**Team Member**: Jingyu Peng (jp550), Libo Zhang (lz200), Shucheng Zhang(sz255)

**Project**: Poisoning Attacks to Federated Learning (18)—— Lead TA: Hongyu He

**Heielmier cataclysm**

### -What are you trying to do?

1. Understand federated learning techniques, implement federated learning on CIFAR-10 dataset.
2. Implement two poisoning attacks (data poisoning attack and model poisoning attack) to the federated learning to examine the vulnerability of federated learning.
3. Check the intensity of two poisoning attacks and the model's robustness.
4. Discuss ways to defend against poisoning attacks.

### -How is it done today?

Federated Learning is investigated to solve the training problems of those privacy or large-in-quantity data[1]. However, the Federated Learning is sensitive to data and model poisoning attacks and the robustness is challenged[2][3]. Research on defense methods keeps going to help the Federated Learning model become robust.

### -Your approach and why do you think it will be successful?

1. Use different models (basic CNN, ResNet-20, Inception Network) as different clients for federated learning. Change the number of clients and the input data batch. Describe the choice of different local models.
2. Apply data poisoning attacks to training data collection, apply model poisoning attacks during the learning process. Compare testing accuracies before and after poisoning attacks, and provide detailed analysis.
3. Discuss potential methods to defend against poisoning attacks, explain why they could be effective.

**Why it will be successful**: We can observe the testing accuracy change to check whether we successfully apply poisoning attacks (accuracy decreases) or whether we find reliable defense strategies (accuracy increases back). As long as federated training, poisoning attacks and defense methods are correctly implemented, we can observe what should be expected.

### -What are the risks?

1. The number of clients is small (small ratio), try to change the number of clients (different ratio).
2. Batch of data may be insufficient, try to change the data. Consider non-IID and IID data.
3. The model may show very different robustness against data and model poisoning attacks, try to introduce more client devices during federated training.

### -How long will it take?

Schedule for 5 weeks. Start from November 4 and end in early December. First two weeks for federated learning, then two weeks for poisoning attacks. The last week for discussing defense methods, finalizing project report, and preparing for poster presentation.

### -What are the final “exams” to check for success?

1. **Code** implementing federated learning, applying data poisoning attacks to training data collection, applying model poisoning attacks to the learning process, and applying defense methods against poisoning attacks.
2. **Table** and **Figure** showing comparison results, especially the testing accuracies of the original federated learning model, the poisoning attacked model, and the model with defenses.
3. **Analysis** explaining all results and demonstrating the way of applying poisoning attacks and corresponding defense methods.

### -What is the task distribution?

Jingyu Peng will implement federated learning. Libo Zhang will apply data poisoning attacks to the training data collection. Shucheng Zhang will apply local model poisoning attacks to the learning process.

All team members should thoroughly understand the theory of federated learning and comprehensively discuss defense methods against poisoning attacks. Team members can help each other.

### References:

[1] Communication-Efficient Learning of Deep Networks from Decentralized Data

[2] Data Poisoning Attacks on Federated Machine Learning

[3] Local Model Poisoning Attacks to Byzantine-Robust Federated Learning