Modeling and Analysis of Intermittent Federated Learning Over Cellular-Connected UAV Networks

Chun-Hung Liu‡, Di-Chun Liang†, Rung-Hung Gau†, and Lu Wei*

Institute of Communications Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan†
Department of Electrical and Computer Engineering, Mississippi State University, USA‡
Department of Computer Science, Texas Tech University, Lubbock TX, USA*
e-mail: chliu@ece.msstate.edu‡; ldc.cm02g@nctu.edu.tw†; gaurunghung@nycu.edu.tw†; luwei@ttu.edu*

Abstract—Federated learning (FL) is a promising distributed learning technique particularly suitable for wireless learning scenarios since it can accomplish a learning task without raw data transportation so as to preserve data privacy and lower network resource consumption. However, current works on FL over wireless networks do not profoundly study the fundamental performance of FL over wireless networks that suffers from communication outage due to channel impairment and network interference. To accurately exploit the performance of FL over wireless networks, this paper proposes a novel intermittent FL model over a cellular-connected Unmanned Aerial Vehicle (UAV) network, which characterizes communication outage from UAV (clients) to their server and data heterogeneity among the datasets at UAVs. We propose an analytically tractable framework to derive the uplink outage probability and use it to devise a simulation-based approach so as to evaluate the performance of the proposed intermittent FL model. Our findings reveal how the intermittent FL model is impacted by uplink communication outage and UAV deployment. Extensive numerical simulations are provided to show the consistency between the simulated and analytical performances of the proposed intermittent FL model.

Index Terms—Federated learning, deep learning, unmanned aerial vehicle network, outage probability, point process.

I. INTRODUCTION

In the recent years, we have witnessed that machine learning (ML) techniques have been dramatically advanced and successfully applied to tackle many real-world problems. The remarkable success of ML is mainly attributed to two key factors – highly powerful computing and extremely efficient data analytics, yet such a remarkable success in ML significantly relies on whether or not there are enough data to support ML algorithms so as to make them work satisfactorily, which becomes a crucial issue in many ML applications. Due to the proliferation of smart mobile devices, collecting data through them becomes much feasible and easier such that a mobile cellular network has gradually been a huge live database abounding with real-time information, which can be utilized by ML to optimize network operations and managements. Proper and efficient utilization of ML techniques based on data distributed over a massive mobile network becomes an important

issue. This is especially true when transporting raw data from all mobile devices to a server in a massive network because it causes many issues, such as network congestion, energy consumption, privacy, security, etc. To avoid transporting a huge amount of distributed data to a server for conducting centralized ML, a distributed learning methodology without raw data transportation, such as federated learning (FL) [1], [2], [3], becomes a viable solution.

A number of the existing FL algorithms were developed with uniformly compressible data and shown to achieve convergence based on the assumption of error-free and reliable data communications between a server and clients. For example, reference [4] proposed a compression protocol that inherits the compression techniques of top-k sparsification and quantization for uplink and downlink communication in FL. In [5], FL-based multi-access edge computing was studied with limited network resources and it adopted a gradient descent approach to find the optimal trade-off between local update at clients and global aggregation at a server. There are also a number of works in the recent years studying the problem of FL over wireless communication, where many of them approached the problem from the perspective of signal processing. The authors of reference [6], for instance, devised a compressive sensing approach for FL over single-antenna communication systems. The authors of reference [7] proposed a compressive sensing approach for FL over a MIMO communication system, where the server recursively finds the linear minimum-mean-square-error estimate of the transmitted signal by exploiting the sparsity of the signal. In [8], the authors studied the over-the-air computation (AirComp, proposed in [9]) problem with one-bit broadband digital aggregation. Furthermore, very few works studied FL over UAV networks, such as [10].

In these prior works, a fundamental issue of FL over wireless network is far from being fully resolved, that is, data communication between clients and a server may fail due to unreliable wireless transmissions, which leads to communication outage and degrades the convergence performance of FL accordingly. Another crucial issue that was not addressed much in the prior works is the heterogeneity of datasets among different mobile clients. Namely, most of the existing works focus on developing wireless FL algorithms by assuming
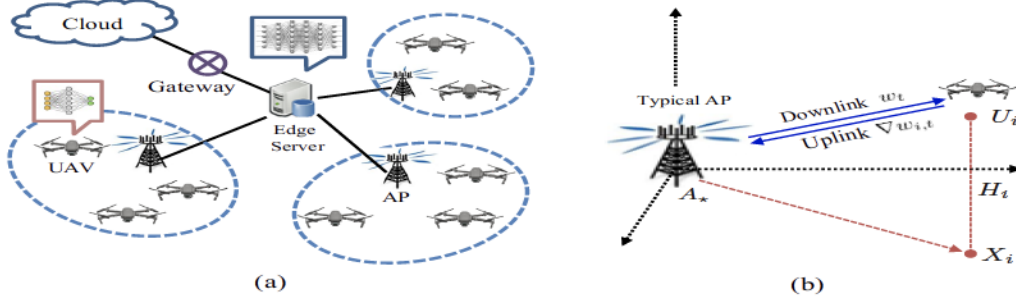
Fig. 1. (a) A cellular-connected UAV network consisting of UAVs (clients), APs, an edge server, and a cloud. Each UAV associates with an AP in order to jointly conduct the proposed intermittent FL with the edge server. (b) A schematic diagram to illustrate a UAV association scenario in which UAV U_i associates with the typical AP located at the origin, denoted by $A_\star$, which sends the global model vector w_t to U_i and receives $\nabla w_{i,t}$ from U_i.

that all mobile clients possess independent and identically distributed (i.i.d.) datasets. To tackle these two issues, we first propose a more realistic FL model over a cellular-connected UAV network that characterizes uplink communication outage from UAV (clients) to an edge server, which is our first contribution. Our second contribution is to propose a 3D random deployment model of UAVs and use it to develop a tractable framework of analyzing the uplink outage probability of a UAV in the network. Our third contribution is to analytically show that the performance of the proposed intermittent FL over a UAV network can be significantly degraded by communication outage from UAVs to the edge server due to the improper deployments of access points (APs) and UAVs in the network. In addition, intensive numerical simulations are conducted to validate our analytical findings.

II. SYSTEM MODEL

A. Model of a Cellular-Connected UAV Network

In this paper, we consider a cellular network consisting of an edge server, a tier of APs, and a tier of flying users, i.e., UAVs, which collect data for learning. The edge server is connected to a cloud through a gateway and it sends data to the cloud whenever it needs the cloud to do large-scale data processing and learning. An illustration of the cellular-connected UAV network is shown in Fig. 1(a). All the UAVs in the network are assumed to be distributed according to the following 3D point process[1]:

$$\Phi_u \triangleq \{U_i \in \mathbb{R}^2 \times \mathbb{R}_+ : U_i = (X_i, H_i), i \in \mathbb{N}_+\}, \quad (1)$$

where U_i, denotes UAV i and its 3D location, $X_i \in \mathbb{R}^2$ is the projection of U_i on the ground, and $H_i \in \mathbb{R}_+$ is the (random) altitude of U_i. All the H_i's are i.i.d. and independent of all the X_i's. Fig. 1(b) shows a typical AP located at the origin and a UAV U_i associating with it. The set of the projections of all the UAVs, i.e., $\{X_i\}$, are assumed to form a 2D independent Homogeneous Poisson Point Process (HPPP) of density λ_u, whereas all the APs also form a 2D independent HPPP of λ_a, which can be expressed as

$$\Phi_a \triangleq \{A_j \in \mathbb{R}^2 : j \in \mathbb{N}_+\}, \quad (2)$$

[1]This 3D point process is a generalization of the 3D point process proposed in our previous work [11] by considering a general distribution of the altitude of each UAV.

where A_j denotes AP j and its location.

Due to the 3D position of a UAV, a wireless link between a UAV and a ground AP can be line-of-sight (LoS) or non-LoS (NLoS). A wireless LoS link between two spatial points means that the link is not visually blocked from one point to the other. For a low-altitude platform of UAV communications, the LoS probability of a wireless link between a UAV and a ground point was proposed in [12]. We adopt it in this paper and express it by using the coordinate system in Fig. 1(b) where a typical AP is located at the origin (denoted by $A_\star$) and a UAV is located at U_i as follows:

$$\rho(\Theta_{i\star}) \triangleq \frac{1}{1 + c_2 \exp(-c_1 \Theta_{i\star})}, \quad (3)$$

where $\Theta_{i\star} \triangleq \tan^{-1}(H_i/\|X_i\|)$ is the elevation angle from the typical AP to UAV U_i, $\|X_i\|$ denotes the distance between X_i and the typical AP, c_1 and c_2 are the environment-related positive coefficients (for rural, urban, etc.).

Each UAV associates with an AP in the network that provides it with the strongest signal power on average. For example, if UAV U_i in Fig. 1(b) associates with the typical AP located at the origin, i.e., $A_\star$, which satisfies the following expression:

$$
\begin{aligned}
A_\star &= \arg \max_{j:A_j \in \Phi_a} L_{ij} \|U_i - A_j\|^{-\alpha} \\
&= \arg \max_{j:A_j \in \Phi_a} L_{ij} \left(\|X_i - A_j\|^2 + \|H_i\|^2\right)^{-\frac{\alpha}{2}} \\
&\overset{(*)}{=} \arg \min_{j:A_j \in \Phi_a} L_{ij}^{-\frac{1}{\alpha}} \|X_i - A_j\| \equiv 0,
\end{aligned}
\quad (4)
$$

where $\alpha > 2$ denotes the path loss exponent, $L_{ij} \in \{1, \ell\}$ is the LoS link gain between U_i and A_j, and $\overset{(*)}{=}$ is due to the fact that $\|H_i\|$ does not affect the result of associating with an AP. L_{ij} is a Bernoulli random variable that is one if the link between U_i and A_j is LoS and $\ell \in (0, 1)$ otherwise. Note that the transmit power of the APs and the fading effect in each wireless link are not considered in (4) because all the APs are assumed to have the same transmit power and the fading effect in each wireless link is averaged out on the receiver side. Moreover, we assume a densely distributed scenario of UAVs (i.e., $\lambda_u \gg \lambda_a$) in the network such that each AP is almost surely associated with at least one UAV, and thereby each AP is able to deliver the signals between its UAVs (clients) and

the edge server for federated learning. A more realistic model of federated learning between the edge server and the UAVs will be proposed in the following subsection.

B. Model of Intermittent Federated Learning

In the UAV network proposed in Section II-A, we aim to study the *realistic* performance of FL over the UAV network that undergoes *communication outages* from a UAV to its (associating) AP, which lead to *non-continuous* FL between the edge server and the UAVs. Thus, we propose an intermittent FL model over the network in Fig. 1(a) as follows. Suppose there are K UAVs distributed on average in the network and they would like to jointly learn a global model with the edge server in T training rounds. To simply characterize the impact of communication outages on FL over the network, we assume FL is much severely impacted by the communication outages in the uplink direction, i.e., from a UAV to its AP. Such an assumption is reasonable since the communication outages in the downlink direction can be significantly mitigated by multiple broadcasting attempts from an AP to its UAVs during the training process of FL. As such, we propose the model of intermittent FL over the cellular-connected UAV network, as detailed in Algorithm 1.

In Algorithm 1, each UAV U_i is assumed to possess a dataset $\mathcal{D}_i$ given by

$$\mathcal{D}_i \triangleq \{D_{ij} = (x_{ij}, y_{ij}), j \in \mathbb{N}_+\}, \; i \in \{1, 2, \ldots, K\}, \quad (5)$$

where D_{ij} denotes data point j in dataset $\mathcal{D}_i$, x_{ij} is the input data vector with an appropriate dimension, y_{ij} is the labeled scalar output corresponding to x_{ij}. We assume that all the datasets $\mathcal{D}_i$'s are non-i.i.d and all the UAVs are able to synchronously update their local learning model. At the tth round of training, UAV U_i updates its local model vector $w_{i,t}$ according to the following algorithm:

$$\begin{cases} w_{i,t} \leftarrow \text{download}_{i \leftarrow S}(w_{t-1}) \\ \nabla w_{i,t} \leftarrow \sum_{j:D_{ij} \in \mathcal{D}_i} \text{SGD}(D_{ij}, w_{i,t}) - w_{i,t} \end{cases}, \quad (6)$$

where $\text{SGD}(D_{ij}, w_{i,t})$ stands for the calculation of Stochastic Gradient Descent with data point D_{ij} and model vector $w_{i,t}$. After receiving the global model vector w_{t-1} from the edge server broadcasted by the AP, UAV U_i updates its local model vector $w_{i,t}$ by SGD so as to find the gradient $\nabla w_{i,t}$. Afterwards, UAV U_i uploads $\nabla w_{i,t}$ to the edge server through the typical AP. To characterize the communication outage from U_i to the typical AP, a Bernoulli random variable $\beta_{i,t} \in \{0, 1\}$ is used to model the uplink communication outage from UAV U_i to the typical AP at the tth round[2]. As a result, the data aggregation algorithm at the edge server can be written as

$$\begin{cases} \nabla w_t \leftarrow \frac{1}{\sum_i \beta_{i,t}|\mathcal{D}_i|} \sum_i \beta_{i,t}|\mathcal{D}_i|\nabla w_{i,t} \\ w_t \leftarrow w_{t-1} + \nabla w_t \end{cases}. \quad (7)$$

Namely, the edge server proportionally combines the received gradients into ∇w_t based on the sizes of the datasets at the

---

[2]Note that the distribution of $\beta_{i,t}$ is affected by the density of the UAVs transmitting at the same time, which will be elaborated in Section III.

Algorithm 1 Intermittent Federated Learning

1: **inputs:** initial model vector w_o
2: **outputs:** improved model vector w
3: **initialize:** the global model is initialized as $w_0 \leftarrow w_o$. Each client holds non-i.i.d. dataset $\mathcal{D}_i$ with equal size. $\nabla w_0, \nabla w_{i,0} \leftarrow 0$.
4: **for** $t = 1, \cdots, T$ **do**
5:     **for** $i \in \{1, \cdots, K\}$ **do in parallel**[3]
6:        client i does:
7:        · $w_{i,t} \leftarrow \text{download}_{i \leftarrow S}(w_{t-1})$
8:        · $\nabla w_{i,t} \leftarrow \sum_{j:D_{ij} \in \mathcal{D}_i} \text{SGD}(D_{ij}, w_{i,t}) - w_{i,t}$
9:        · $\text{upload}_{S \leftarrow i}(\nabla w_{i,t})$
10:     **end for**
11:     edge server does:
12:     · $\nabla w_t \leftarrow \frac{1}{\sum_i \beta_{i,t}|\mathcal{D}_i|} \sum_i \beta_{i,t}|\mathcal{D}_i|\nabla w_{i,t}$
13:     · $w_t \leftarrow w_{t-1} + \nabla w_t$
14:     · $w \leftarrow w_t$
15:     · $\text{broadcast}_{i \leftarrow S}(w_t)$
16: **end for**
17: **return** w

UAVs [2]. As previously pointed out, some local gradients may not be successfully uploaded to the edge server because of uplink communication outage, and thereby the edge server may not be able to aggregate all the local gradients transmitted by all the UAVs. The global model vector w_t at the tth round is updated and then broadcast to all the UAVs in the next round. Such an intermittent FL process between the edge server and the K UAVs proceeds until the predesignated number T of training rounds is reached.

To illustrate how the proposed intermittent FL in Algorithm 1 is impacted by uplink communication outage, we will first provide the analyses of the outage probability of the uplink communications from a UAV to its AP in the following section. Next, some simulation results regarding intermittent FL will be provided in Section IV-A to numerically demonstrate how the performance of intermittent FL is impacted by the uplink communication outages.

III. ANALYSIS OF THE UPLINK OUTAGE PROBABILITY

Suppose the network is interference-limited and consider the uplink communication scenario from U_i to the typical AP $A_\star$ shown in Fig. 1(b). As such, the Signal-to-Interference Ratio (SIR) at $A_\star$ can be defined as

$$\gamma_{i\star} \triangleq \frac{G_{i\star}L_{i\star}\|A_\star\|^{-\alpha}}{I_\star}, \quad (8)$$

where $G_{i\star}$ is the fading channel gain from U_i to $A_\star$. Accordingly, the uplink outage probability of a UAV is defined as

$$p_{out} \triangleq \mathbb{P}[\gamma_{i\star} \leq \eta] = \mathbb{P}[\beta_{i,t} = 0], \; \text{for all } i \text{ and } t, \quad (9)$$

where $\eta > 0$ is the SIR threshold for successful decoding. In this section, we focus on the analysis of p_{out} that affects the performance of the intermittent FL model proposed in Section II-B. To facilitate the derivation of p_{out}, we first need

to introduce two related theorems. The first theorem stated in the following is about the distribution of the path loss of a wireless link from a UAV to its AP.

Theorem 1. *Suppose UAV U_i associates with the typical AP located at the origin. If $R_{i\star}(r) = \mathbb{P}[L_{i\star}\|U_i\|^{-\alpha} \leq r]$, then for UAV U_i with a given altitude H_i it can be found as given by*

$$R_{i\star}(r) = \exp\left[-\pi\lambda_a \Upsilon_i(r)\right], \quad (10)$$

where $\Upsilon_i(\cdot)$ is defined as

$$\Upsilon_i(r) \triangleq \int_{((\frac{\ell}{r})^{-\frac{2}{\alpha}} - H_i^2)^+}^{(r^{-\frac{2}{\alpha}} - H_i^2)^+} \rho\left(\vartheta_i(y)\right) \mathrm{d}y + \left(\left(\frac{\ell}{r}\right)^{-\frac{2}{\alpha}} - H_i^2\right)^+ \quad (11)$$

in which $(x)^+ \triangleq \max\{0, x\}$ and $\vartheta_i(y) \triangleq \tan^{-1}(H_i/\sqrt{y})$.

Proof: See Appendix A. ∎

Note that $\Upsilon_i(r)$ is a decreasing function of r when the upper limit of the integral in (11) is not zero. From the expression in (11), we thus know how $R_{i\star}(r)$ varies with H_i, λ_a, and ℓ.

Next, we need to analyze the interference received by the typical AP. To make the analysis tractable, we assume all the UAVs associating with the same AP do not use the same resource blocks in the uplink so that the UAVs using the same resource blocks in the network also form a 3D point process. Let $I_\star$ be the interference received by the typical AP and it can be defined as

$$I_\star \triangleq \sum_{k: U_k \in \widetilde{\Phi}_u \backslash U_i} G_{k\star} L_{k\star}\|U_k\|^{-\alpha}, \quad (12)$$

where $\widetilde{\Phi}_u$ is the set of all the UAVs using the same resource block as U_i and $G_{k\star} \sim \exp(1)$ that denotes an exponential random variable with unit mean is the fading channel gain from UAV $U_k \in \widetilde{\Phi}_u$ to the typical AP and and independent of any other random variables in $I_\star$ for all $k \in \mathbb{N}_+$. Hence, all $G_{k\star}$'s are i.i.d. In the following theorem, we specifically show the Laplace transform of $I_\star$, which is defined as $\mathcal{L}_{I_\star}(s) \triangleq \mathbb{E}[\exp(-sI_\star)]$ for $s > 0$.

Theorem 2. *The Laplace transform of $I_\star$ can be found as*

$$\mathcal{L}_{I_\star}(s) = \exp\left(-\pi\lambda_a \int_0^\infty \mathcal{I}_G\left(sy^{-\frac{\alpha}{2}}, \vartheta(y)\right) \mathrm{d}y\right), \quad (13)$$

where $\vartheta(y) \triangleq \tan^{-1}(\frac{H_k}{\sqrt{y}})$ and $\mathcal{I}_G(u, w)$ for $u, w > 0$ is defined as

$$\mathcal{I}_G(u, w) \triangleq \rho(w)[1 - \mathbb{E}\{\mathcal{L}_G(u\cos^\alpha(w))\}]$$
$$+ (1 - \rho(w))[1 - \mathbb{E}\{\mathcal{L}_G(\ell u\cos^\alpha(w))\}] \quad (14)$$

in which $\mathcal{L}_G(s) \triangleq \mathbb{E}[\exp(-sG_{k\star})]$.

Proof: See Appendix B. ∎

From Theorem 2, we are able to learn how the statistical properties of the uplink interference is affected by the deployments of UAVs and APs. Note that (13) is a function of λ_a since the density of the UAVs using the same uplink resource blocks is also the same as that of the APs.

Parameter	Value
UAV Density λ_u (UAVs/m^2)	1×10^{-5}
AP Density λ_a (APs/m^2)	$\lambda_u/[10, 300]$
UAV Height H_i (m)	100
SIR Threshold η	0.5
Path-loss Exponent α	2.75
(c_1, c_2) in (3) for urban	(0.1581, 43.9142)
Attenuation Gain of NLoS Channels ℓ	0.25
Size of Training Dataset at each UAV $|D_i|$	20
Average Number of UAVs (clients) K	10, 30, 50
Number of Training Rounds T	200

Using the results in Theorems 1 and 2, we can derive the uplink outage probability as shown in the following theorem.

Theorem 3. *According to the uplink outage probability defined in (9), it can be explicitly found as*

$$p_{out} = 1 - \mathbb{E}\left\{\exp\left(-\pi\lambda_a \int_0^\infty \mathcal{I}_G\left(\frac{\eta y^{-\frac{\alpha}{2}}}{R'_{i\star}}, \vartheta_i(y)\right) \mathrm{d}y\right)\right\}, \quad (15)$$

where $R'_{i\star}(r) \triangleq \frac{\mathrm{d}R_{i\star}(r)}{\mathrm{d}r} = -\pi\lambda_a \Upsilon'_i(r) \exp\left(-\pi\lambda_a \Upsilon_i(r)\right)$.

Proof: According to p_{out} in (9), it can be rewritten as

$$p_{out} = \mathbb{P}\left[G_{i\star} \leq \frac{\eta I_\star}{L_{i\star}\|A_\star\|^{-\alpha}}\right] = 1 - \mathcal{L}_{I_\star}\left(\frac{\eta}{L_{i\star}\|A_\star\|^{-\alpha}}\right).$$

By employing the results in Theorems 1 and 2 to the above expression of p_{out}, the explicit result of $\mathcal{L}_{I_\star}(\cdot)$ can be found so that p_{out} in (15) is readily obtained. ∎

Theorem 3 clearly indicates how p_{out} relates to λ_a and ℓ, which reveals the fact that the distribution of $\beta_{i,t}$ is affected by λ_a and ℓ as well. In other words, the realistic performance of FL over a UAV network depends on how densely UAVs are distributed in the network and how they are positioned in the sky. In the following section, some numerical results will be provided to demonstrate this observation.

IV. NUMERICAL RESULTS AND DISCUSSIONS

This section provides some simulation results of the proposed intermittent FL by considering the MNIST datasets of handwritten digits stored at the UAVs. We numerically evaluate the performance of the proposed intermittent FL by using the metric of *learning accuracy*, which is defined as the rate of using the global model w learned by the proposed intermittent FL to successfully classify the images of handwritten digits in the entire dataset stored in the network[4]. We first present and discuss the numerical results of the learning accuracy of the proposed intermittent FL over the network and afterwards we show the numerical results of how the learning accuracy is influenced by the deployment densities of the APs and UAVs. The values of the network parameters used for simulation are listed in Table I.

---

[4]Thus, the entire dataset in the network is the union of all the local datasets stored at the UAVs.
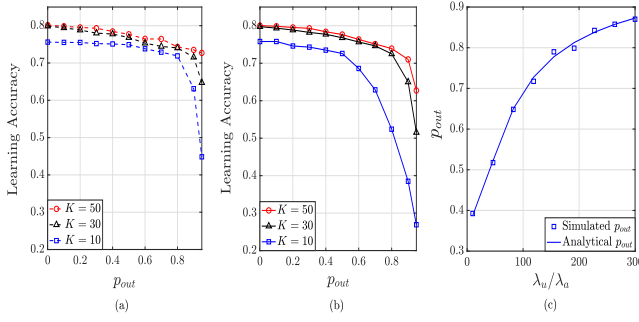
Fig. 2. The numerical results of the proposed intermittent FL: (a) Uplink outage probability versus learning accuracy for different numbers of UAVs with i.i.d. datasets; (b) Uplink outage probability versus learning accuracy for different numbers of UAVs with non-i.i.d. datasets; (c) Uplink outage probability versus the ratio of the UAV density to the AP density.

Fig. 3. Simulation results of the learning accuracy of the proposed FL model for the case of non-i.i.d. datasets at the UAVs.

A. Numerical Results of the Proposed FL Model

In this subsection, we provide numerical results regarding how the performance of the proposed intermittent FL model is influenced by uplink communication outages. All the uplink channels are assumed to experience independent block fading such that all $\beta_{i,t}$'s are i.i.d. for all i and t. Also, the datasets at the UAVs are assumed to be of the same quality and size. To make the learning processes at different UAVs consistent, all the UAVs adopt the same architecture of a deep neural network to perform local learning, that is, the neural networks at different UAVs adopt the same batch sizes, the same number of the neurons in each layer, and the same number of hidden layers. The learning cases of i.i.d. and non-i.i.d. datasets are both considered in the simulation. Each UAV collects a dataset where the ratio of training data size to testing data size is 2:1. To make each learning case fairly compared, all the ten image classes are uniformly collected by the UAVs. For example, if there are 50 UAVs involved in the training processing of FL, the image class of handwritten digit "3" is collected by exactly 5 UAVs among the 50 UAVs, which happens to the other nine image classes likewise.

The numerical results of the proposed intermittent FL are shown in Fig. 2. Specifically, Figs. 2 (a) and (b) show how the learning accuracy varies with p_{out} for the different numbers of the UAVs with i.i.d. and non-i.i.d. datasets, respectively. As can be seen in Fig. 2(a), the learning accuracy reduces as p_{out} increases, yet it improves as the average number of the UAVs in the network increases. As p_{out} increases, the uplink communication outage is more likely to happen and thus the edge server more likely does the global model aggression with less local learning outcomes, which essentially slows down the convergence process of FL. As a result, the global model vector w_t is less likely to converge to a stable vector within T training rounds. This is why the learning accuracy reduces as p_{out} increases. When more UAVs with i.i.d. datasets join FL, more local training outcomes with a similar statistical distribution are likely to be aggregated at the edge, which improves the convergence rate of FL. Therefore, increasing the average number of the UAVs with i.i.d. datasets in the
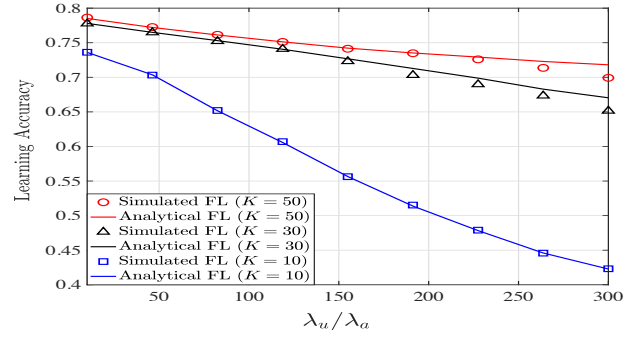
network helps to improve the learning accuracy because it mitigates the negative impact of the uplink communication outage on the convergence process of FL. Moreover, Fig. 2(b) reveals a phenomenon different from Fig. 2(a), that is, the learning accuracy is more sensitive to the average number of the UAVs with non-i.i.d. datasets in the network. For example, when $p_{out} = 0.6$, increasing K from 10 to 50 significantly improves the learning accuracy, whereas we cannot observe this in Fig. 2(a). This phenomenon stems from the fact that *every* local learning outcome is crucial to the global model aggregation at the edge server in that the local learning outcomes obtained from non-i.i.d. datasets may have a very distinct statistical distribution. Hence, the negative impact of non-i.i.d. datasets on the convergence process of FL can be effectively mitigated by increasing the average number of the UAVs. Fig. 2(c) presents how p_{out} increases along with λ_u/λ_a. When λ_u/λ_a increases, more UAVs are in the network and more uplink interference is generated, thereby making uplink communication outage more likely occur. The results in this figure provide a fundamental relationship between p_{out} and λ_u/λ_a and they can be used together with the results in Figs. 2(a) and (c) to provide some insight into how to deploy APs and UAVs in order to achieve a desired learning accuracy for a given uplink outage probability. For instance, we can deploy APs and UAVs with a ratio of $\lambda_u/\lambda_a \approx 100$ in order to achieve a learning accuracy about 0.75 by deploying the average number of the UAVs with non-i.i.d. datasets greater than 30.

B. Numerical Results of the Learning Accuracy vs. λ_u/λ_a

Fig. 3 shows how the learning accuracy varies with λ_u/λ_a when all the UAVs have non-i.i.d. datasets. Two cases of simulated FL and analytical FL are presented in the figure. The results of the simulated FL are completely obtained by running numerical simulations, yet the results of the analytical FL are obtained by first calculating p_{out} based on the expression in (15) for different values of λ_u/λ_a and then using Fig. 2(b) to find the values of learning accuracy corresponding to the calculated values of p_{out}. As can be seen in the figure, the results of analytical FL almost coincide with their corresponding results of simulated FL. This reveals the correctness and accuracy of the uplink outage probability found in (15).

Furthermore, the results in Fig. 3 demonstrate the fact that the performance of FL over a UAV network can indeed be impacted by how densely the APs and UAVs are deployed in the network. This is an important and interesting finding worth further investigation.

V. CONCLUSION

In the literature, the studies of FL over wireless network were mainly conducted based on a unrealistic assumption, i.e., no communication outage between clients and a server when conducting FL. Such studies cannot practically reflect the accurate performance of FL over wireless networks. To understand how communication outage impacts FL, this paper proposed an intermittent FL model that is able to characterize uplink communication outages in a cellular-connected UAV network. A tractable approach to analyzing the uplink outage probability was proposed and the uplink outage probability of a UAV was explicitly derived in a neat form. We found that the performance of FL over a UAV network can be significantly impacted by the uplink outage probability that depends on how the APs and UAVs are deployed in the network. Numerical results not only validate the accuracy of the analyses of the uplink outage probability, but also support the finding on how the uplink outage probability degrades the the performance of FL over unreliable wireless networks.

APPENDIX

A. Proof of Theorem 1

According to the definition of $R_{i\star}(r)$, $R_{i\star}(r)$ for a given H_i can be further written as

$$
\begin{aligned}
R_{i\star}(r) &\overset{(a)}{=} \mathbb{P}\left[\max_{j:A_j\in\Phi_a}\left\{\frac{L_{ij}}{[\|X_i - A_j\|^2 + H_i^2]^{\frac{\alpha}{2}}}\right\} \le r\right] \\
&\overset{(b)}{=} \mathbb{E}\left\{\prod_{j:A_j\in\Phi_a}\mathbb{P}\left[\frac{L_{ij}}{[\|A_j\|^2 + H_i^2]^{\frac{\alpha}{2}}} \le r\right]\right\} \\
&\overset{(c)}{=} \exp\left(-2\pi\lambda_a\int_0^\infty \mathbb{P}\left[\frac{L}{[x^2 + H^2]^{\frac{\alpha}{2}}} > r\right]x\,\mathrm{d}x\right),
\end{aligned}
$$

where (a) is obtained based on the UAV association scheme in (4), (b) is obtained by considering X_j as the origin and the fact that all $L_{ij}/(\|A_j\|^2 + H_j^2)^{\alpha/2}$'s are independent, and (c) is obtained by applying probability generating functional (PGFL) of an HPPP to Φ_a. The subscript ij of L_{ij} is dropped for simplifying notation. Replacing x^2 with y and letting $\vartheta_y = \tan^{-1}(H_j/\sqrt{y})$ yield the following:

$$
\begin{aligned}
&\mathbb{P}\left[\frac{L}{[y + H_i^2]^{\frac{\alpha}{2}}} > r\right] = \rho(\vartheta_i(y))\mathbb{P}\left[\frac{1}{(y + H_i^2)^{\frac{\alpha}{2}}} > r\right] \\
&+ [1 - \rho(\vartheta_i(y))]\mathbb{P}\left[\frac{\ell}{(y + H_i^2)^{\frac{\alpha}{2}}} > r\right] = \rho(\vartheta_i(y))\times \\
&\mathbb{P}\left[y < r^{-\frac{2}{\alpha}} - H_i^2\right] + [1 - \rho(\vartheta_i(y))]\mathbb{P}\left[y < \left(\frac{\ell}{r}\right)^{\frac{2}{\alpha}} - H_i^2\right].
\end{aligned}
$$

Thus, $\Upsilon_i(r) = \int_0^\infty \mathbb{P}\left[L(y + H_i^2)^{-\frac{\alpha}{2}} > r\right]\mathrm{d}y$ and $R_{i\star}(r)$ are obtained accordingly.

B. Proof of Theorem 2

According to the definition of $I_\star$ and $\|U_k\|^2 = \|X_k\|^2 + \|H_k\|^2$, the Laplace transform of $I_\star$ can be further written as

$$
\begin{aligned}
\mathbb{E}\left[e^{-sI_\star}\right] &= \mathbb{E}\left[\prod_{i:U_i\in\widetilde{\Phi}_u\setminus U_\star}\exp\left(-\frac{sG_{k\star}L_{k\star}}{(\|X_k\|^2 + H_k^2)^{\frac{\alpha}{2}}}\right)\right] \\
&\overset{(a)}{=} \exp\left(-\pi\lambda_a\int_0^\infty\left\{1 - \mathbb{E}\left[e^{-\frac{sGL}{(y+H^2)^{\alpha/2}}}\right]\right\}\mathrm{d}y\right) \\
&\overset{(b)}{=} \exp\left(-\pi\lambda_a\int_0^\infty\left\{1 - \mathbb{E}\left[\mathcal{L}_G\left(\frac{sL}{(y + H^2)^{\frac{\alpha}{2}}}\right)\right]\right\}\mathrm{d}y\right),
\end{aligned}
$$

where (a) is obtained by applying the PGFL of an HPPP to the projections of $\widetilde{\Phi}_u$ and using G instead of $G_{k\star}$. Since $\vartheta(y) = \tan^{-1}(H/\sqrt{y})$, we thus have

$$
\begin{aligned}
&\mathbb{E}\left[\mathcal{L}_G\left(\frac{sL}{(y + H^2)^{\frac{\alpha}{2}}}\right)\right] = \mathbb{E}\left[\mathcal{L}_G\left(sLy^{-\frac{\alpha}{2}}\sec^\alpha(\vartheta(y))\right)\right] = \\
&\rho(\vartheta_i(y))\mathbb{E}\left[\mathcal{L}_G\left(sy^{-\frac{\alpha}{2}}\sec^\alpha(\vartheta(y))\right)\right] + [1 - \rho(\vartheta_i(y))] \\
&\times \mathbb{E}\left[\mathcal{L}_G\left(sy^{-\frac{\alpha}{2}}\ell\sec^\alpha(\vartheta(y))\right)\right].
\end{aligned}
$$

REFERENCES

[1] X. Wang, Y. Han, C. Wang, Q. Zhao, X. Chen, and M. Chen, "In-Edge AI: Intelligentizing mobile edge computing, caching and communication by federated learning," *IEEE Netw.*, vol. 33, no. 5, pp. 156–165, Jul. 2019.

[2] Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, and H. Yu, *Federated Learning*, 1st ed. Morgan and Claypool, 2019.

[3] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities, and challenges," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 46–51, Jun. 2020.

[4] F. Sattler, S. Wiedemann, K. R. Müller, and W. Samek, "Robust and communication-efficient federated learning from non-i.i.d. data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 9, pp. 3400–3413, Sep. 2020.

[5] S. Wang, T. Tuor, T. Salonidis, K. K. Leung, C. Makaya, T. He, and K. Chan, "Adaptive federated learning in resource constrained edge computing systems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1205–1221, Mar. 2019.

[6] M. M. Amiri and D. Gündüz, "Federated learning over wireless fading channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3546–3557, Feb. 2020.

[7] Y.-S. Jeon, M. M. Amiri, J. Li, and H. V. Poor, "A compressive sensing approach for federated learning over massive MIMO communication systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1990–2004, Nov. 2021.

[8] G. Zhu, Y. Du, D. Gündüz, and K. Huang, "One-bit over-the-air aggregation for communication-efficient federated edge learning: Design and convergence analysis," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 2120–2135, Nov. 2021.

[9] G. Zhu, Y. Wang, and K. Huang, "Broadband analog aggregation for low-latency federated edge learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 491–506, Oct. 2020.

[10] Q.-V. Pham, M. Zeng, R. Ruby, T. Huynh-The, and W.-J. Hwang, "UAV communications for sustainable federated learning," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3944–3948, Mar. 2021.

[11] C.-H. Liu, D.-C. Liang, and R.-H. Gau, "A 3D tractable model for UAV-Enabled cellular networks with multiple antennas," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3538–3554, Jun. 2021.

[12] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 569–572, Dec. 2014.