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

To cite this version:

HAL Id: hal-01956155
https://hal.science/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

Motivation:
- Central Supervision Authority

Data Mining for fraud detection

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
- Data privacy protection

Privacy Preservation techniques
- Cryptographic techniques (SMC/HE, GC, OT)
- Original data protection
- Distributed [4]
- Outourced [5]

Design principles
- Cryptographic based protection (data model, training data, classification queries and responses)
- Partial homomorphic encryption (PHE) based building blocks
- Combine PHE with cryptographic binding (DTPKC cryptosystem) [6]
- We implemented the VFDT incremental decision tree learning algorithm [7]

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