ASE 2020

CrossPriv: User Privacy Preservation Model for Cross-Silo Federated Software

Harshita Diddee *Guru Gobind Singh Indraprastha University* harshitadd@gmail.com
Bhrigu Kansra *Guru Gobind Singh Indraprastha University* bhrigukansra98@gmail.com

Unique Paper ID is: ase20-src-14

GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY

Abstract

CrossPriv aims to dictate preliminary norms of SaaS based collaborative software - including the client and server side characteristics of the software deployed on each side. The efficacy of the proposed system is demonstrated by training a convolution neural network on distributed data of two different silos to detect pneumonia using X-Rays.

Challenges

1. Generation of a covariate and representive centralized data repository to train deep learning algorithms is practically difficult to achieve given legal constraints around private user data. This problem is exacerbated for silos which have additional legal constraints due to their nature of collective representation i.e They represent their client's data as well have organizations cooperation guidelines.
2. Constant relay of data to central server for generating inference on cloud deployed deep learning models increases the susceptibility of the data to malicious attacks.

Motivating Question

1. Is there a definitive terminology that defines the stakeholders and parameters in a cross-silo federated learning setup ?
2. Are the characteristics of both, client silos and server delineated for the cross silo FL setup ?

Server Characteristics

- Identifying clients that are apt for participation in FedAvg
- Aggregation of Weights
- Decrypting/Unmasking weights received by silos
- Redistribution of Weights

Approach

Traditional Federated Iteration Cycle is described as follows -

Train model on user's local data → Weights aggregated on the central server → Replacement and Redistribution of Weights → Central Averaging Server

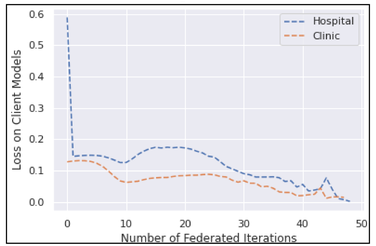
CrossPriv's Federated Iteration Cycle is described as follows -

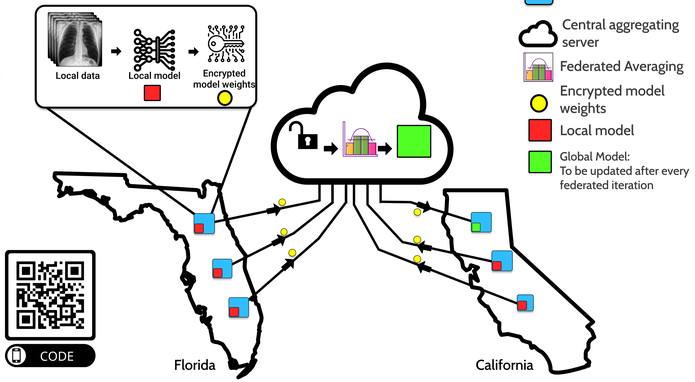
User Identity Anonymization: Masking any personally identifiable data (Eg - In the case of DICOM Images - Age or Gender) → Train model on user's local data → If participant is active → Weight Aggregation and Redistribution → Central Averaging Server

Note - Here user indicates the user who is a part of the silo that is participating in the federated iteration cycle.

Experimental setup and Results

CrossPriv Model - 2+3 layer CNN
Number of silos - 2
Number of epochs on each silo - 2
Number of federated Iterations - 50
Total Accuracy - 81%





CODE

Client or Silo characteristics

- Report Inability to participate in the aggregation due to resource inadequacy
- Mask the personally identifiable user data
- Report possible Byzantine Behaviour
- Identify if resource dissipation is allowed

Conclusion and Impact

We propose CrossPriv, a User-Privacy Sensitive Model that enlists:

a. Characteristics of the entities - Silo (the client) and the central server deployed across a cross-silo federated setup.

b. Simulate the efficacy of the model by demonstrating with a medical use case of training a model to detect Pneumonia using X-Rays.