DSFL: Decentralized Satellite Federated Learning for Energy-Aware LEO Constellation Computing

Chenrui Wu[1,2], Yifei Zhu[5], Fangxin Wang[2,1,3,4,*]

[1]FNii, The Chinese University of Hong Kong, Shenzhen

[2]SSE, The Chinese University of Hong Kong, Shenzhen

[3]The Guangdong Provincial Key Laboratory of Future Networks of Intelligence

[4]Peng Cheng Laboratory

[5]University of Michigan-Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University

Email: chenruiwu@link.cuhk.edu.cn, yifei.zhu@sjtu.edu.cn, wangfangxin@cuhk.edu.cn

***Abstract*—The dense constellation of Low Earth Orbit (LEO) satellites plays a significant role in the sixth generation mobile network (6G) and Terrestrial-Aerial-Space network. With the communication network composed of the LEO satellites, ground stations and terminals, global network services can be accessed anywhere and anytime. The global communication coverage and precious data resource embraces new opportunities to integrate the computing resources and data in LEO satellite constellations for intelligent learning tasks, like carbon estimation, transportation surveillance, forest fire detection, etc. Federated learning (FL) stands out as a promising paradigm towards this goal with a balance in privacy preservation and data utilization. Traditional FL employs a centralized deployment, i.e., regarding the ground station as the server and satellites as clients, which however faces two challenges: 1) The risk of single point failure from the centralized server. 2) Slow convergences resulting from satellites' intermittent communication. To tackle the challenges, in this paper, we propose a decentralized satellite federated learning, considering the limited communication traffic in satellite communication, the privacy of data and the efficiency of machine learning. Satellites collaborate in a decentralized way to reach a consensus on model parameters, overcoming the effects of data heterogeneity between satellites. Considering the scarce power capacity, we further design an energy-aware communication strategy to prolong satellites life and avoid communication congestion. Experiments on real-world datasets demonstrate that our framework can speed training process by 5-10 times, and requires only 1/3 to 1/6 communication energy cost.**

***Index Terms*—Federated Learning, Satellite Computing, Low Earth Orbit (LEO), Distributed Deep Learning.**

I. Introduction

With the rapid development of space technology and the falling cost of satellites entering space, the constellation of large numbers of Low Earth Orbit (LEO) satellites emerges as a promising communication paradigm [1]. Numerous satellites in LEO can provide global seamless communication coverage, even for remote areas such as desert and polar regions. These LEO constellations can effectively alleviate terrestrial traffic and real-time service pressure, promoting the fusion of global sensor networks. These satellites simultaneously conduct observations of the Earth, enabling such applications as transportation surveillance, agricultural resource monitoring, earthquake/ tsunami alert, forest fire detection, etc [2].

Considering the excellent communication coverage and the precious data resource of LEO satellites from the global perspective, integrating the computing resources and data in LEO satellite constellations for learning tasks brings many new opportunities. However, the limited bandwidth resource and uncertain communication connections prevent the raw data downloading to a central node, calling for a distributed learning framework. Federated learning (FL) [3] serves as a popular distributed learning paradigm where each node only shares the model weight to generate an aggregated model. Therefore, FL's properties of preserving privacy and reducing communication overhead, making it a perfect candidate for distributed model training in satellite constellation.

Existing works [4]–[6] mainly focus on the architecture with ground base stations as the FL central server. This architecture however exists two challenges: 1) *The first one comes from the risk of single point failure as a centralized system.* The central server of FL will have extensive communication load during aggregation, which is easy to cause traffic congestion, especially when the bandwidth of communication link is limited. Meanwhile, if the ground base station fails, the entire federated learning process will be suspended. 2) *The second challenge lies in the slow model convergence speed.* Unlike traditional terrestrial edge devices that can connect to the central server at any time, a satellite can only connect to the ground base station within a short orbit window. So waiting for the model aggregation of every satellite in a synchronous mechanism will be quite time-consuming. Even if asynchronous FL is employed, the numerous and superimposed orbit windows will lead to huge congested traffic, which is prone to cause imbalanced waiting time and model timeout, if without a sophisticated schedule.

Decentralized federated learning (DFL) emerges in recent

*Fangxin Wang is the corresponding author.

The work was supported in part by the National Key R&D Program of China with grant No. 2018YFB1800800, the Basic Research Project No. HZQB-KCZYZ-2021067 of Hetao Shenzhen-HK S&T Cooperation Zone, by National Natural Science Foundation of China with Grant No.62102342, by Shenzhen Outstanding Talents Training Fund 202002, by Guangdong Research Projects No. 2017ZT07X152 and No. 2019CX01X104, and by the Guangdong Provincial Key Laboratory of Future Networks of Intelligence (Grant No. 2022B1212010001).

years as a promising approach that can jointly satisfy privacy preservation, system robustness, and fast convergence. Without a centralized node, DFL allows each node to exchange their local models with neighbors and finally achieve model consensus after several rounds of communication. Communication between satellites relies on inter-satellite links (ISLs), including Intra-plane ISL and Inter-plane ISL. In this way, the communication overhead is dispersed and the communication congestion of the central server can be avoided. The most common DFL approach is gossip averaging [7], [8], where each node progressively averages its model with neighbors using an iterative averaging method. However, such DFL mechanism can be negatively affected by data heterogeneity, and the multi-round process will cost higher bandwidth resource.

Distinguished with terrestrial edge computing, satellite networks are limited by time-varying topology, dynamic bandwidth, and limited battery power. To extend the satellite network's life, an energy-efficient communication strategy of model exchanging is essential. Communication strategy should also concern with avoiding congestion.

In this paper, we propose a decentralized satellite federated learning (DSFL) framework that enables LEO satellite networks to achieve collaborative model training. In DSFL, every satellite plays as a relay, transferring its model parameters to the whole network without decaying its weight. The specific process contains two steps: *Intra-plane transmission and Inter-plane transmission*. In the first step, after finishing the local update, the satellites in the same orbit will communicate with each other and average their weight. In the second step, neighbor orbits will an optimal Inter-plane ISL link to exchange model parameters to reach a global consensus. We next design an energy-aware algorithm to solve the communication route problem for the Inter-plane model exchange. We summarize our main contributions as follows:

- We propose the DSFL, the a novel decentralized satellite federated learning framework, which allows LEO satellites cooperatively and efficiently train machine learning models without excessive communication burden.
- We propose a communication strategy to handle the path selection problem of model sharing in order to minimize communication overhead, considering communication between Intra-plane ISL and Inter-plane ISL in a real satellite constellation scenario.
- Extensive experiments on multiple real-world traces demonstrate that the DSFL outperforms state-of-the-art solutions with a significant improvement in convergence speed and great saving in communication cost.

The rest of this paper is organized as follows. We introduce the related work of federated satellite learning and decentralized federated learning background in Section II. The system model and problem formulation are given in Section III. The main algorithm is presented in Section IV. In Section V, experiments are conducted to evaluate the performance of our method. Finally, Section VI concludes this paper.

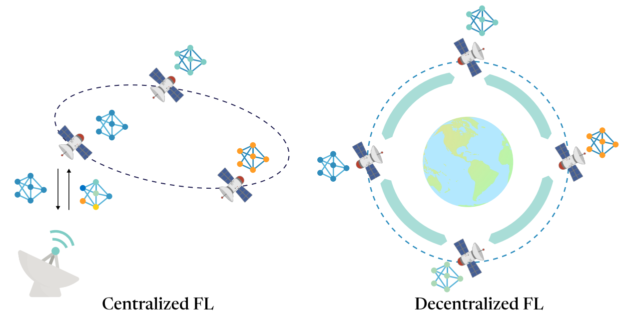

Fig. 1. Illustration of Centralized FL and Decentralized FL for Satellites

II. Related Work

A. Federated Learning on Satellites

Deploying federated learning tasks in LEO satellites is a challenging and promising application. Some pioneer works have explored satellite computing. Most of the existing work follows a centralized federated learning architecture. Fadlullah *et al.* [2] introduce a Terrestrial-Aerial-Space network federated learning with an asynchronous setting. Razmi *et al.* [4] propose FedSat as an improvement of FedAsync [9]. In [9], the server and worker updates asynchronously with Non-blocking communication. Older models should have lower merge coefficients on newer server models. So *et al.* [5] present their FedSpace to optimize the trade-offs between satellite idleness and local model staleness. They train a linear model to determine the model aggregation scheduler. These centralized architectures face the potential of a single point of failure especially in the unstable communications scenario of satellites between space and ground. The intermittent communication with the ground station limits their communication and aggregation efficiency. To overcome the defect of the communication bottleneck of centralized federated learning in LEO satellite computing, we present our decentralized satellite federated learning framework.

B. Decentralized Federated Learning

In contrast to widely studied server-coordinated federated learning, decentralized federated learning can be used as an alternative to centralized federated learning. Existing decentralized federated learning mainly utilizes gossip average (or called DP-SGD [10]) [7], [8]. Hegedűs *et al.* [8] compare the efficiency of federated learning and gossip learning. Hu *et al.* [7] divide the model into segments. The clients conduct segments pulling and sending to aggregate the mixed model. Meng *et al.* [11] apply DP-SGD for a low-bandwidth satellite constellation environment. They implement parameter sparsification to reduce communication bandwidth.

These gossip learning methods consider frequent gossip among neighbor clients so that each client's update spreads across the network before the models become too stale. However, the client in naive gossip learning may asymptotically receive its uniform share of the update, resulting in inefficient training. Besides, unlike terrestrial edge computing, energy is more precious in satellite computing. Slower convergence and more communication steps are not suitable for this environment. From the perspective of data heterogeneity, clients are more influenced by neighbours leading to unsatisfactory performance.

Therefore, we propose our DSFL to deploy decentralized federated learning in LEO satellite constellation. The core is to forward clients' weights across the network without attenuating their weights at each hop. We further optimize the communication strategy for weight exchange to reduce communication costs.

III. System Models and Problem Formulation

A. Satellite Constellation Model

We firstly introduce the LEO satellite constellation network. In this paper, we consider a Walker Star constellation as depicted in Fig. 2. A typical Walker Star LEO constellation consists of M orbital planes, $\mathcal{M} = \{1, 2, \ldots, M\}$; they are evenly separated by π/M radians, and each plane contains the same number of satellites. The set of satellites is $\mathcal{N} = \{1, 2, \ldots, N\}$. Each orbital plane $p \in \mathcal{M}$ is deployed at a given longitude ϵ_p, in h_p km from the Earth's surface. Each satellite's latitude is denoted as θ_i. We use $p(i)$ to describe the satellite i belonging to an orbital plane p. Hence, the coordinates of satellite i in an orbital plane are written as $(h_{p(i)} + R_E, \epsilon_{p(i)}, \theta_i)$, where R_E is the radius of the Earth. Satellites communicate through Intra-plane ISLs and Inter-plane ISLs.

Among the orbital planes, satellites on some planes move north (ascending), while in other orbits move south (descending). Such inter-plan ISLs with early-opposite directions are called cross-seam ISLs. Due to the rapid movement of the satellites in the opposite direction of orbital plans, cross-seam ISLs may suffer from Doppler shift.

We model the LEO satellite network as an undirected weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = N$, where $\mathcal{V}$ is the set of satellites and $\mathcal{E}$ is the set of ISLs. The weight $w(u, v)$ is described as the communication cost between linked satellite u and v, where $u, v \in \mathcal{V}$.

B. Satellite Energy Model

As mentioned above, inter-satellite links are the approach to communicating between satellites. We consider traditional communication protocol is OFDMA or CDMA.

Due to the earth's obstruction, the maximum line-of-sight distance of two satellites i and j to build ISL can be calculated as:

$$\hat{l}(v_i, v_j) = \sqrt{h_{p(i)}(h_{p(i)} + 2\mathrm{R_E})} + \sqrt{h_{p(j)}(h_{p(j)} + 2\mathrm{R_E})}, \tag{1}$$

where $h_{p(i)}$ and $p(j)$ are the heights of plane to which satellites i and j belong. To establish ISL, the distance between two satellites should be shorter than the maximum line-of-sight

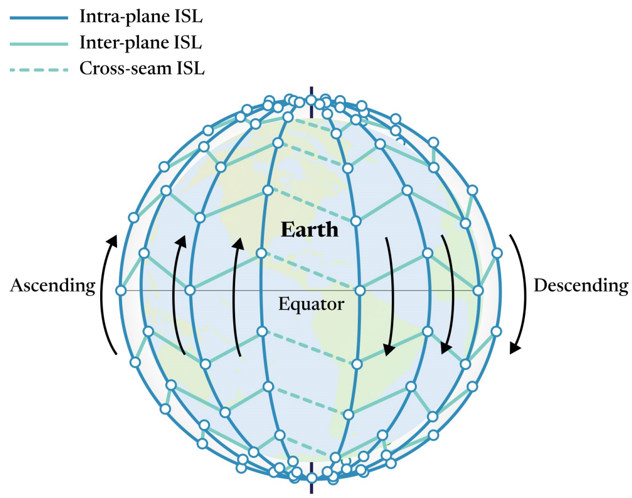

Fig. 2. Model of a typical Walker Star constellation

distance. Hence, the Euclidean distance between these two satellites in written as:

$$\begin{aligned}\|v_i v_j\| = \Big(&(h_{p(i)} + \mathrm{R_E})^2 + (h_{p(j)} + \mathrm{R_E})^2 \\ &- 2\left(h_{p(i)} + \mathrm{R_E}\right)\left(h_{p(j)} + \mathrm{R_E}\right) \\ &\times \left(\cos\theta_i \cos\theta_j + \cos(\epsilon_{p(i)} - \epsilon_{p(j)}) \sin\theta_i \sin\theta_j\right) \Big)^{1/2}.\end{aligned} \tag{2}$$

Let R_{ij} to be the data rate in the radio link. The energy consumed by sending data packets in satellite link (i, j) is defined as:

$$E_s(v_i, v_j) = P_s t_b = \frac{\left(2^{\frac{1}{B \cdot t_b}} - 1\right) \cdot B \cdot N_0}{l^2(v_i, v_j)} \cdot t_b, \tag{3}$$

where P_s is transmission power. The transmission time is t_b. B is the bandwidth capacity. N_0 is the spectral density of noise. The receiving communication cost is calculated with reverse vertical. According to the energy consumption, we formulate the weight of an edge in the network as:

$$w(v_i, v_j) = E_s(v_i, v_j) + E_r(v_i, v_j). \tag{4}$$

C. Decentralized Federated Learning Model

We first illustrate the decentralized satellite federated learning architecture. In the parallel SGD algorithm, the decentralized network contains N satellite as mentioned above. In the network $\mathcal{G}$, a satellite is enabled to communicate with neighbor nodes. Each satellite stores distributed dataset $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_N$. In order to distillate the feature information hidden in the distributed dataset, the goal of decentralized federated learning is to optimize the global loss function $\mathcal{L}(\omega)$ by minimizing the weighted average of every client i 's local loss function $\mathcal{L}(\omega_i)$. The problem can be expressed as:

$$\mathbf{P}1: minimize\ \mathcal{L}(\omega) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}(\omega_i). \tag{5}$$

where an optimal global model parameters ω^* is expected to be found. To reach the optimization goal, our DSFL aims to train an aggregate models in a ways by sharing the weights of all clients to the whole network in each round.

D. Problem Formulation

The communication problem is to craft a communication strategy to transfer every satellite's weight to the entire network, instead of a typical point-to-point routing problem with a start node and an end node. This design requires Intra-plane ISL transmission and Inter-plane ISL transmission. Through Intra-plane ISLs, the model consensus can be reached for each palne. In Inter-plane ISLs, the optimization goal is to find the most energy-efficient communication strategy among planes to exchange planes' models in each round.

$$\mathbf{P}2: \quad minimize\ E(\mathcal{E}^*) = \sum\nolimits_{\mathcal{E}_{ij} \in \mathcal{E}*} w(v_i, v_j), \tag{6}$$

$$\text{s.t.} \quad 0 \le f_{ij} \le B_{ij}, \forall (v_i, v_j) \in \mathcal{E}^*, \tag{a}$$

$$\|v_i v_j\| \le \hat{l}(v_i, v_j), \forall (v_i, v_j) \in \mathcal{E}, \tag{b}$$

$$\sum\nolimits_{v_i \in \mathcal{V}} f_{ij} = \sum\nolimits_{v_i \in \mathcal{V}} f_{ji}, \tag{c}$$

$$E_i \le \hat{E}_i, \forall v_i \in \mathcal{E}^*, \tag{d}$$

where $f_{j,i}$ is the expected maximum flow. $B_{i.j}$ is the link maximum bandwidth limit. Constraint (a) limits traffic flow over any communication link that could not exceed the link's capacity. Constraint (b) restricts the distance of two satellites to be less than the maximum slant range. Constraint (c) is the flow conservation constraint. That is, the inflow flow of any relay satellite is equal to the outflow flow. The total energy is restricted under one satellite's battery power in Constraint (d).

IV. Decentralized Satellite Federated Learning

A. Design of DSFL

We present the DSFL, a decentralized SGD method for boosting mixing in distributed deep learning. It is proposed to improve clients' performance with heterogeneous data distribution, adapted to the satellite computing environment.

The procedure of the DSFL includes two stages: Intra-plane update and Inter-plane update. The key strategy to improve learning performance by transmitting weight data packets through the whole topology without decaying at every hop. It is described in Algorithm 1.

In the Intra-plane update phase, every satellite updates its local model ω_i^{t-1} by private data $\mathcal{D}_i$, denoted as:

$$\omega_i^{t+1/3} = \omega^{t-1} - \eta \nabla \mathcal{L}(\omega^{t-1}, \mathcal{D}_i). \tag{7}$$

After completing the local update, the satellites in the same orbital plane start Intra-plane communication to reach a consensus model. In the ring topology of orbital planes, $2N-2$ communication steps are required with one and a half circle route. Each client sends the data packet $\mathcal{P}_{ij}^t$ forward. The data packets contain the client's own and received weights. Hence, satellites in the same plane own the model parameters from all others. Intra-plane averaging is conducted at present, written as:

$$\omega_i^{t+2/3} = \frac{1}{N} \sum_{j=1}^{N} \omega_j^{t+1/3}. \tag{8}$$

Algorithm 1: Decentralized Satellite Federated Learning (DSFL)

Input: Network $\mathcal{G}$, satellite set $\mathcal{N}$, plane set $\mathcal{M}$, total number of iterations T, learning rate η
Output: ω_i^{t+1}

Initialize $\mathcal{M}^t \leftarrow \emptyset$;
Initialize the model parameter $\omega_i^0 \leftarrow \omega^0$;
for $t \leftarrow 0$ **to** T **do**
  `// Intra-plane Update`
  for *plane* $p \in \mathcal{M}$ ***in parallel*** **do**
    Determine a start satellite and the direction;
    for *satellite* i ***in*** *plane* p ***in parallel*** **do**
      $\omega_i^{t+1/3} \leftarrow \omega^{t-1} - \eta \nabla \mathcal{L}(\omega^{t-1}, \mathcal{D}_i)$;
      Send $\mathcal{P}_{ij}^t \leftarrow \omega_i^{t+\frac{1}{3}} + \sum_{k \in \mathcal{N}_p} \mathcal{P}_{kj}^t$ forward to neighbor j as soon as receive $\mathcal{P}_{ki}^t$;
      if *receive all* $\omega_j^{t+1/3}$ *in* p **then**
        $\omega_i^{t+2/3} \leftarrow \frac{1}{N} \sum_{j=1}^{N} \omega_j^{t+1/3}$;
      end
    end
    Obtain the uniform weight $\omega_p^{t+2/3}$ of plane p;
  end
  `// Inter-plane Update`
  $\mathcal{E}^* \leftarrow$ EACS($\mathcal{G}$);
  Transfer weight set $\{\omega_p^{t+2/3}\}$ through $\mathcal{E}^*$;
  Synchronize $\{\omega_p^{t+2/3}\}$ in each plane;
  for *satellite* $i \in \mathcal{N}$ ***in parallel*** **do**
    $\omega_i^{t+1} \leftarrow \frac{1}{M} \sum_{j=1}^{M} \omega_j^{t+2/3}$;
  end
end

When every plane reaches its Intra-plane consensus, the model sharing among planes is followed as Inter-plane update.

We consider the Inter-plane update in the unique topology of the satellite orbital plane. Due to the stability of Inter-plane ISL, cross-plane communication requires a higher data rate to reduce data delivery time, leading to more energy consumption. We propose an energy-aware communication strategy in Section IV-B to relay the models to the whole constellation. After cross-plane communication, each orbit obtains all the gradients. One additional round Intra-plane exchange is conducted to ensure that every satellite obtains these model parameters. Thus, the satellite constellation achieves the round's model consensus in this learning strategy. Each satellite is able to aggregate all gradients as FedAvg.

$$\omega_i^{t+1} = \frac{1}{M} \sum_{j=1}^{M} \omega_j^{t+2/3}. \tag{9}$$

Continually, federated learning rounds repeat this process until the model reaches convergence.

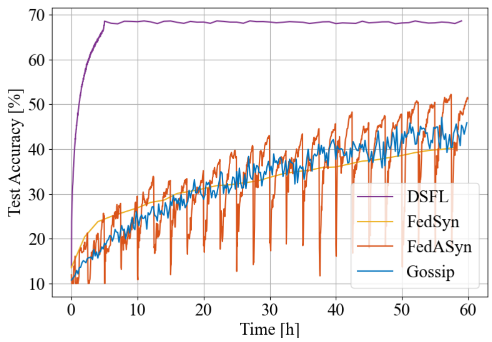

Fig. 3. Test accuracy on CIFAR-10

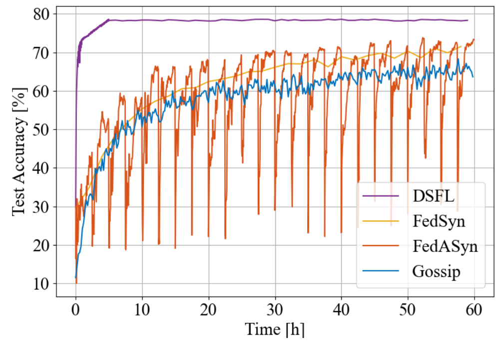

Fig. 4. Test accuracy on Fashion-Mnist

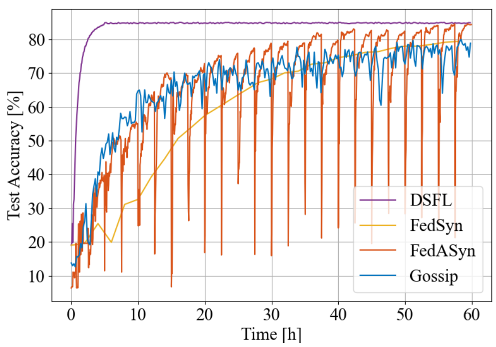

Fig. 5. Test accuracy on SVHN

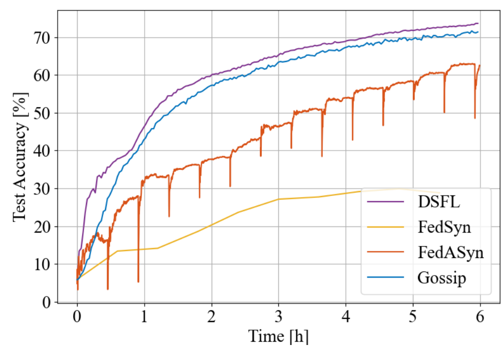

Fig. 6. Test accuracy on Femnist

B. Energy-Aware Communication Strategy for DSFL

In Algorithm 1, models need to be shared across planes. Hence, we propose an energy-aware communication strategy (EACS) to realize energy efficiency and communication load balance, as shown in Algorithm 2.

In the first phase of the DSFL, the model exchange is done within each orbital plane. To handle the cross-plane exchange, we modify the classic Dijkstra algorithm. Starting from a random plane, the satellites in this plane compare the weight w of feasible Inter-plane ISLs to neighbor planes. Select the most energy-efficient route $\mathcal{E}_{ij}$, which satisfies traffic flow and battery constraint. Add the optimal path $\mathcal{E}_{ij}$ into the selected path set $\mathcal{E}^*$, and adds the plane of optimal path into the selected plane set $\mathcal{M}'$. Update satellites' expected traffic flow and energy consumption. Regard selected plane set as a whole, select next path from all neighbors of selected planes. Repeat the above process until all plane are covered $\mathcal{M}' = \mathcal{M}$. A route is established to link all planes. The model parameters can be exchanged in this route to be transferred to all planes.

Considering the algorithm complexity, the sorting complexity is $O(Nlog_2N)$. The sort action will be repeated $M-1$ times. Therefore, the total complexity of Algorithm 2 is $O((M-1)Nlog_2N)$, where M and N are the number of orbits and satellites.

Algorithm 2: Energy-Aware Communication Strategy (EACS) Algorithm

Input: Network $\mathcal{G}$
Output: $\mathcal{E}^*$

Initialize $\mathcal{E}^* \leftarrow \emptyset$, $\mathcal{M}' \leftarrow \emptyset$;
Randomly select a starting plane $P_i \in \mathcal{M}$;
$\mathcal{M}' \leftarrow \mathcal{M}' \cup \{P_i\}$;
while $\mathcal{M}' \neq \mathcal{M}$ **do**
  Selected planes $\mathcal{M}'$ have neighbor planes $\mathcal{N}$;
  Sort all $\mathcal{E}(v_a, v_b) \in \mathcal{G}$ in ascending order of $w(\mathcal{E})$, where $v_a \in P_i, v_b \in P_j$, $P_j \in \mathcal{N}$;
  while $f_a + f_{a,b} \leq B_a$ *and* $f_b + f_{a,b} \leq B_b$ *and* $E_{a,b} \leq \hat{E}_a, \hat{E}_b$ **do**
    Find the minimum $w(\mathcal{E}(v_a, v_b))$;
    $\mathcal{E}^* \leftarrow \mathcal{E}^* \cup \{P_j\}, v_b \in P_j$;
    $\mathcal{M}' \leftarrow \mathcal{M}' \cup \{P_j\}$;
    $f_a \leftarrow f_a + f_{a,b}$, $f_b \leftarrow f_b + f_{a,b}$;
    $E_a \leftarrow E_a - E_{a,b}$, $E_b \leftarrow E_b - E_{a,b}$;
    break;
  end
end

V. Evaluation

A. Experimental Setup

In this section, we build the DSFL framework in Pytorch and STK (Satellite Tool Kit) satellite simulator.

Satellite Constellation: We consider a Walker Star with 50 satellites. The satellite constellation consists of 5 polar orbits, each with 10 satellites. The satellite cycle is about 115 minutes. The satellite's battery capacity is set to 117 KJ. Transmission power is 7-10 w. Reception power is 3-5 w. ISL's bandwidth is 10 Mbps.

Models and Datasets: We adopt convolutional neural network (CNN) to conduct local training. The implemented model is trained for four classic image classification tasks:

Mnist, Fashion-Mnist, Femnist, SVHN and CIFAR-10. The distribution of data is under a Non-IID setting. Non-IID data is formulated by dividing dataset into $2N$ shards, where each shard contains $\frac{len(dataset)}{2N}$ images. Each client will be randomly assigned 2 shards that are not duplicated.

Benchmarks: We evaluate our method with two typical centralized FL and a decentralized FL. The centralized FL methods need a ground station to receive and aggregate weights. The evaluation of centralized FLs are based on a test dataset on the ground station. For decentralized methods, the test accuracy is calculated as an average of all clients.

- **Synchronous FL with GS (FedSyn) [3]:** The majority of vanilla FL algorithms. It requires a ground station to aggregate model until all models are received.
- **Asynchronous FL with GS (FedAsyn) [9]:** FedAsyn allows ground station to aggregate the model as long as any weight uploaded.
- **Gossip averaging FL without GS (Gossip) [10]:** In each round, every satellite updates its model by aggregating weights from direct-connected neighbors.

B. Experimental Results

We firstly explore the model performance for different algorithms, as shown in Fig. 3-6. In Fig. 3, we can observe that the DSFL only needs less than 10 hours to converge. However, the FedSyn takes additional training time of FL rounds multiplied by the orbital period. The convergence trends of FedAsyn and gossip are slow at a speed of less than $1/10$. The accuracy after convergence shows that DSFL holds 20%-30% advantages for overcoming Non-IID over gossip and FedAsyn. We can obtain that gossip averaging fails to overcome data heterogeneity, influenced by their neighbors, leading to undesired model drift. The large staleness of the Asynchronous FL model prevents its fast convergence.

Fig. 4 shows that DSFL provides substantial speedup over the FedAsyn by 30 hours and gossip learning by more than 60 hours in an easier task of Fashion-Mnist. Moreover, the final accuracy of the DSFL is 10%-20% higher than gossip learning, FedSyn and FedAsyn. The results in Fig. 5 also show the convergence speed of DSFL is about 12 times faster than benchmarks. However, in a tougher task Femnist in Fig. 6, which is naturally Non-IID, the gap between DSFL and gossip learning is smaller. The DSFL still improves performance 2 times faster than FedAsyn, 3 times faster than FedSyn.

We further evaluate the energy consumption of each approach to reach the target accuracy and present the results in Fig. 7. The results are generated by the energy consumption to reach the target accuracy of specific datasets among different algorithms. Although our DSFL holds nearly double communication steps in each round, it can reach satisfactory accuracy with fewer rounds and energy-aware transmission, which leads to reduced total energy consumption. In most tasks, the DSFL holds comparable costs with synchronous FL. As motioned above, although synchronous FL consumes normal-level energy, it remarkably lacks convergence efficiency. Compared with gossip averaging and FedAsyn, the DSFL only consumes $1/6$ to $1/3$ of the energy to reach the target accuracy.

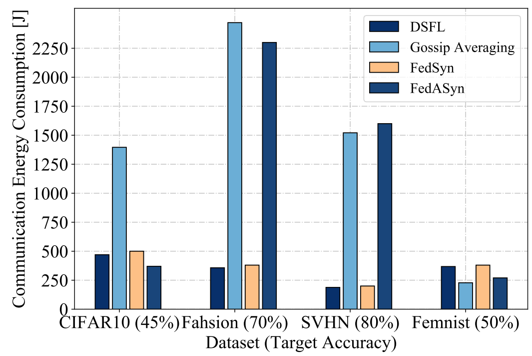

Fig. 7. Energy consumption evaluation

VI. Conclusion

In this paper, we propose the DSFL, a novel decentralized federated learning enabling LEO satellites collaboratively train intelligence models. We design a decentralized SGD mechanism to avoid a single point of failure and communication congestion. Our aggregation policy is also robust in overcoming heterogeneous data collected by satellites. Particularly, we formulate an energy-aware communication strategy to exchange models in the whole constellation to reach a model consensus with less communication cost. Extensive experiments demonstrate the effectiveness of the proposed mechanism.

References

[1] F. Tang, H. Zhang, and L. T. Yang, "Multipath cooperative routing with efficient acknowledgement for leo satellite networks," *IEEE Transactions on Mobile Computing*, vol. 18, no. 1, pp. 179–192, 2019.

[2] Z. M. Fadlullah and N. Kato, "On smart iot remote sensing over integrated terrestrial-aerial-space networks: An asynchronous federated learning approach," *IEEE Network*, vol. 35, no. 5, pp. 129–135, 2021.

[3] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 2017.

[4] N. Razmi, B. Matthiesen, A. Dekorsy, and P. Popovski, "Ground-assisted federated learning in leo satellite constellations," *IEEE Wireless Communications Letters*, vol. 11, no. 4, pp. 717–721, 2022.

[5] J. So, K. Hsieh, B. Arzani, S. Noghabi, S. Avestimehr, and R. Chandra, "Fedspace: An efficient federated learning framework at satellites and ground stations," *arXiv preprint arXiv:2202.01267*, 2022.

[6] N. Razmi, B. Matthiesen, A. Dekorsy, and P. Popovski, "On-board federated learning for dense leo constellations," *arXiv preprint arXiv:2111.12769*, 2021.

[7] C. Hu, J. Jiang, and Z. Wang, "Decentralized federated learning: A segmented gossip approach," *arXiv preprint arXiv:1908.07782*, 2019.

[8] I. Hegedűs, G. Danner, and M. Jelasity, "Gossip learning as a decentralized alternative to federated learning," in *Proceedings of International Conference on Distributed Applications and Interoperable Systems (IFIP)*, 2019.

[9] C. Xie, S. Koyejo, and I. Gupta, "Asynchronous federated optimization," in *Proceeding of 12th Annual Workshop on Optimization for Machine Learning*, 2020.

[10] X. Lian, C. Zhang, H. Zhang, C.-J. Hsieh, W. Zhang, and J. Liu, "Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent," in *Proceedings of Advances in Neural Information Processing Systems*, vol. 30, 2017.

[11] Q. Meng, M. Huang, Y. Xu, N. Liu, and X. Xiang, "Decentralized distributed deep learning with low-bandwidth consumption for smart constellations," *Space: Science & Technology*, vol. 1, no. 1, pp. 56–65, 2021.