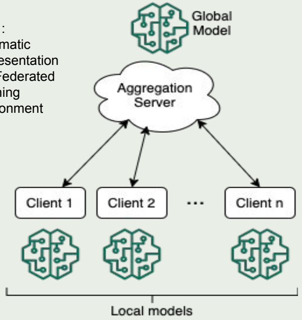
A SURVEY OF TWO OPEN PROBLEMS OF PRIVACY PRESERVING FEDERATED LEARNING: VERTICALLY PARTITIONED DATA AND VERIFIABILITY

1. Federated Learning

Fig. 1: Schematic Representation of a Federated Learning Environment



2. Preliminaries

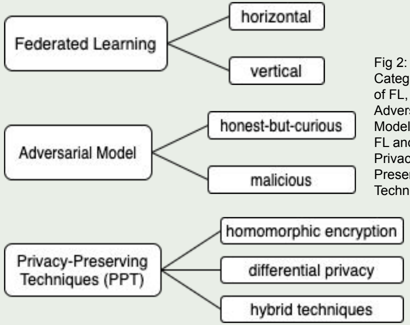


Fig 2: Categories of FL, Adversarial Models in FL and Privacy Preserving Techniques

3. Research Questions

I. What are the privacy preserving schemes available for vertical federated learning, and how do they compare?

II. What are the privacy preserving schemes providing aggregation verifiability, and how do they compare?

4. Methodology

Individual analysis:

1. workflow
2. computational complexity
3. communication overhead
4. accuracy impact
5. security guarantees

Comparative analysis:

1. benefits and downsides
2. potential improvements

5. Results

Framework	FedV	SecureBoost	MP-FEDXGB
Computational Complexity at Aggregator	$O(en(ks+f))$	$O(nN+2^dTn)$	$O((n+fs)T2^d)$
Computational Complexity at Client	$O(e(ks+f))$	$O(N+2^dT)$	$O((n+fs)T2^d)$
Communication Overhead of Aggregator	$O(en)$	$O(nN+n2^dT)$	$O(nT2^d)$
Communication Overhead of Client	$O(e)$	$O(N+2^dT)$	$O(nT2^d)$
Accuracy Impact	lossless	lossless	lossless
Security Model	honest-but-curious aggregator, malicious and colluding users	honest-but-curious aggregator and clients	honest-but-curious colluding auxiliary parties, honest active party
Machine Learning Model	linear and non-linear models supporting updates through SGD	Classification and Regression Trees	Classification and Regression Trees
Security Mechanisms	Inference Prevention Module, Batch Randomization, FEIP encryption (MIFE+SIFE)	Paillier Encryption	Secret-Sharing, First-Layer-Mask

Fig. 3: Comparison between Vertical Federated Learning Privacy Preserving Techniques

Fig. 4: Comparison between Federated Learning Privacy Preserving Techniques providing Aggregation Verifiability

Framework	VFL	VerifyNet	Secure Verifiability
Computational Complexity at Aggregator	$O(egn)$	$O(emn^2)$	$O(enD)$
Computational Complexity at Client	$O(em^2g)$	$O(emn)$	$O(eD)$
Communication Overhead at Aggregator	$O(egn)$	$O(enm)$	$O(enD)$
Communication Overhead at Client	$O(eg)$	$O(e(n+m))$	$O(eD)$
Accuracy Impact	94% on MNIST (95% for [38])	not assessed	97% on MNIST (98% for [38])
Security Model	malicious server, honest-but-curious clients, colluding up to $n-2$	honest-but-curious clients (colluding up to $t-1$) and server (malicious aptness)	honest-but-curious clients and a malicious server
Machine Learning Model	neural networks	neural networks	neural networks
Security Mechanisms	Lagrange Interpolation, Pseudo-random technology, Chinese Remainder Theorem	Diffie-Hellman, Homomorphic Hash Functions, Double-Masking, Secret-Sharing, Pseudo-random technology	Paillier encryption, Chinese Remainder Theorem, Bilinear aggregate signature

[38] H. Brendan McMahan et al. "Federated Learning of Deep Networks using Model Averaging".In:CoRRabs/1602.05629 (2016). arXiv: 1602.05629.URL: http://arxiv.org/abs/1602.05629.

6. Conclusion

Takeaways:

- complementing classical PPT improves data privacy
- there is no universal best PPT within FL

Future direction of research:

- integration of verifiability within vertical FL
- security enhancements via alternative methods (e.g. DP)

student: Horia-Claudiu Culea (4779126)
email: h.c.culea@student.tudelft.nl
responsible professor: Dr. Kaitai Liang
supervisor: Rui Wang

TU Delft