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

2018 ACM/IFIP International Middleware Conference, Doctoral Symposium,
December 10-14th 2018 – Rennes, France

Context and Motivation
M: Company i
A: Fraudulent company
C: Fraudulent company
X: Data Mining for fraud detection

Objective
- 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 Ciphermed framework [8].

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

Different Privacy-preservation objectives
- ML output protection
- Original data protection
- Privacy Preservation techniques

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 VDIT incremental decision tree learning algorithm [7]

References

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON
SIBIL-Lab