Perfectly Privacy-Preserving AI
What is it and how do we achieve it?

Patricia Thaine, Gerald Penn
University of Toronto
{pthaine,gpenn}@cs.toronto.edu

ABSTRACT

Many AI applications need to process huge amounts of sensitive information for model training, evaluation, and real-world integration. These tasks include facial recognition, speaker recognition, text processing, and genomic data analysis. Unfortunately, one of the following two scenarios occur when training models to perform the aforementioned tasks: either models end up being trained on sensitive user information, making them vulnerable to malicious actors, or their evaluations are not representative of their abilities since the scope of the test set is limited. In some cases, the models never get created in the first place.

There are a number of approaches that can be integrated into AI algorithms in order to maintain various levels of privacy. Namely, differential privacy, secure multi-party computation, homomorphic encryption, federated learning, secure enclaves, and automatic data de-identification. We will briefly explain each of these methods and describe the scenarios in which they would be most appropriate.

Recently, several of these methods have been applied to machine learning models. We will cover some of the most interesting examples of privacy-preserving ML, including the integration of differential privacy with neural networks to avoid unwanted inferences from being made of a network's training data. We will also discuss the work we have done on privacy-preserving language modeling and on training neural networks on obfuscated data.

Finally, we will discuss how the privacy-preserving machine learning approaches that have been proposed so far would need to be combined in order to achieve perfectly privacy-preserving ML.

PrivateNLP '20, February 7, 2020, Houston, TX, USA

Perfectly Privacy-Preserving AI

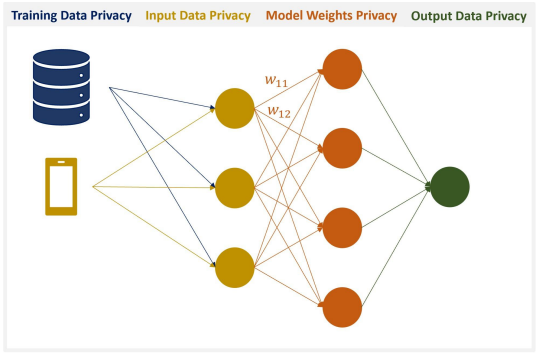
What is it and how do we achieve it?

Patricia Thaine, Gerald Penn

{pthaine,gpenn}@cs.toronto.edu

Computer Science
University of Toronto

Four Pillars Perfectly Privacy-Preserving AI



Consumer and provider privacy are put at risk by:

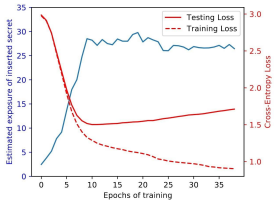
- criminal hackers,
- incompetent employees,
- autocratic governments,
- manipulative companies.

But privacy protection does not have to be about preventing access to sensitive data.

Privacy-preserving methods allow scientists and engineers to use otherwise inaccessible data due to privacy-concerns (e.g., for genomic data analysis (Jagadeesh et al., 2017)).

Data privacy and data utility are positive-sum features of effective ML models.

Training Data Vulnerabilities



Architecture	Layers	Units	Test Loss	Exposure
GRU	1	370	1.18	36
GRU	2	235	1.18	37
LSTM	1	320	1.17	38
LSTM	2	200	1.16	35
CNN	1	436	1.29	24
CNN	2	188	1.28	19
CNN	4	122	1.25	22
WaveNet	2	188	1.24	18
WaveNet	4	122	1.25	20

Left: Results for 5% of PTB. Loss at a minimum after 10 epochs, when estimated exposure peaks. **Right**: Estimated exposure of inserted secret. 620K(+/-5K) parameters / model. (Carlini et al., 2018)

Some Alternative Solutions

- Automatic data de-identification (e.g., El Emam et al. (2009))
- Data synthesis (e.g., Triastcyn et al. (2018))
- Secure Enclaves (Intel SGX, Keystone Project, AMD-SP)

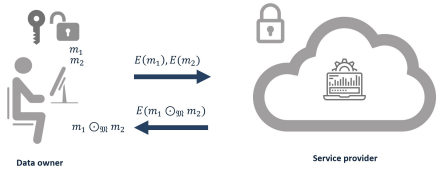
Differential Privacy

Definition. A random algorithm $\mathcal{A}$ is (ϵ, δ)-differentially private if

$$\mathbf{Pr}[\mathcal{A}(D) \in \mathcal{S}] \leq \exp(\epsilon) \times \mathbf{Pr}[\mathcal{A}(D') \in \mathcal{S}] + \delta$$

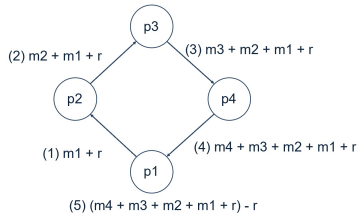
for any set $\mathcal{S}$ of possible outputs of $\mathcal{A}$, and any two data sets D, D' that differ in at most one element. (Carlini et al., 2018)

Homomorphic Encryption



Secure Multi-Party Computation

A very simple (and not truly secure) example:



Federated Learning

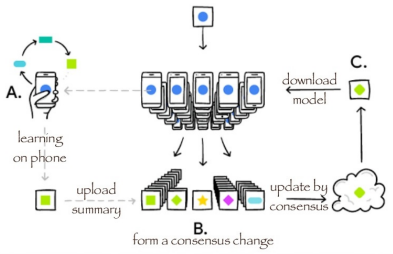


Image source: https://www.slideshare.net/MindosCheng/federated-learning

Creating Perfectly Privacy-Preserving AI

We can achieve perfectly privacy-preserving AI by *combining* different privacy-preserving methods. For example:

- Differential Privacy + Homomorphic Encryption or Secure Enclaves
- Secure Multi-Party Computation + Federated Learning + Differential Privacy + Secure Enclaves or Homomorphic Encryption

Resources

Differential Privacy

- TensorFlow Privacy: `https://github.com/tensorflow/privacy`
- Blog post explaining DP + ML: `http://www.cleverhans.io/privacy/2018/04/29/privacy-and-machine-learning.html`
- "The Promise of Differential Privacy. A Tutorial on Algorithmic Techniques" by Cynthia Dwork (2011)

Homomorphic Encryption

- PALISADE Library: `https://git.njit.edu/palisade/PALISADE`
- Microsoft SEAL Library: `https://github.com/Microsoft/SEAL`
- Intro to HE: "Homomorphic Encryption for Beginners: A Practical Guide"
- "Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy" by Gilad-Bachrach, Ran, et al. (2016)

Federated Learning and Secure Multi-Party Computation

- Code: `https://github.com/OpenMined`
- Florian Hartmann's Blog: `https://florian.github.io/`
- "Practical Secure Aggregation for Privacy Preserving Machine Learning" by Bonawitz et al. (2017)

References

Carlini, Nicholas, et al. "The secret sharer: Measuring unintended neural network memorization extracting secrets." arXiv preprint arXiv:1802.08232 (2018).

El Emam, Khaled, et al. "A globally optimal k-anonymity method for the de-identification of health data." Journal of the American Medical Informatics Association 16.5 (2009): 670-682.

Jagadeesh, Karthik A., et al. "Deriving genomic diagnoses without revealing patient genomes." Science 357.6352 (2017): 692-695.

Thaine, P., Gorbunov, S., Penn, G. (2019). Efficient Evaluation of Activation Functions over Encrypted Data. In Proceedings of the 2nd Deep Learning and Security Workshop, 40th IEEE Symposium on Security and Privacy, San Francisco, USA.

Thaine, P., Penn, G. (2019). Privacy-Preserving Character Language Model. In Proceedings of the Privacy-Enhancing Artificial Intelligence and Language Technologies AAAI Spring Symposium, PAL 2019, Stanford University, Palo Alto, USA.

Triastcyn, Aleksei, and Boi Faltings. "Generating differentially private datasets using gans." arXiv preprint arXiv:1803.03148 (2018).