Towards Scalable, Efficient and Privacy Preserving Machine Learning
Rania Talbi, Sara Bouchenak

To cite this version:
Rania Talbi, Sara Bouchenak. Towards Scalable, Efficient and Privacy Preserving Machine Learning, Middleware ’18 Doctoral Symposium, Dec 2018, Rennes, France. hal-01956155

HAL Id: hal-01956155
https://hal.archives-ouvertes.fr/hal-01956155
Submitted on 14 Dec 2018
Context and Motivation

- Minimize the computational costs incurred by privacy preservation.
- Provide an end-to-end privacy preserving outsourced data classification service.
- Enable a set of mutually untrusted data owners to have a global vision on the union of their data without breaching the privacy of each one of them.
- Enable dynamic data model updates when new training data samples are available.

Related work

Different ML algorithms
- Clustering [1]
- Classification [2]
- Association Rule Mining [3]

Different Privacy-preservation objectives
- ML output protection [2]
- Data protection

Privacy Preservation techniques
- Cryptographic techniques (SMC/HE, GC, OT)
- Non-cryptographic techniques (PP-Data Publishing techniques)

Design principles
- Decent privacy and utility levels
- Efficient runtime
- Entirely outsourced ML computations over encrypted data

Objectives

- Use Naive Bayes classification over encrypted data [4]
- Entirely outsourced machine learning classification over encrypted data [5]
- Naive approach: a combination of low level PP-building blocks
- 1st optimization: use inline building blocks
- 2nd optimization: Parallel computing

Preliminary results

- We have used a synthetic dataset for fraud detection in a B2B network.
- This dataset contains 1000 bank transactions with 9 attributes each.
- We compare our work to the Ciphermed framework [8].

References