Uplink Time Constrained Federated Learning over Wireless Networks

Ji Ho Choi and Dong In Kim, *Fellow*, *IEEE*
Department of Electrical and Computer Engineering
Sungkyunkwan University (SKKU), Suwon, Korea
Email: {klepy, dongin}@skku.edu

***Abstract*—Federated learning (FL) is an emerging distributed learning paradigm that collaboratively trains a shared model while preserving their data privacy. However, clients participating in the FL process have heterogeneous computation power and communication resources. Therefore, the central server needs to wait for the slowest client finishes uploading its local model update, which is known as *straggler effect*. In this paper, we propose a novel FL method with uplink time constraint set by the server to mitigate the straggler effect over wireless networks. First, we study the impact of uplink time constraint on the number of clients with successful upload in Rayleigh fading channel. Since acquiring perfect channel state information (CSI) is challenging in practical FL applications, the server exploits channel statistics instead. We then formulate an optimization problem for the error convergence speed with respect to uplink time constraint. Finally, we confirm that the proposed scheme, namely FL with *Uplink Time Constraint* (FedUTC), achieves faster convergence speed than the baseline FL method.**

***Index Terms*—Federated learning (FL), straggler effect, uplink time constraint, convergence time, fading channel.**

I. Introduction

With the rapid growth of the Internet of things (IoT), massive amount of data are generated from the large-scale IoT devices. To create valuable information from the data, machine learning (ML) has been widely used to train data-driven models for enabling intelligent IoT applications [1]. However, conventional centralized learning method, which collects data in a central server, is not feasible for the future wireless network applications due to heavy traffic and privacy issue of the data. To overcome these challenges, distributed learning framework like federated learning (FL) has been proposed as a means to enable collaborative training based on the model information exchange through the server. Specifically, clients train a shared global model with own data by sending local updates to a central server each round, rather than sending raw data for training. This allows for models to be trained on a tremendous volume of dataset generated from the large-scale IoT networks, while preserving the privacy of the data.

Unfortunately, there are still persisting challenges for the implementation of FL over the wireless networks. In general, clients participating in FL have heterogeneous computation and communication resources due to system heterogeneity [2]. Therefore, each client varies in time to finish its gradient computation and the uplink communication time to upload its local model update to the server, like the base station (BS). Since the power budget of the BS is assumed sufficient enough to ignore its downlink communication time [3] [4], uplink communication time is considered to be a critical factor of communication overhead. To mitigate the communication overhead, the authors in [5] studied the optimization error upper bound for different communication periods and proposed adaptive local update scheme that improves the convergence speed. Meanwhile, the study in [6] has jointly optimized the local update and gradient compression level to accelerate learning convergence speed. But both works do not consider realistic channel model and assume fixed data rate.

To implement FL over the wireless networks, the random variation of wireless channels due to fading should be taken into account. For example, the time-varying channel capacity of clients in wireless fading channels decides whether they can successfully upload local update with arbitrarily small probability of error or not. However, most existing works focus on the communication-efficient aspect while assuming perfect channel state information (CSI). In practice, such random variation likely causes the outage of not receiving the local update within a designated time, resulting in the fluctuation of the number of clients participating in the FL process.

The authors in [7] studied the convergence time of FL over the wireless networks under imperfect CSI condition and proposed FL with fixed data rate. They showed that assigning a global constant rate to all participating clients decreases the convergence time. However, they analyzed the convergence time in terms of the round duration only, which is not sufficient to identify the effect of the average number of clients with fixed global rate on the model convergence as shown in [5]. For accurate analysis, various elements related to FL, such as computation-communication trade-off with local update, should be considered with respect to the wall-clock time. For example, performing multiple local updates addresses a communication-computation trade-off.

In this paper, we propose a new FL method called *FedUTC* (federated learning with uplink time constraint), in which the BS neglects the local update of some clients in outage, if they fail to upload within the uplink time constraint set by the BS, to mitigate the *straggler effect* due to channel fading. First, we analyze the trade-off between the uplink time constraint and the average number of clients with successful upload by exploiting the cumulative distribution function (CDF) of the fading channel. Herein, we also consider the computation time and local update coefficient as well. We then formulate an optimization problem about the uplink time constraint for convergence performance with a given wall-clock time. We conduct the experiments using a multilayer-perceptron (MLP) neural network model on independent and identically distributed (i.i.d) MNIST datasets to evaluate the performance of *FedUTC*. The experiment results show that FL convergence performance can be improved by applying the optimal uplink time constraint where no perfect CSI is available.

The rest of the paper is organized as follows. Section II describes the system model of a general FL process with uplink time constraint. In Section III, we address how the uplink time constraint affects the performance of FL. We then formulate an optimization problem about the uplink time constraint to improve the learning convergence. Experiment results are provided in Section IV to show that the optimal uplink time constraint does speed up the learning convergence with best accuracy with respect to the wall-clock time. Lastly in Section V, we summarize and conclude our work.

II. System Model

In this section, we describe the fundamental federated learning procedure and study the impact of fading channel to the FL process.

A. Federated Learning

In FL, a global model is trained by sharing the local model updates from clients over the networks. Here, the total dataset $\mathcal{D}$ is distributed among M clients and each client has a disjoint dataset $\mathcal{D}_j \subset \mathcal{D}$ for $j = 1, 2, \ldots, M$, where $\cup_{j=1}^{M} \mathcal{D}_j = \mathcal{D}$ and $\mathcal{D}_i \cap \mathcal{D}_j = \emptyset$ for $i \neq j$. The goal of FL is to minimize the global loss function $F(\boldsymbol{w}_k, \mathcal{D})$ which is the weighted average of the local loss functions $f(\boldsymbol{w}_k^j, \mathcal{D}_j)$, that is given by

$$F(\boldsymbol{w}_k, \mathcal{D}) = \sum_{j=1}^{M} \frac{|\mathcal{D}_j|}{|\mathcal{D}|} f(\boldsymbol{w}_k^j, \mathcal{D}_j). \tag{1}$$

Unlike the centralized stochastic gradient descent (SGD), finding an optimal global model parameter $\boldsymbol{w}_k$ is done by minimizing the local loss functions of clients since the data is distributed over multiple clients.

In the kth round, the BS broadcasts the global model $\boldsymbol{w}_k$ to each clients. Then, each client performs local updates using SGD based on the local dataset. After the ℓth local update is finished, the local model parameter is updated as follows:

$$\boldsymbol{w}_k^{j,\ell} = \boldsymbol{w}_k^{j,\ell-1} - \eta \nabla f(\boldsymbol{w}_k^{j,\ell-1}, \xi_j) \tag{2}$$

for $k = 1, 2, \cdots, K$, and $l = 1, 2, \cdots \tau$, where η is the learning rate and ξ_j is a randomly selected training sample from $\mathcal{D}_j$. In this paper, we assume that the number of local updates τ is the same for all clients and fixed during the whole training. After the local updates are finished, clients upload updated local model information to the BS. In this paper, we adopt the global model averaging method of *FedAvg* [8] where the BS averages all local updates and updates the global model $\boldsymbol{w}_{k+1}$ as follows:

$$\boldsymbol{w}_{k+1} = \sum_{j=1}^{M} \frac{|\mathcal{D}_j|}{|\mathcal{D}|} \boldsymbol{w}_{k+1}^j. \tag{3}$$

Then the updated global model is broadcast to all clients and the training process repeats until the performance metrics, such as accuracy and convergence speed, of the global model satisfy the target requirements. Here we focus on the convergence performance of FL over the wireless networks. Therefore, the computation and communication time analysis can be provided similarly to those in [9]. For simplicity, we assume the computation time to be homogeneous among clients while the communication time varies due to the fading channel.

B. FL over Wireless Channels

For the communication time analysis of FL, we assume an orthogonal frequency division multiplexing (OFDM) system. Here the downlink time is negligible compare to the uplink time considering the sufficient power budget and bandwidth resource of the BS [3]. For the uplink time, each client is assigned the same bandwidth B^j, and we adopt an i.i.d. slow fading *Rayleigh channel*, i.e., $h_k^j \sim \mathcal{CN}(0, 1)$, which remains constant in each round. Following the work in [7], we assume that the BS has only the knowledge of the channel statistics like the CDF of the channel gain instead of perfect CSI assumption. Then the channel capacity of the jth client in the kth round can be evaluated as

$$C_k^j = B^j \log_2 \left(1 + \frac{p_k^j \phi_k^j |h_k^j|^2}{B^j N_0}\right) = B^j \log_2 \left(1 + g_k^j A\right) \tag{4}$$

where g_k^j represents the channel power gain associated with the jth client and kth round, namely $g_k^j = |h_k^j|^2$, p_k^j the transmit power, ϕ_k^j the large-scale propagation effect like path loss, and N_0 the one-sided noise power spectral density. We examine the scenario where clients exercise the open-loop power control such that the path loss is compensated to maintain a constant value A during the whole training. In this situation, a client with low channel capacity due to poor channel gain requires longer time to upload its local update to the BS, which results in a straggler effect that slows down the overall training time. To mitigate this, we propose the concept of uplink time constrained FL in the next section.

III. FL with Uplink Time Constraint

In this section, we introduce a novel FL method with uplink time constraint set by the BS to deal with the straggler issue in a fading channel and provide the analysis on the convergence performance of the proposed scheme. We then formulate the problem to obtain the optimal uplink time constraint.

A. Uplink Time Constraint

We consider a delay-sensitive FL framework that neglects the clients who fail to upload their local updates within the uplink time constraint t_{ul}. Fig. 1 illustrates the FL mechanism with uplink time constraint. In conventional synchronous FL, the BS waits until the slowest (e.g., straggler) user finishes its uploading the local model parameter, denoted as t_{sync}. When we employ the uplink time constraint t_{ul}, the BS stops receiving the local model updates from clients. Therefore, the communication time becomes the uplink time constraint itself

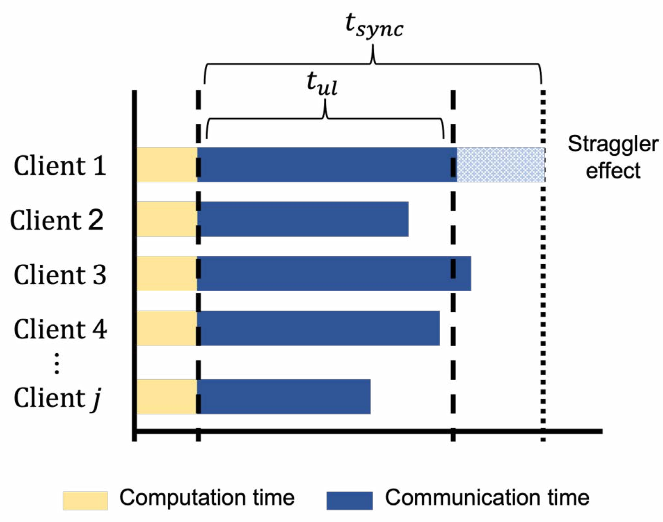


Fig. 1. Illustration of FL with uplink time constraint.

and the extra time that would have been wasted for waiting the straggler is saved. Allowing for longer uplink time can lead to obtaining the updated model information from a larger number of clients, but the training time increases. Therefore, we should be able to address the trade-off between the number of clients with successful upload and the training time for convergence analysis. Here we assume that a client experiences the outage if the channel capacity is lower than the target data rate $R(t_{ul})$ for a given t_{ul}. We express the CDF of the channel power gain as $F_{cdf}(g_k^j)$, which is given by

$$F_{cdf}\left(g_k^j\right) = 1 - \exp\left(-g_k^j/2\right), \quad g_k^j \geq 0. \tag{5}$$

Then the outage probability can be evaluated as

$$\begin{aligned} p_{out}(t_{ul}) &= \mathbb{P}\left[C_k^j < R(t_{ul})\right] && (6)\\ &= \mathbb{P}\left[g_k^j < \frac{2^{R(t_{ul})/B^j} - 1}{A}\right] && (7)\\ &= F_{cdf}\left(\frac{2^{Z/(B^j t_{ul})} - 1}{A}\right) && (8)\end{aligned}$$

where Z is the size of model parameter with $R(t_{ul}) = Z/t_{ul}$. Hence, the average number of clients with successful upload can be evaluated as

$$\begin{aligned} \mathbb{E}\left[M_{ul}\right] &= M\left[1 - p_{out}(t_{ul})\right] && (9)\\ &= M\exp\left(-\frac{2^{Z/(B^j t_{ul})} - 1}{2A}\right). && (10)\end{aligned}$$

In the above, the expectation $\mathbb{E}$ has been taken with respect to the channel fading varying in each round for the clients.

B. Convergence Analysis

For the i.i.d. data distribution, the convergence rate improves as the number of clients participating in FL increases [10]. To investigate the impact of the uplink time constraint to the model convergence, we study the learning error convergence with respect to a given wall-clock time. As for the indicator of convergence in FL, the expected gradient norm was used since the objective function in FL is non-convex [11]. Based on this observation, we invoke the upper bound analysis of the expected gradient norm, namely the error upper bound analysis conducted in [5] to study the trade-off between the uplink time constraint and the convergence performance. To make the analysis tractable, the following assumptions are made: (i) the global loss function $F(\boldsymbol{w}_k, \mathcal{D})$ is differentiable, and L-Lipschitz smooth, i.e., $\|\nabla F(\boldsymbol{x}, \mathcal{D}) - \nabla F(\boldsymbol{y}, \mathcal{D})\| \leq L\|\boldsymbol{x} - \boldsymbol{y}\|$ where L is the Lipschitz constant, and the function value is bounded below by a scalar F_{inf}; (ii) the stochastic gradient evaluated on a mini-batch ξ is an unbiased estimator of the full batch gradient $\mathbb{E}\left[\nabla F(\boldsymbol{w}_k, \xi)\right] = \nabla F(\boldsymbol{w}_k, \mathcal{D})$ for all k; (iii) the variance of stochastic gradient evaluated on a mini-batch ξ is bounded as $\mathbb{E}\left[\|\nabla F(\boldsymbol{w}_k, \xi) - \nabla F(\boldsymbol{w}_k, \mathcal{D})\|\right] \leq \beta\|\nabla F(\boldsymbol{w}_k, \mathcal{D})\|^2 + \sigma^2$, where β and σ^2 are non-negative constants and in inverse proportion to the mini-batch size.

Let Y_k and D_k denote the computation and communication time for the kth round. Then, if the learning rate satisfies $\eta L + \eta^2 L^2 \tau(\tau - 1) \leq 1$, and all clients are initialized at the same point $\boldsymbol{w}_0$, the error upper bound after total T wall-clock time can be derived as [5]

$$\frac{2\left[F(\boldsymbol{w}_0) - F_{inf}\right]}{\eta T}\left(Y_k + \frac{D_k}{\tau}\right) + \frac{\eta L \sigma^2}{M} + \eta^2 L^2 \sigma^2 (\tau - 1) \tag{11}$$

where L is the Lipschitz constant of the loss function and σ^2 is the variance bound of mini-batch stochastic gradients.

Now we replace some terms in (11) to fit the analysis for the proposed FL scheme. First, we replace the communication time D_k by t_{ul} and assume that the computation time remains constant in each round, namely $Y_k = Y$. Also, we change the number of participants M in the denominator of the second term to $\mathbb{E}\left[M_{ul}\right]$ as given in Section III-A. Moreover, the total wall-clock time can be quantified in terms of the local update coefficient τ and uplink time constraint t_{ul}, which is given by $T = K(Y\tau + t_{ul})$ for $k = 1, 2, \cdots, K$, assuming that t_{ul} is fixed during the whole FL process.

Then, the error upper bound in (11) can be reformulated as a function of the uplink time constraint t_{ul}, that is

$$\begin{aligned}\psi(t_{ul}) = &\frac{2\left[F(\boldsymbol{w}_0) - F_{inf}\right]}{\eta T}\left(Y + \frac{t_{ul}}{\tau}\right) + \frac{\eta L \sigma^2}{M}\\ &\times \exp\left(\frac{2^{Z/(B^j t_{ul})} - 1}{2A}\right) + \eta^2 L^2 \sigma^2 (\tau - 1). \quad (12)\end{aligned}$$

This upper bound in (12) represents the trade-off between the uplink time constraint and the volume of data access from the clients with successful upload, namely effective participants. Longer time constraint increases the number of effective participants and the upper bound decreases since the exponential function in the second term of (12) is a decreasing function of t_{ul}. However, employing too long t_{ul} causes the first term to increase which makes the round completion time longer. Moreover, the local update coefficient τ in the first term also affects the upper bound. Unlike fully synchronous

FL where the local update is performed once ($\tau = 1$), using multiple local updates helps the global model converge faster. But τ in the third term of (12) causes the same trade-off as t_{ul}. In this work, we do not jointly optimize the local update coefficient and uplink time constraint in each round. Instead, we briefly analyze how the local update coefficient affects the FL performance by the experiments in Section IV.

To minimize the global loss function given the wall-clock time T, we need to select an optimal value of t_{ul} by balancing the trade-off mentioned above. To this end, we formulate an optimization problem to minimize the error-runtime upper bound in (12) as follows:

$$t_{ul}^* = \arg\min_{t_{ul}} \psi(t_{ul}). \tag{13}$$

We first prove that the convexity of $\psi(t_{ul})$ by showing that the second derivative of $\psi(t_{ul}) > 0$ for $t_{ul} > 0$, and then present the overall FL algorithm of *FedUTC*. Let $g(t_{ul})$ denote the exponential function in the second term of (12) as

$$g(t_{ul}) = \exp\left(\frac{2^{Z/(B^j t_{ul})} - 1}{2A}\right). \tag{14}$$

Note that $g(t_{ul})$ is a decreasing function of t_{ul} whereas the first term of (12) is in proportion to t_{ul}. Then, we have $g'(t_{ul}) < 0$ and $g''(t_{ul}) > 0$ for $t_{ul} > 0$. We confirm that (12) is a convex problem as the sign of the second derivative of $\psi(t_{ul})$ is non-negative. Thus, the optimal uplink time constraint t_{ul} can be calculated by setting the first derivative of (12) to zero.

Algorithm 1 presents *FedUTC* between the BS and clients over the wireless channels. The goal of *FedUTC* is to mitigate the straggler effect by limiting the time for receiving local updates from the clients with weak channel gains. First, the BS initializes a global model $\boldsymbol{w}_0$ and broadcasts it to all clients. In each round, the clients train the shared global model and transmit the updated local model to the BS. The clients will stop uploading their local updates if the communication time t exceeds t_{ul}. Then, the BS aggregates the received local updates within the uplink time constraint. Finally, the BS broadcasts the updated global model to the clients and continues training subject to the wall-clock time T or total rounds K.

Algorithm 1 FedUTC

```
BS set the uplink time constraint t_ul;
BS initializes and broadcasts w_0 and t_ul;
Clients receive w_0 and t_ul;
for k = 1, 2, ..., K do
    for j ∈ M do
        for ℓ = 1, 2, ..., τ do
            Clients compute ∇f(w_k^{j,ℓ}, ξ_j);
            Clients update w_k^{j,ℓ} as in (2);
        end for
        while t < t_ul do
            Clients transmit w_{k+1}^j to the BS;
        end while
    end for
    The BS receives w_{k+1}^j from the clients for t_ul;
    The BS aggregates all local updates;
    The BS updates the global model as in (3);
    The BS broadcast w_{k+1} to clients;
end for
```

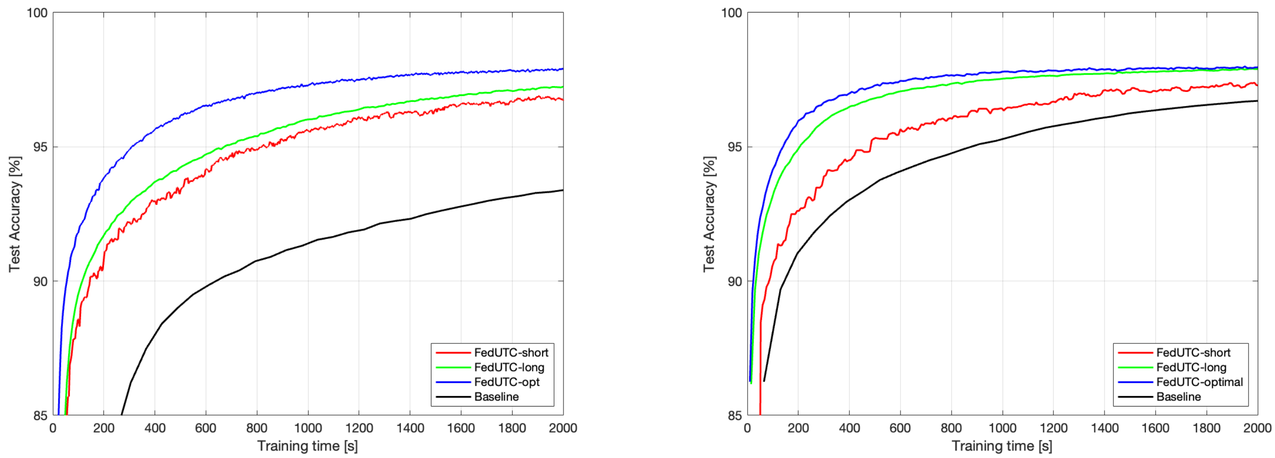
IV. Experiments

In this section, we describe the experiment setup of the proposed *FedUTC* scheme and provide numerical results to validate the theoretical analysis.

A. Experiment Setup

We consider a FL network scenario where $M = 50$ wireless clients collaboratively train the global model. Each client is assigned the same bandwidth $B^j = 1$ MHz during the whole training process. Also, we consider the Rayleigh fading channel with the normalized value $A = 1$ as assumed in [7]. Further, to focus on the effect of the uplink time constraint on the convergence performance, the computation time of all clients is assumed identical and fixed, i.e., $Y = 1$. For the FL task, we choose the image classification one using MNIST dataset. The dataset includes 60,000 and 10,000 images of handwritten digits, each for training and testing the task. Here, we consider only the i.i.d. data distribution where the classes and training samples are uniformly distributed over the clients. We train a MLP network model used in [8] which consists of two hidden layers with 200 units with ReLu activations (199,210 total parameters). Throughout the experiments, we set the learning rate to 0.01 with SGD optimizer and the mini-batch size to 32 for the local update. Note that the mini-batch size acts as a critical factor to FL. Adopting too small mini-batch size leads to too little computation per communication, while too large mini-batch size cripples and slows the learning due to excessive computation per communication [12].

To implement FL for model training over wireless edge networks, achieving the best performance within a limited time is required. In our experiments, we plot the performance of test accuracy against wall-clock time instead of using rounds as in [6]. Therefore, we define the wall-clock time as the sum of the computation time and uplink communication time t_{ul}, whereas the downlink communication time is ignored as mentioned in Section II-B. We conduct experiments with two different local update coefficients of $\tau = 1$ (known as fully synchronous FL) and $\tau = 5$. When we look at the first term of the upper bound in (12), the local update coefficient τ is combined with the uplink time constraint t_{ul}. Hence, when the uplink time constraint is determined, it is necessary to look into the effect of the number of local updates on the learning convergence performance. To verify that the optimal uplink time exists, we evaluate the learning performance given the limited wall-clock time $T = 2,000$ sec with two cases: (i) short uplink time constraint ($t_{ul} = 2.5$ sec) and (ii) long uplink time constraint ($t_{ul} = 10$ sec). In case of short t_{ul}, the global model remains the same as the previous round and only the training time is added if all clients fail to upload their local updates. As for the baseline, we use sufficient uplink time constraint with 60 sec for comparison with the proposed scheme.



(a) (b)

Fig. 2. Test accuracy with different local update coefficients (a) $\tau = 1$ and (b) $\tau = 5$.

B. Experiment results

In this subsection, we demonstrate the experiment results of the proposed *FedUTC* scheme. We plot the test accuracy of the global model to compare the performance. In Fig. 2(a), we plot the test accuracy when the local update is performed once. For both experiments with short and long t_{ul}, we see that the proposed scheme outperforms the baseline, and when the optimal uplink time constraint $t^*_{ul} = 4$ sec is applied, the performance reaches the highest accuracy during the whole training time. Meanwhile, the case with long uplink time $t_{ul} = 10$ sec shows slightly higher accuracy than the one with short uplink time $t_{ul} = 2.5$ sec.

The number of effective participants with long uplink time is larger than that with short uplink time, but the longer communication time leads to slower convergence speed than the optimal case. The case with short uplink time, however, yields poor convergence performance due to the lack of participation from clients because of high outage probability, despite using less communication time. But thanks to the i.i.d. data distribution and randomness of the fading channel, the global model can be trained over various clients, thereby enabling the model to achieve an accuracy of over 95% by utilizing the local updates from various clients during training, despite having a smaller number of clients than the case with long uplink time. In terms of best accuracy, the optimal case reaches the highest at 97.92%, followed by 97.28% for the one with long uplink time, and the lowest accuracy at 96.92% for the one with short uplink time. These observations confirm that the experiment results correspond with the theoretical analysis made in Section III, and applying the optimal uplink time constraint does improve the convergence performance of FL under no perfect CSI.

Fig. 2(b) shows the test accuracy when we perform multiple local updates with $\tau = 5$. We notice that the optimal uplink time has increased to $t^*_{ul} = 5$ sec. Also, we observe that the accuracy gap between the optimal case and the one with long uplink time has been significantly narrowed in the early training stage compared to Fig. 2(a). This is because the local update technique reduces the communication overhead by performing multiple local updates instead of the single update. However, the accuracy gap between the optimal case and the one with short uplink time is still noticeable, despite multiple local updates. Given the same wall-clock time T, the best accuracy has increased for all three cases thanks to multiple local updates. Especially, we see that the best accuracy with long uplink time showed 97.91% which is almost same as the optimal case of 98.04%.

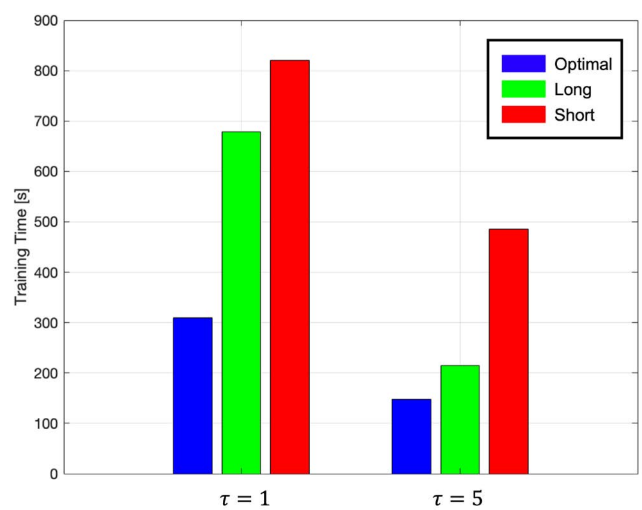


Fig. 3. Time to reach the accuracy of 95% for the three cases with different uplink time constraints.

We present the bar graph in Fig. 3 for comparison of overall convergence performance. Here we measure the training time

TABLE I
SUMMARY OF THE EXPERIMENT RESULTS

	Number of Participants	Best Accuracy [%]	
		$\tau = 1$	$\tau = 5$
Baseline	46.3	93.48	96.75
Long uplink time	28.7	97.28	97.91
Short uplink time	$<$1	96.92	97.38
Optimal uplink time	6.6	**97.92**	**98.04**

to reach the accuracy of 95% as the performance metric. As anticipated, the optimal case achieves the fastest convergence for both cases of different multiple local updates. Fig. 3 shows that the optimal case yields the best convergence speed, which attains almost twice faster than the one with long uplink time when the local update is performed once. When multiple local updates are employed, we notice that the training time has reduced significantly for all three cases. Especially, we see that the time taken to converge has decreased by 68.3% for the case with long uplink time.

We summarize the experiment results with best accuracy in Table I. The number of participants represents the average effective clients who successfully uploaded their local updates within the uplink time. The optimal uplink time achieves the highest accuracy in terms of the best test accuracy during the whole training time. As for the baseline, almost all clients succeed in uploading because of sufficient uplink time, but various channel conditions make the straggler effect worse, showing about 4% lower accuracy than the optimal case given the single update ($\tau = 1$). Also, we observe that about one client can upload in each round with short uplink time while the one with long uplink time enables over half of the total clients. It can be seen that this difference affected not only the convergence speed but also the best accuracy. Further, the best accuracy has increased for all cases thanks to multiple local updates ($\tau = 5$). As a result, we confirm that applying the optimal uplink time constraint to neglect the clients with poor channel conditions and employing multiple local updates both improve the convergence speed of FL.

V. CONCLUSION

In this paper, we have proposed the concept of uplink time constraint for FL over the wireless networks to mitigate the straggler effect. We considered a scenario where the BS knows only the CDF of the fading channel instead of perfect CSI and analyzed the impact of limiting the uploading time to the convergence performance of FL with fixed wall-clock time. To evaluate the convergence performance, we measured the training time in terms of both the computation and uplink time constraint. For this, we formulated an optimization problem to determine the optimal uplink time constraint that minimizes the error upper bound. The experiment results demonstrated that applying the optimal uplink time constraint for a given FL environment does improve the convergence performance while achieving the highest accuracy during the whole training time. Furthermore, we observed that employing multiple local updates accelerates the convergence speed even with the long uplink time constraint.

ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ICT Creative Consilience Program(IITP-2023-2020-0-01821) supervised by the IITP(Institute for Information communications Technology Planning Evaluation).

REFERENCES

[1] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. V. Poor, “Federated learning for Internet of Things: A comprehensive survey,” *IEEE Commun. Surveys & Tutorials*, vol. 23, pp. 1622-1658, Apr. 2021.

[2] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, “Federated learning: Challenges, methods, and future directions,” arXiv:1908.07873, 2019.

[3] N. H. Tran, W. Bao, A. Zomaya, M. N. H. Nguyen, and C. S. Hong, “Federated learning over wireless networks: Optimization model design and analysis,” in *Proc. 2019 IEEE Conf. Comput. Commun. (IEEE INFOCOM)*, 2019, pp. 1387-1395.

[4] C. T. Dinh et al., “Federated learning over wireless networks: Convergence analysis and resource allocation,” *IEEE/ACM Trans. Netw.*, vol. 29, no. 1, pp. 398-409, 2021.

[5] J. Wang and G. Joshi, “Adaptive communication strategies to achieve the best error-runtime trade-off in local update SGD,” in *Proc. Conf. Mach. Learn. Syst.*, vol. 1, 2019, pp. 126-145.

[6] M. K. Nori, S. Yun, and I.-M. Kim, “Fast federated learning by balancing communication trade-offs,” *IEEE Trans. Commun.*, vol. 69, no. 8, pp. 5168-5182, Aug. 2021.

[7] F. Pase, M. Giordani, and M. Zorzi, “On the convergence time of federated learning over wireless networks under imperfect CSI,” in 2021 *IEEE International Conference on Communications Workshops (ICC Workshops)*, Montreal, QC, Canada, 2021, pp. 1-7.

[8] H. Brendan McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, “Communication-efficient learning of deep networks from decentralized data,” *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS*, 2017.

[9] P. Liu, J. Jiang, G. Zhu, L. Cheng,W. Jiang,W. Luo, Y. Du, and Z.Wang, “Training time minimization in quantized federated edge learning under bandwidth constraint,” in *Proceedings of 2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 530-535, 2022.

[10] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, “On the convergence of FedAvg on non-IID data,” in *Proc. ICLR*, Apr. 2020, pp. 1-26.

[11] J. Wang and G. Joshi, “Cooperative SGD: A unified framework for the design and analysis of communication-efficient SGD algorithms,” *arXiv*:1808.07576, 2018.

[12] F. Sattler, S. Wiedemann, K. R. Müller, and W. Samek, “Robust and communication-efficient federated learning from non-IID data,” *IEEE Trans. Neural Networks and Learning Systems*, vol. 31, no. 9, pp. 3400-3413, Sep. 2020.