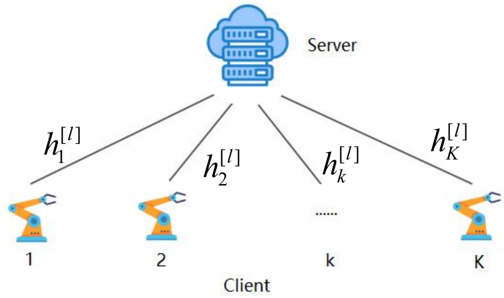
Our FL model consists of a cloud server and K client devices.



Basic structure of federated learning

A. Model broadcast

Considering a specific l-th round, the global parameter $w^{[l]}$ is broadcasted to each client k, $k \in \{1,2,\ldots,K\}$.

B. Local model update

Based on the received model parameter $w^{[l]}$, the client k will compute the local gradient $g_k^{[l+1]}$ and update the local model parameter, given by

$$g_k^{[l+1]} = \nabla F_k(w^{[l]})$$

$$w_k^{[l+1]} \leftarrow w_k^{[l]} - \eta g_k^{[l+1]}$$

where η is the learning rate and $\nabla F_k(w^{[l]})$ denotes the gradient of loss function $F_k(\cdot)$ at $w = w^{[l]}$.

C. Model aggregation (FedAvg)

The clients will transmit the updated model parameters to the server. The server takes the average of the model parameters from all the clients. Finally, the server gets an updated global model parameter $w^{[l+1]}$ and starts the next FL round.

$$w^{[l+1]} = \frac{1}{K}\sum_{k=1}^{K} w_k^{[l+1]}$$

Wireless transmission

Then we take the wireless transmission part into account. In the first step, the symbol client k receives, denoted by y_k, can be simply represented as

$$y_k^{[l]} = h_k^{[l]} w^{[l]} + n^{[l]}$$

where $h_k^{[l]} \in C$ denotes the channel from client k to the cloud server and $n^{[l]} \in C$ is the additive white Gaussian noise following $n^{[l]} \sim CN(0,\sigma^2)$. We assume that the channel of client

k follows the distribution as $h_k^{[l]} \sim \sqrt{\frac{\alpha}{\alpha+1}} + \sqrt{\frac{1}{\alpha+1}} CN(0,1)$ where $\alpha = 1,\ 10, \ldots$. Then the amplitude of $h_k^{[l]}$, denoted as $\left|h_k^{[l]}\right|$, follows the Rician distribution. The downlink and uplink will use the same channel at different time slots, which means the channel parameter is the same when broadcast and uploaded.

We will not consider the noise of the downlink in the first step. Since the transmission power of the server is much higher than that of the client, the noise is negligible. There is no wireless transmission in the second step. In the third step, however, the distortion and noise from the channel will affect the model aggregation.

Before transmitting, we need to normalize the weights in model parameters since the numbers are very small. We compute the first-order and second-order statistics of the local model parameters by

$$\overline{w}_k^{[l]} = \frac{1}{d} \sum_{i=1}^{d} w_{k,i}^{[l]}$$

$$\delta_k^{[l]} = \sqrt{\frac{1}{d} \sum_{i=1}^{d} (w_{k,i}^{[l]} - \overline{w}_k^{[l]})^2}$$

Then the two parameters $\overline{w}_k^{[l]}$ and $\delta_k^{[l]}$ will be uploaded to the server. The weights can be normalized in the server as

$$\widetilde{w}_k^{[l]} = \frac{w_k^{[l]} - \overline{w}_k^{[l]}}{\delta_k^{[l]}}$$

The symbol y received by the server can be represented as

$$y_i^{[l]} = \sum_{k=1}^{K} h_k^{[l]} b_k^{[l]} \widetilde{w}_{k,i}^{[l]} + n^{[l]}$$

where $b_k^{[l]} \in C$ is the corresponding transmit equalization coefficient of client k, and i denotes the i-th symbol of $\widetilde{w}_k^{[l]}$. So the symbol client k really transmits is $s_{k,i}^{[l]} = b_k^{[l]} \widetilde{w}_{k,i}^{[l]}$. Since the client has its maximum transmission power $P_{\max}$, the power of each symbol should satisfy the power constraint as

$$\frac{1}{d} \sum_{i=1}^{d} \left|s_{k,i}^{[l]}\right|^2 = \frac{1}{d} \sum_{i=1}^{d} \left|b_k^{[l]} \widetilde{w}_{k,i}^{[l]}\right|^2 \le P_{\max} .$$

where d means the length of $\widetilde{w}_k^{[l]}$.

We assume that $P_{\max} = 1$ here. So the SNR between the symbol and noise can be written as

$$SNR = \frac{P_{\max}}{\sigma^2} = \frac{1}{\sigma^2}$$

And the white Gaussian noise (WGN) $n^{[l]}$ will follow the complex Gaussian distribution $CN\left(0, \frac{1}{SNR}\right)$.

What the server needs to do next is to recover the normalized model parameter $\widetilde{w}^{[l]} = \frac{1}{K}\sum_{k=1}^{K} \widetilde{w}_k^{[l]}$ from $y^{[l]}$.

However, after the antenna of the server receives the signal, it will be first processed by the hardware circuit, and we express the effect of the circuit as $m^{[l]}$. Thus, the final received signal can be expressed as

$$r_i^{[l]} = m^{[l]} y_i^{[l]} = m^{[l]} \sum_{k=1}^{K} h_k^{[l]} b_k^{[l]} \widetilde{w}_{k,i}^{[l]} + m^{[l]} n^{[l]}$$

Since we need to get the average of the model parameter, the recovered symbol can be written as

$$\hat{w}_i^{[l]} = \frac{1}{K} r_i^{[l]} = \frac{1}{K} m^{[l]} y_i^{[l]} = \frac{1}{K} m^{[l]} \sum_{k=1}^{K} h_k^{[l]} b_k^{[l]} \widetilde{w}_{k,i}^{[l]} + \frac{1}{K} m^{[l]} n^{[l]}$$

If we set $b_k^{[l]} = \frac{(m^{[l]} h_k^{[l]})^*}{\left|m^{[l]} h_k^{[l]}\right|^2}, \forall k$, where $(\cdot)^*$ means the conjugation, and then

$$\hat{w}_i^{[l]} = \frac{1}{K} \sum_{k=1}^{K} m^{[l]} h_k^{[l]} \frac{(m^{[l]} h_k^{[l]})^*}{\left|m^{[l]} h_k^{[l]}\right|^2} \widetilde{w}_{k,i}^{[l]} + \frac{1}{K} m^{[l]} n^{[l]} = \frac{1}{K} \sum_{k=1}^{K} \widetilde{w}_{k,i}^{[l]} + \frac{1}{K} m^{[l]} n^{[l]} = \widetilde{w}_i^{[l]} + \frac{1}{K} m^{[l]} n^{[l]}$$

From the equation above, we can find the error or the noise between the recovered parameter and the original parameter can be denoted as $e_i^{[l]} = \hat{w}_i^{[l]} - \widetilde{w}_i^{[l]} = \frac{1}{K} m^{[l]} n^{[l]}$

There are three variables in this expression, and K is already known. The WGN $n^{[l]}$ also follows the complex Gaussian distribution $CN(0, \frac{1}{SNR})$. The problem is to define $m^{[l]}$ under the circumstance that the mean-square-error (MSE) can be minimized.
The MSE can be written as

$$MSE = E\left[\left|e^{[l]}\right|^2\right] = E\left[\left|\frac{1}{K} m^{[l]} n^{[l]}\right|^2\right] = \frac{\left|m^{[l]}\right|^2}{K^2 \cdot SNR}.$$

To minimize MSE, we need to minimize $\left|m^{[l]}\right|^2$.

In the wireless transmission process, the signal will be modulated. Take QPSK modulation, for example, the channel can transmit a complex symbol at a time slot. Since $\widetilde{w}_k^{[l]}$ has length *d*, the *i*-th complex symbol $\left[w_k^C\right]_i^{[l]}$ has the length of $\frac{d}{2}$. The average power of a complex symbol can be defined as

$$\beta_k^{[l]} \stackrel{\Delta}{=} \frac{2}{d}\sum_{i=1}^{\frac{d}{2}}\left|[w_k^C]_i^{[l]}\right|^2$$

Following the equation $\frac{1}{d}\sum_{i=1}^{d}\left|s_{k,i}^{[l]}\right|^2 = \frac{1}{d}\sum_{i=1}^{d}\left|b_k^{[l]}\widetilde{w}_{k,i}^{[l]}\right|^2 \leq P_{\max}$, we can derive that

$$\beta_k^{[l]}\left|b_k^{[l]}\right|^2 \leq P_{\max} \Rightarrow \frac{\beta_k^{[l]}}{\left|m^{[l]}h_k^{[l]}\right|^2} \leq 1$$

Then we formulate an optimization problem, given by

$$\min_{m^{[l]}} \quad \left|m^{[l]}\right|^2$$
$$s.t. \quad \frac{\beta_k^{[l]}}{\left|m^{[l]}h_k^{[l]}\right|^2} \leq 1$$

which can be written as $\left|m^{[l]}\right|^2 \geq \frac{\beta_k^{[l]}}{\left|h_k^{[l]}\right|^2}, \forall k$. So $\left|m^{[l]}\right|^2 = \max_{k=1,..K}\frac{\beta_k^{[l]}}{\left|h_k^{[l]}\right|^2}$ and we already know the value of $\beta_k^{[l]}$ and $\left|h_k^{[l]}\right|^2$.

So $MSE = \frac{1}{K^2 \cdot SNR}\max_{k=1,..K}\frac{\beta_k^{[l]}}{\left|h_k^{[l]}\right|^2}$.

The recovered symbol $\hat{w}_i^{[l]} = \widetilde{w}_i^{[l]} + \frac{\beta_k^{[l]}n^{[l]}}{K\left|h_k^{[l]}\right|^2}$.

And then $\hat{w}_i^{[l]}$ will be inverse normalized as $W_k^{[l]} = \hat{w}^{[l]}\delta_k^{[l]} + \overline{w}_k^{[l]}$ and sent $W_k^{[l]}$ back to the corresponding client *k*.

Then, the expression of noise in the wireless transmission is found.