# POSTER: The Price of Privacy in Collaborative Learning

Balázs Pejó
SnT, University of Luxembourg
Esch-sur-Alzette, Luxemburg
balazs.pejo@uni.lu

Qiang Tang
Luxembourg Institute of Science and Technology
Esch-sur-Alzette, Luxemburg
qiang.tang@list.lu

Gergely Biczók
CrySyS Lab, Dept. of Networked Systems and Services, BME
Budapest, Hungary
biczok@crysys.hu

## ABSTRACT

Machine learning algorithms have reached mainstream status and are widely deployed in many applications. The accuracy of such algorithms depends significantly on the size of the underlying training dataset; in reality a small or medium sized organization often does not have enough data to train a reasonably accurate model. For such organizations, a realistic solution is to train machine learning models based on a joint dataset (which is a union of the individual ones). Unfortunately, privacy concerns prevent them from straightforwardly doing so. While a number of privacy-preserving solutions exist for collaborating organizations to securely aggregate the parameters in the process of training the models, we are not aware of any work that provides a rational framework for the participants to precisely balance the privacy loss and accuracy gain in their collaboration.

In this paper, we model the collaborative training process as a two-player game where each player aims to achieve higher accuracy while preserving the privacy of its own dataset. We introduce the notion of *Price of Privacy*, a novel approach for measuring the impact of privacy protection on the accuracy in the proposed framework. Furthermore, we develop a game-theoretical model for different player types, and then either find or prove the existence of a Nash Equilibrium with regard to the strength of privacy protection for each player.

## CCS CONCEPTS

• **Security and privacy** → *Data anonymization and sanitization*;

## KEYWORDS

Privacy; Game Theory; Machine Learning; Recommendation Systems

## 1 INTRODUCTION

As data has become more valuable than oil, everybody wants a slice of it; Internet giants (e.g., Netflix, Amazon, etc.) and small businesses alike would like to extract as much value from it as possible. Machine Learning (ML) has received much attention over the last decade, mostly due to its vast application range. For machine learning tasks, it is widely known that more training data will lead to a more accurate model. Unfortunately, most organizations do not possess a dataset as large as Netflix's or Amazon's. In such a situation, to obtain a relatively accurate model, a natural solution would be to aggregate all the data from different organizations on a centralized server and train on the aggregated dataset. This approach is efficient, however, data owners have a valid privacy concern about sharing their data, particularly with new privacy regulations such as the European General Data Protection Regulation (GDPR). Therefore, in real-life scenarios, improving ML via straightforward data aggregation is likely undesirable and potentially unlawful.

In this paper, we are interested in a scenario with two participants, each of whom possesses a significant amount of data and would like to obtain a more accurate model than what they would obtain if training was carried out in isolation. It is clear that the players will only be interested in collaboration if they can actually benefit from each other. To this end, we assume that the players have already evaluated the quality of each other's datasets to make sure training together is beneficial for both of them before the collaboration.

### 1.1 Problem Statement

In the literature, Privacy Preserving Distributed Machine Learning [2, 5, 6] have been proposed to mitigate the above privacy concern by training the model locally, and then aggregating all the local updates securely. However, these approaches' efficiency depend on the number of participants and the sample sizes. Moreover, in most of them, the players are not provided with the option of choosing their own privacy parameters.

To bridge this gap, we consider the parties involved as rational players and model their collaboration as a game similarly to [1, 4], but focusing on the two-player case. In our setting, players have their own trade-offs with respect to their privacy and expected utility and can flexibly set their own privacy parameters. Hence, the central problem of this research is to propose a general game theoretical model, and to find a Nash Equilibrium. Our goal is to answer the following fundamental questions given a specific ML task:

- What are the potential ranges for privacy parameters that make the collaborative ML model more accurate than training alone?
- What is the optimal privacy parameter (which results in the highest payoff)?
- With this optimal parameter, how much accuracy is lost overall due to the applied privacy-preserving mechanisms?

## 2 GAME THEORETIC MODEL

The Collaborative Learning (CoL) game captures the actions of two privacy-aware data holders in the scenario of applying an arbitrary privacy preserving mechanism and training algorithm on their datasets. At a high level, the players' goal in the CoL game is to maximize their utility, which is a function of the model accuracy and the privacy loss:


*CCS '18, October 15–19, 2018, Toronto, ON, Canada*

ACM ISBN 978-1-4503-5693-0/18/10.
https://doi.org/10.1145/3243734.3278525

Definition (Collaborative Learning game). *The CoL game is a tuple $\langle \mathcal{N}, \Sigma, \mathcal{U} \rangle$, where the set of players is $\mathcal{N} = \{1, 2\}$, their actions are $\Sigma = \{p_1, p_2\}$ where $p_1, p_2 \in [0, 1]$ while their utility functions are $\mathcal{U} = \{u_1, u_2\}$ such that for $n \in \mathcal{N}$:*

$$u_n(p_1, p_2) = B_n \cdot b(\theta_n, \Phi_n(p_1, p_2)) - C_n \cdot c(p_n)$$

The variables of the CoL game are listed in Tab. 1. Accuracy is measured as the prediction error of the trained model. Maximal privacy protection is represented via $p_n = 1$, while $p_n = 0$ means no protection for player $n$. As the benefit and the privacy loss are not on the same scale, we introduce weight parameters $B_n > 0$ and $C_n \geq 0$.

| Variable | Meaning |
|---|---|
| $p_n$ | Privacy parameter for player $n$ |
| $C_n$ | Privacy weight for player $n$ |
| $B_n$ | Accuracy weight for player $n$ |
| $\theta_n$ | Accuracy by training alone for player $n$ |
| $\Phi_n(p_1, p_2)$ | Accuracy by training together for player $n$ |
| $b(\theta_n, \Phi_n)$ | Benefit function for player $n$ |
| $c(p_n)$ | Privacy loss function for player $n$ |

**Table 1: Parameters of the CoL game.**

## 2.1 Assumptions

We only rely on a couple of basic assumptions to keep the model as general as possible. We assume

- that the privacy loss function is monotone, and the maximal potential privacy leakage is 1 which corresponds to no protection at all, while maximal privacy protection corresponds to zero privacy loss.
- that the benefit function is monotone, and there is no benefit if the accuracy of training together is lower than the accuracy of training alone.
- that the function of the accuracy by training together is monotone, and maximal privacy protection cannot result in higher accuracy than training alone, while training together with no privacy corresponds to higher accuracy than training alone.

Based on these assumptions, no collaboration is a trivial NE of the CoL game.

## 2.2 Price of Privacy

By definition, in a Nash Equilibrium (NE) no player can do better by changing strategies unilaterally [3]. However, NE is not necessarily a social optimum. Price of Stability measures the ratio between these two: how the efficiency of a system degrades due to the selfish behavior of its players. Inspired by this, we define *Price of Privacy* to measure the accuracy loss due to the privacy mechanism: $PoP \in [0, 1]$ where 0 corresponds to the highest possible accuracy which can be achieved via collaboration with no privacy; on the contrary, 1 corresponds to the lowest possible accuracy which can be achieved by training alone.

## 2.3 Remarks

$\Phi_n$ plays a crucial role in the CoL game. However, since it is determined by the privacy mechanism, the complex training algorithm and the used datasets, the exact form in general is unknown. Given that the actual value of $\Phi_n$ is required to compute the optimal strategies, it has to be numerically evaluated for putting the CoL game to practical use. Computing this function precisely requires access to the joint dataset; thus, it raises the very privacy concern which we want to mitigate in the first place.

To break this loop, we propose an approach called Self-Division in [7]. Via this approximation, the players determine $\widetilde{\Phi_n}$, which can be used with the CoL game to find the optimal privacy parameter $p_n^*$. The actual collaboration takes place only if these parameters correspond to positive utilities for both players. The process diagram summarizing the steps above can be seen in Fig. 1; the rest of the parameters are determined via an external setup mechanism.

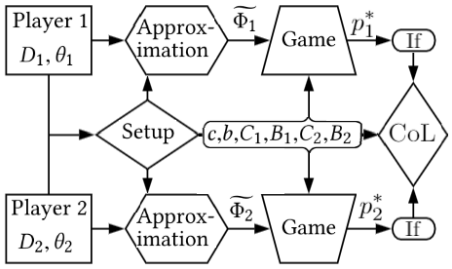


**Figure 1: Process Diagram for Collaborative Learning.**

# 3 THEORETICAL RESULTS

Based on the properties of CoL game, two natural expectations arise:

Lemma. *$\exists \alpha_n \geq 0$ such that if $\frac{C_n}{B_n} \leq \alpha_n$ for player n than its best response is to train together without any privacy protection. Similarly, $\exists \beta_n \geq 0$ such that if $\frac{C_n}{B_n} \geq \beta_n$ for player n then its best response is to train alone.*

The questions we are interested in answering are: *what are the exact values of $\alpha_n$ and $\beta_n$* and *what is the NE in case $\alpha_n \leq \frac{C_n}{B_n} \leq \beta_n$*.

We separate our theoretical analysis into two case based on the privacy weight parameter of the players:

- **Unconcerned**: This type of player cares only about accuracy, i.e., $C_n = 0$.
- **Concerned**: This player is more privacy-aware, as the privacy loss is present in its utility function (i.e., $C_n > 0$).

## 3.1 One Player is Privacy Concerned

We derive symbolic NEs in closed form for the case where exactly one of the players is privacy-concerned.

Theorem (Training as an unconcerned player). *If player n is unconcerned ($C_n = 0$) then the NE is to collaborate without any privacy protection: $p_n^* = 0$.*

When both players are unconcerned ($C_1 = C_2 = 0$), $(p_1^*, p_2^*) = (0, 0)$ is a NE. As a result, the unconcerned player does not apply any privacy-preserving mechanism. Without loss of generality we assume Player 2 is unconcerned, so its best response is $p_2 = 0$. This allows us to make the following simplifications: $\Phi(p_1) := \Phi_1(p_1, 0)$, $b(p_1) := b(\theta_1, \Phi(p_1, 0))$ and $u(p_1) := u_1(p_1, 0)$ while $f' = \partial_{p_1} f$ and $f'' = \partial^2_{p_1} f$.

Theorem (Training with an unconcerned player). *A NE of the CoL game when Player 1 is concerned ($C_1 > 0$) while Player 2 is unconcerned ($C_2 = 0$) is $(p_1^*, p_2^*) = (\rho, 0)$ where*

$$\rho = \begin{cases} & & u''(\rho) < 0 \\ \left[\frac{b'\Phi'}{c'}\right]^{-1}(r) & if & \rho \in [0,1] \\ & & u(\rho) > 0 \\ 0 & if & b(0) > r \\ 1 & & otherwise \end{cases}$$

*where* $[\cdot]^{-1}$ *is the inverse function of* $[\cdot]$ *and* $r = \frac{C_1}{B_1}$.

## 3.2 Both Players are Privacy Concerned

We prove the existence of a pure strategy NE in the general case, where both players are privacy-concerned to a given degree.

Theorem (Main). *The CoL game has at least one non-trivial pure-strategy NE if*

$$\partial_\Phi^2 b \cdot (\partial_{p_1}\Phi_1 - \partial_{p_2}\Phi_2) = \partial_\Phi b \cdot (\partial_{p_1}\partial_{p_2}\Phi_2 - \partial_{p_1}\partial_{p_2}\Phi_1)$$

Corollary. *If we assume* $\partial^i_{p_1}\Phi_1 = \partial^i_{p_2}\Phi_2$ *for* $i \in \{1,2\}$ *then the Main theorem holds.*

The condition on the derivatives of $\Phi_n$ in the Corollary means that the player's accuracy changes the same way in relation to their own privacy parameter, independently from the other player's privacy parameter. We demonstrate this empirically using a recommendation system use case.

# 4 USE CASE: RECOMMENDATION SYSTEM

In our example recommender use case, we specify the training algorithm to be matrix factorization via stochastic gradient descent; moreover, we use the Movilens (1M) and Netflix (NF10) dataset. As privacy-preserving mechanisms we consider both suppression (*Sup*) and differential privacy (*DP*). For details, please consult [7].

## 4.1 Empirical Results

*4.1.1 Alone vs. together.* In Fig. 2 we compare the achieved accuracy with and without the other player's data. Training together is superior to training alone for both datasets and all size ratios. Moreover, the owner of the smaller dataset benefits more from collaboration.

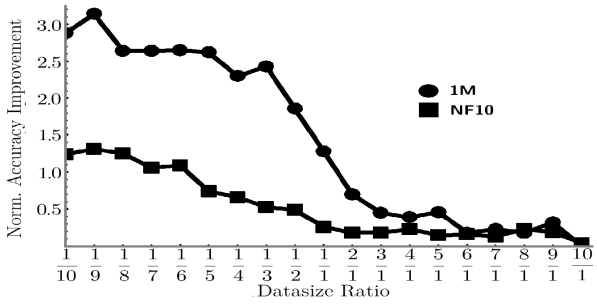


**Figure 2: Accuracy improvement achieved by training together instead of alone using 1M/NF10 datasets where one player has $x$ times more (less) data than the other.**

*4.1.2 Training with an unconcerned.* Fig. 3 represents the case when Player 2 is unconcerned (i.e., $p_2 = 0$). Player 1's options are either to set $p_1$ for *Sup* (the percentage of discarded data) or $\varepsilon_1$ for *DP* (the level of the injected noise). The main observation is that as the dataset size ratio increases, the accuracy improvement decreases, independently from the dataset ratios. Moreover, it is visible that the dataset size only effects the accuracy through a constant factor, i.e., the Corollary which ensures the existence of a non-trivial pure strategic NE does hold.

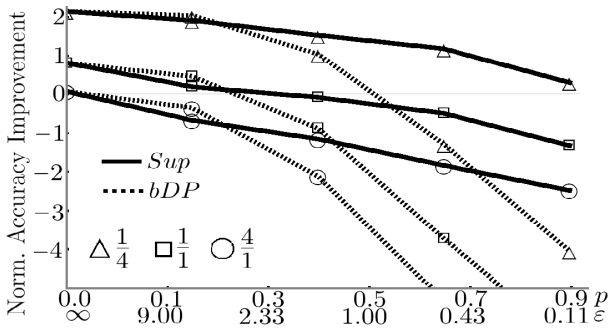


**Figure 3: Accuracy improvements of training together for different privacy levels using the 1M dataset divided with data size ratios 0.25, 1 and 4.**

*4.1.3 Training with a concerned.* The experiment corresponding to the case when both players can apply privacy preservation can be found in [7].

# 5 CONCLUSION

In this paper we designed a Collaborative Learning process among two players. We defined two player types and modeled the training process as a two-player game. We proved the existence of a Nash Equilibrium with a natural assumption about the privacy-accuracy trade-off function in the general case, while provided the exact NE when one player is privacy unconcerned. We also defined *Price of Privacy* to measure the overall degradation of accuracy due to the player's privacy protection.

On the practical side, we studied a Recommendation System use case: we applied two different privacy-preserving mechanisms on two real-world datasets and confirmed via experiments that the assumption which ensures the existence of a NE holds. Complementary to the CoL game, we interpolated $\Phi$ for our use case, and devised a possible way to approximate it in real-world scenarios [7].

We find that privacy protection degrades the accuracy heavily for its user, making Collaborative Learning practical only if the players have similar dataset sizes and their privacy weights are low.

# REFERENCES


[1] Michela Chessa, Jens Grossklags, and Patrick Loiseau. 2015. A game-theoretic study on non-monetary incentives in data analytics projects with privacy implications. In *Computer Security Foundations Symposium (CSF), 2015 IEEE 28th*. IEEE.

[2] Jihun Hamm, Yingjun Cao, and Mikhail Belkin. 2016. Learning privately from multiparty data. In *International Conference on Machine Learning*.

[3] John C Harsanyi, Reinhard Selten, et al. 1988. A general theory of equilibrium selection in games. *MIT Press Books* (1988).

[4] Stratis Ioannidis and Patrick Loiseau. 2013. Linear regression as a non-cooperative game. In *International Conference on Web and Internet Economics*. Springer.

[5] H Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, et al. 2016. Communication-efficient learning of deep networks from decentralized data. *arXiv preprint arXiv:1602.05629* (2016).

[6] Jeffrey Pawlick and Quanyan Zhu. 2016. A Stackelberg game perspective on the conflict between machine learning and data obfuscation. In *Information Forensics and Security (WIFS), 2016 IEEE International Workshop on*.

[7] Balazs Pejo, Qiang Tang, and Gergely Biczok. 2018. Together or Alone: The Price of Privacy in Collaborative Learning. *arXiv preprint arXiv:1712.00270* (2018).