

Optimizing Federated Learning in LEO Satellite Constellations via Intra-Plane Model Propagation and Sink Satellite Scheduling

Mohamed Elmahallawy, Tie Luo*

Computer Science Department, Missouri University of Science and Technology, USA

E-mail: {meqxk, tluo}@mst.edu

Abstract—The advances in satellite technology developments have recently seen a large number of small satellites being launched into space on Low Earth orbit (LEO) to collect massive data such as Earth observational imagery. The traditional way which downloads such data to a ground station (GS) to train a machine learning (ML) model is not desirable due to the bandwidth limitation and intermittent connectivity between LEO satellites and the GS. Satellite edge computing (SEC), on the other hand, allows each satellite to train an ML model onboard and uploads only the model to the GS which appears to be a promising concept. This paper proposes FedLEO, a novel federated learning (FL) framework that realizes the concept of SEC and overcomes the limitation (slow convergence) of existing FL-based solutions. FedLEO (1) augments the conventional FL’s star topology with “horizontal” intra-plane communication pathways in which *model propagation* among satellites takes place; (2) optimally schedules communication between “sink” satellites and the GS by exploiting the *predictability* of satellite orbiting patterns. We evaluate FedLEO extensively and benchmark it with the state of the art. Our results show that FedLEO drastically expedites FL convergence, without sacrificing—in fact it considerably increases—the model accuracy.

I. INTRODUCTION

With the fast developments of satellite technology, a large number of low Earth orbit (LEO) satellites have been launched into space. Among the active players in LEO satellite deployment are government agencies such as NASA and ESA, and large companies such as DigitalGlobe, SpaceX (Starlink), and OneWeb. They offer 1) Earth observational and monitoring services or 2) broadband Internet services. For the first category which this paper focuses on, a massive amount of data such as high-resolution images are continuously collected by LEO satellites. Performing data analytics on such big data using AI techniques in order to derive insights and inform decision-makers has huge value and implications, for example in disaster early warning [1] and urban planning [2]. In the traditional school of thought, we can download all these images to a central server such as a ground station (GS), and then train some complex machine learning (ML) models to perform data analytics. However, this is not desirable in satellite communication (Satcom) because of 1) the network bandwidth limitation [2], 2) the large propagation delay between satellites and the GS [3], 3) the short and irregular visibility period between satellites and the GS [4], and 4) the need for low-latency tracking of space missions [5].

Federated Learning (FL) [6] offers a promising solution to this problem. Instead of downloading raw data, FL realizes the concept of satellite edge computing (SEC) by enabling each satellite to train a local ML model onboard and upload only the model parameters to the GS to aggregate into a global model. Although this solves the aforementioned bandwidth issue, the learning

process of FL will incur a large delay in Satcom because FL is an iterative process that requires many communication rounds between clients (LEO satellites) and the central server (GS) for training. In Satcom, LEO satellites and the GS move in distinct directions¹ and speeds, which leads to highly *intermittent and irregular visibility* between them, and thus each communication round takes significantly longer and the entire training process takes several days or even weeks to converge [7].

In this paper, we propose FedLEO, a novel FL framework for LEO satellite constellations that accelerates the FL training process in Satcom. FedLEO consists of two novel components:

- 1) We propose a *model propagation* scheme that augments the conventional “vertical” star topology between the FL server and LEO satellites with “horizontal” intra-plane communication pathways among satellites;
- 2) We propose a *distributed scheduling* algorithm that selects an optimal “sink” satellite on each orbit to generate a *partial global model*, and optimally schedules communication between sink satellites and the GS by exploiting the *predictability* of satellite orbiting patterns.

In addition, this paper also makes the following contributions:

- 3) FedLEO addresses the challenging issue of data discrepancy between satellites on different orbits by taking care of non-independent and identically distributed (non-IID) data.
- 4) We demonstrate via extensive simulations that FedLEO drastically accelerates FL convergence by *more than an order of magnitude* than state-of-the-art FL-Satcom methods. Moreover, despite the much shorter time, FedLEO achieves much higher accuracy than those benchmarks.

II. RELATED WORK

In general, FL can be categorized into synchronous and asynchronous approaches. We discuss these two categories in the context of Satcom.

In synchronous FL, the central server must wait to receive all the clients’ models and then aggregate them into a global model. As a result, slow clients, often referred to as “stragglers”, will become the bottleneck of the training process. A few recent studies have attempted this synchronous approach in Satcom. For example, [8] simulated the vanilla FL (i.e., FedAvg [6]) in LEO constellations and show that FL is more advantageous than directly downloading raw data to GS to perform centralized training. Razmi et al. proposed FedISL [3] to leverage inter-satellite links (ISL) to reduce delay. They use an ideal setup in which the server is either a GS located at the North Pole (NP) or a

¹A GS always rotates on the 0° plane (in reference to the Equator), while a satellite flies in an orbit of inclination angle between 0-90°. Thus, there are very irregular and large delays between successive visits of a satellite to the same GS.

*Corresponding author.

medium Earth orbit (MEO) satellite located above the Equator (at an altitude of 20,000 km), so that each satellite visits the server at regular intervals (i.e., no longer irregularly) and much more often. This ideal setup is rarely available in practice. A most recent approach called FedHAP [2] was proposed which substitutes the traditional GS by one/multiple high altitude platforms (HAPs) floating at 20-30 km above the Earth's surface to act as servers. Although the performance improves, it requires extra hardware.

On the other hand, an asynchronous FL approach allows the server to proceed to the next training round with just a *subset* of the clients' model updates received, instead of waiting for all. This mitigates the bottleneck problem existing in synchronous FL, but it faces a *staleness* problem where some received model updates in a certain round may come from earlier rounds (due to stragglers). FedSat [9] is such an asynchronous FL approach applied to Satcom, and it again assumes the above ideal setup where the GS is located at NP to avoid the challenge of irregular satellite visit patterns. Another work [10] attempts to reduce model staleness by calculating whether the visible interval of each satellite is sufficient for global model downloading, local model training, and uploading; if not, it will schedule local model uploading to the next communication round so that the satellite can train local model during the (long) invisible interval. However, this has only a limited effect on efficiency, which still requires several days to finish training a global model. So et al. proposed another asynchronous FL algorithm called FedSpace [7], which collects client models into a buffer with a predicted size, and down-weights stale models. However, it requires each satellite to upload a small portion of local data to the GS for aggregation scheduling, which violates the FL principle of privacy protection by avoiding raw data sharing. Recently, we proposed AsyncFLEO [4], which tackles the model staleness issue and achieves fast convergence within a few hours. However, it still requires HAPs to serve as servers like FedHAP [2]. Different from all the above, FedLEO does not assume ideal setups, violate any FL principle, or introduce any extra hardware (e.g., HAP). While it is a synchronous protocol, FedLEO achieves the lowest FL convergence delay in Satcom yet improves accuracy at the same time, in a realistic setting including irregular satellite visiting patterns, bandwidth limitation, and non-IID data.

III. SYSTEM MODEL

As illustrated in Fig. 1, we consider a generic LEO satellite constellation \mathcal{K} consisting of L orbits with each having K equally distributed satellites. An orbit $l \in \mathcal{L} = \{l_1, l_2, \dots, l_L\}$ contains its own set of satellites \mathcal{K}_l , and is characterized by its inclination angle α_l and altitude h_l . Each satellite in orbit l travels at the same speed v_l and has the same orbital period T_l . Here, $v_l = \sqrt{\frac{GM}{(R_E + h_l)}}$ and $T_l = \frac{2\pi}{\sqrt{GM}}(R_E + h_l)^{3/2}$, where G denotes the gravitational constant; M denotes the mass of the Earth; and $R_E = 6371$ km is the radius of the Earth. Satellites and the GS typically use a radio frequency (RF) link that is more reliable than free-space optical (FSO) for long-distance communication. Thus, any satellite $k \in \mathcal{K}$ will be visible to a GS g when the line of sight (LoS) link is not blocked by the Earth. This can be translated mathematically to the condition $\vartheta_{k,g}(t) \triangleq \angle(r_g(t), (r_k(t) - r_g(t))) \leq \frac{\pi}{2} - \vartheta_{min}$, where $r_k(t)$

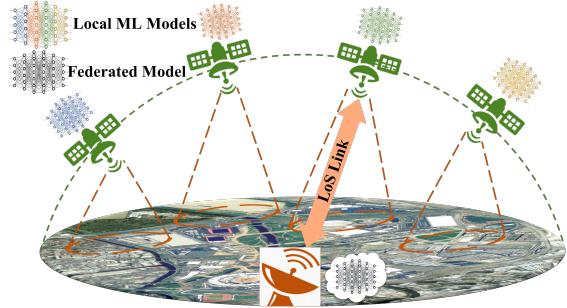


Fig. 1: FL in the context of an LEO satellite constellation, with a GS serving as the federation server. Only one orbit is shown for clarity.

and $r_g(t)$ represent the trajectory of a satellite k and a GS g , respectively, and ϑ_{min} represents the minimum elevation angle, a constant depending on the GS position.

A. Computation Model

Consider an FL-LEO system that collects data such as Earth observational imagery. Each satellite k in the constellation \mathcal{K} captures a set of images, which we assume to be non-IID due to their different orbits and positions. Let $\mathcal{D}_k = \{(X_{k,1}, y_{k,1}), (X_{k,2}, y_{k,2}), \dots, ((X_{k,m_k}, y_{k,m_k})\}$ represents the local dataset collected by satellite k , where $X_{k,j}$ denotes the feature vector of the j -th sample, and $y_{k,j}$ denotes its corresponding label (for notation purposes only; labels are not required), and $m_k = |\mathcal{D}_k|$ is the dataset size. The objective of the FL system is to train an ML model by the GS and the satellites collaboratively; specifically, to solve

$$\arg \min_{w \in \mathbb{R}^d} F(w) = \sum_{k \in \mathcal{K}} \frac{m_k}{m} F_k(w), \quad (1)$$

where $F(w)$ is the overall loss function of the model; w denotes the model weights; and $m = \sum_{i \in \mathcal{K}} m_k$ is the total data size. $F_k(w)$ below is the loss function of satellite k

$$F_k(w) = \frac{1}{m_k} \sum_{j=1}^{m_k} f_k(w; X_{k,j}, y_{k,j}), \quad (2)$$

where $f_k(\cdot)$ is the training loss on a sample $X_{k,j}$.

In a sync FL system such as FedAvg [6], the training process takes multiple communication rounds $t = 0, 1, 2, \dots, T$. In each round, the GS broadcasts the global model w^t to all satellites that are in its visible zone. Each satellite k then applies a local optimization method such as stochastic gradient descent (SGD) for I local epochs as

$$w_k^{t,i+1} = w_k^{t,i} - \eta \nabla F_k(w_k^{t,i}; X_k^i, y_k^i), \quad i = 0, 1, 2, \dots, I-1 \quad (3)$$

where η is the learning rate. Then, the satellites transmit their updated model weights (or gradients) to the GS, when entering its visible zone. Once the GS receives all satellite models from all the orbits, it lumps them into a global model as follows:

$$w^{t+1} = \sum_{k \in \mathcal{K}} \frac{m_k}{m} w_k^{t,I}, \quad (4)$$

The above procedure iterates with an incremental t until the FL model converges (e.g., a target loss, target accuracy, or the maximum number of training rounds T is reached).

A key problem with this learning process is that all the communications (upload/download) can only happen when a

satellite transiently comes into the GS's visible zone, which causes the entire process to take days or even weeks. This also implies that the learning speed is not able to catch up with the data generation or collection speed. As a result, the model will always be out-dated.

B. Communication Model

The signal-to-noise ratio (SNR) between any satellite k and the GS on a symmetric RF channel with additive white Gaussian noise (AWGN) can be defined as [4]

$$SNR(k, GS) = \frac{P_t G_k G_{GS}}{K_B T B L_{k, GS}}, \quad (5)$$

where P_t is the sender power, G_k and G_{GS} are the total antenna gains of the k -th satellite and the GS, respectively, K_B is the Boltzmann constant, T is the noise temperature at the receiver, B is the channel bandwidth, and $L_{k, GS}$ denotes the free-space pass loss between a satellite k and the GS. As long as the LoS link between the satellite k and the GS is established (i.e., Earth does not block the communication between them), then we can express $L_{k, GS}$ as follows:

$$L_{k, GS} = \left(\frac{4\pi \|k, GS\|_2 f}{c} \right)^2 \quad (6)$$

where $\|k, GS\|_2$ is the Euclidean distance between the k -th satellite and the GS if LoS is achieved, f is the carrier frequency, and c is the speed of the light. For exchanging local or global model weights (w_k or w) between a satellite k and the GS, the total required time t_c can be calculated as

$$t_c = t_t + t_p + t_k + t_{GS}, \quad t_t = \frac{z|\mathcal{N}|}{R}, \quad t_p = \frac{\|k, GS\|_2}{c}, \quad (7)$$

where t_t and t_p are the transmission and propagation times, respectively, t_k and t_{GS} are the processing delay at k -th satellite and the GS, respectively (we omit them in our simulation since they are much smaller than t_t and t_p), $|\mathcal{N}|$ is the number of data samples, z is the number of bits in each sample, and R is the maximum achievable data rate, which can be approximated by the Shannon formula as

$$R \approx B \log_2(1 + SNR(k, GS)) \quad (8)$$

IV. FEDLEO FRAMEWORK

FedLEO consists of two main components, model propagation, and distributed scheduling, which we describe below.

A. Model Propagation

The model propagation scheme augments the conventional “vertical” star topology between the FL server and LEO satellites with “horizontal” intra-plane ISL pathways among satellites to improve efficiency. Fig. 2 depicts this scheme. On any orbit l , once a visible satellite k receives the global model w^t from the GS, the satellite k will send the model to k 's neighbors via intra-plane ISL², and each of these neighbors will further forward this model to its next-hop neighbor. In the meantime, each satellite on l also starts to train its local copy w^t upon receiving it (so there

²A LEO satellite has four antennas, two on the pitch axis and two on the roll axis for inter-plane and intra-plane communications, respectively.

will be multiple concurrent training processes). During model propagation on an orbit, a satellite might receive w^t twice from its neighbors (e.g., when more than one satellite is visible to the GS), which does not matter since it can simply drop the duplicate.

After training, Fig. 2c depicts the process of relaying the trained local models (represented by different colors) to a “sink” satellite via one or multiple ISL hops. The sink is the only one on each orbit that is responsible for *partial aggregation of models* and *sending the partial model back to the GS*, and is selected by a scheduler running on each satellite (so every satellite knows who is the sink) and explained in the next subsection. Upon having collected all the trained models from the same orbit l , the sink will aggregate them to generate a *partial global model* $w_{\mathcal{K}_l}$ as follows:

$$w_{\mathcal{K}_l} = \sum_{k \in \mathcal{K}_l} \frac{m_k}{m_{\mathcal{K}_l}} w_k^I, \quad \text{where } m_{\mathcal{K}_l} = \sum_{k \in \mathcal{K}_l} m_k. \quad (9)$$

After that, the sink satellite uploads $w_{\mathcal{K}_l}$ along with the satellites' data distribution (in terms of the number of class labels, which is piggybacked onto the model propagation process) to the GS.

In the conventional star topology, where each satellite individually communicates with the GS, the total time T_{sum} required for all the satellites on an orbit \mathcal{K}_l to exchange their models with the GS can be defined in two intervals as: $T_{sum} =$

$$\begin{cases} \sum_{k=1}^{\mathcal{K}_l} (2t_c(k) + t_{wait}(k) + t_{train}(k)), & \text{if } t_{train}(k) < t_{visible}(k) \\ \sum_{k=1}^{\mathcal{K}_l} (2t_c(k) + 2t_{wait}(k) + t_{train}(k)), & \text{if } t_{train}(k) \geq t_{visible}(k) \end{cases} \quad (10)$$

where $t_{wait}(k)$ is the time that the GS must wait before the k -th satellite enters its visible zone and $t_{visible}(k)$ is the visibility period of a satellite k with the GS. The below $t_{train}(k)$ is the time required for a satellite k to train a local model

$$t_{train}(k) = \frac{I(n_k b_k c_k)}{f_k}, \quad (11)$$

where n_k is the number of mini-batches, b_k is the size of each mini-batch, c_k is the average number of processing cycles needed for training a data sample on k satellite, and f_k is the CPU/GPU frequency. In (10), we consider two possible scenarios of calculating T_{sum} : (1) a satellite is able to complete its local model training within its visible period, or (2) a satellite requires an *additional round of waiting* for becoming visible again (to upload its trained model to the GS). The above represents a largely sequential procedure (note the accumulation of t_{wait} and t_c). In contrast, by introducing model propagation, t_{wait} and t_c will occur to the sink satellite *only*. In other words, we can significantly reduce the total time to

$$T_{sum}^* = 2t_c + t_{wait}^* + t_{train}(\mathcal{K}_l), \quad (12)$$

where t_{wait}^* is the time waiting for the sink node to be visible to the GS, and $t_{train}(\mathcal{K}_l) = \max_{k \in \mathcal{K}_l} t_{train}(k)$ is the parallel training time of all satellite models on orbit l . Note that we take advantage of the fact that the process of relaying the model to the sink happens in parallel with t_{wait}^* . In fact, this relay is very fast on its own because ISL typically uses FSO links, which support a much higher data rate (Gbps to Tbps) than RF links. However, in our simulation, we *forgo this benefit* of using FSO links so that *all our performance advantages over baseline approaches come from our architecture and algorithm design, rather than being mixed with hardware benefits*.

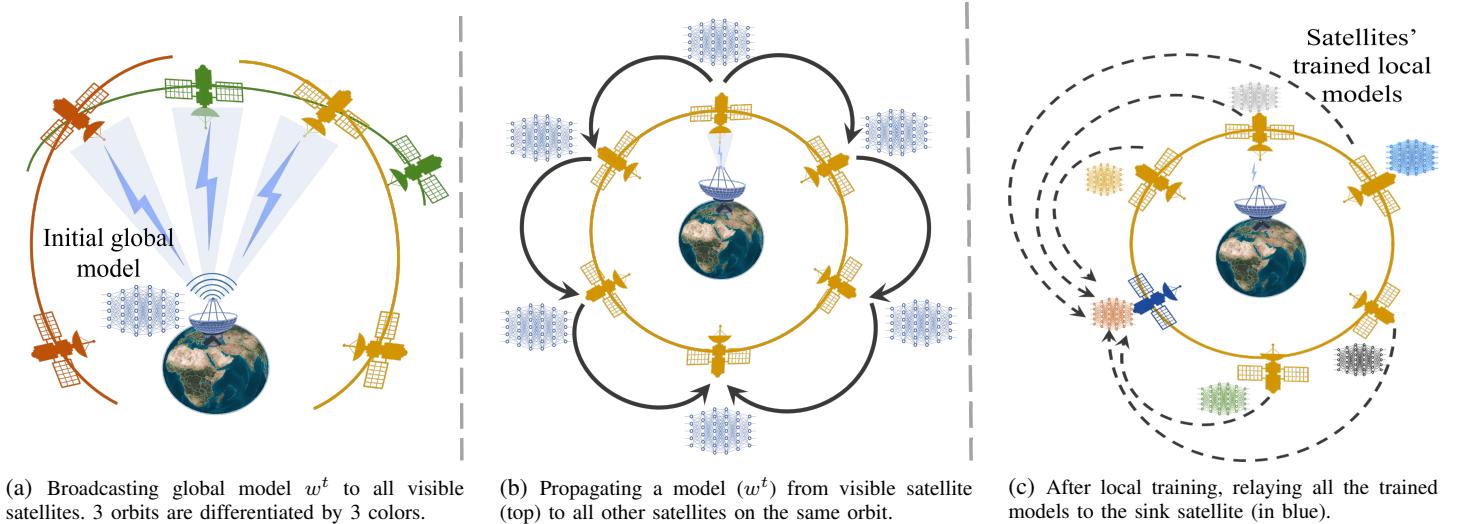


Fig. 2: Illustration of model propagation scheme.

B. Distributed Scheduling

We propose a distributed scheduling algorithm to select an optimal “sink” satellite on each orbit to generate a partial global model, and optimally schedule communication between each sink satellite and the GS by exploiting the predictability of the satellite orbiting pattern. Fig. 3 shows a real visiting pattern of 16 satellites evenly distributed over four orbits (i.e., 4 on each orbit) communicating with a GS. Although these satellites travel at the same speed v_l with the same orbital period T_l and are equally spaced on each orbit, they have different and irregular visibility periods and varying numbers of visits to the same GS; even each satellite’s own visits are not equally spaced and have different durations. This is because 1) the trajectory of the Earth differs from that of a satellite (which has an inclination angle between $0\text{--}90^\circ$) and 2) LEO satellites travel at very high speeds (≈ 7.8 km/s) in comparison to Earth’s rotational speed (≈ 0.45 km/s).

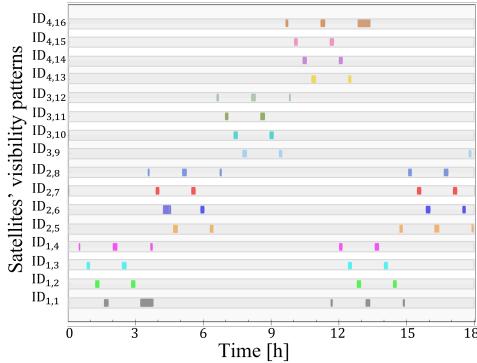


Fig. 3: A visiting pattern over 18 hours of a Walker-delta constellation consisting of 16 satellites that communicate with a GS in Rolla, MO, USA ($ID_{l,k}$ is the k -th satellite in an orbit l).

To tackle the high irregularity of the satellite-GS visiting pattern, we propose a distributed scheduling algorithm that selects a *sink satellite* for each orbit via prediction of its visibility period (and the sink changes over time). This scheduler runs at each satellite once the satellite completes its local training of model w^t , in order to determine which satellite on the same orbit will

be the first next satellite to visit the GS for a visiting period longer than the minimum interval of sending all $|\mathcal{K}_l|$ models to the GS. This satellite is referred to as the sink satellite (one per orbit). We predict the satellite visibility using a method introduced in [11]. Once the sink is determined, each satellite sends its locally trained model to the sink using our propagation algorithm discussed in Section IV-A. The sink k_s then bags all these local models into a *partial global model* and uploads it to the GS when k_s first becomes visible to the GS. In other words, our distributed scheduling algorithm leverages the predictable satellite-GS visibility despite its sporadic nature.

Suppose there are N resource blocks (RBs) of bandwidth B^D each, for exchanging models between satellites and the GS. To ensure the smooth operation of the iterative learning process, the GS needs to allocate the entire bandwidth B (i.e., $B = NB^D$) to all visible satellites for broadcasting the global model via the uplink. On the other hand, sink satellites from multiple orbits will compete for an RB to send their partial global models to the GS via the downlink (using a multiple access protocol such as ALOHA). As a result, SNRs for uplink and downlink channels between a satellite k and the GS can be expressed as

$$SNR(GS)^U = P_t(GS) + G_k + G_{GS} - L_{(k,GS)} - K_B TB \quad (dB) \quad (13)$$

$$SNR(k)^D = P_t(k) + G_k + G_{GS} - L_{(k,GS)} - K_B TB^D \quad (dB) \quad (14)$$

Thus, the overall communication latency given in (7) can be reformulated as

$$t_c^U \approx \frac{z|\mathcal{N}|}{B \log_2(1+SNR(GS)^U)} + \frac{\|k,GS\|_2}{c} + t_k + t_{GS} \quad (15)$$

$$t_c^D \approx \frac{z|\mathcal{N}|}{B^D \log_2(1+SNR(k)^D)} + \frac{\|k,GS\|_2}{c} + t_k + t_{GS} \quad (16)$$

Accordingly, the total time required by each sink satellite to exchange models with a GS can be determined by substituting (15, 16) in (12), yielding

$$T_{sum}^* = t_c^U + t_c^D + t_{wait}^* + t_{train}(\mathcal{K}_l), \quad (17)$$

In addition, we denote the start time and end time of the access window $AW(k, GS)$ between a satellite k and the GS by $t_{start}(k, GS)$ and $t_{end}(k, GS)$, respectively. The start time is the time when a satellite k enters the visibility zone of the GS, and the end time is the time when k leaves the visibility zone. Consider all the K satellites,

$$T_{start} = \left\{ \{t_{start,(k=1,GS)}^r\}_{r \in \mathcal{R}_1}, \{t_{start,(k=2,GS)}^r\}_{r \in \mathcal{R}_2}, \dots, \{t_{start,(k=K,GS)}^r\}_{r \in \mathcal{R}_K} \right\} \quad (18)$$

$$T_{end} = \left\{ \{t_{end,(k=1,GS)}^r\}_{r \in \mathcal{R}_1}, \{t_{end,(k=2,GS)}^r\}_{r \in \mathcal{R}_2}, \dots, \{t_{end,(k=K,GS)}^r\}_{r \in \mathcal{R}_K} \right\},$$

where $t_{start,(k,GS)}^r$ and $t_{end,(k,GS)}^r$ refer to $t_{start,(k,GS)}$ and $t_{end,(k,GS)}$ for the r -th visit, respectively, i.e., k enters and leaves the GS's visibility zone for the r -th time. $\mathcal{R}_k = \{1, 2, \dots, R_k\}$ is the index set of all the visits of satellite k . Thus, the AW can be written as

$$AW(k, GS) = \bigcup_{k \in \mathcal{K}} \left\{ [T_{start}, T_{end}] \right\}$$

$$= \left\{ \left[[t_{start(k=1,GS)}^r, t_{end(k=1,GS)}^r] \right]_{r \in \mathcal{R}_1}, \dots, \left[[t_{start(k=K,GS)}^r, t_{end(k=K,GS)}^r] \right]_{r \in \mathcal{R}_K} \right\} \quad (19)$$

Suppose that each ISL hop h between any two adjacent satellites is allocated one RB with bandwidth B^h . Accordingly, the time required for exchanging a model between two adjacent satellites $(k, k+1)$ via a single hop h is

$$t_h(k, k+1) = \frac{z|\mathcal{N}|}{B^h \beta_h} \quad (20)$$

where β_h is the spectral efficiency of the communication link.

Consider a set of candidate sink satellites $\mathcal{C}_l = \{1, 2, \dots, C_l\}$ on an orbit l , each of which satisfies the requirements on training time $t_{train}(\mathcal{K}_l)$ and communication latency t_c^U and t_c^D (see eqn. 12, 15, 16). On each orbit, to determine the optimal sink satellite c_{opt} for exchanging models with the GS, we formulate an optimization problem that minimizes a sum that involves (1) the waiting time t_{wait}^* for the sink satellite to enter the GS's visible zone, and (2) the model relaying time $t_h^*(k, \mathcal{C}_l)$ for each other satellite k to reach the sink, where

$$t_h^*(i, j) = \max_{i \in \mathcal{K}_l; i \neq j; \forall j \in \mathcal{C}_l} \frac{h z |\mathcal{N}|}{B^h \beta_h}, \quad h = 1, 2, \dots, H \quad (21)$$

Therefore, the optimal sink satellite for each orbit should minimize the overall latency as

$$T_{sum}^* = \min \{t_c^D + t_c^U + t_{wait}^* + t_{train}(\mathcal{K}_l) + t_h^*\} \quad (22)$$

Given this, the scheduler selects the optimal sink satellite c_{opt} to be the one whose AW satisfies the time constraint $AW(c_{opt}, GS) \geq T_{sum}^*$. In the case of more than one, the scheduler selects the one that will visit the GS the first to be the optimal sink.

V. PERFORMANCE EVALUATION

A. Simulation setup

LEO Constellation. We consider a Walker-delta constellation \mathcal{K} consisting of 40 LEO satellites distributed evenly on five orbits. Each orbit l is located at an altitude h_l of 1500 km above the Earth's surface with an inclination angle of 80 degrees. A GS is located in Rolla, MO, USA (but can be anywhere on the Earth) with a minimum elevation angle of 10 degrees. Communication link parameters as described in Section III-B are

TABLE I: Simulation Parameters (Comms: upper; Learning: lower)

Parameters	Values
Transmission power (satellite & GS) P_t	40 dBm
Antenna gain of (satellite & GS) G_k, G_{GS}	6.98 dBi
Carrier frequency f	2.4 GHz
Noise temperature T	354.81 K
Transmission data rate R	16 Mb/sec
Average number of processing cycles c_k	10^3 [12]
CPU/GPU frequency f_k	10^9 [12]
Number of local training epochs I	100
Learning rate η	0.001
Mini-batch size b_k	32

assigned according to the values in Table I (upper part). We use the Systems Tool Kit (STK), a software tool for analyzing satellite constellations, to compute the visiting pattern of LEO satellites with regard to the GS. To obtain each set of results, we run each simulation over a period of three days.

Baselines. We compare FedLEO with state-of-the-art (SOTA) approaches proposed recently and reviewed in Section II: *sync* approaches including FedAvg [6], FedISL [3] and FedHAP [2], and *async* approaches including FedAsync [13], FedSpace [7], FedSat [9], FedSatSched [10], and AsyncFLEO [4].

TABLE II: FedLEO vs. SOTA FL under non-IID.

FL Approaches	Accuracy (%)		Conv time (h)	Remark
	MNIST	CIFAR-10		
FedAvg [6]	85.84	63.26	48	GS at any location
FedISL (<i>ideal setup</i>) [3]	83.67	75.31	4	GS at NP; MEO above Equator
FedISL [3]	63.45	54.33	72	GS at any location
FedHAP [2]	83.94	80.89	40	GS at any location
FedHAP [2]	80.45	77.67	5 (89.83) (84.31)	using two HAPs as server not a GS
FedAsync [13]	81.63	59.18	48	GS at any location
FedSpace [7]	53.12	40.69	72	Sats need to upload raw data to the GS
FedSat (<i>ideal setup</i>) [9]	88.65	69.19	24	Sats visit GS at regular interval
FedSatSched [10]	76.32	70.95	48	GS at any location
AsyncFLEO [4]	80.62	77.23	6	GS at any location (ignores sink's visible period)
FedLEO	89.37	82.13	15	GS at any location

Dataset and ML models. To compare FedLEO with the SOTA FL-Satcom approaches, we use the same datasets they used (MNIST and CIFAR-10). But in addition, we also use a real satellite dataset, DeepGlobe Road Extraction [14]. This DeepGlobe dataset contains 6,226 colored satellite images of size 1024×1024 where each pixel represents a resolution of 50 cm, and these images have been annotated for road extraction. We augment the DeepGlobe dataset via flipping, rotating, varying lighting conditions, etc. to generate more training data. For the ML model, we use a deep CNN for MNIST and CIFAR-10, and a U-Net model for DeepGlobe. We consider both IID and non-IID data distributions among satellites (except for DeepGlobe which is non-IID by nature). In the IID setting, training images are randomly shuffled and equally distributed across all satellites, with each satellite having all 10 classes. In the non-IID setting, satellites in two orbits are trained on 4 classes, while satellites in the remaining three orbits are trained on the other 6 classes. Table I (the lower part) summarizes the training hyperparameters.

B. Results

Comparison with SOTA. Table II shows a comparison of FedLEO's performance with SOTA using the MNIST and CIFAR-10 datasets in the non-IID setting (results for IID have a

similar trend and are omitted due to space constraints). According to the results for the MNIST dataset, FedLEO achieves an accuracy of 89.37% within only 15 hours (including the waiting time for sink satellites to enter the visibility zone of GS). With the ideal assumption of FedISL [3] where the GS is located at the NP or is an MEO satellite orbiting above the Equator, FedISL converges within 4 hours with an accuracy of 83.67%. However, when we omit this assumption (GS at any location), the FedISL's accuracy is reduced to 63.45% within 72 hours of training. The accuracy of FedSat [9] and FedHAP [2] are marginally close to FedLEO, however, their FL convergence processes are considerably slower. In addition, FedSat presumes a very similar setup to FedISL, limiting its application, and FedHAP requires extra hardware (HAP) to achieve better results. When FedSatSched [10] removes FedSat's ideal assumption, its accuracy drops to 76.32%, and the convergence time doubles. AsyncFLEO [4] converges in only 6 hours with an accuracy of 80.62%; however, it does not consider whether the sink node will have sufficient time to train and exchange models with the GS.

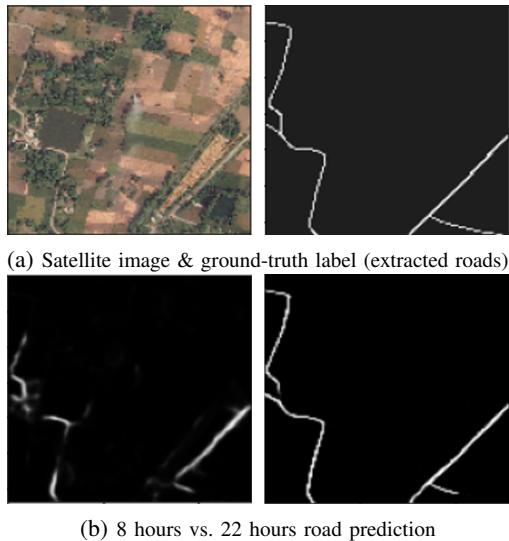


Fig. 4: Sample of FedLEO's performance on the DeepGlobe dataset.

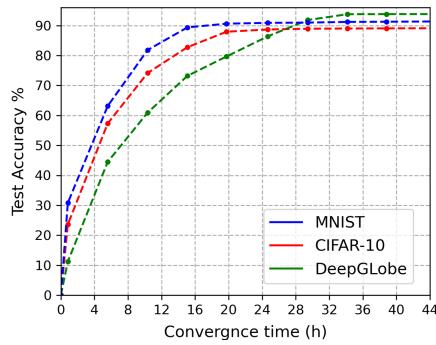


Fig. 5: FedLEO's accuracy vs. convergence on all datasets.

When the CIFAR-10 dataset is used, all the FL approaches fail to achieve the same accuracy on the MNIST dataset within the same convergence timeframe (saliently for FedAvg [6], FedAsync [13] and FedSat [9]), while FedLEO still maintains the highest accuracy 82.13% within only 15 hours, proving its robustness. **Evaluating FedLEO on DeepGlobe.** DeepGlobe is a more sophisticated and realistic dataset for LEO satellites. Fig. 4 shows

how FedLEO recognizes the roads from a DeepGlobe image at two different times of convergence. The accuracy was 52.37% in 8 hours and then increased to 82.77% after 22 hours.

Fig. 5 compares the accuracy vs. convergence time of FedLEO. It can be seen that FedLEO can achieve similar or even better accuracy on DeepGlobe than on MNIST and CIFAR-10, but requires extra time due to the complexity of the dataset. Other SOTA FL approaches take much longer convergence time with lower accuracy under the same setting.

VI. CONCLUSION

We have proposed FedLEO, an efficient FL framework designed for LEO satellite constellations. FedLEO tackles the challenging problem of highly intermittent and irregular patterns in which LEO satellites visit GSs. FedLEO consists of a model propagation algorithm and a distributed scheduling algorithm to achieve the goal. In our quantitative study with several SOTA benchmarks, we find that FedLEO drastically reduces FL training/convergence time; in the meantime, it achieves an accuracy of 89.37% even under the tough non-IID setting, outperforming the benchmarks by a wide margin. We also evaluate FedLEO on a real satellite dataset (DeepGlobe), which no prior work in this area has done before.

REFERENCES

- [1] P. Barmpoutis, P. Papaioannou, K. Dimitropoulos, and N. Grammalidis, "A review on early forest fire detection systems using optical remote sensing," *Sensors*, vol. 20, no. 22, p. 6442, 2020.
- [2] M. Elmahallawy and T. Luo, "FedHAP: Fast federated learning for LEO constellations using collaborative HAPs," in *14th International Conference on Wireless Communications and Signal Processing (WCSP)*. IEEE, 2022.
- [3] N. Razmi, B. Matthiesen, A. Dekorsy, and P. Popovski, "On-board federated learning for dense LEO constellations," in *IEEE International Conference on Communications (ICC)*, May 2022.
- [4] M. Elmahallawy and T. Luo, "AsyncFLEO: Asynchronous federated learning for LEO satellite constellations with high-altitude platforms," in *2022 IEEE International Conference on Big Data (Big Data)*, 2022, pp. 5478–5487.
- [5] D. Vasisht, J. Shenoy, and R. Chandra, "L2D2: Low latency distributed downlink for LEO satellites," in *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*, 2021, pp. 151–164.
- [6] B. McMahan *et al.*, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.
- [7] J. So *et al.*, "FedSpace: an efficient federated learning framework at satellites and ground stations," *arXiv:2202.01267*, 2022.
- [8] H. Chen, M. Xiao, and Z. Pang, "Satellite-based computing networks with federated learning," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 78–84, 2022.
- [9] N. Razmi *et al.*, "Ground-assisted federated learning in leo satellite constellations," *IEEE Wireless Communications Letter*, 2022.
- [10] ———, "Scheduling for ground-assisted federated learning in LEO satellite constellations," in *2022 30th European Signal Processing Conference (EUSIPCO)*, 2022, pp. 1102–1106.
- [11] I. Ali, N. Al-Dahir, and J. E. Hershey, "Predicting the visibility of LEO satellites," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 4, pp. 1183–1190, 1999.
- [12] T. Zeng *et al.*, "Federated learning in the sky: joint power allocation and scheduling with UAV swarms," in *IEEE International Conference on Communications (ICC)*. IEEE, 2020, pp. 1–6.
- [13] C. Xie *et al.*, "Asynchronous federated optimization," in *12th OPT Workshop on Optimization for Machine Learning*, 2020.
- [14] I. Demir *et al.*, "DeepGlobe 2018: a challenge to parse the earth through satellite images," in *Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.