Improving Federated Learning Protection with Digital Envelopes

Mario Dib[1], Pedro Prates[2,3] and Bernardete Ribeiro[4]

1 - Univ Coimbra, CISUC, Coimbra, Portugal
2 - Univ Coimbra, CEMMPRE, Department of Mechanical Engineering, Coimbra, Portugal
3 - Department of Mechanical Engineering, Centre for Mechanical Technology and Automation (TEMA), University of Aveiro, Aveiro, Portugal
4 - Univ Coimbra, CISUC, Department of Informatics Engineering, Coimbra, Portugal

Introduction

In the field of machine learning, Federated Learning (FL) brings a new concept that includes a privacy-preserving approach regarding the datasets used in the model's training, allowing multiple participants to collaborate with their data, without giving up their private information. This allows to solve common issues that would be more challenging if done alone, opening a new way to approach issues' solutions in the industrial field [1]. However, there are some concerns regarding data protection, since the FL approach is susceptible to cyber-attacks, that includes model poisoning [2], which is the one analyzed in this work. These attacks can compromise the models' results. So, this work proposed an approach to handle the data poisoning attack before the machine learning training, in order to prevent the models to be compromised, by combining the federated learning approach with the digital envelopes (DE) [3]. That way, its possible to verify the authenticity of the client trying to participate in the training phase and the integrity of the datasets, rather than looking for malicious patterns in the data.

Objective

To propose a method combining Federated Learning and Digital Envelopes that can identify possible malicious data before the machine learning training procedure, in order to prevent its negative influence on results.

Why?

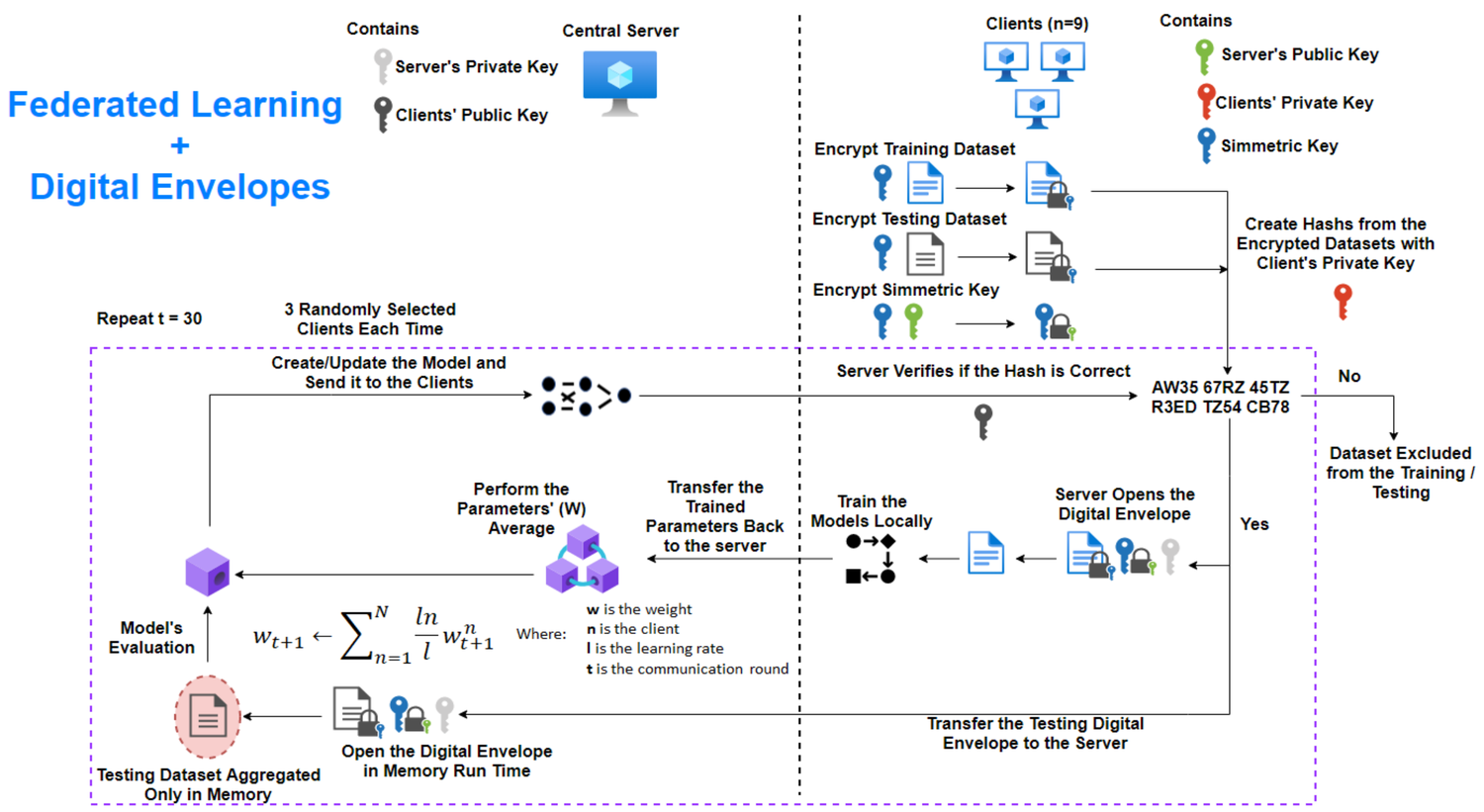
Federated Learning

- ✚ Allows Collaboration
- ✚ Provides Data Privacy
- ✚ Distributed Training
- ▬ **Susceptible to Cyber-attacks**
- ▬ **Less Accuracy**

Digital Envelopes

- ✚ Data Integrity
- ✚ Data Authenticity
- ✚ Data Confidentiality
- ✚ Prevention Instead of Correction

How?

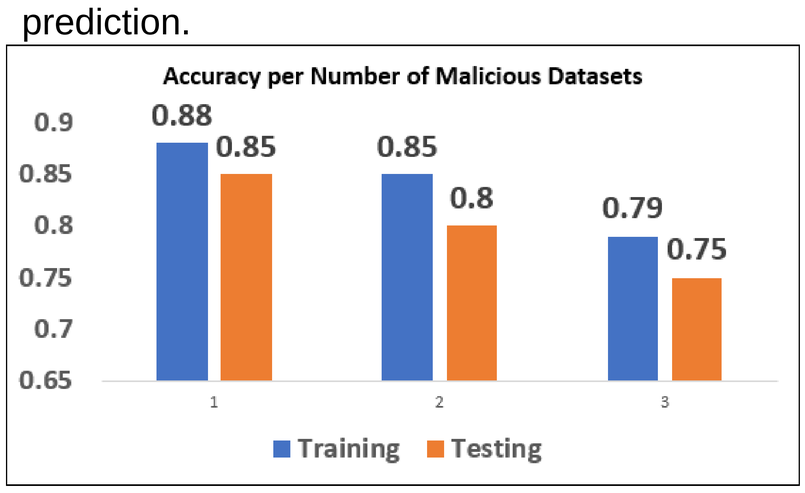


Use Case

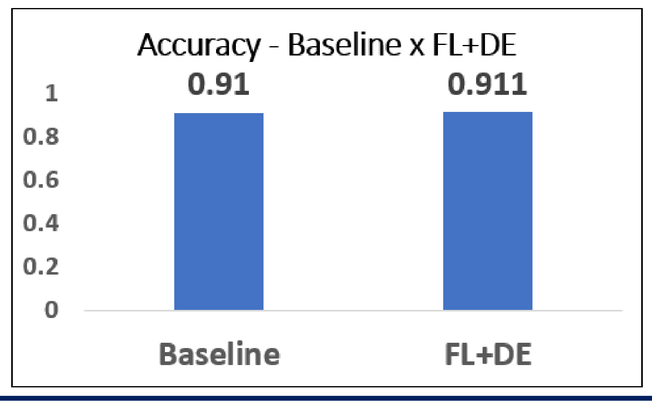
The selected use case was based on sheet metal forming processes, which are widely used in the automotive manufacturing sector to produce parts with complex geometries at high cadences. In this sector, process productivity is often impaired by sources of variability, which may lead to forming defects and subsequent high scrap rates. Such challenges can be tackled in a collaborative manner with FL-based decision support tools, despite the strong data privacy policies in the sector.

Results

The more malicious datasets participating in communication rounds, the worse is the model‘s performance regarding defect prediction.



When the proposed approach is used, the result was virtually the same as the baseline result, correctly removing the malicious datasets from the training procedure



Conclusion

- The proposed approach was able to correctly identify compromised datasets and excluded them from the training and the result achieved indeed was on par with the baseline model.
- This approach would be suitable to help handling other data attacks, since it verifies if the data is corrupted, instead of trying to correct the data for a specific threat.
- The proposed approach improves authenticity, confidentiality and integrity of the used datasets.
- Although promising, there are more work to be done since this experiment is quite restrictive and may be challenging in a real environment.
- More work will be done in order to include protection to the models.

References

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas. Communication-efficient learning of deep networks from decentralized data. arXiv, 2016.

[2] G. Sun, Y. Cong, J. Dong, Q. Wang, and J. Liu. Data poisoning attacks on federated machine learning. arxiv, 2020.

[3] S. Perez, J. L. Hernandez-Ramos, D. Pedone, D. Rotondiand, L. Straniero, and A. F. Skarmeta. A digital envelope approach usingattribute-based encryption for secure data exchange in iot scenarios. Global Internet of Things Summit (GIoTS), pages 1–6, 2018.

Acknowledgements

Research funded by FEDER and by FCT under the projects UIDB/00285/2020, UIDB/00326/2020, UIDB/00481/2020 and UIDP/00481/2020 and co-funded by POCI under the projects PTDC/EME-EME/31243/2017 (RDFORMING) and PTDC/EME-EME/31216/2017 (EZ-SHEET).

Cofinanciado por: