbb{P}\left[\mathcal{A}(d) \in S\right] \leq e^{\epsilon}\mathbb{P}\left[\mathcal{A}(d') \in S\right] + \delta.$$



- In general, DP is achieved by adding some random **noise**.
- **Post-processing** does not impact the DP guarantees.

Problem

Scenario :

- Several entities owning sensitive data want to collaborate in order to train a model.
  - e.g., hospitals that have patients' data and wish to train a model to detect a specific disease.

Problem

Scenario :

- Several entities owning sensitive data want to collaborate in order to train a model.
  - e.g., hospitals that have patients' data and wish to train a model to detect a specific disease.

- The data owners do not want to share or outsource their data.

  → **approach** : collaborative learning, with use of an **aggregation server**.

Problem

Scenario :

- Several entities owning sensitive data want to collaborate in order to train a model.
  - e.g., hospitals that have patients' data and wish to train a model to detect a specific disease.

- The data owners do not want to share or outsource their data.

  → **approach** : collaborative learning, with use of an **aggregation server**.

Problem :

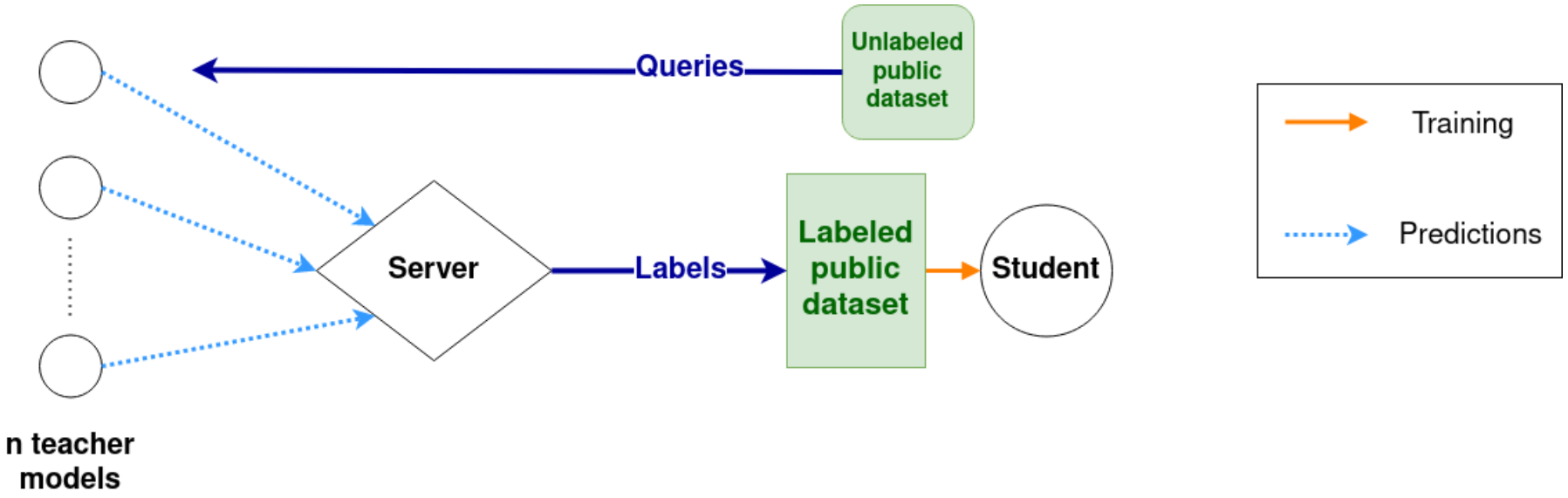
The data are sensitive and anyone is a potential adversary.

- **SPEED: Secure, PrivatE, and Efficient Deep learning,** Grivet Sébert, A., Pinot, R., Zuber, M., Gouy-Pailler, C., & Sirdey, R. (2021). *ECML-PKDD 'Machine learning'*

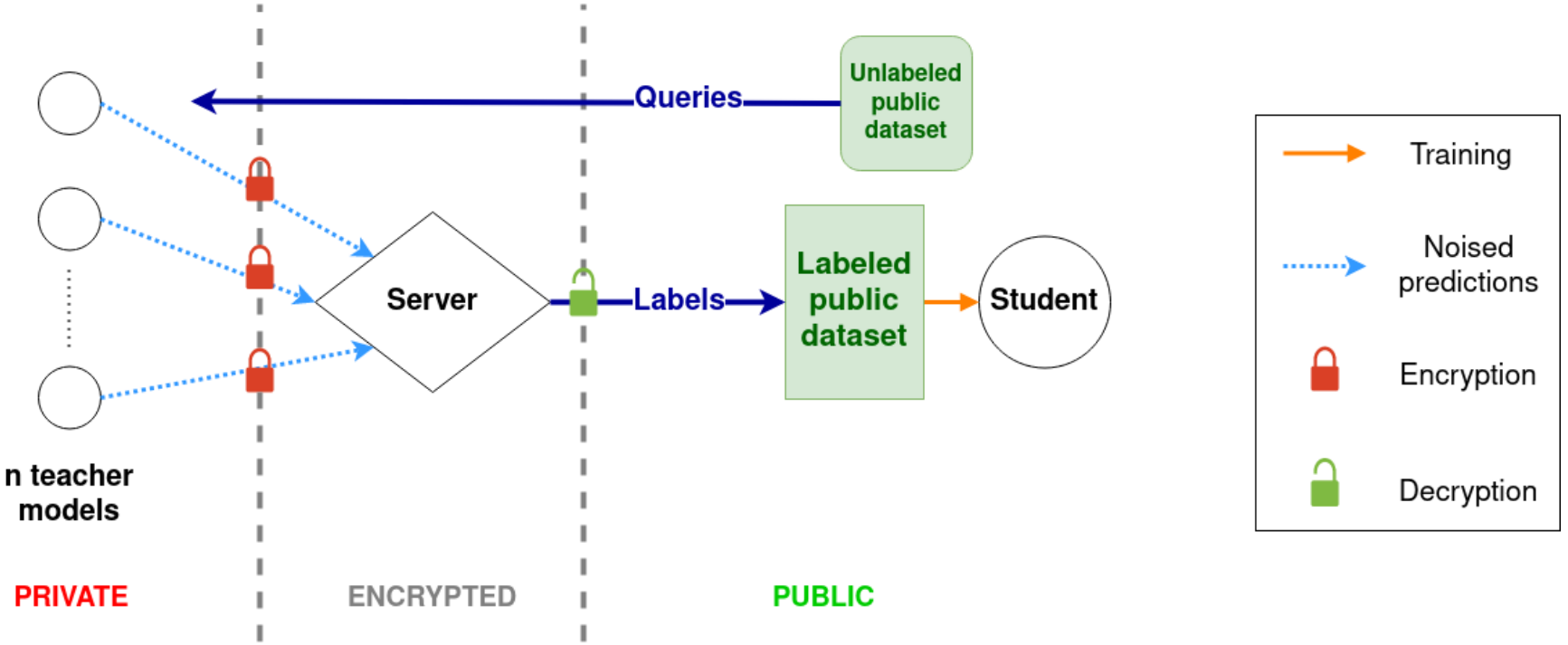
II. SPEED

SPEED - architecture

- Collaborative architecture inspired from ***Papernot et al., 2016***.
- Task of labeling a public unlabeled dataset to train a **student model**.
- The data owners (**teachers)** pretrained **local models** using their sensitive data.
- For each student's query, each teacher votes to label the sample using its local model.
- The aggregation outputs the **most frequent class**.
- Agnostic to the type of teacher and student models.



Assumption : the server is **honest-but-curious**.



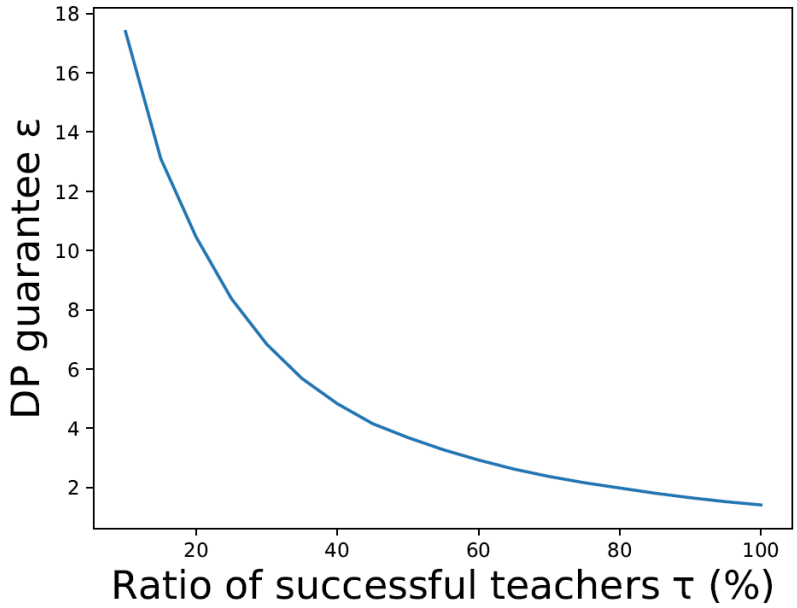
Source of threat	End-users	Server
Countermeasure	Differential privacy (DP)	Homomorphic encryption (HE)

DP analysis

- Distributed generation of a Laplacian noise among the teachers

- We determine the privacy cost per query and derive the overall privacy cost using the **moments accountant** method (***Abadi et al., 2016***).

- SPEED is **robust** against colluding teachers: DP guarantee per query as a function of the ratio τ of teachers whose noise is kept secret
  - ➢ DP guarantees from the point of view of a colluding teacher, an honest teacher or any entity who has access to the final model



$\delta = 10^{-5}$

- **Protecting Data from all Parties: Combining Homomorphic Encryption and Differential Privacy in Federated Learning,** Grivet Sébert, A., Sirdey, R., Stan O. & Gouy-Pailler, C. (2022). *arXiv preprint arXiv:2205.04330*

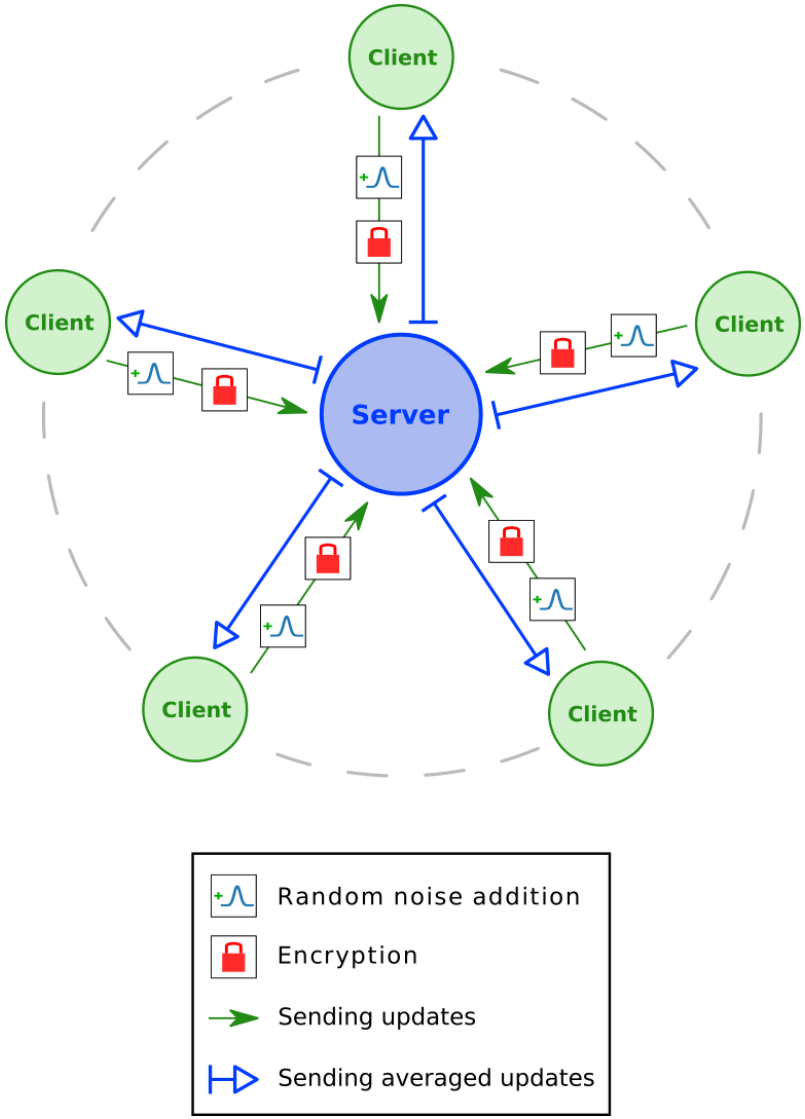
III. Private federated learning

A secure framework

- **Federated learning** (FL) architecture

- **DP**: distributed addition of **Gaussian noise** (the server is not trusted for adding the noise). This also requires **clipping**.

- **HE**: the noised updates are **encrypted** before they are sent to the server
  - the noised updates have to be discretised



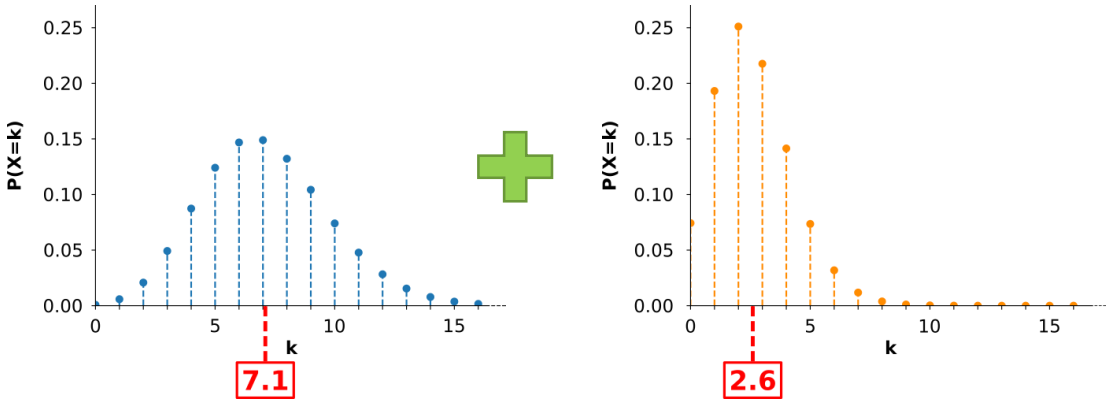
Poisson quantisation

- New stochastic quantisation operator

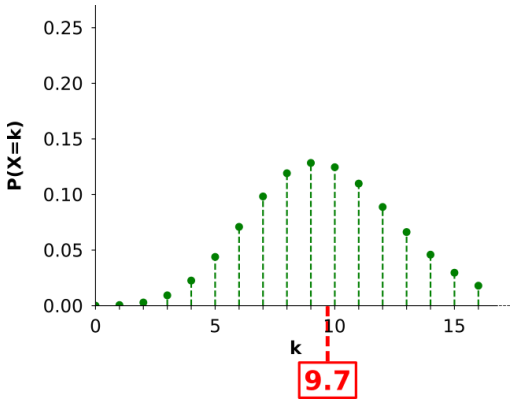
$$Q_{s,\mu} : x \in]\mu; +\infty[\mapsto sY + \mu$$

where $\mu \in s\mathbb{Z}$

and $Y \sim \mathcal{P}\left(\frac{x-\mu}{s}\right)$



- Commutes with the sum (the sum of the quantised values has the same distribution as the quantised sum)
  - quantisation can be viewed as a post-processing -> no impact on the DP guarantees



Solving problems

- Lower bound of the Gaussian noise « thanks to » the imperfection of the sampling algorithms (Box-Muller in cartesian and polar forms, ziggurat)

- Poisson distribution is not bounded but the encryption automatically applies a **modulo** operation
  - the modulo operation does not affect the DP guarantees (post-processing) and, in practice, does not affect the model accuracy either

$$\sum_{i=1}^{K}(x_i \mod N) \mod N = \sum_{i=1}^{K} x_i \mod N$$

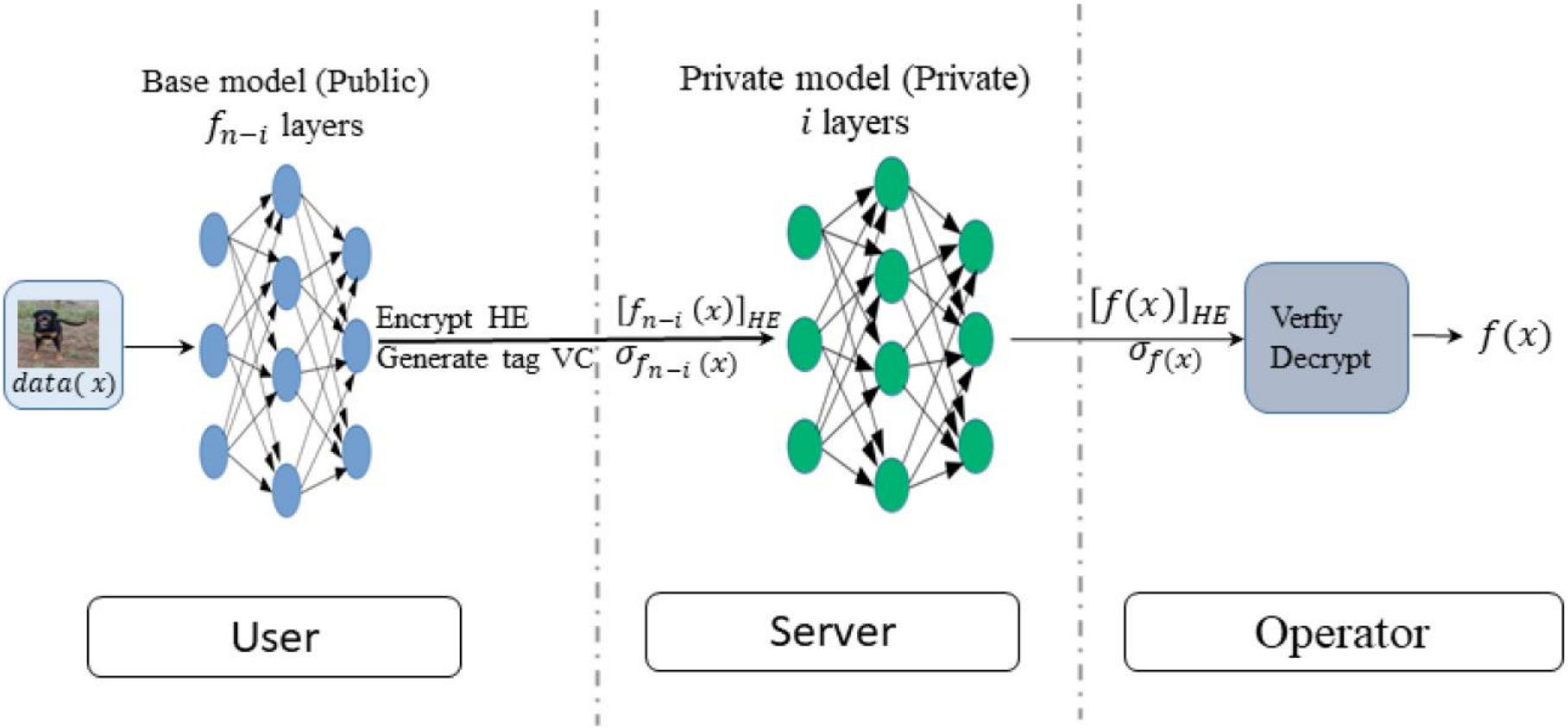
Advantages

- Same DP guarantees as the **Gaussian mechanism**, even for **colluding participants**, just by changing the value of the noise standard deviation in the analysis
- Only homomorphic addition -> mild computational overhead (+3,6% in time)
- Side-effect: the parameters of the **model** are protected from the server
- The distributed noise generation would allow to use verifiable computing techniques, as in **A secure federated learning framework using homomorphic encryption and verifiable computing**, Madi et al.(2021) (which lacks DP)

- **SecTL: Secure and Verifiable Transfer Learning-based inference**, Madi, A., Stan, O., Sirdey, R., & Gouy-Pailler, C. (2022). In *ICISSP* (pp. 220-229).

IV. SecTL

- Secure and verified **inference** thanks to **homomorphic encryption** (HE) and **verifiable computing** (VC)
- Leverages **transfer learning** (and dimensionality reduction) to simplify computations for HE and VC



V. Conclusion and perspectives

Conclusion and perspectives

Two full workflows for **collaborative** learning addressing a large scope of threats, coming from **any honest-but-curious entity**.

A secure and verifiable inference process going beyond the honest-but-curious assumption.

Perspectives :

- Lighter argmax operator for HE
- **Robust** aggregation operator (against **Byzantine** attacks) easily implementable in HE (no comparison nor division)
  - adapt median, Krum
- **Verifiable computing** for integrity

Thank you for your attention

Arnaud Grivet Sébert

Illustration credits:

- Data is the new oil: *The Myth of China's Big A.I. Advantage*, Justin Sherman and Samm Sacks (Slate)
- Differential privacy: *Privacy and machine learning: two unexpected allies ?*, Papernot et Goodfellow (blog post, 2018)

Bibliography

- Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). **Deep learning with differential privacy**. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*.
- Grivet Sébert, A. G., Pinot, R., Zuber, M., Gouy-Pailler, C., & Sirdey, R. (2020). **SPEED: Secure, PrivatE, and Efficient Deep learning**. *ECML-PKDD 'Machine learning'*.
- Grivet Sébert, A. G., Sirdey, R., Stan O. & Gouy-Pailler, C. (2022). **Protecting Data from all Parties: Combining FHE and DP in Federated Learning**. *arXiv preprint arXiv:2205.04330*.
- Madi, A., Stan, O., Mayoue, A., Grivet-Sébert, A., Gouy-Pailler, C., & Sirdey, R. (2021). **A secure federated learning framework using homomorphic encryption and verifiable computing**. In *2021 Reconciling Data Analytics, Automation, Privacy, and Security: A Big Data Challenge (RDAAPS)* (pp. 1-8). IEEE.
- Madi, A., Stan, O., Sirdey, R., & Gouy-Pailler, C. (2022). **SecTL: Secure and Verifiable Transfer Learning-based inference**. In *ICISSP* (pp. 220-229).
- Narayanan & Shmatikov (2008). **Robust de-anonymization of large sparse datasets**.
- Papernot, N., Abadi, M., Erlingsson, U., Goodfellow, I., & Talwar, K. (2016). **Semi-supervised Knowledge Transfer for Deep Learning from Private Training Data**. In *Proceedings of the 5th International Conference on Learning Representations*, Toulon, France.
- Zuber, M., Carpov, S., & Sirdey, R. (2019). **Towards real-time hidden speaker recognition by means of fully homomorphic encryption**. *IACR Cryptol. ePrint Arch.*, *2019*, 976.