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>
Towards Scalable, Efficient and Privacy Preserving Machine Learning
Rania Talbi, Sara Bouchenak
INSA Lyon, France
{firstname.lastname}@insa-lyon.fr

Context and Motivation

COVID-19 pandemic has forced many of us to work remotely, making data privacy and security concerns more pressing. In this work, we focus on developing an efficient and privacy-preserving machine learning framework for fraud detection in a B2B network.

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.

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 Ciphered framework [8].

Related work

- Different ML algorithms
  - Clustering
  - Classification
  - Association Rule Mining
- Different Privacy-preserving objectives
  - ML output protection
  - Data protection
- Different architectures
  - Distributed
  - Outsourced

Design principles

- 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]

References