NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

School of Computer Science and Engineering
College of Engineering

# Peer-to-peer Federated Learning

## How-to: Maximizing communication efficiencies

Student: Yeow Wei Liang, Brandon

Supervisor: Assoc Prof Anupam Chattopadhyay

## Project Objectives

This project seeks to design an algorithm to determine the best action a peer-to-peer client can take to optimize efficiency of communication by maximising local model inference accuracy while minimising communication costs. The algorithm will consider other peer client's data set size, accuracy, CPU speed, RAM available, bandwidth available, connected peers and privacy constraints of sharing data.

## Experiment Setup

Figure 1 illustrates the setup and parameters tweaked to isolate and recreate scenarios. Dataset Handler induces various non-i.i.d-ness into the dataset through image filters, missing labels, and skewed quantities. Coloured parameters are exhaustively tested.

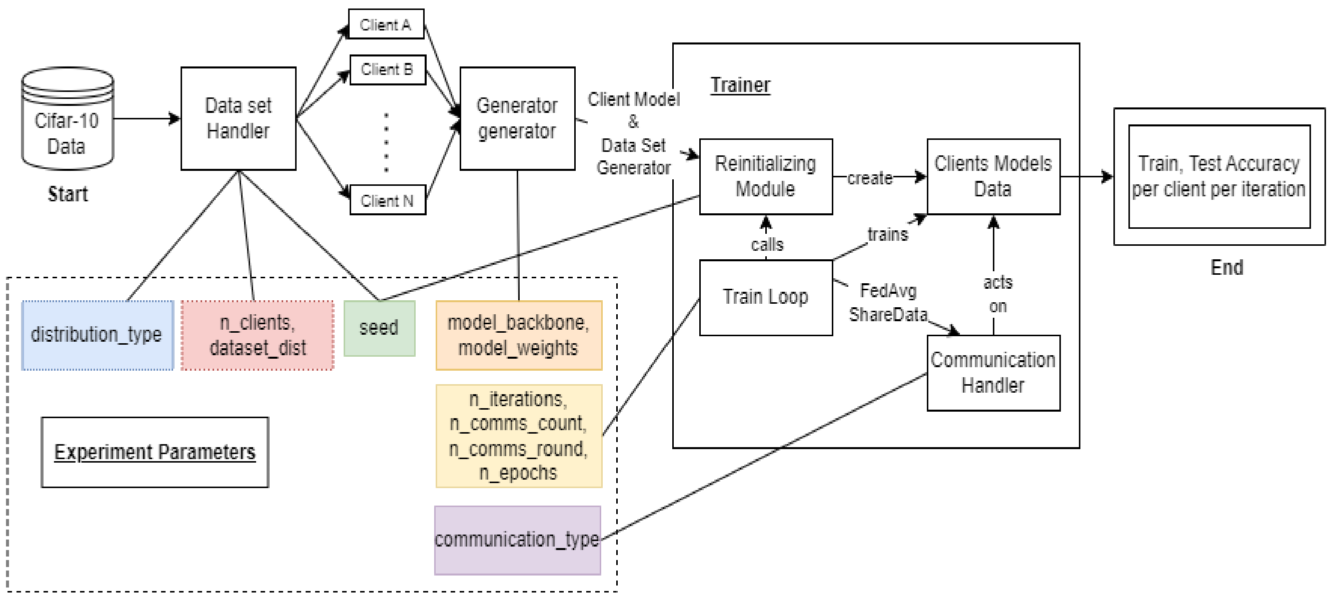


Figure 1: Overall Experiment Setup

Conditional VAEs are used to preserve some privacy when sharing data. Various latent spaces are also tested to vary quality of reconstructed images. All 4 VAEs showed better performance than just simply sharing data with little differences in inference accuracy between each of them.

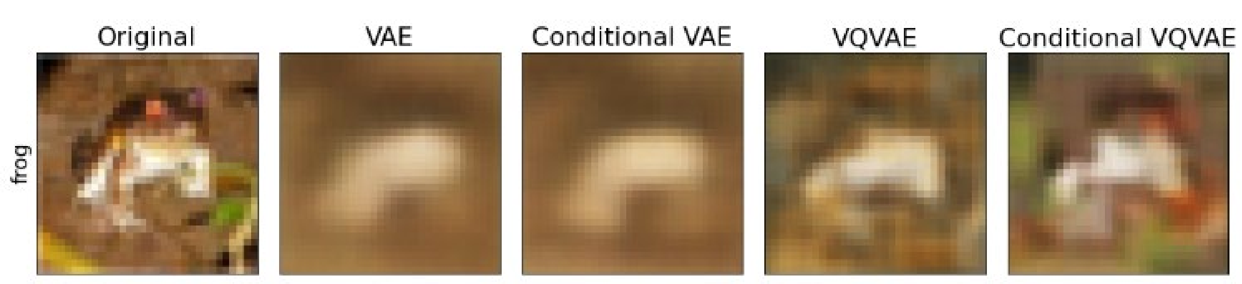


Figure 2: Variational Autoencoder Image Quality

## Results and Decision Tree

The following are the insights derived:

1. Ask for data when there is high local compute availability and idle cycle times
2. Send trusted peers reconstructed data, non-trusted peers minimal sampled data
3. FedAvg with clients that have a higher accuracy
4. Synthetic data sharing and successive FedAvg is effective in highly skewed non-i.i.d clients
5. Synthetic data sharing is best in cases of missing class labels

Along with empirical results, the following decision tree is formed to determine the best action.

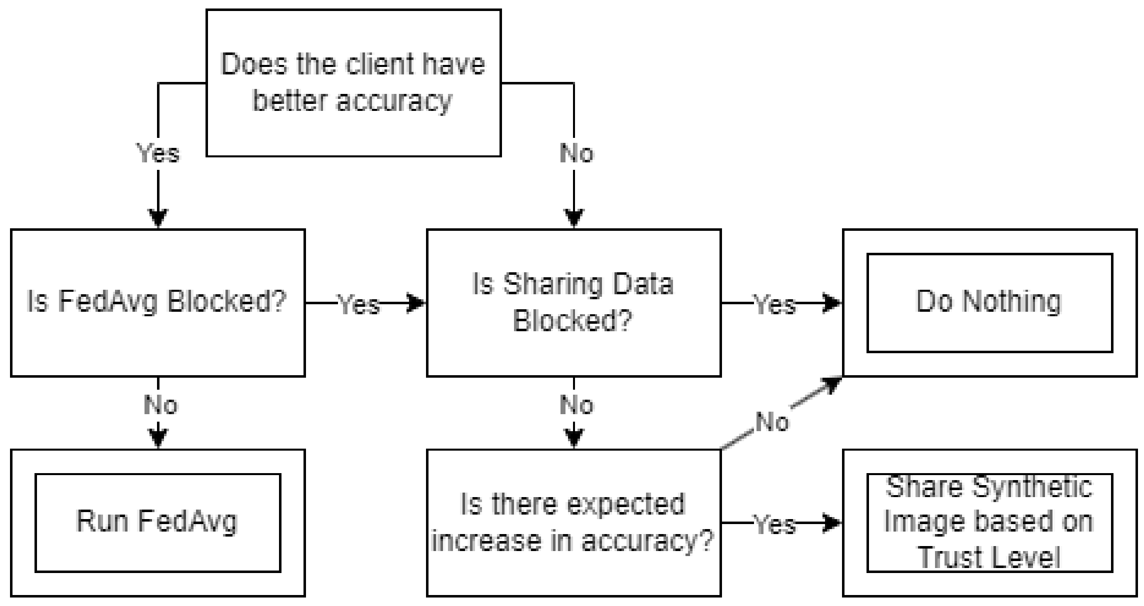


Figure 3: Decision Tree based on Insights

## Validation Experiments

8 nodes are setup with a pre-trained VGG-16 and has equal dataset size. A 200 Mbps bandwidth constraint is applied. The algorithm is compared against running only FedAvg.

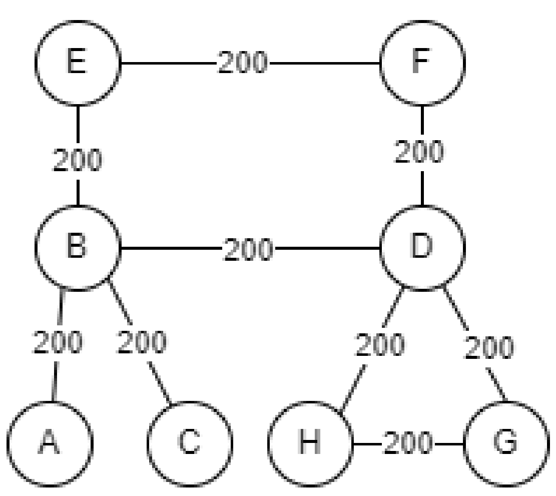


Fig 4: Node Setup

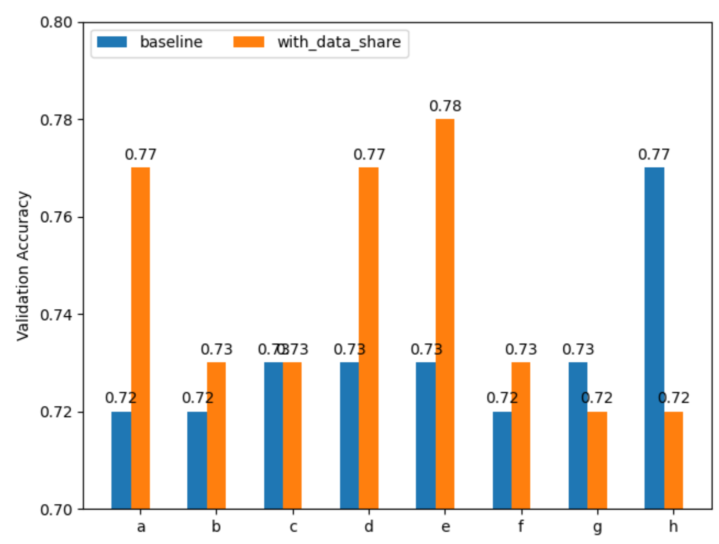


Fig 5: Validation Results