Capstone Project: Adversarial Federated Learning

Team Members:- Piyush Tiwari(ptiwari), Kavita Kumari(kkumari), Prasanth Yadla(pyadla2)

Motivations:-

Users want to keep their data private.
The server wants to offload computation.

Threat - Malicious agent targets the misclassification of an auxiliary set of examples for the global model.

Aim- We will explore the vulnerability of the update aggregation step of federated learning and devise strategies to prevent or mitigate those attacks.

Data:-

We will use MNIST data for our experiment.

Execution Plan:-

- We will use the Convolutional LeNet model to classify the images.

  We intend to use 15 to 20 agents in which there would be 1 to 3 malicious clients.

- Attack strategies -
  - Updates from other agents could render malicious agent's update ineffective
    Boosting malicious update, no local training
    Boosting malicious update, with local training
  - The server may detect based on the effect on accuracy on validation data or weight update statistics
    - Improve on the baseline by adding benign training and distance constraints
    - Alternating minimization of benign and malicious objectives, no distance constraints
    - Alternating minimization of benign and malicious objectives, with distance constraints

- Prevent Strategies -
  Choose randomly among clients.
  Averaging with other agents

Avoid detection

Planned roles for each Team Member :-

Piyush Tiwari - will work on federated learning and adversarial malicious and algorithm.
Prasanth - will work on hyperparameter tuning and plots of loss/accuracy, attacking success rate.
Kavita - will work on Related work and the comparison between the non-attacked model and the attacked model.

Time Phases-

literature review 1 Week
data set searching and engineering 0.5 Week
methodology derivation 0.5 Week
development, and implementation 2 Week
drafting a final report 1 Week

References:-

Analyzing Federated Learning through an Adversarial Lens https://arxiv.org/pdf/1811.12470.pdf
FEDERATED ADVERSARIAL DOMAIN ADAPTATION
https://openreview.net/pdf?id=HJezF3VYPB
Can You Really Backdoor Federated Learning? https://arxiv.org/pdf/1911.07963.pdf
TOWARDS FEDERATED LEARNING AT SCALE: SYSTEM DESIGN
(https://arxiv.org/pdf/1902.01046.pdf)