Cyprus University of Technology

Faculty of Engineering and Technology

Master's Thesis

Towards privacy-preserving cybersafety tools by using Federated Learning

Pantelitsa Leonidou

Limassol, August 2021

Cyprus University of Technology

Faculty of Engineering and Technology

Department of Electrical Engineering
Computer Engineering and Informatics

Towards privacy-preserving cybersafety tools by using Federated Learning

Pantelitsa Leonidou

Limassol, August 2021

Approval Form

Cyprus University of Technology

Towards privacy-preserving cybersafety tools by using Federated Learning

Presented by

Pantelitsa Leonidou

Advisor ______________________________

Dr. Michael Sirivianos

Committee member ______________________________

Dr. Fragkiskos Papadopoulos

Committee member ______________________________

Dr. Sotirios Xatzis

Cyprus University of Technology

Limassol, August 2021

Copyrights

The approval of the thesis by the Department of Electrical Engineering, Computer Engineering and Information does not imply necessarily the approval by the Department of the views of the writer

Acknowledgements

I would like to express my gratitude to those who contributed to this work by sharing their valuable knowledge and ideas. More specifically, I want to thank Dr. Michael Sirivanos for supervising and guiding me through the whole thesis process, from finding a topic to completing it. I want to thank Nicolas Kourtellis for introducing the Federated Learning technology to me and for all the great ideas on how to apply it in the context of cybersafety tools. It would be my omission not to thank Kostantinos Papadamou for his knowledge sharing and the technical support he provided. Lastly, I owe a lot to my family and friends who stood by me and encourage me through the whole thesis process.

Abstract

Living in the digital era, people can access a huge amount of online content daily. Online services might benefit humanity by easing many everyday life tasks. However, people and especially minor users can encounter many threats while they are online. Despite various cybersafety tools and applications, the number of minors experiencing online threats is not decreasing. This work focuses on cybersafety tools that use Machine Learning algorithms for automatic detection of inappropriate content. Such tools require the collection of big of data that are often sensitive. Additionally, keeping these datasets up-to-date and retraining the models can be challenging. We propose using Federated Learning (FL) training to overcome these challenges. FL allows training a model on distributed data without transferring data to a central unit. We provide a conceptual mapping between the components of a cybersafety framework architecture and the actors in the FL communication protocol to explain how FL can be applied in the context of cybersafety tools. We design and implement a TensorFlow-Federated simulation to explore FL training on a text classification model that detects aggressive text. We experimented with a centralized dataset of aggressive tweet posts to assess the performance of the model trained in the FL approach compared to a model trained in the centralized approach and explore how the number of clients participating in FL affects the model's performance. Additionally, we experimented with a local model training to assess the client device's CPU utilization, memory consumption, and execution time. The results show that the model's performance when trained in FL settings can approach the model's performance when trained in the traditional approach. Also, the model's performance improves when more clients participate in the FL training. Regarding the performance of a client's device, the results show that the execution time for a local model's training is short and does not over-consume the device resources. The findings show that cybersafety tools are an applicable use case for FL training.