Towards Scalable, Efficient and Privacy Preserving Machine Learning

Rania Talbi, Sara Bouchenak

To cite this version:

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

\( C_1 \): Company 1
\( B_1 \): Local bank transactions of \( C_1 \)
\( C_2 \): Fraudulent company
\( A \): Central Supervision Authority
\( M \): Data Mining for fraud detection

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
- Data protection
- Utility
- Runtime
- Distributed
- Non-cryptographic techniques

Privacy Preservation techniques

Cryptographic techniques
( SMC/HE, GC, OT)

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]

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].

References

[3] L Liu et al.: Privacy-Preserving Mining of Association Rule on Outsourced Data from Multiple Parties. ACSI2018: 413-415