Towards Scalable, Efficient and Privacy Preserving Machine Learning

Rania Talbi, Sara Bouchenak

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

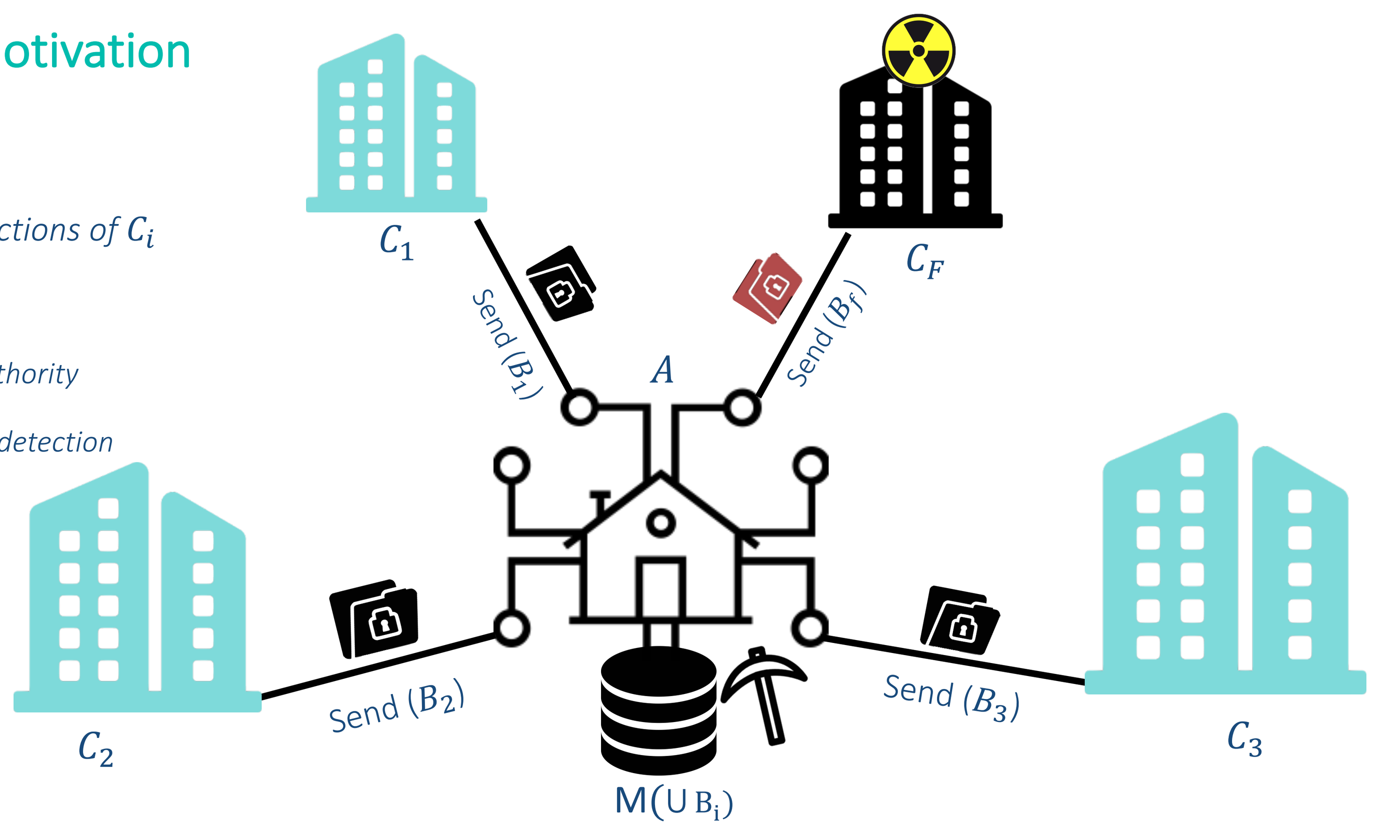
C_i : Company i

B_i : Local bank transactions of C_i

C_F: Fraudulent company

A: Central Supervision Authority

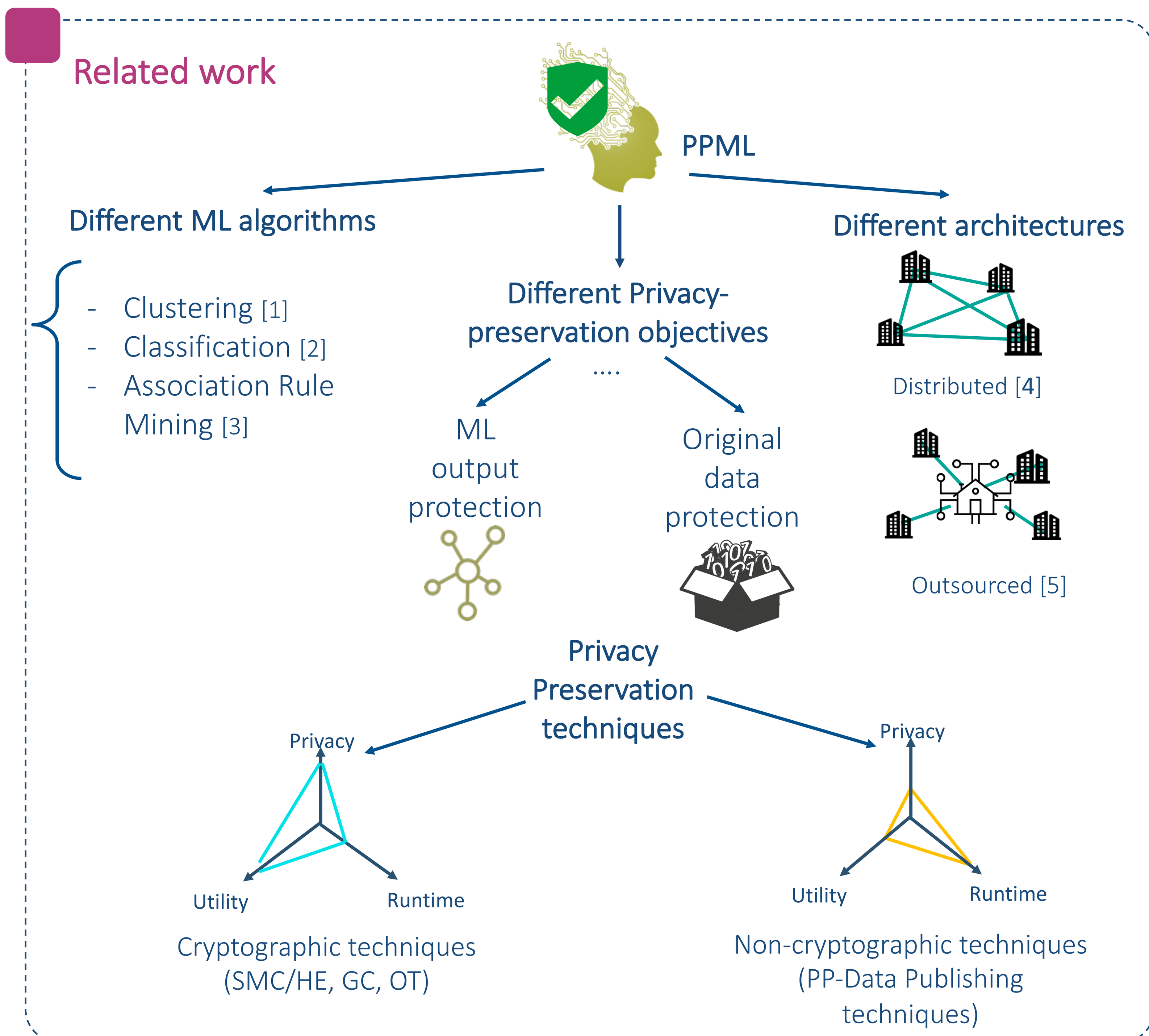
M: Data Mining for fraud detection



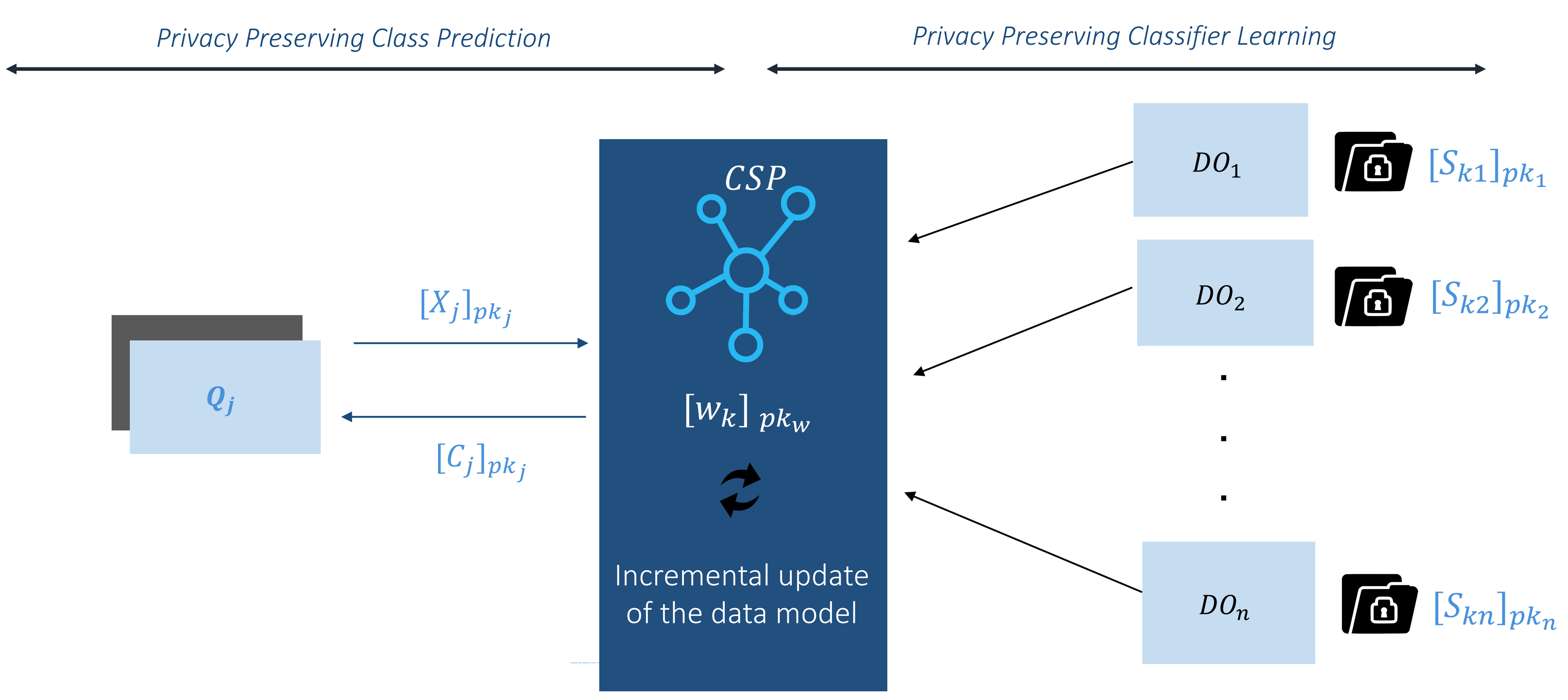
Objectives

- 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



DynAmic Privacy Preserving machine Learning Framework (DAPPLE)



DO_i : Data Owner i

Q_j : Classification Qeurier j

CSP: *Classification Service Provider*

$[w_k]_{pk_w}$: Encrypted data model

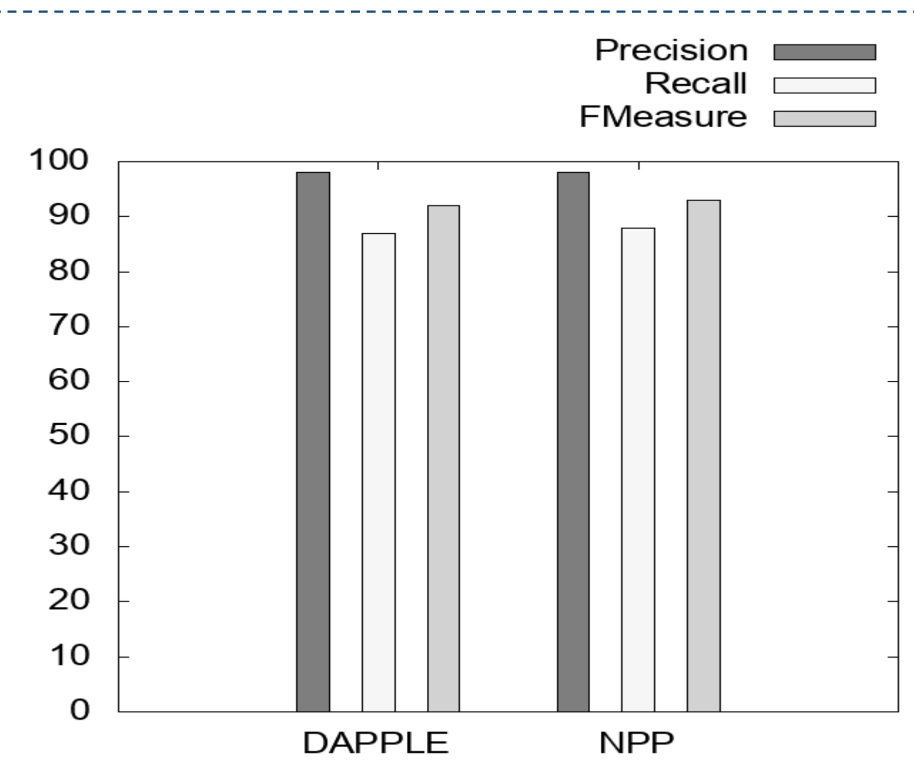
$[X_j]_{pk_j}$: Encrypted classification query

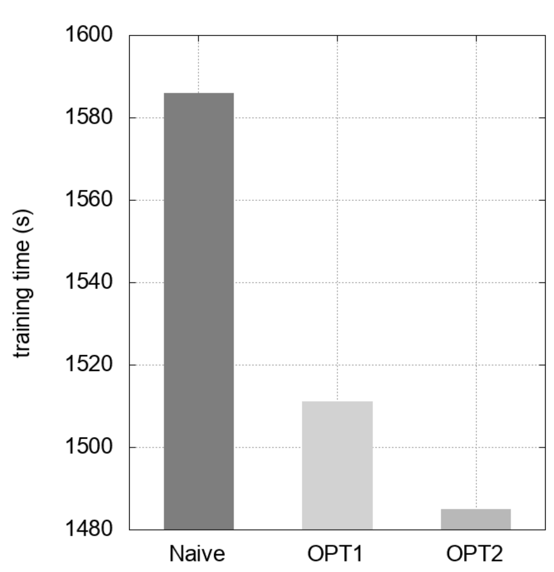
$[C_j]_{pk_j}$: Encrypted classification response

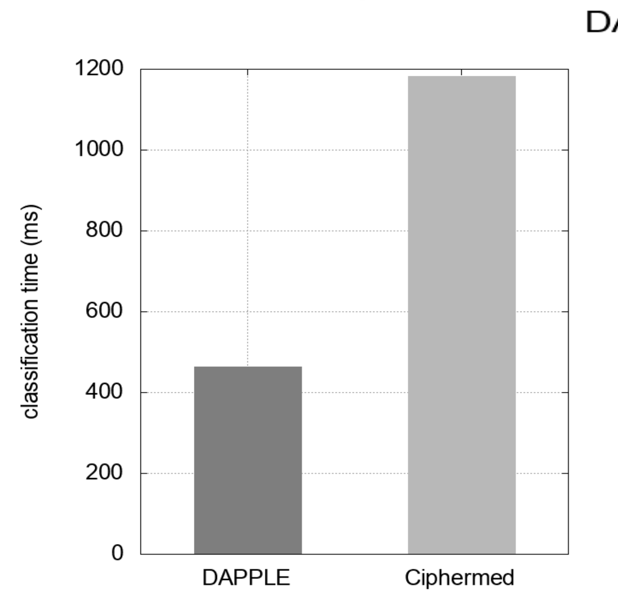
$[S_{ki}]_{pk_i}$: Encrypted local training data chunk from data owner DO_i

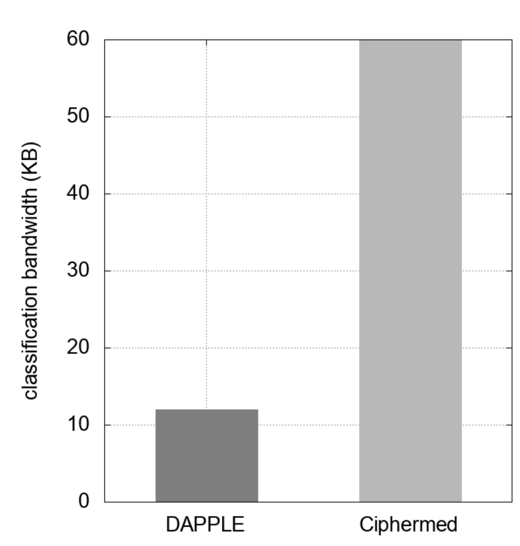
Preliminary results

- 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].



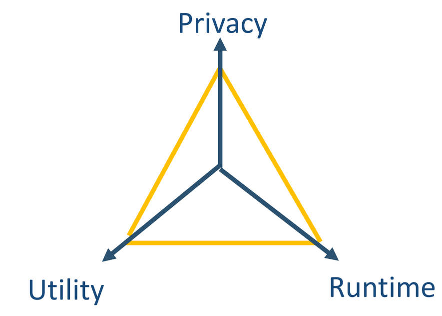






Design principles

- 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]



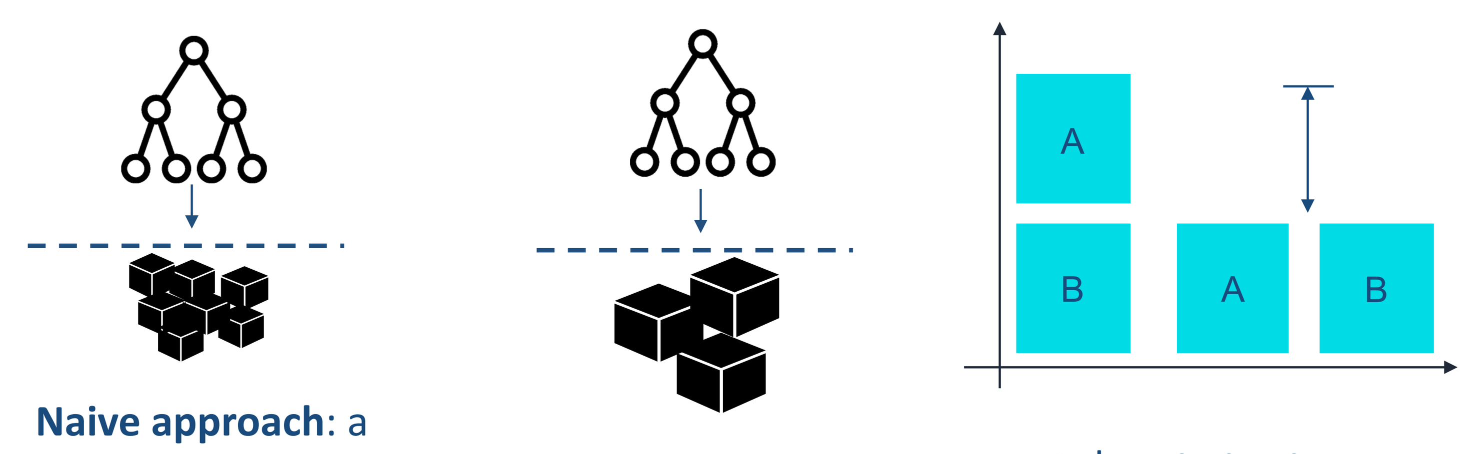
- Decent privacy and utility levels
- Efficient runtime
- Entirely outsourced ML computations over encrypted data

- **(1)** Blind inputs
- **(2)** Partially decrypt blinded values

$ex:\ [x]_{pk} \otimes [r]_{pk} = [x \oplus r]_{pk}$

- **(4)** remove blinding from the result

- **(3)** Decrypt blinded values
- **(4)** Run operation over blinded values



Naive approach: a combination of low level PP-building blocks

1st optimization : use inline building blocks

2nd optimization : Parallel computing

References

- [1] X. Hu, et. al: Privacy-Preserving K-Means Clustering Upon Negative Databases. ICONIP (4) 2018.
- [2] S. Kim et al. Privacy-Preserving Naive Bayes Classification Using Fully Homomorphic Encryption. ICONIP (4)2018: 349-358
- [3] L.Liu et al : Privacy-Preserving Mining of Association Rule on Outsourced Cloud Data from Multiple Parties. ACISP2018: 431-451
- [4] H.Yu et al.: Privacy-Preserving SVM Classification on Vertically Partitioned Data. PAKDD 2006: 647-656
- [5] T.Li et al. : Outsourced privacy-preserving classification service over encrypted data. J. Network and Computer Applications 106: 100-110 (2018)
- [6] X.Liu et al. : An Efficient Privacy-Preserving Outsourced Calculation Toolkit With Multiple Keys. IEEE Trans Information Forensics and Security 11(11): 2401-2414 (2016)
- [7] M. Domingos et al.: Mining high-speed data streams. KDD 2000: 71-80
- [8] R.Bost et al. : Machine Learning Classification over Encrypted Data. NDSS 2015

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

SIBIL-Lab

LIRIS