STEVENS
INSTITUTE OF TECHNOLOGY
1870

Ph.D. DISSERTATION DEFENSE

Candidate: Mengjiao Zhang
Degree: Doctor of Philosophy
School/Department: Charles V. Schaefer, Jr. School of Engineering and Science / Computer Science
Date: Tuesday, September 19th, 2023
Time/Location: 10:00 am EST / https://stevens.zoom.us/my/xujia
Title: Privacy in Federated Learning

Chairperson:

Dr. Jia Xu, Department of Computer Science, Stevens Institute of Technology

Committee Members:

Dr. Shusen Wang, Xiaohongshu, Xingyin Information Technology Shanghai Ltd.

Dr. Yi Guo, Department of Electrical and Computer Engineering, Stevens Institute of Technology

Dr. Jie Shen, Department of Computer Science, Stevens Institute of Technology

Dr. Shucheng Yu, Department of Electrical and Computer Engineering, Stevens Institute of Technology

ABSTRACT

The rise of Artificial Intelligence technology has raised concerns about the potential compromise of privacy due to the handling of personal data. Private AI prevents cybercrimes and falsehoods and protects human freedom and trust. While Federated Learning offers a solution by model training across decentralized devices or servers, thereby preserving data localization, it may leak client information through communicated gradients and parameters. Conventional defenses, such as dropout, GAN, and adversarial training, fail to either obstruct these attacks or significantly hamper model performance.

This thesis centers on defending against gradient-based attacks in Federated Learning while upholding model efficiency and performance. Our first contribution introduces the pragmatic defense mechanism of Double-Blind Collaborative Learning (DBCL), which employs random matrix sketching on parameters and repeated sketching generation, achieving enhanced privacy without substantial computational overhead or lowering accuracy. Our primary investigation delves into byte coding for privacy in Natural Language Processing (NLP). This novel approach involves random-byte mapping with a subword fusion strategy, yielding promising experimental outcomes characterized by fortified privacy, memory efficiency, and accuracy. Notably, our approach obstructs an attacker's ability to reconstruct text token candidates for a batch of input, thus fortifying the resilience of private text in Federated Learning against potential recovery attempts, making the recovery of private data in federated learning much harder - paving a way to a safer environment in both the real and virtual worlds.