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

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Context and Motivation

Objectives

- 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 (1)
- Data protection (2)

Different architectures

- Distributed (4)
- Outsourced (5)

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 on encrypted data

Objective

- Cryptographic based protection (data model, training data, classification queries and responses)
- Partial homomorphic encryption (PHE) based building blocks
- Combine PHE with cryptographic blinding (DTPKC cryptosystem [6])

We implemented the VFDT incremental decision tree learning algorithm [7]

1. (1) Blind inputs
2. (2) Partially decrypt blinded values
3. (3) Decrypt blinded values
4. (4) Run operation over blinded values

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