Karlsruhe Institute of Technology

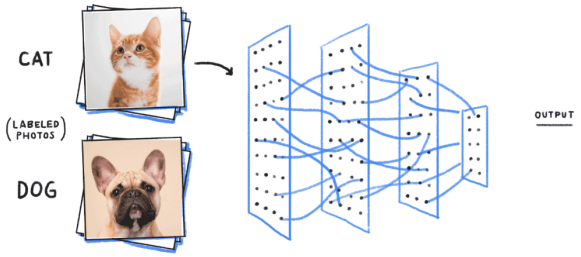
KASTEL

Institute of Information Security and Dependability
Chair of Privacy and Security
Prof. Thorsten Strufe
Àlex Miranda-Pascual
Ana-Maria Crețu
https://ps.kastel.kit.edu

Master's Thesis

How Data Sampling Affects Differentially Private Stochastic Gradient Descent

Machine learning (ML) models are increasingly used to automate decision making in companies, public services, and the healthcare domain. Yet, researchers have shown that adversaries having access to a model can infer private information about individuals in the training dataset. The most successful approach to prevent these attacks is to train models with **differential privacy (DP)** guarantees.



Differential privacy ensures that the output of the training algorithm does not change too much regardless of the inclusion or exclusion of any person's data, guaranteeing strong privacy. In ML, DP is implemented by modifying the standard training algorithm, stochastic gradient descent (SGD), into an algorithm called DP-SGD [1]. A key step in DP-SGD is the sampling of mini-batches. Previous research showed that **sampling** [2] **is detrimental to utility in non-ML algorithms**, meaning that these algorithms achieve better utility for the same privacy level when sampling is not applied at all. **These results raise the question of whether sampling is detrimental in DP-SGD as well, and if it reduces the inference accuracy of the models.**

The goal of this project is to empirically analyze the DP-SGD algorithm in order to understand the impact of sampling on accuracy. You will empirically evaluate the impact by implementing neural networks in Pytorch and training them using Opacus, an open-source implementation of DP-SGD [3]. If your results provide evidence that sampling does benefit accuracy, you will investigate potential reasons why this is the case, building on previous work on non-ML algorithms. If you find that sampling does not benefit accuracy, you will attempt to design an algorithm that provides DP guarantees without sampling, with improved performance.

Requirements:

- Motivation in privacy, machine learning, and attacks.
- Basic knowledge of probabilities and machine learning (e.g., gradient descent).
- Coding skills in PyTorch and some knowledge of neural networks.
- English communication skills.

If you are interested in this topic or have further questions, please contact Àlex Miranda-Pascual (alex.pascual@kit.edu) and Ana-Maria Crețu (ana-maria.cretu@epfl.ch). *16th July 2024*

[1] M. Abadi et al., "Deep learning with differential privacy", in CCS '16.
[2] B. Balle et al., "Privacy amplification by subsampling: tight analyses via couplings and divergences", in NIPS'18.
[3] A. Yousefpour et al., "Opacus: User-friendly differential privacy library in PyTorch", arXiv preprint (2021)