FLSTRA：平流层联邦学习

Amin Farajzadeh,

Animesh Yadav、Omid Abbasi、Wael Jaafar 和 Halim Yanikomeroglu

arXiv:2302.00163v3 [cs.NI] 9 Jun 2023

摘要

我们提出一种平流层联邦学习（FL）系统（FLSTRA），其中高空平台站（HAPS）可促使大量地面客户端在不共享训练数据的情况下协作学习全局模型。FLSTRA 克服了联邦学习在地面网络中面临的挑战，例如，由于客户端参与有限以及多跳通信而导致收敛缓慢、通信时延较高。HAPS 利用其高度和尺寸，通过视距（LOS）链路支持更多客户端参与，并搭载功能强大的服务器。然而，同时处理大量客户端会带来计算时延和传输时延。因此，我们旨在为 FLSTRA 实现时延-准确率权衡。具体而言，我们首先针对上行链路和下行链路设计联合客户端选择与资源分配算法，在满足能量和服务质量（QoS）约束条件的前提下，将联邦学习时延降至最低。其次，我们提出通信与计算资源感知（CCRA-FL）算法，以达到目标联邦学习准确率，同时推导其收敛率的上界。所构建的问题是非凸的，因此我们提出一种迭代算法来求解。仿真结果表明，与地面基准方法相比，所提出的 FLSTRA 系统在联邦学习时延和准确率方面表现有效。

索引词

高空平台站、联邦学习、资源分配、时延最小化、准确率。

I. 引言

RAPID 社会变化正在重塑个人和社区对网络服务、用例和关键绩效指标 (关键绩效指标) 的需求与预期 [1]。一种高度自适应、实时、智能、保护隐私且可自我演进的机制似乎

A. Farajzadeh、A. Yadav、O. Abbasi 和 H. Yanikomeroglu 隶属于加拿大渥太华卡尔顿大学系统与计算机工程系非地面网络（NTN）实验室，邮编 K1S 5B6（电子邮箱： {aminfarajzadeh、animeshyadav、omidabbasi、halim}@sce.carleton.ca）。

Wael Jaafar 任职于加拿大魁北克省蒙特利尔市 H3C 1K3 的 ´高等技术学院（´ETS）软件与信息技术工程系，´（电子邮箱：wael.jaafar@etsmtl.ca）。

下一代无线网络需要具备应对这些不同需求和要求的能力 [3]。去中心化机器学习（ML）框架近来受到关注，被视为一种高效解决方案，可在无线网络中实现端到端智能和实时决策 [4]。具体而言，Google 于 2016 [5], 开发的联邦学习（FL）技术为去中心化模型训练提供了分布式智能。联邦学习的基本思路是邀请具有足够计算能力（即配备 CPU、GPU 和 RAM 等计算模块）且拥有数据的网络活跃用户参与网络参数的预测或估计，也就是模型训练 [6]。联邦学习具有内在的隐私保护机制，可保障网络用户的数据隐私 [7]。事实上，它允许协作训练模型，而无需显式共享本地数据。在联邦学习中，每位用户都会将其模型训练参数发送给中心化服务器；服务器收到所有用户的参数后将其聚合，并把聚合后的参数广播回用户。这个过程会持续多个轮次（也称为通信轮次），直到实现收敛 [8]。

A. 相关工作

近年来，地面无线网络中的联邦学习受到了广泛关注。研究人员一直在探索各种方法，以提升无线网络中联邦学习的性能。已有多项研究致力于提升无线网络中联邦学习的性能。例如，在 [9], 中，作者提出了一种异步联邦学习算法，根据客户端的到达顺序选择部分客户端，并在联邦学习服务器上进行聚合，从而提高通信效率。在 [10], 中，作者研究了无线通信网络中联邦学习的节能传输与计算资源分配问题。他们将联合学习与通信问题建模为一个优化问题，旨在满足时延约束的同时最小化能耗，并提出了一种迭代算法来求解该问题。

此外， [11] 中的作者提出了一种由同时无线信息与能量传输（SWIPT）辅助的联邦学习，联邦学习服务器在广播全局模型的同时向无线设备提供无线能量传输，从而提高系统的能量效率。在 [12], 中，作者研究了采用多个边缘服务器的联邦学习，并利用位于不同边缘服务器重叠区域的客户端加速训练。 [13] 中的作者提出了一种新颖的分层联邦学习框架，可将部分模型聚合从云端迁移至边缘服务器。

B. 研究动机与贡献

上述现有方法显著提升了联邦学习在地面无线网络中的性能。然而，仍存在一些挑战，包括参与方数量受限、无线环境异构、非视距（NLOS）信道故障以及客户端计算资源异构。研究仍在持续应对这些挑战。在地面网络中，由于覆盖范围和带宽等网络资源有限、客户端可用性不稳定以及信道故障，每个联邦学习轮次中通常只有部分客户端能够可靠地参与 [14]。考虑到联邦学习的迭代特性，限制参与方数量往往会导致计算延迟较高、训练准确率较低 [15]。此外，在地面环境中的联邦学习系统通常采用多个基站（BS）作为中继，与云端的云服务器通信；由于原始数据分布在接入网络中的边缘设备上，因此需要经过多跳无线链路，联邦学习不可避免地面临较高的通信时延。当前文献对联邦学习时延问题进行了研究 [16], ，而现有方法的效率会随着通信时延增加而下降。因此，严格的时延要求使得地面网络中联邦学习的集中式配置不适用于智能电网、自动驾驶汽车、电子健康和增强现实等未来应用 [17]。

为应对传统网络架构所带来的挑战，需要采用创新设计，以扩大覆盖范围并提供更可靠的通信信道。一种潜在解决方案是将非地面平台用于联邦学习，在传统地面基站覆盖范围有限或通信可靠性不足的情况下，这类平台能够提供更高效、更可靠的通信方式。不过需要注意的是，合适平台的选择取决于多种因素，而且非地面平台未必适用于或适合所有场景。鉴于地面基站的覆盖范围有限，且通信可靠性受环境影响，采用高空平台站（HAPS）等非地面平台似乎颇具前景 [18]。事实上，近期已有研究提出将HAPS作为6G网络的潜在候选方案，以实现泛在通信 [19]。HAPS可在约20 km的平流层高度运行，提供强视距（LOS）通信和广泛的覆盖范围，覆盖半径为50至500 km [20]。HAPS拥有较大的有效载荷，可搭载稳健的通信、计算和存储设备，以及续航时间长的电池 [21]。鉴于

这些优势，HAPS能够与大量用户可靠通信，并让他们参与联邦学习过程。此外，预计得益于无线信道质量的提升，以及大量用户参与联邦学习训练，基于HAPS的联邦学习能够比地面方案实现更快的收敛速度和更高的模型准确率。近期， [22], 研究利用HAPS的战略位置，为低地球轨道（LEO）卫星星座实现快速、高效的联邦学习。具体而言，HAPS被用于汇聚高速轨道运行的卫星客户端所传回的本地模型。相比之下，我们的工作聚焦于一种联邦学习系统，其中单个高空平台站作为服务器，促进大量地面客户端之间的协同学习。因此，我们的方法旨在服务更多地面客户端，而FedHAP则面向移动中的卫星。这一网络架构上的根本差异带来了不同的设计与优化挑战，而这正是我们工作所要解决的问题。

事实上，将联邦学习融入HAPS网络可带来潜在的应用场景。例如，在智慧城市中，高空平台站可用于交通管理和监控，联邦学习模型则可利用全市各处传感器和摄像头采集的数据进行训练。自动驾驶汽车和无人机也能受益于通过HAPS网络开展的联邦学习，从而基于本地数据进行实时决策。在医疗保健领域，通过HAPS网络开展的联邦学习可用于远程患者监测和诊断：利用分布在广大地理区域的可穿戴设备采集患者数据，并发送至高空平台站进行聚合，以训练隐私保护型联邦学习模型。在增强现实和虚拟现实应用中，通过HAPS网络开展的联邦学习能够实时处理传感器和摄像头采集的大量数据，从而打造大规模沉浸式体验。此外，通过HAPS网络开展的联邦学习有望为各种应用场景提供更高效、更可靠的解决方案。以自动驾驶汽车为例，使用HAPS网络有助于改善基站之间的切换，从而在切换过程中实现更高效、更可靠的通信。此外，HAPS网络可为覆盖广大地理区域的大规模联邦学习系统提供平台，这对于灾害响应等应用场景尤为重要，因为大量参与方可能分散在广阔区域内。

另一方面，即使参与方数量众多，非独立同分布（non-IID）数据分布也会减缓联邦学习的收敛速度，并对准确率产生负面影响 [23]。为解决这一问题，需要一种新机制，根据数据分布控制选定客户端的数量。例如，在医疗保健应用中，不同的

医院的患者人口统计特征和疾病状况可能存在巨大差异，从而导致非独立同分布（non-IID）数据分布 [24]。我们提出的机制能够优化客户端选择，在联邦学习时延/收敛速度与准确率之间实现尽可能好的平衡，即使在 non-IID 程度很高的情况下也不例外。通过谨慎控制选定客户端的数量，可以显著提高联邦学习模型的准确率，同时最大限度地降低端到端联邦学习时延。据我们所知，本研究首次针对高空平台辅助联邦学习系统提出联合参与方选择与资源分配设计策略，同时考虑时延和准确率。本文的主要贡献可总结如下:

- 我们讨论了地面环境中联邦学习面临的当前挑战，例如覆盖范围有限、无线环境异构以及客户端可用性不稳定，并着重说明高空平台站在帮助联邦学习解决这些问题方面所能带来的独特优势。高空平台站凭借广阔的覆盖区域和直接的视距链路，可以缓解这些挑战，提升联邦学习性能。

- 我们提出了一种新颖的资源感知联邦学习算法，在达到所需联邦学习精度的同时，将通信开销和计算开销降至最低。我们还推导出所提联邦学习算法的收敛率上界，该上界是选定客户端数量、通信轮次和位置迭代次数的函数。该算法有助于克服无线客户端计算资源有限的问题，并降低通信开销。

- 我们构建了一个联合通信与计算优化问题，以最小化上行链路和下行链路的端到端联邦学习时延。为解决这一问题，我们提出了一种迭代算法，并获得了闭式解。该优化问题综合考虑高空平台站和客户端的资源，旨在实现联邦学习精度与时延之间的平衡。

- 我们的仿真结果表明，与地面网络相比，高空平台站能够显著提升联邦学习性能。高空平台站的大覆盖范围和视距链路有利于联邦学习系统；我们提出的联合优化高空平台站与客户端资源的方案，能够降低联邦学习时延，同时权衡联邦学习准确率与时延。此外，我们还表明，高空平台站发生非预期位移对联邦学习时延性能的影响可以忽略不计。

C. 组织结构

本文其余部分安排如下。第二节介绍系统模型，并提出一种通信与计算资源感知联邦学习算法。第三节

表 I：符号列表。

符号	定义	符号	定义
K	可访问客户端总数	E_k^{cp}	客户端 k 的总能耗
$\mathcal{K}$	选定客户端数量	p_k^{cp}	客户端 k的计算功率
$\mathcal{J}_k$,$\mathbf{x}$ y	客户端 k 的本地数据集、输入向量、输出	t_k^{cp}	客户端 k 的计算时间
$\mathbf{w}$	全局联邦学习模型	p_k^{bc}	客户端 k的上传功率
f	每个客户端的损失函数	$\mathcal{N}_0$	噪声功率
F_k	客户端 k 的总损失函数	t_k^{up}	客户端 k 的上传时间
F	联邦学习总损失函数	E_k^{up}	客户端 k 上传时的能耗
J_k	客户端 k的数据样本	b_k	分配给客户端 k的带宽
$\mathbf{z}$	全局模型与本地更新之间的偏差	B	总网络带宽
i	本地迭代次数	r_k	客户端 k 的上行数据速率
n	通信轮数	r_H	下行数据速率
s	本地更新的固定大小	L_H	高空平台站服务器的计算密度
η	目标本地精度	p_H^{bc}	高空平台站服务器的广播功率
ϵ	目标全局精度	t_H^{bc}	高空平台站服务器的广播时间
v,M,u	正的常数值	a_k	客户端选择二元变量
δ	随机梯度下降的步长	E_H	高空平台站服务器的广播能量
C_k	客户端 k 的 CPU 周期数	p_H^{cp}	高空平台站服务器的计算功率
h_k	客户端 k 的信道功率增益	t_H^{cp}	高空平台站服务器的计算时间
f_k	客户端 k的 CPU处理能力	$E_k^{\max}$	客户端 k的最大可用能量
d_k	客户端 k 与高空平台站之间的距离	$E_H^{\max}$	高空平台站服务器的最大可用能量
g_k	莱斯信道功率增益	$f^{\min}$,$f^{\max}$	客户端的最小和最大计算能力
			客户端
$\hat{h}_0$	参考自由空间信道功率增益	τ^{FL}	（给定轮次的）端到端联邦学习时延
Δd	高斯分布随机位移	τ^{DL}	下行链路时延
	HAPS		
d_0	参考距离	τ^{UL}	上行时延
f_H	高空平台站服务器的计算能力	x_0	与泰勒展开相关的实数
ζ_k	与客户端硬件相关的系数 k	θ,$\boldsymbol{\gamma}$,$\boldsymbol{\lambda}$,ψ,ω	拉格朗日乘子
ζ_H	与高空平台站硬件相关的系数	l_k	客户端 k 的路径损耗模型
	服务器		

将构建相关优化问题，并介绍用于端到端联邦学习时延最小化的迭代算法。随后，第四节将展示并讨论仿真结果。最后，第五节给出结论。

II. 系统模型

A. 网络拓扑模型

我们考虑一种部署在综合地面—空中异构[1]网络上的FLSTRA系统，如图1所示。该综合网络由一个与平流层中的高空平台站共址的联邦学习服务器，以及 K 个位于地面高空平台站大范围覆盖区域内、随机分布的单天线地面设备组成。高空平台站负责计算、数据聚合、广播和飞行等操作，而地面设备负责计算，并将训练得到的本地更新传输至高空平台站上的联邦学习服务器。在FLSTRA系统中，从这些设备中选择一个子集，记为 $\mathcal{K}$，满足 $|\mathcal{K}| \leq K$，作为联邦学习客户端与高空平台站服务器连接，其中 $|\cdot|$ 为基数算子。每个客户端 k 拥有一个本地数据集 $\mathcal{J}_k$，其中包含 $|\mathcal{J}_k| = J_k$ 个数据样本。数据集$\mathcal{J}_k = \{\mathbf{x}_{kl}, y_{kl}\}_{l=1}^{J_k}$，其中（$\mathbf{x}_{kl}$,$y_{kl}$） $\in \mathbb{R}^q \times \mathbb{R}$ 表示大小为 q 的输入向量与对应于客户端 k的标量输出所组成的对。需要注意的是，我们提出的框架适用于任何高空平台站；但在本文中，我们聚焦于平流层平台，以展示其独特的特性和优势，例如较大的覆盖区域、稳定且长时间的飞行，以及视距通信信道。尽管如此，该框架也可轻松扩展至具有不同高度和覆盖区域的其他类型高空平台站。本文所提出的分析和解决方案与所采用的具体高空平台站类型无关，因此该框架可根据应用需求适配不同的高空平台站。

B. 联邦学习模型

采用传统联邦学习模型 [25], ，我们定义一个大小为 q 的向量 $\mathbf{w}$，用于收集 FLSTRA 的全局训练模型或参数，并引入损失函数 $f(\mathbf{w},\mathbf{x}_{kl},y_{kl})$ 来评估客户端 k 上学习的性能。我们假设所有客户端都接收相同的全局模型 $\mathbf{w}$，并使用相同的损失函数 f。对于不同的学习任务，损失函数的定义可能不同。例如，线性回归的损失函数为 $f(\mathbf{w},\mathbf{x}_{kl},y_{kl}) = \frac{1}{2}(\mathbf{x}_{kl}^T\mathbf{w} - y_{kl})^2$，而逻辑回归的损失函数为 $f(\mathbf{w},\mathbf{x}_{kl},y_{kl}) = -\log(1+\exp(-y_{kl}\mathbf{x}_{kl}^T\mathbf{w}))$。由于客户端 k 拥有 J_k 个数据样本，因此客户端 k 的总损失函数可定义为

$$F_k(\mathbf{w}) = \frac{1}{J_k}\sum_{l=1}^{J_k} f(\mathbf{w},\mathbf{x}_{kl},y_{kl}). \tag{1}$$

[1]异构是指网络中的设备可能拥有不同的计算或通信资源，以及不同的数据分布。

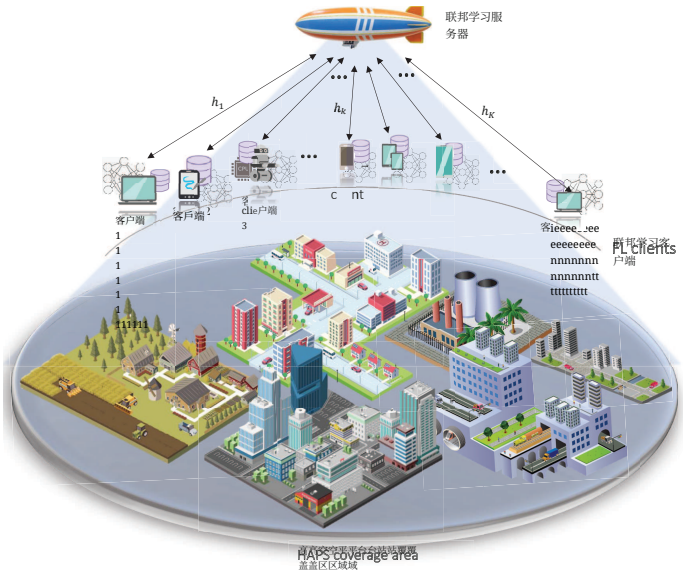


图1：系统模型。

要部署基于联邦学习的训练，必须邀请一组客户端参与训练过程，以训练其全局模型w。训练由所有受邀客户端共同执行，各客户端无需共享其原始数据，即可生成新的共同全局模型。因此，FLSTRA学习问题可表述为

$$\min_{\mathbf{w}\in\mathbb{R}^q}\left\{F(\mathbf{w})\triangleq\sum_{k\in\mathcal{K}}\frac{J_k}{J}F_k(\mathbf{w})=\frac{1}{J}\sum_{k\in\mathcal{K}}\sum_{l=1}^{J_k}f(\mathbf{w},\mathbf{x}_{kl},y_{kl})\right\},\tag{2}$$

其中，$J=\sum_{k=1}^{K}J_k=|\cup_{k\in\mathcal{K}}\mathcal{J}_k|$是所有参与客户端或选定客户端的数据样本总数。

为解决问题 (2)，我们通过改进联邦平均（FedAvg） [5] 算法，开发了通信与计算资源感知（CCRA-FL）算法。FedAvg由两部分交替组成：客户端执行若干次本地梯度更新，随后服务器执行一次模型平均更新。具体而言，CCRA-FL 使用随机梯度下降（SGD）作为优化器，在每个客户端上生成本地更新。批量梯度下降的一个缺点是，在大型数据集上，每次参数更新之前都会针对相似样本重复计算梯度，从而造成冗余计算。另一方面，随机梯度下降（SGD）通过每次执行一次更新来消除这种冗余 [26]。因此，随机梯度下降（SGD）通常比梯度下降快得多，更适用于对计算效率要求很高的大规模数据集。此外，在CCRA-FL 算法中，FedAvg 参数通过求解网络

资源优化问题来进行优化，该问题将在第 III 节中提出，并受通信与计算约束条件限制。

CCRA-FL算法以迭代方式求解问题（2），直到全局模型达到所需的全局目标精度 $\epsilon > 0$。我们用 n表示收敛所需的总迭代次数（即全局联邦学习迭代或通信轮次的数量）。具体而言，在通信轮次 n:

1）每个客户端 $k \in \mathcal{K}$ 在本地执行随机梯度下降，以获得其本地更新。2）覆盖范围内的所有选定客户端 $|\mathcal{K}|$ 使用频分多址（FDMA）方案，将其更新后的模型参数上传至高空平台站的联邦学习服务器。3）高空平台站的联邦学习服务器接收到所有本地更新后，对其进行聚合，并形成对应于通信轮次 n的新更新全局模型参数$\mathbf{w}^{(n)}$。4）高空平台站服务器将更新后的全局模型参数广播回所有客户端。5）每个客户端在每个通信轮次以给定的本地目标精度 η 求解局部优化问题。

在本文余下部分中，我们将联邦学习的全局迭代称为通信轮次，将本地联邦学习迭代称为本地迭代。

我们进一步假设，全局模型与本地更新的大小相同，并且在所有通信轮次中保持不变，即 $|\mathbf{w}^{(n)}| = s$比特，$\forall n$。每个客户端 $k \in \mathcal{K}$ 在每个通信轮次中求解以下局部优化问题:

$$\min_{\mathbf{z}_k^{(n)} \in \mathbb{R}^q} \left\{ G_k(\mathbf{w}^{(n)}, \mathbf{z}_k^{(n)}) \triangleq F_k(\mathbf{w}^{(n)} + \mathbf{z}_k^{(n)}) - (\nabla F_k(\mathbf{w}^{(n)}) - \xi \nabla F(\mathbf{w}^{(n)}))^T \mathbf{z}_k^{(n)} \right\}, \tag{3}$$

其中，ξ 是一个常数，$\mathbf{z}_k^{(n)}$ 是通信轮次 n 时全局模型与客户端 k 的本地更新之间的偏差项。具体而言，$\mathbf{w}^{(n)} + \mathbf{z}_k^{(n)}$ 是通信轮次 n 时客户端 k 的本地更新。$\nabla F_k(\mathbf{w}^{(n)})$ 是函数 $F_k(\mathbf{w}^{(n)})$ 在点 $\mathbf{w}^{(n)}$ 处的梯度。每个客户端采用随机梯度下降法求解 (3)。为分析随机梯度下降法的收敛率，我们对损失函数 $F_k(.)$ 作出以下标准假设:

假设1（[12]）。若满足以下条件，则 $F_k(\cdot)$ 是 M-利普希茨连续且 u-强凸:

$$F_k(\mathbf{w}) \leq F_k(\mathbf{w}') + \langle \nabla F_k(\mathbf{w}'), \mathbf{w} - \mathbf{w}' \rangle + \frac{M}{2} \|\mathbf{w} - \mathbf{w}'\|^2, \tag{4}$$

$$F_k(\mathbf{w}) \geq F_k(\mathbf{w}') + \langle \nabla F_k(\mathbf{w}'), \mathbf{w} - \mathbf{w}' \rangle + \frac{u}{2} \|\mathbf{w} - \mathbf{w}'\|^2, \tag{5}$$

其中，利普希茨常数 M 和 u 为正数，$\langle \cdot, \cdot \rangle$ 表示内积运算，$\|\cdot\|$表示欧几里得范数，而 $\nabla F_k(\mathbf{w}')$表示 $F_k(\cdot)$ 关于 $\mathbf{w}'$ 的梯度。

根据假设1，如果最大随机梯度下降迭代次数 i 的下界满足 [10]，则问题 (3) 可以达到给定的目标本地精度 η （即， η的值越低，本地精度越高，反之亦然）。

$$i \geq v \log_2\left(\frac{1}{\eta}\right), \tag{6}$$

其中， $v = \frac{2}{(2-M\delta)\delta u}$, $\delta < \frac{2}{M}$ 是随机梯度下降法中的步长，且 η 满足CCRA-FL算法中随机梯度下降法的以下收敛条件:

$$G_k(\mathbf{w}^{(n)}, \mathbf{z}_k^{(n,i)}) - G_k(\mathbf{w}^{(n)}, \mathbf{z}_k^{(n)*}) \leq \eta \left[G_k(\mathbf{w}^{(n)}, \mathbf{z}_k^{(n,0)}) - G_k(\mathbf{w}^{(n)}, \mathbf{z}_k^{(n)*}) \right], \tag{7}$$

其中，$\mathbf{z}_k^{(n)*}$ 是局部优化问题（3）的最优解。有关上述界（6）和（7）的证明的更多细节，请参见 [10]。（6）表明，要获得高精度的本地更新（即较低的 η值），就必须增加本地迭代次数。

在上述假设下，我们在定理1中给出了CCRA-FL算法收敛率的一个上界。具体而言，定理1给出了CCRA-FL算法的最终损失与最优全局模型之间差值的上界。该差值表示联邦学习系统收敛率的一个上界，是衡量此类系统性能的常用指标 [27]。值得注意的是，最优全局模型的损失函数始终小于或等于任何全局模型的损失函数，因为最优全局模型是使全局损失函数最小化的模型。因此，定理1中的最终损失与最优损失之差始终为正，代表CCRA-FL算法收敛率的上界。

定理1. 若运行CCRA-FL算法 n 个通信轮次，在每个通信轮次中有 $|\mathcal{K}^{(\cdot)}|$ 个选定客户端（处于高空平台站的覆盖范围内），且给定目标本地精度 η，则其收敛率的上界为

$$F(\mathbf{w}^{(n)}) - F(\mathbf{w}^*) \leq \epsilon, \tag{8}$$

其中

$$\epsilon = \left(1 - \frac{(1-\eta)u^2\xi}{2M^2}\right)^n \left[F(\mathbf{w}^{(0)}) - F(\mathbf{w}^*)\right] + \frac{M\xi - u}{2\xi} \times \left[\sum_{n'=0}^{n-1} \frac{1}{|\mathcal{K}^{(n')}|} \left(1 - \frac{(1-\eta)u^2\xi}{2M^2}\right)^{n-n'-1} \sum_{k \in \mathcal{K}^{(n')}} \|\mathbf{z}_k^{(n')}\|^2\right], \tag{9}$$

是满足条件 $u \leq M \min\left\{\zeta, \sqrt{\frac{2}{\zeta}}\right\}$ 时联邦学习过程的目标收敛上界，可保证 $\epsilon \geq 0$。$\mathbf{z}^{(n')}{}_k$ 是通信轮次 n', $n' \in \{0,...,n\}$ 中客户端 k 的全局模型与本地更新之间的偏差项。

算法1 CCRA-FL算法。

1: 初始化 $\rightarrow$ 全局模型 $\mathbf{w}^{(0)}$ 和 $n=0$。

循环过程

2：优化 $\rightarrow$ 高空平台站使用算法2寻找 $\eta^{(n)}$、$a_k^{(n)}$ 和资源分配策略。 3：广播 $\rightarrow$ 高空平台站向选定客户端传输 $\mathbf{w}^{(n)}$ 。 4：本地计算 $\rightarrow$ 每个选定客户端 $k\in\mathcal{K}^{(n)}$、$|\mathcal{K}^{(n)}|=\sum_{k=1}^{K}a_k^{(n)}$，计算 $\nabla F_k(\mathbf{w}^{(n)})$ 并将其上传至高空平台站服务器。 5：聚合$\rightarrow$ 高空平台站服务器聚合所有梯度，并计算$\nabla F(\mathbf{w}^{(n)}) =\frac{1}{|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\nabla F_k(\mathbf{w}^{(n)})$。 并行循环 $k\in\mathcal{K}^{(n)}$ 6：初始化 $\rightarrow$ $i=0$ 和 $\mathbf{z}_k^{(n,0)}=0$。 循环过程（随机梯度下降算法）7：更新 $\rightarrow$ $\mathbf{z}_k^{(n,i+1)}=\mathbf{z}_k^{(n,i)}-\delta\nabla G_k(\mathbf{w}^{(n)},\mathbf{z}_k^{(n,i)})$ 和 $i=i+1$。 8：直到 $i=\upsilon\log_2(\frac{1}{\eta^{(n)}})$，即达到本地精度 $\eta^{(n)}$。 9：本地更新 $\rightarrow$ 将 $\mathbf{z}_k^{(n)}=\mathbf{z}_k^{(n,i)}$ 记为并上传 $\mathbf{z}_k^{(n)}$ 至高空平台站服务器。

结束循环

10：新的全局模型 $\rightarrow\mathbf{w}^{(n+1)}=\mathbf{w}^{(n)}+\frac{1}{|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\mathbf{z}_k^{(n)}$。 11：设置 $n=n+1$。12：直到在（8）中达到目标收敛上界 ϵ 。

证明。请参见附录A。

我们观察到，定理1在式（9）的右侧包含一个新项，而 [10] 中给出的结果并未包含这一项。这个附加项反映了选定客户端数量对基于高空平台的系统中联邦学习收敛率的影响。我们的分析表明，每个通信轮次中选定客户端数量的增加 $|\mathcal{K}^{(n)}|$ 会使偏差项消失，从而加快收敛速度。然而，这种改进是以增加网络资源利用率、提高时延和增加能耗为代价的。值得注意的是，对选定客户端数量的这种依赖是基于高空平台的联邦学习所独有的，这是因为其覆盖范围广，使得每轮通信都能灵活选择不同数量的参与方。

C. 计算模型

FLSTRA 流程包括客户端的本地梯度计算，以及由高空平台上的联邦学习服务器聚合本地更新。因此，高空平台站服务器和客户端都要执行计算任务。下面将介绍它们的计算模型。

1）客户端的计算模型：对于客户端 k 的本地计算，我们将 f_k （单位为每秒 CPU 周期）记为 CPU计算能力，将 C_k 记为处理一个数据样本所需的 CPU周期数。在任意通信轮次中，处理 J_k 个数据样本并进行 i 次本地迭代所需的计算时间为

$$t_k^{cp} = \frac{iC_k J_k}{f_k}, \forall k. \tag{10}$$

在一个通信轮次中，客户端 k 的相应能耗可计算如下 [28, 引理 1]:

$$E_k^{cp} = t_k^{cp} p_k^{cp}, \forall k, \tag{11}$$

其中， $p_k^{cp} = \zeta_k f_k^3$ 是客户端 k 的计算功率 [2]， ζ_k 是取决于客户端 k 的硬件和芯片架构的系数。

2）高空平台站上联邦学习服务器的计算模型：在每个通信轮次中，高空平台站接收的本地更新会被解码并聚合，以生成新的全局模型。该任务的处理时间可计算为

$$t_H^{cp} = \frac{L_H Q_H}{f_H}, \tag{12}$$

其中， $Q_H = s\sum_{k=1}^{K} a_k$ 是在给定通信轮次中高空平台站需要处理的总比特数， s 是高空平台站接收到的单个客户端本地更新的大小（单位为比特）。我们使用二元变量 a_k 来表示客户端 k 是否被选中进行联邦学习训练，并在给定通信轮次期间将其本地更新上传至高空平台站服务器。L_H 是计算密度（单位为 CPU周期/比特）， f_H 是高空平台站联邦学习服务器的 CPU计算能力 [30]。

D. 信道与通信模型

我们同时考虑大尺度和小尺度衰落效应，以对地面客户端与高空平台站之间的通信信道进行建模。令 d_k 表示客户端 k之间的距离

[2]根据 [28, 引理 1], ，每个客户端在每个通信轮次的能耗可计算为$E_k^{cp} = i\zeta_k C_k J_k f_k^2$， $\forall k$。因此，计算功耗可确定为 $p_k^{cp} = E_k^{cp}/t_k^{cp}$，使得 $p_k^{cp} = \zeta_k f_k^3$， $\forall k$.

与高空平台站之间的距离，其中 $k \in \{1,...,K\}$。因此，可以计算给定全局轮次下对应的信道功率增益：

$$h_k = g_k \hat{h}_0 \left(\frac{d'_k}{d_0} \right)^{-2}, \forall k, \tag{13}$$

其中， $d'_k = d_k + \Delta d$ 是客户端 k 与高空平台站之间的有效距离， Δd 表示由于平流层风等自然现象 [31]的影响，高空平台站偏离其原始位置的随机位移。我们假设 Δd 服从均值为零、方差为 σ^2 的高斯分布。$\hat{h}_0$ 表示 $d_0 = 1$ 公里处的参考自由空间信道功率增益。 g_k 刻画了小尺度衰落效应，并遵循莱斯块衰落模型。我们采用块衰落信道模型，即信道增益在一个通信轮次内保持不变，但在不同轮次之间独立变化。跨轮次的信道变化由莱斯模型（用于刻画小尺度衰落）和平均路径损耗（用于描述大尺度衰落）共同表征 [32]。该模型能够有效适应不同轮次之间的信道条件变化，并使我们能够在真实条件下评估系统性能 [33]。

1）上行传输：为减轻在时分多址（TDMA）中客户端依次上传其更新所造成的时延，我们在系统模型中为上行传输采用频分多址（FDMA）。采用 FDMA 时，本地计算完成后，多个客户端可以在不同频段上同时将其本地更新上传至高空平台站服务器进行聚合，从而降低整体时延并提高系统效率 [34]。因此，客户端 k 的瞬时可达上行速率为

$$r_k = b_k \log_2 \left(1 + \frac{h_k p_k^{up}}{b_k \mathcal{N}_0} \right), \forall k, \tag{14}$$

其中， b_k 是分配给客户端 k 的带宽， p_k^{up} 是发射功率， $\mathcal{N}_0$ 是以 dBm/Hz 为单位的噪声功率谱密度。此外，为确保每个选定客户端在上传时间 t_k^{up} 内上传其本地更新，必须满足以下约束：

$$r_k t_k^{up} \geq s, \forall k. \tag{15}$$

因此，客户端 k 上传本地更新所消耗的相应能量可表示为

$$E_k^{up} = t_k^{up} p_k^{up}, \forall k. \tag{16}$$

2）下行传输：聚合所有本地更新并生成新的全局模型后，高空平台站服务器会将新模型广播给所有客户端。高空平台站服务器可以根据选定客户端的信道质量调整广播功率，从而调整数据速率。具体而言，服务器会根据瞬时信噪比（SNR）最差的客户端调整数据速率。因此，每次通信轮次或全局联邦学习轮次中相应的瞬时可达下行速率可确定为

$$r_H = \min_{k \in \mathcal{K}} B \log_2\left(1 + \frac{p_H^{bc} h_k}{B \mathcal{N}_0}\right), \tag{17}$$

其中， p_H^{bc} 是分配用于广播全局更新的功率， B 是可用网络带宽总量。下行链路速率方程确保高空平台站服务器始终以最高可能的数据速率进行广播，同时保证其覆盖范围内的所有选定客户端均可连接。因此，为确保在广播时间t_H^{bc}内将新的全局模型 $\mathbf{w}^{(n)}$ 传输给所有选定客户端，应满足以下约束：

$$r_H t_H^{bc} \geq s, \tag{18}$$

其中， s 是在给定通信轮次中全局更新的大小。因此，高空平台站的能耗可表示为

$$E_H = p_H^{cp} t_H^{cp} + p_H^{bc} t_H^{bc}, \tag{19}$$

其中，$p_H^{cp} = \zeta_H f_H^3$ 是高空平台站处理接收到的本地更新并生成新的全局模型所需的计算功耗， ζ_H 是取决于高空平台站服务器 [35]的硬件和芯片架构的系数。由于我们假设飞行功率在所有通信轮次 [36], 中保持不变，因此可以在高空平台站的总能耗中忽略它。

III. D 延迟最小化问题表述与所提解决方案

根据定理1以及高空平台站固有的广阔覆盖范围，可以通过单跳通信邀请大量可访问客户端参与 FLSTRA 过程，从而提高联邦学习收敛率。此外，定理1表明，通信（和计算）效率与联邦学习收敛率之间存在权衡。因此，在本研究中，我们旨在通过优化选定客户端的数量和资源分配，在每个通信轮次最小化端到端联邦学习时延的同时，提高联邦学习收敛率。该时延包括每个通信轮次中上行链路和下行链路的通信与计算时间。本节首先针对 FLSTRA 建立端到端时延最小化问题。其次，由于相应问题具有非凸性，我们设计了一种迭代算法来求解。

A. 问题表述

我们的目标是设计联合资源分配与客户端选择策略，在达到目标全局精度ϵ的同时，最小化一个通信轮次内的端到端联邦学习时延。设 τ^{FL} 为给定通信轮次中联邦学习过程的端到端时延，其定义为 $\tau^{FL}=\tau^{UL}+\tau^{DL}$。 τ^{UL} 和 τ^{DL} 分别为上行链路时延和下行链路时延，其确定方式如下：

$$\tau^{UL}=\max_{k}\ a_k(t_k^{up}+t_k^{cp}), \tag{20}$$

$$\tau^{DL}=t_H^{bc}+t_H^{cp}, \tag{21}$$

其中，t_H^{cp} 和 t_H^{bc}分别是高空平台站为计算（即聚合接收到的本地更新）以及在给定通信轮次 n向参与客户端广播全局模型参数 $\mathbf{w}^{(n)}$ 所分配的时间限制。

现在，在给定的通信轮次下，同时考虑上行链路和下行传输，将端到端时延最小化问题表述如下：

$$\min_{p_H^{bc},t_H^{bc},f_H,\eta,\mathbf{p},\mathbf{f},\mathbf{b},\mathbf{a},\mathbf{t}} \quad \tau^{FL} \tag{22a}$$

$$\text{s.t.} \quad a_k\left[t_k^{up}b_k\log_2(1+\frac{p_k^{up}h_k}{b_k\mathcal{N}_0})-s\right]\geq 0,\forall k, \tag{22b}$$

$$a_k\left[t_H^{bc}B\log_2(1+\frac{p_H^{bc}h_k}{B\mathcal{N}_0})-s\right]\geq 0,\forall k, \tag{22c}$$

$$a_k\left[t_k^{up}p_k^{up}+p_k^{cp}t_k^{cp}\right]\leq E_k^{\max},\forall k, \tag{22d}$$

$$p_H^{cp}t_H^{cp}+p_H^{bc}t_H^{bc}\leq E_H^{\max}, \tag{22e}$$

$$\frac{vC_kJ_k\log_2(\frac{1}{\eta})}{f_k}\leq t_k^{cp},\forall k, \tag{22f}$$

$$\frac{L_HQ_H}{f_H}\leq t_H^{cp}, \tag{22g}$$

$$\sum_{k=1}^{K}a_kb_k\leq B, \tag{22h}$$

$$a_k\in\{0,1\},\forall k, \tag{22i}$$

$$0\leq p_k^{up}\leq p_k^{\max},\forall k, \tag{22j}$$

$$f_k^{\min}\leq f_k\leq f_k^{\max},\forall k, \tag{22k}$$

$$0<\eta\leq 1,\forall k, \tag{22l}$$

$$t_k^{up},t_H^{bc}\geq 0,b_k\geq 0,\forall k, \tag{22m}$$

算法 2：所提出的迭代算法。

1：初始化 $\rightarrow$ 高空平台站参数（$p_H^{bc(0)}$，$t_H^{bc(0)}$，$f_H^{(0)}$）、资源分配参数（$\mathbf{f}^{(0)}$,$\mathbf{t}^{(0)}$,$\mathbf{p}^{(0)}$,$\mathbf{b}^{(0)}$）、用户选择 $\mathbf{a}^{(0)}$、目标本地精度 $\eta^{(0)}$, $l^{\max}$ 和 $l=0$。循环过程 2：求解 (23) $\rightarrow$ 对于给定的（$p_H^{bc(l)}$,$t_H^{bc(l)}$,$f_H^{(l)}$,$\mathbf{t}^{(l)}$,$\mathbf{p}^{(l)}$,$\mathbf{b}^{(l)}$,$\mathbf{a}^{(l)}$），求得最优的（$\eta^{(l+1)}$,$\mathbf{t}^{(l+1)}$）。3：求解 (27) $\rightarrow$ 给定（$\eta^{(l+1)}$,$\mathbf{t}^{(l+1)}$），求得最优的 ($\mathbf{f}^{(l+1)}$,$\mathbf{a}^{(l+1)}$)。4：求解 (29) $\rightarrow$ 对于给定的（$\eta^{(l+1)}$,$\mathbf{t}^{(l+1)}$,$\mathbf{f}^{(l+1)}$,$\mathbf{a}^{(l+1)}$），求得最优的 ($\mathbf{p}^{(l+1)}$,$\mathbf{b}^{(l+1)}$)。5：求解 (32) $\rightarrow$ 给定（$\eta^{(l+1)}$,$\mathbf{t}^{(l+1)}$,$\mathbf{f}^{(l+1)}$,$\mathbf{a}^{(l+1)}$, $\mathbf{p}^{(l+1)}$,$\mathbf{b}^{(l+1)}$），求得最优的高空平台站参数（$p_H^{bc(l+1)}$, $t_H^{bc(l+1)}$, $f_H^{(l+1)}$）。6：设置 $l=l+1$。7：直到收敛=为真或 $l=l^{\max}$

其中 $\mathbf{b}=[b_1,...,b_K]$、$\mathbf{f}=[f_1,...,f_K]$、$\mathbf{a}=[a_1,...,a_K]$、$\mathbf{t}=[t_1^{up},...,t_K^{up}]$ 和 $\mathbf{p}=[p_1^{up},...,p_K^{up}]$。(22b) 和 (22c) 分别是上行链路和下行链路数据传输约束。约束条件 (22d) 确保每个选定客户端在每个通信轮次中的本地能耗不超过客户端 k 的最大可用能量（即 $E_k^{\max}$）。约束条件 (22e) 确保高空平台站每个通信轮次的总能耗不超过 $E_H^{\max}$。(22f) 和 (22g) 分别表示客户端和高空平台站服务器的计算时间约束。约束条件 (22h) 确保客户端之间的带宽分配不超过 B Hz。约束条件 (22i) 将客户端选择决策变量限制为只能取二进制值，而约束条件 (22j)-(22k) 分别表示客户端的平均发射功率限制和最大本地计算能力。$p_k^{\max}$ 是客户端 k 可用于上传的最大功率。最后，目标本地精度约束由 (22l) 给出。

B. 所提出的迭代算法

我们注意到，问题 (22) 是一个混合整数非线性规划（MINLP），无法通过标准凸优化方法在快速收敛的同时获得最优解。我们提出了一种高效的迭代算法，通过交替优化四个子问题，直至算法收敛。下面将讨论求解这四个子问题所涉及的步骤。

1) 最优客户端上传时间和本地精度 ($\mathbf{t}^*$,η^*)：第一步，我们固定与高空平台站相关的变量（即 $\{p_H^{bc},t_H^{bc},f_H\}$）、资源分配变量和用户选择变量

（即 (p,f,b,a,t)），然后通过求解以下第一个子问题，联合优化客户端上传时间和目标本地精度：

$$\min_{\mathbf{t},\eta} \quad \Big[\max_k \; a_k\Big(t_k^{up}+\frac{vC_kJ_k\log_2(\frac{1}{\eta})}{f_k}\Big)\Big] \tag{23a}$$

$$\text{s.t.} \quad t_k^{up}\geq t_k^{\min},\forall k, \tag{23b}$$

$$a_k\Big[t_k^{up}p_k^{up}+\zeta_k vC_kJ_k\log_2(\frac{1}{\eta})f_k^2\Big]\leq E_k^{\max},\forall k, \tag{23c}$$

$$0<\eta\leq 1, \tag{23d}$$

其中

$$t_k^{\min}=\frac{a_k s}{b_k\log_2(1+\frac{p_k^{up}h_k}{b_k\mathcal{N}_0})},\forall k. \tag{24}$$

在问题 (23) 中，我们旨在最小化 τ^{UL}，这体现为最小化参与客户端中最差客户端的上行时延。从时延最小化的角度来看，将本地上传时间分配为最短总是最有效的。因此，可以利用以下定理 2 得到 (23a) 的闭式最优解。

定理 2. 问题 (23) 的最优解 ($\mathbf{t}^*=[t_k^{up^*}]_{1\times K},\eta^*$) 满足

$$t_k^{up^*}=t_k^{\min},\forall k, \tag{25}$$

and

$$\eta^*=\frac{\beta_1+\beta_2}{\theta\ln 2}, \tag{26}$$

其中 $\beta_1=\max_k\frac{a_kvC_kJ_k}{f_k}$ and $\beta_2=\sum_{k\in\mathcal{K}}\lambda_k a_k\zeta_k vC_kJ_kf_k^2$。此外，$\boldsymbol{\lambda}=[\lambda_k]_{1\times K}$ and θ 分别是与约束条件 (23c) 和 (23d) 相关联的拉格朗日乘子。

证明. 见附录 B。

2）最优客户端选择与计算能力 ($\mathbf{a}^*$, $\mathbf{f}^*$)：在第二步中，我们旨在通过求解以下子问题，在给定 ($\mathbf{t}$ η,p,b) 的情况下，同时确定最优客户端选择策略（即 a）和本地计算能力（即 f）：

$$\min_{\mathbf{a},\mathbf{f}} \quad \tau^{FL} \tag{27a}$$

$$\text{s.t.} \quad a_k t_k^{up} b_k \log_2(1+\frac{p_k^{up} h_k}{b_k \mathcal{N}_0}) \geq a_k s, \forall k, \tag{27b}$$

$$a_k \left[t_k^{up} p_k^{up} + \zeta_k v C_k J_k \log_2(\frac{1}{\eta}) f_k^2 \right] \leq E_k^{\max}, \forall k, \tag{27c}$$

$$a_k t_H^{bc} B \log_2(1+\frac{p_H^{bc} h_k}{B \mathcal{N}_0}) \geq a_k s, \forall k, \tag{27d}$$

$$\sum_{k=1}^{K} a_k \leq \frac{f_H t_H^{cp}}{L_H s}, \tag{27e}$$

$$\sum_{k=1}^{K} a_k b_k \leq B, \tag{27f}$$

$$a_k \in \{0,1\}, \forall k, \tag{27g}$$

$$f_k^{\min} \leq f_k \leq f_k^{\max}, \forall k. \tag{27h}$$

为了最小化给定通信轮次的时延，需要最小化本地计算时间，这等价于最大化本地计算能力 f。因此，利用 (27c) 和 (27h)，可推导出最优的 $\mathbf{f}^* = [f_k^*]_{1\times K}$ 为

$$f_k^* = \min\left\{ f_k^{\max}, \sqrt{\frac{a_k (E_k^{\max} - t_k^{up} p_k^{up})}{\zeta_k v C_k J_k \log_2(\frac{1}{\eta})}} \right\}, \forall k = 1,...,K. \tag{28}$$

对于给定的 $\mathbf{f}^*$，(27) 仍然是一个混合整数非线性规划。为了解决该问题，我们将二元变量 a_k近似为连续变量，使得 $0 \leq a_k \leq 1, \forall k$ [38]。随后，问题 (27) 转化为关于 $= [a_k]_{1\times K}$ 的线性规划。因此，可以使用内点法等标准凸优化方法高效求解。

3）最优功率与带宽分配（$\mathbf{p}^*$、$\mathbf{b}^*$）：为找到使端到端上行链路时延最小的最优客户端功率与带宽分配，我们求解以下第三个子问题：

$$\min_{\mathbf{p},\mathbf{b}} \quad \tau^{UL} \tag{29a}$$

$$\text{s.t.} \quad a_k b_k \log_2(1+\frac{p_k^{up} h_k}{b_k \mathcal{N}_0}) \geq \frac{a_k s}{t_k^{up}}, \forall k, \tag{29b}$$

$$p_k^{up} \leq \frac{a_k \left[E_k^{\max} - \zeta_k v C_k J_k \log_2(\frac{1}{\eta}) f_k^2 \right]}{t_k^{up}}, \forall k, \tag{29c}$$

$$\sum_{k=1}^{K} a_k b_k \leq B, \tag{29d}$$

$$0 \leq p_k^{up} \leq p_k^{\max}, \forall k, \tag{29e}$$

$$b_k \geq 0, \forall k. \tag{29f}$$

利用以下定理 3，可以得到（29）的闭式最优解。

定理 3. 问题 (29) 的最优解 $(\mathbf{p}^* = [p_k^{up^*}]_{1\times K}, \mathbf{b}^* = [b_k^*]_{1\times K})$ 满足

$$p_k^{up^*} = \min \left\{ p_k^{\max}, \frac{a_k \left[E_k^{\max} - \zeta_k v C_k J_k \log_2(\frac{1}{\eta}) f_k^2 \right]}{t_k^{up}} \right\}, \forall k, \tag{30}$$

and

$$b_k^* = \frac{\beta_1 + \beta_2(x_0 - \pi_k)}{2\beta_2} + \frac{\sqrt{(\beta_1 + \beta_2(x_0 - \pi_k))^2 + 4\beta_2 \pi_k (\beta_1 + \beta_2 x_0 + 1)}}{2\beta_2}, \forall k, \tag{31}$$

其中 $\pi_k = \frac{a_k h_k p^{up^*}}{\mathcal{N}_0}$, $\beta_1 = \ln(1+\frac{\pi_k}{x_0})$, $\beta_2 = \frac{x_0}{(x_0+\pi_k)}$, 且 x_0 是与泰勒展开相关的实数值。

证明。参见附录 C。

4) 最优 HAPS 资源分配 $(p_H^{bc^*}, t_H^{bc^*}, f_H^*)$：在所提出的迭代算法的最后一步，我们需要求解以下第四个子问题，以优化下行链路中的 HAPS 资源分配：

$$\min_{t_H^{bc}, p_H^{bc}, f_H} \quad \tau^{DL} \tag{32a}$$

$$\text{s.t.} \quad t_H^{bc} \geq t_H^{\min}, \tag{32b}$$

$$\zeta_H f_H^2 L_H Q_H + p_H^{bc} t_H^{bc} \leq E_H^{\max}, \tag{32c}$$

其中

$$t_H^{\min} = \frac{s}{\min_{k\in\mathcal{K}} B \log_2(1+\frac{p_H^{bc} h_k}{B\mathcal{N}_0})}. \tag{33}$$

在这个子问题中，我们假设与客户端相关的参数已知。我们进一步将该问题分解为两个新的子问题，分别优化 f_H 和 (t_H^{bc}, p_H^{bc})。

因此，在给定 HAPS 广播时间和功率分配策略的情况下，(32) 关于 f_H 是凸的。根据 Karush-Kuhn-Tucker（KKT）条件，可以确定 HAPS 的闭式最优计算能力，如下所示：

$$f_H^* = \sqrt[3]{\frac{1}{2\omega\zeta_H}}, \tag{34}$$

其中，ω 是与约束 (32c) 相关的拉格朗日乘子，其计算方法见附录D。

现在，对于给定的 f_H^*，我们求解以下问题，以确定高空平台站的最优发射功率和时间分配：

$$\min_{t_H^{bc}, p_H^{bc}} \quad \tau^{DL} \tag{35a}$$

$$\text{s.t.} \quad t_H^{bc} \geq t_H^{\min}, \tag{35b}$$

$$\zeta_H {f_H^*}^2 L_H Q_H + p_H^{bc} t_H^{bc} \leq E_H^{\max}. \tag{35c}$$

问题 (35) 的闭式解可通过以下定理4得到：

定理4. 给定 f_H^* 时，问题 (35) 的最优解 $(t_H^{bc^*}, p_H^{bc^*})$ 满足

$$t_H^{bc^*} = t_H^{\min}, \tag{36}$$

and

$$p_H^{bc^*} = \sqrt{\frac{B\mathcal{N}_0}{\psi \min_{k\in\mathcal{K}} h_k}}, \tag{37}$$

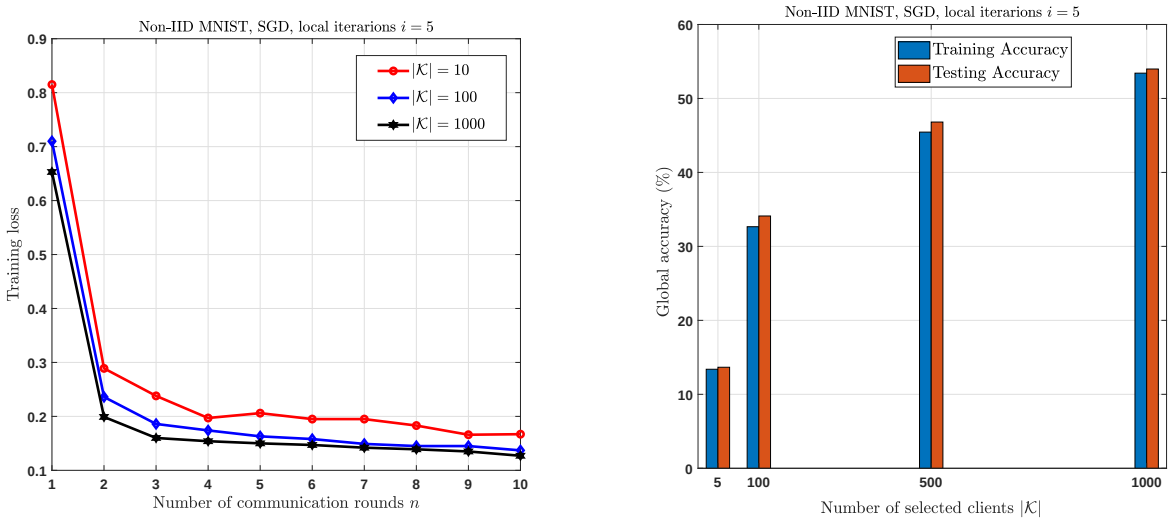
其中，ψ 是与约束 (32c) 相关的拉格朗日乘子。

证明. 见附录D。

伪　　代码总结了所提出的迭代算法，详见算法2。

C. 收敛性与复杂度分析

1) 收敛性分析：子问题 (23)、(29) 和 (32) 均为凸问题，而子问题 (27) 经二进制松弛后，就各自的优化变量块而言也变为凸问题。此外，每个子问题的目标函数均以零为下界，因为客户端上行链路时间、通信时延和高空平台站下行链路时间均不能为负值。在算法2的每次迭代 l 中，由于凸性，子问题都有唯一的最小值，并且目标函数值总是低于上一次迭代的值，



(a) 训练损失与通信轮数的关系 n。(b) 全局准确率与选定客户端数量的关系$|\mathcal{K}|$。

图2：不同选定客户端数量下的全局联邦学习准确率（即 $1-\epsilon$）和训练损失$|\mathcal{K}|$。

即 $\tau^{FL,\ (l)} < \tau^{FL,\ (l-1)} < \ldots < \tau^{FL,\ (0)}$，其中 $\tau^{FL,\ (l)}$表示算法2在第 l 次迭代时的目标函数值。该算法返回一个以零为下界的单调递减目标函数值序列，因此该序列收敛。2) 复杂度分析：我们分析算法2的复杂度。使用算法2求解时延最小化问题 (22) 时，每一步的主要复杂度在于求解问题 (23) 和问题 (27)。求解问题 (23) 的复杂度在于根据定理 3 得到最优的 η^* 和 t* ，这涉及复杂度 $\mathcal{O}(K\log_2(1/\epsilon_1))$，准确率为 ϵ_1。求解问题 (27) 时，总体复杂度主要取决于使用内点法求解凸优化问题以得到 a 的时间复杂度。因此，求解该问题的复杂度为 $\mathcal{O}(K^3)$。由此，所提出的算法2的总复杂度为 $\mathcal{O}(l_{it}K^2(K+\log_2(1/\epsilon_1))$，其中 l_{it} 是算法2中迭代优化变量的迭代次数。与近期工作 [10]，相比，我们提出的方法计算复杂度更低。

D. 框架与解决方案总结

我们的提案 所提出的联邦学习框架、问题和求解方法总结如下 s:

- 我们提出的联邦学习框架 CCRA-FL 在启用高空平台的网络架构中，采用联合通信与计算资源优化问题。根据算法 1，为解决与联邦学习过程相关的局部优化问题，我们对 FedAvg 算法进行了调整，并使用随机梯度下降在每个客户端上生成本地更新。CCRA-FL 通过求解满足通信和计算约束的网络资源优化问题来优化 FedAvg 参数。

- 在定理1中，我们给出了CCRA-FL算法收敛率的上界，这表明可以通过单跳通信邀请大量可访问客户端参与联邦学习过程，从而提高联邦学习收敛率。不过，我们也认识到通信效率与联邦学习收敛率之间存在权衡，因此我们着重于优化客户端选择和网络资源，同时最小化每个通信轮次的端到端联邦学习时延。

- 为解决端到端时延最小化问题，我们开发了一种迭代算法，即算法2，通过交替优化四个子问题，直至算法收敛到给定阈值。该问题属于混合整数非线性规划，无法通过标准凸优化方法求得最优解。所提出的迭代算法对高空平台站和客户端的资源分配、本地精度以及客户端选择策略进行优化。

E. 扩展至多高空平台场景

针对单个高空平台的 FLSTRA 框架可扩展至多高空平台场景，用于处理地理位置相距甚远的海量数据库。考虑这样一种场景：由多个高空平台组成的网络覆盖一组不同的地理区域，每个高空平台都采用算法1进行联邦学习。每个高空平台在学习到区域专属全局模型后，会与其他高空平台交换其模型参数，以学习最终全局模型。要生成最终全局模型，各高空平台需要相互协作，对所有高空平台的区域专属全局模型参数进行聚合。聚合可以通过去中心化方式进行。因此，无需共享或泄露海量数据库，所有地理区域都能获得最终全局模型。FLSTRA 在自动驾驶车辆、现代医疗保健和智慧城市等应用中极具价值，因为独立数据库可能分布于多个大型城市。对于多高空平台场景，可在每个高空平台上便捷地采用算法1来实现 FLSTRA。然而，要获得最终全局模型，我们需要开发一种端到端、去中心化、资源感知的高空平台间学习算法，以最大限度地降低时延。

IV. 仿真结果与讨论

本节通过数值分析评估并讨论所提出的 CCRA- 联邦学习算法的性能。在仿真中，我们考虑 $K \in \{10,100,1000\}$ 个客户端均匀分布在半径为 50 km 的圆形区域内，高空平台位于区域中心上方 25 km 处。假设高空平台会偏离其原始位置，偏移方差范围为 $[0.01,3]$ km。以 dB 为单位的平均空地自由空间路径损耗模型为 [40]

$$l_k = 128.1 + 20\log_{10}(d_k), \tag{38}$$

其中， l_k和 $d_k \in [25,55]$ km分别表示高空平台站与客户端 k之间通信链路的路径损耗和距离。此外，噪声功率谱密度为 $\mathcal{N}_0 = -174$dBm/Hz。我们还假设每个客户端从非独立同分布MNIST数据集 [41], 中以相同概率随机选取 $J_k = 500$ 个数据样本。客户端 k的本地计算能力参数C_k在 $[1,3] \times 10^4$cycles/sample范围内均匀分布。类似地，HAPS计算能力参数 L_H 设为 3×10^4。本地计算和HAPS计算中的有效开关电容分别为 $\kappa = 10^{-28}$ 和 $\zeta_H = 10^{-27}$。此外，我们将上行链路最大发射功率设为 $p_k^{\max} = 10$ dBm $\forall k$，而HAPS最大广播功率设为 50 dBm。最大计算容量为 $f^{\max} = 2$ GHz，本地更新大小为 $s = 28.1$ kbits，系统总带宽设为 $B = 20$ MHz。

为评估所提出的联邦学习算法，我们使用 PyTorch 实现，在非独立同分布MNIST数据集上进行实验。具体而言，MNIST数据集包含 60,000 张训练图像，分布在 1200 个数据分片中，每个分片包含 50 张图像。本地批次大小设为 10 张图像。每个客户端训练一个卷积神经网络（CNN），该网络由两个 5×5 卷积层组成（第一层有 32 个通道，第二层有 64 个通道，每层后接 2×2 最大池化）。采用的损失函数为交叉熵。为进行比较，我们采用以下三种基线算法:

1) 无选择的TERR-FL：该算法不采用客户端选择策略（即所有客户端均参与学习），且联邦学习服务器与地面云服务器共置。

2) 采用RAN-SEL的TERR-FL [42]: 在该算法中，客户端随机选取，每个选定客户端都会获得无线电资源的最优分配。“无线电资源”涵盖多种可用的无线通信资源，例如发射功率、带宽和计算能力，客户端利用这些资源来

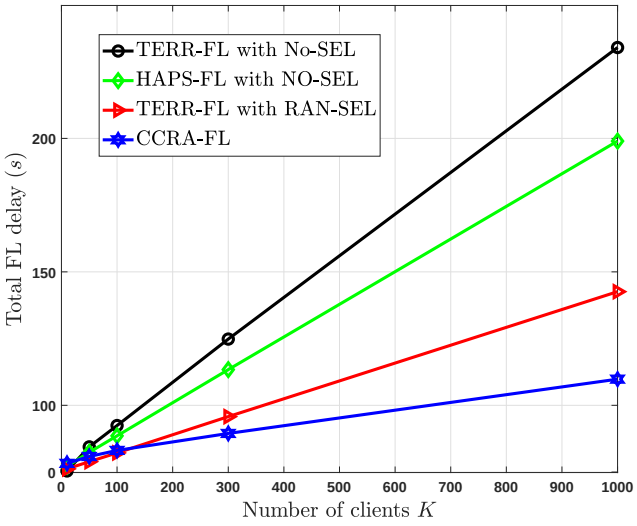


图3：在达到相同目标训练准确率的情况下，不同客户端数量对应的联邦学习总时延表现，并与基准方法进行比较。

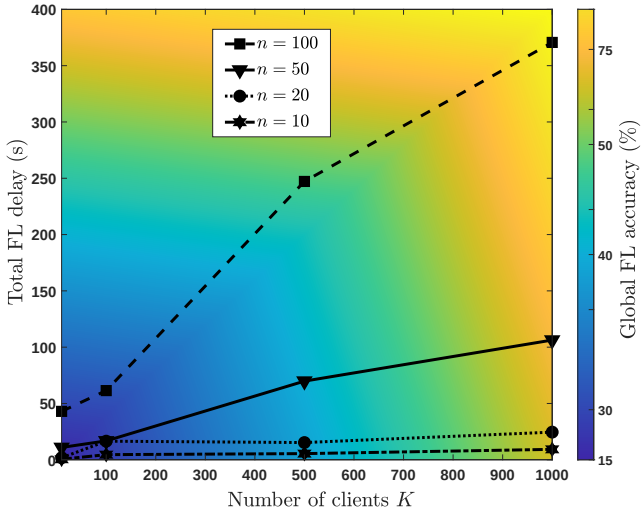


图4：不同通信轮次 n 下，不同客户端数量对应的联邦学习总时延和全局准确率表现。

向高空平台站传输本地模型更新。对无线电资源进行优化，使客户端能够更好地利用可用的无线资源，并提高联邦学习过程的效率。此外，值得注意的是，在该算法中，联邦学习服务器与地面云服务器部署在同一位置，这进一步增强了系统的处理能力和计算能力。

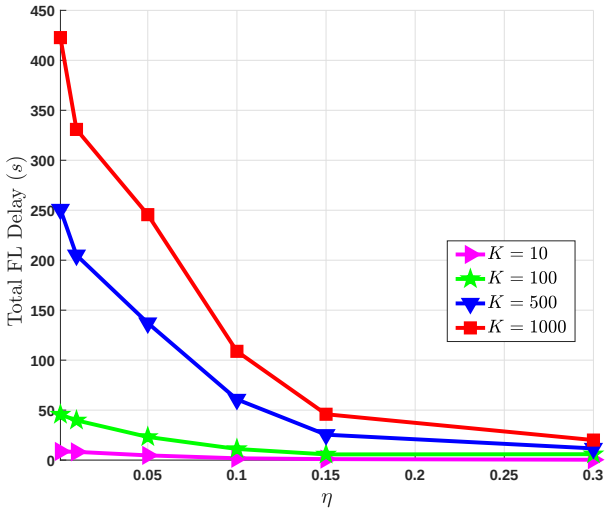


图5：不同客户端数量 K下，总联邦学习时延性能随局部目标精度 η 的变化。

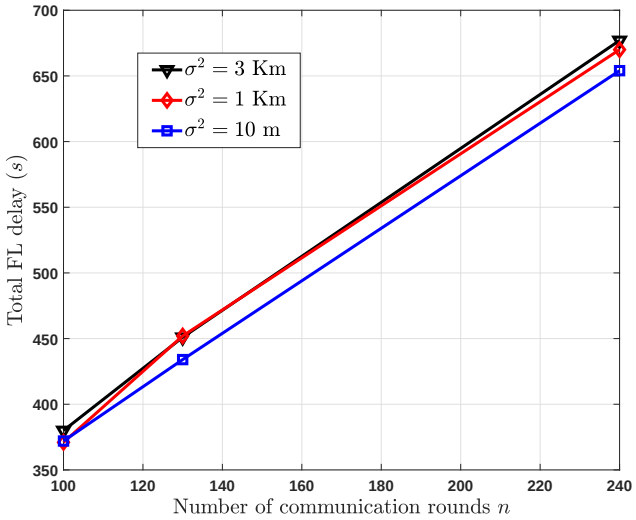


图6：考虑不同高空平台位移方差 σ^2 并满足 $K=1000$时，总联邦学习时延性能随通信轮数 n 的变化。

3) 采用无客户端选择（NO-SEL）的高空平台站联邦学习：该算法不采用客户端选择策略（即所有客户端都参与训练过程）。此外，联邦学习服务器与高空平台站共址。

对于地面联邦学习场景，我们考虑一个多跳场景，其中有 5 个宏基站，每个宏基站的覆盖半径最大为 10 km（城区），以提供与高空平台站相同的覆盖范围 [42]。这些宏基站将本地更新中继至云服务器上的服务器，以便进行聚合和模型

在联邦学习过程中生成。路径损耗指数设为 4。此外，在地面场景下，本地更新通过双跳通信到达宏基站上的服务器。

在图2中，我们通过改变选定客户端的数量 $|\mathcal{K}|$，评估所提出的CCRA-FL算法在全局准确率和训练损失方面的表现。本地迭代次数固定为 $i=5$。具体而言，图2a比较了在选定客户端数量的三种不同设置 $|\mathcal{K}| \in \{10,100,1000\}$ 下，联邦学习训练损失随通信轮数 n 变化的情况。可以观察到，随着参与联邦学习过程的选定客户端数量增加，CCRA-FL算法的准确率显著提升，最高可达15%。此外，图2b考察了联邦学习准确率随选定客户端数量 $|\mathcal{K}|$ 变化的情况。可以观察到，随着参与客户端数量的增加，全局训练准确率显著提升，最高可达300%。这些结果也与定理1的结论相符。

在图3中，我们评估并比较了所提出的CCRA-FL算法与上述基准方法的性能。可以观察到，在给定通信轮次下，要达到相同的全局联邦学习准确率，CCRA-FL算法产生的端到端联邦学习时延更低。这是通过求解时延最小化问题，对联邦学习参数（即 n、$|\mathcal{K}|$、i）进行次优选择的结果。所有算法的时延均呈上升趋势，主要有两个原因。首先，随着客户端参与度提高，高空平台站服务器需要处理更多数据，因此计算时延增加。其次，随着选定客户端数量增加，分配给客户端的带宽会减少。因此，上传速率降低，上传时间增加。然而，所提出的FLSTRA算法的时延增幅最小。具体而言，与采用RAN-SEL的TERR-FL算法相比，得益于视距链路以及客户端与高空平台站服务器之间的单跳通信，所提出的CCRA-FL算法的端到端联邦学习时延最多降低20%。

在图4中，对于固定的本地迭代次数 $i=5$，我们研究了端到端联邦学习时延对所选参数对（n，$|\mathcal{K}|$）的依赖关系。可以观察到，通信轮数 n 的增加会显著影响所有通信轮次中的端到端联邦学习时延，并且在联邦学习参数（即（n，$|\mathcal{K}|$，i））中起主导作用。此外可以推知，为达到期望的目标精度，存在一组次优参数，即通信轮数 n 、选定客户端数量 $|\mathcal{K}|$和本地迭代次数 i，使联邦学习时延最小。例如，在每个通信轮次中以最小联邦学习时延达到 $\epsilon=0.5$时，对于固定的 $i=5$，最佳的（n，$|\mathcal{K}|$）参数对为（10，500）。在图5中，我们评估

表II：FLSTRA 总延迟、训练损失和准确率表现 e.

选定客户端数量（$\|\mathcal{K}\|$）	通信轮次数量（n）	No. of 本地迭代次数（i）	联邦学习总时延（s）	全局训练 loss	全局准确率（$1-\epsilon$）$\times 100$
10	10	5	0.458	7.50%	15.33%
10	10	50	2.347	7.01%	21.13%
10	20	5	3.72	6.14%	32.27%
100	10	5	10.76	2.19%	29.84%
100	10	50	17.61	1.80%	35.75%
100	20	5	20.06	1.77%	50.03%
1000	10	5	23.51	1.61%	45.37%
1000	10	50	55.95	1.41%	51.22%
1000	20	5	61.43	1.38%	60.73%

所提出的CCRA-FL算法在 $K \in \{10,100,500,1000\}$ 和 $n = 20$条件下，随着本地精度 η 变化时的端到端联邦学习时延性能。结果表明，当 η大于0.15时，给定通信轮次中的端到端总时延小于50秒。在固定 $\eta = 0.1$（即本地精度更高）的情况下，将客户端数量从 $K = 500$ 增加到 $K = 1000$，可以观察到，总联邦学习时延最多增加38%，这是由于客户端上传时间和高空平台站计算时间增加所致。在图6中，我们评估了高空平台站发生非预期位移或漂移对联邦学习时延性能的影响。可以观察到，随着位移方差 σ^2 增大，端到端总时延最多增加4%。总体而言，由于位移，一些选定客户端在当前通信轮次中会遇到较差的信道质量，因此联邦学习过程需要更长时间才能在所有通信轮次中达到期望的目标精度。不过，所提出的CCRA-FL算法通过采用客户端选择策略（即算法1中的步骤2）缓解了信道质量较差的问题。

最后，表II揭示了所提出的联邦学习设置中一个重要且实用的权衡：服务器可以在每个通信轮次选择更多客户端，同时让每个客户端减少通信量，从而达到与选择较少客户端、但让每个客户端增加通信量相同的准确率。当可用客户端较多，但每个客户端的上行带宽非常有限时（这在实际中很常见），前一种方案可能更可取。

V. 结论

本文提出了FLSTRA这一新颖理念，利用高空平台站的独特特性实现大规模联邦学习。FLSTRA提高了收敛率和模型准确率，但也增加了时延。因此，我们开发了CCRA-FL算法，旨在实现时延-准确率权衡。具体而言，我们构建了一个时延最小化问题，对通信与计算资源分配（包括上行链路和下行链路资源）、客户端选择以及联邦学习参数进行联合优化。由于该问题是非凸的，我们将其分解为四个子问题，并采用迭代方式求解。对于其中一些子问题，我们还推导出了闭式解。仿真结果表明，在达到相同模型准确率的情况下，与地面基准相比，FLSTRA系统可显著降低联邦学习时延。此外，结果还表明，所提出的适用于FLSTRA的CCRA-FL算法能够容忍高空平台站非期望位移。进一步研究数据分布（即非独立同分布程度）对高空平台辅助联邦学习系统的影响，并采用CINIC-10和CIFAR-10等更复杂、规模更大的数据集以及更精细的信道模型，是一个值得探索的未来方向。此外，将多天线设备纳入考虑，并在问题表述中优化高空平台站服务器的三维部署和波束成形，将有助于提升联邦学习系统的性能并降低总端到端时延。

附录A

定理1的证明

利用二阶泰勒展开，我们首先将 $F(\mathbf{w}^{(n+1)})$ 重写为

$$F(\mathbf{w}^{(n+1)}) = F(\mathbf{w}^{(n)}) + (\mathbf{w}^{(n+1)} - \mathbf{w}^{(n)})\nabla F(\mathbf{w}^{(n)}) + \frac{1}{2}(\mathbf{w}^{(n+1)} - \mathbf{w}^{(n)})^T \nabla^2 F(\mathbf{w}^{(n)})(\mathbf{w}^{(n+1)} - \mathbf{w}^{(n)})$$
$$\leq F(\mathbf{w}^{(n)}) + (\mathbf{w}^{(n+1)} - \mathbf{w}^{(n)})^T \nabla F(\mathbf{w}^{(n)}) + \frac{M}{2}\|\mathbf{w}^{(n+1)} - \mathbf{w}^{(n)}\|^2, \tag{A.1}$$

其中，不等式源于假设1。根据联邦学习算法，在每一轮通信中，新的全局模型生成为 $\mathbf{w}^{(n+1)} = \mathbf{w}^{(n)} + \frac{1}{|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}} \mathbf{z}_k^{(n)}$。因此，我们有

$$F(\mathbf{w}^{(n+1)}) \leq F(\mathbf{w}^{(n)}) + \frac{1}{|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}} \nabla F(\mathbf{w}^{(n)})\mathbf{z}_k^{(n)} + \frac{M}{2|\mathcal{K}^{(n)}|^2}\left\|\sum_{k\in\mathcal{K}} \mathbf{z}_k^{(n)}\right\|^2, \tag{A.2}$$

根据式（3）中局部优化问题的目标函数定义，我们有

$$F(\mathbf{w}^{(n+1)})$$

$$\leq F(\mathbf{w}^{(n)})+\frac{1}{\xi|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\Big[G_k(\mathbf{w}^{(n)},\mathbf{z}_k^{(n)})F_k(\mathbf{w}^{(n)}+\mathbf{z}_k^{(n)})+\nabla F_k(\mathbf{w}^{(n)})\mathbf{z}_k^{(n)}\Big]+\frac{M}{2|\mathcal{K}^{(n)}|^2}\Big\|\sum_{k\in\mathcal{K}^{(n)}}\mathbf{z}_k^{(n)}\Big\|^2, \tag{A.3}$$

根据假设1，我们有：

$$F(\mathbf{w}^{(n+1)})\leq F(\mathbf{w}^{(n)})+\frac{1}{\xi|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\Big[G_k(\mathbf{w}^{(n)},\mathbf{z}_k^{(n)})-F_k(\mathbf{w}^{(n)})-\frac{u}{2}\|\mathbf{z}_k^{(n)}\|^2\Big]+\frac{M}{2|\mathcal{K}^{(n)}|^2}\Big\|\sum_{k\in\mathcal{K}^{(n)}}\mathbf{z}_k^{(n)}\Big\|^2. \tag{A.4}$$

根据三角不等式和均值不等式，我们有

$$\Big\|\frac{1}{|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\mathbf{z}_k^{(n)}\Big\|^2\leq\Big[\frac{1}{|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\|\mathbf{z}_k^{(n)}\|\Big]^2\leq\frac{1}{|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\|\mathbf{z}_k^{(n)}\|^2. \tag{A.5}$$

将 (A.5) 代入 (A.4)，我们有

$$F(\mathbf{w}^{(n+1)})\leq F(\mathbf{w}^{(n)})+\frac{1}{\xi|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\Big[G_k(\mathbf{w}^{(n)},\mathbf{z}_k^{(n)})-F_k(\mathbf{w}^{(n)})-\frac{(u-M\xi)}{2}\|\mathbf{z}_k^{(n)}\|^2\Big], \tag{A.6}$$

根据 (3)， $F(\mathbf{w}^{(n)})=G_k(\mathbf{w}^{(n)},0)$。因此，我们有$F(\mathbf{w}^{(n+1)})$

$$\leq F(\mathbf{w}^{(n)})+\frac{1}{\xi|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\Big[G_k(\mathbf{w}^{(n)},\mathbf{z}_k^{(n)})-G_k(\mathbf{w}^{(n)},\mathbf{z}_k^{(n)*})-(G_k(\mathbf{w}^{(n)},0)-G_k(\mathbf{w}^{(n)},\mathbf{z}_k^{(n)*}))-\frac{(u-M\xi)}{2}\|\mathbf{z}_k^{(n)}\|^2\Big]$$

$$\overset{(7)}{\leq}F(\mathbf{w}^{(n)})-\frac{1}{\xi|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\Big[(1-\eta)(G_k(\mathbf{w}^{(n)},0)-G_k(\mathbf{w}^{(n)},\mathbf{z}_k^{(n)*}))+\frac{(u-M\xi)}{2}\|\mathbf{z}_k^{(n)}\|^2\Big]$$

$$\overset{(3)}{=}F(\mathbf{w}^{(n)})-\frac{1}{\xi|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\Big[(1-\eta)\big(F_k(\mathbf{w}^{(n)})$$

$$-F_k(\mathbf{w}^{(n)}+\mathbf{z}_k^{(n)*})+(\nabla F_k(\mathbf{w}^{(n)})-\xi\nabla F(\mathbf{w}^{(n)}))^T\mathbf{z}_k^{(n)*}\big)+\frac{(u-M\xi)}{2}\|\mathbf{z}_k^{(n)}\|^2\Big]. \tag{A.7}$$

问题 (3) 的最优解 $\mathbf{z}_k^{(n)*}$ 始终满足一阶导数条件，即 $\nabla G_k(\mathbf{w}^{(n)},\mathbf{z}_k^{(n)*})=0$。因此，我们有

$$\nabla F_k(\mathbf{w}^{(n)}+\mathbf{z}_k^{(n)*})=\nabla F_k(\mathbf{w}^{(n)})-\xi\nabla F(\mathbf{w}^{(n)}). \tag{A.8}$$

将 (A.8) 代入 (A.7)，我们有

$$F(\mathbf{w}^{(n+1)})\leq F(\mathbf{w}^{(n)})+\frac{1}{\xi|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\Big[(1-\eta)\big(F_k(\mathbf{w}^{(n)})$$

$$-F_k(\mathbf{w}^{(n)}+\mathbf{z}_k^{(n)*})+\nabla F_k(\mathbf{w}^{(n)}+\mathbf{z}_k^{(n)*})^T\mathbf{z}_k^{(n)*}\big)+\frac{(u-M\xi)}{2}\|\mathbf{z}_k^{(n)}\|^2\Big]. \tag{A.9}$$

在假设1下，可以得到以下不等式 [10]:

$$F_k(\mathbf{w}^{(n)}) \geq F_k(\mathbf{w}^{(n)}+\mathbf{z}_k^{(n)*}) - \nabla F_k(\mathbf{w}^{(n)}+\mathbf{z}_k^{(n)*})^T \mathbf{z}_k^{(n)*} + \frac{u}{2}\|\mathbf{z}_k^{(n)*}\|^2, \tag{A.10}$$

and

$$\|\mathbf{z}_k^{(n)*}\|^2 \leq \frac{1}{M^2}\|F_k(\mathbf{w}^{(n)}+\mathbf{z}_k^{(n)*}) - \nabla F_k(\mathbf{w}^{(n)})\|^2. \tag{A.11}$$

将 (A.10) 和 (A.11) 应用于 (A.9)，可得

$$\begin{aligned}
&F(\mathbf{w}^{(n+1)}) \\
&\leq F(\mathbf{w}^{(n)}) - \frac{1}{\xi|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\Big[\frac{(1-\eta)u}{2}\|\mathbf{z}_k^{(n)*}\|^2 + \frac{u-\xi M}{2}\|\mathbf{z}_k^{(n)}\|^2\Big] \\
&\leq F(\mathbf{w}^{(n)}) - \frac{1}{\xi|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\Big[\frac{(1-\eta)u}{2M^2}\|\nabla F_k(\mathbf{w}^{(n)}+\mathbf{z}_k^{(n)*}) - \nabla F_k(\mathbf{w}^{(n)})\|^2 + \frac{(u-M\xi)}{2}\|z_k^{(n)}\|^2\Big] \\
&\overset{\text{(A.8)}}{=} F(\mathbf{w}^{(n)}) - \frac{1}{\xi|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\Big[\frac{(1-\eta)u\xi^2}{2M^2}\|\nabla F(\mathbf{w}^{(n)})\|^2 + \frac{(u-M\xi)}{2}\|\mathbf{z}_k^{(n)}\|^2\Big].
\end{aligned} \tag{A.12}$$

根据假设1，我们可以推导出以下不等式 [10]:

$$\|\nabla F(\mathbf{w}^{(n)})\|^2 \leq u\Big[F(\mathbf{w}^{(n)}) - F(\mathbf{w}^*)\Big]. \tag{A.13}$$

应用 (A.13)，我们有

$$F(\mathbf{w}^{(n+1)}) \leq F(\mathbf{w}^{(n)}) - \frac{(1-\eta)u^2\xi}{2M^2}\Big[F(\mathbf{w}^{(n)}) - F(\mathbf{w}^*)\Big] - \frac{(u-M\xi)}{2\xi|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\|\mathbf{z}_k^{(n)}\|^2. \tag{A.14}$$

据此，我们可以得到

$$\begin{aligned}
&F(\mathbf{w}^{(n+1)}) - F(\mathbf{w}^*) \leq (1-\frac{(1-\eta)u^2\xi}{2M^2})\Big[F(\mathbf{w}^{(n)}) - F(\mathbf{w}^*)\Big] - \frac{(u-M\xi)}{2\xi|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\|\mathbf{z}_k^{(n)}\|^2 \\
&\leq (1-\frac{(1-\eta)u^2\xi}{2M^2})^{n+1}\Big[F(\mathbf{w}^{(0)}) - F(\mathbf{w}^*)\Big] - \frac{(u-M\xi)}{2\xi}\Big[\frac{1}{|\mathcal{K}^{(0)}|}(1-\frac{(1-\eta)u^2\xi}{2M^2})^n\sum_{k\in\mathcal{K}^{(0)}}\|\mathbf{z}_k^{(0)}\|^2 + \ldots \\
&\quad + \frac{1}{|\mathcal{K}^{(n)}|}\sum_{k\in\mathcal{K}^{(n)}}\|\mathbf{z}_k^{(n)}\|^2\Big] \\
&= (1-\frac{(1-\eta)u^2\xi}{2M^2})^{n+1}\Big[F(\mathbf{w}^{(0)}) - F(\mathbf{w}^*)\Big] + \frac{(M\xi-u)}{2\xi}\Big[\sum_{n'=0}^{n}\frac{1}{|\mathcal{K}^{(n')}|}(1-\frac{(1-\eta)u^2\xi}{2M^2})^{n-n'}\sum_{k\in\mathcal{K}^{(n')}}\|\mathbf{z}_k^{(n')}\|^2\Big].
\end{aligned} \tag{A.15}$$

为了确保收敛上界始终为正，我们需要确保该上界左侧的两项均为正。众所周知，最优全局模型的损失函数值始终小于其他任何模型的损失函数值，这意味着$F(\mathbf{w}^{(0)}) - F(\mathbf{w}^*) \geq 0$。此外，根据公式（6），局部目标精度的界为 $0 \leq \eta \leq 1$，这意味着 $0 \leq 1-\eta \leq 1$。结合这两个事实，我们得到不等式 $0 \leq {}_{(1-\eta)}{}_{2M^2}^{u^2\zeta} \leq \frac{u^2\zeta}{2M^2}$。对于第二项，我们要求 $M\zeta - u \geq 0$。

因此，考虑这两个条件后，我们可以得到关于 u 和M的以下条件，其中二者均为正值:

$$u \leq M \min\left\{\zeta, \sqrt{\frac{2}{\zeta}}\right\}. \tag{A.16}$$

至此，证明完成。

APPENDIX B
PROOF OF THEOREM 2

为最小化时延 τ^{FL}，需要最小化本地上传时间 t_k^{up}。因此，为最小化式 (23c) 中的 t_k^{up} ，有

$$t_k^{up^*} = t_k^{\min}, \forall k. \tag{A.1}$$

给定该解，式 (23a) 关于本地精度 η 是一个凸优化问题。令 $\mathcal{L}_1(\eta,\boldsymbol{\lambda},\theta)$ 表示拉格朗日函数，其定义为

$$\mathcal{L}_1(\eta,\boldsymbol{\lambda},\theta) = \max_k a_k\left(t_k^{up} + \frac{vC_kJ_k\log_2(\frac{1}{\eta})}{f_k}\right) + \sum_{k=1}^{K}\lambda_k\left[a_k\left(t_k^{up}p_k^{up} + \zeta_k vC_kJ_k\log_2(\frac{1}{\eta})f_k^2\right) - E_k^{\max}\right] + \theta(\eta - 1), \tag{A.2}$$

其中， $\boldsymbol{\lambda} = [\lambda_k]_{1\times K}$ 和 θ 分别是与约束条件 (23c) 和 (23d) 相关的拉格朗日乘子。利用 KKT 条件，即 $\frac{\partial \mathcal{L}_1}{\partial \eta} = 0$，可得闭式解 (26)。

APPENDIX C
PROOF OF THEOREM 3

从时延最小化的角度来看，在满足本地能量约束的前提下，每个通信轮次的本地上传功率 p_k^{up} 需要最大化。根据式 (29c) 和 (29e)， p_k^{up} 优化为

$$p_k^{up^*} = \min\{p_k^{\max}, \frac{a_k\left[E_k^{\max} - \zeta_k vC_kJ_k\log_2(\frac{1}{\eta})f_k^2\right]}{t_k^{up}}\}, \forall, k. \tag{C.1}$$

给定 $\mathbf{p} = [p_k^{up^*}]_{1\times K}$，问题 (29a) 关于 b_k 是一个凸优化问题，可使用拉格朗日方法求解。令 $\mathcal{L}_2(\mathbf{b}\ \boldsymbol{\gamma},\psi)$ 表示拉格朗日函数，其定义为

$$\mathcal{L}_2(\mathbf{b},\boldsymbol{\gamma},\psi) = \tau^{FL} - \sum_{k=1}^{K}\gamma_k\left[a_kb_k\log_2(1 + \frac{h_kp_k^{up^*}}{b_kN_0}) - \frac{a_ks}{t_k^{up}}\right] + \psi(\sum_{k=1}^{K}a_kb_k - B), \tag{C.2}$$

其中， $\boldsymbol{\gamma} = [\gamma_k]_{1\times K}$和 ψ 分别是与约束条件 (29b) 和 (29d) 相关的拉格朗日乘子。根据 KKT 条件，我们接下来需要求解以下问题:

$$\frac{\partial \mathcal{L}_2}{\partial b_k} = \gamma_k\left[a_k\log_2(1 + \frac{\pi_k}{b_k}) - \frac{a_k\pi_k}{b_k(1 + \frac{\pi_k}{b_k})\ln 2}\right] - \psi_n a_k = 0, \forall k, \tag{C.3}$$

其中 $\pi_k = \frac{h_k p_k^{up^*}}{\mathcal{N}_0}$。式 (C.3) 难以用解析方法求解。因此，在点 $x_0 \geq \pi_k, \forall k \in \mathcal{K}$ 处进行一阶泰勒展开，可得

$$\beta_1 - \beta_2(b_k - x_0) + \frac{\pi_k}{b_k + \pi_k} - \frac{\psi \ln 2}{\gamma_k} = 0, \tag{C.4}$$

其中 $\beta_1 = \ln(1 + \frac{\pi_k}{x_0})$ 和 $\beta_2 = \frac{\pi_k}{x_0 (x_0 + \pi_k)}$。给定 $\psi = 0$，可通过 (31) 求得该二次方程的解。

APPENDIX D
PROOF OF THEOREM 4

由 (32a) 可知，问题关于 f_H 是凸的，因为其二阶导数为正。因此，我们可以应用拉格朗日方法来求最优解。令 $\mathcal{L}_3(f_H, \omega)$ 表示拉格朗日函数，其定义为

$$\mathcal{L}_3(f_H, \omega) = \left(\frac{L_H \sum_{k=1}^{K} a_k}{f_H} + t_H^{bc}\right) + \omega \left[\left(\zeta_H f_H^2 L_H \sum_{k=1}^{K} a_k + p_H^{bc} t_H^{bc}\right) - E_H^{\max} \right]. \tag{D.1}$$

然后，通过求解 $\frac{\partial \mathcal{L}_3}{\partial f_H} = 0$，可得最优 HAPS 计算能力 f_H^*，如 (34) 所示。给定 f_H^* 后，可求解 (32a)，以找到每个通信轮次的最优 HAPS 广播时间和功率。从最小时延的角度来看，为 HAPS 广播联邦学习全局更新分配最短时间始终是有效的。因此，为了最小化 (32a) 中的广播时间，我们有

$$t_H^{bc^*} = \frac{s}{\min_{k \in \mathcal{K}} B \log_2\left(1 + \frac{p_H^{bc} h_k}{B \mathcal{N}_0}\right)}. \tag{D.2}$$

可以验证，(32a) 关于 p_H^{bc} 是一个凸优化问题，因此可以使用拉格朗日乘子法求解。令 $\mathcal{L}_4(p_H^{bc}, \psi)$ 表示定义如下的拉格朗日函数：

$$\mathcal{L}_4(p_H^{bc}, \psi) = \frac{L_H \sum_{k=1}^{K} a_k}{f_H} + \frac{s}{B \log_2\left(1 + \frac{p_H^{bc} \min_{k \in \mathcal{K}} h_k}{B \mathcal{N}_0}\right)} + \psi \left[\left(\zeta_H f_H^2 L_H \sum_{k=1}^{K} a_k[n] + p_H^{bc} t_H^{bc}\right) - E_H^{\max} \right], \tag{D.3}$$

其中，ψ 是拉格朗日乘子。注意，由于对数函数关于 p_H^{bc} 单调递增，因此有 $\min_{k \in \mathcal{K}} B \log_2\left(1 + \frac{p_H^{bc} h_k}{B \mathcal{N}_0}\right) = B \log_2\left(1 + \frac{p_H^{bc} \text{最小值}_{k \in \mathcal{K}} h_k}{B \mathcal{N}_0}\right)$。然后，通过应用 KKT 条件，可以得到闭式解 (37)。

参考文献

[1] M. Alsabah 等，“6G 无线通信网络：全面综述”，IEEE Access，第9卷，第148191-148243页，2021年。[2] H. Yanikomeroglu，“面向泛在三维超连接的集成式地面/非地面 6G 网络”，载于第21届 ACM 国际无线移动系统建模、分析与仿真会议（MSWIM）论文集，加拿大魁北克省蒙特利尔，2018年，第3-4页。

[3] A. Farajzadeh、M. G. Khoshkholgh、H. Yanikomeroglu 和 O. Ercetin，“自演进集成垂直异构网络”，IEEE Open J. Commun. Soc.，第4卷，第552-580页，2023年。[4] M. Ericson 等，“6G架构趋势与赋能技术”，载于IEEE第4届5G世界论坛（5GWF）论文集，加拿大魁北克省蒙特利尔，2021年，第406-411页。[5] H. B. McMahan、E. Moore、D. Ramage、S. Hampson 和 B. A. Y. Arcas，“基于去中心化数据的深度网络通信高效学习”，载于第20届国际人工智能与统计学会议论文集，美国佛罗里达州劳德代尔，2017年，第1273-1282页。[6] Y. Chen、Y. Ning、M. Slawski 和 H. Rangwala，“面向非独立同分布数据边缘设备的异步在线联邦学习”，载于IEEE国际大数据会议（BigData）论文集，美国亚特兰大，2020年，第15-24页。[7] Q. Yang、Y. Liu、Y. Cheng、Y. Kang、T. Chen 和 H. Yu，“联邦学习”，《人工智能与机器学习综合讲义》，施普林格，第13卷，第3期，第1-207页，2019年。[8] P. Kairouz 等，“联邦学习的进展与开放问题”，《机器学习基础与趋势》，第14卷，第1–2期，第1-210页，2021年。[9] H. Zhu、Y. Zhou、H. Qian、Y. Shi、X. Chen 和 Y. Yang，“兼顾公平性的异步联邦学习在线客户端选择”，IEEE《无线通信汇刊》，第22卷，第4期，第2493-2506页，2023年4月。[10] Z. Yang、M. Chen、W. Saad、C. S. Hong 和 M. Shikh-Bahaei，“无线通信网络中的能量高效联邦学习”，IEEE《无线通信汇刊》，第20卷，第3期，第1935-1949页，2021年3月。[11] Y. Wu、Y. Song、T. Wang、M. Dai 和 T. Q. S. Quek，“基于非正交多址接入的无线信息与能量同传辅助联邦学习”，IEEE《绿色通信与网络汇刊》，第6卷，第3期，第1846-1861页，2022年9月。[12] S. Wang 等，“资源受限边缘计算系统中的自适应联邦学习”，IEEE《精选通信领域期刊》，第37卷，第6期，第1205-1221页，2019年6月。[13] S. Luo、X. Chen、Q. Wu、Z. Zhou 和 S. Yu，“HFEL：面向成本高效分层联邦边缘学习的联合边缘关联与资源分配”，IEEE《无线通信汇刊》，第19卷，第10期，第6535-6548页，2020年10月。[14] D. C. Nguyen、M. Ding、P. N. Pathirana、A. Seneviratne、J. Li 和 H. V. Poor，“物联网联邦学习：全面综述”，IEEE《通信调查与教程》，第23卷，第3期，第1622-1658页，2021年第三季度。[15] L. Qian 等，“无线通信分布式学习：方法、应用与挑战”，IEEE《信号处理精选主题期刊》，第16卷，第3期，第326-342页，2022年4月。[16] W. Gao、Z. Zhao、G. Min、Q. Ni 和 Y. Jiang，“工业物联网中时延感知联邦学习的资源分配”，IEEE《工业信息学汇刊》，第17卷，第12期，第8505-8513页，2021年12月。[17] A. Imteaj、U. Thakker、S. Wang、J. Li 和 M. H. Amini，“资源受限物联网设备联邦学习综述”，IEEE《物联网期刊》，第9卷，第1期，第1-24页，2022年1月。[18] A. Vanelli-Coralli、A. Guidotti、T. Foggi、G. Colavolpe 和 G. Montorsi，“5G及超越5G的非地面网络：趋势与研究挑战”，载于IEEE第3届5G世界论坛（5GWF）论文集，印度班加罗尔，2020年，第163-169页。[19] G. Kurt、M. G. Khoshkholgh、S. Alfattani、A. Ibrahim、T. Darwish、M. S. Alam、H. Yanikomeroglu 和 A. Yongacoglu，“未来高空平台站（HAPS）网络的愿景与框架”，IEEE《通信调查与教程》，第23卷，第2期，第729-779页，2021年第二季度。[20] B. E. Y. Belmekki 和 M. S. Alouini，“释放联网系留飞行平台的潜力：前景、挑战与应用”，IEEE Open J. Veh. Technol.，第3卷，第278-320页，2022年。[21] Y. Xing、F. Hsieh、A. Ghosh 和 T. S. Rappaport，“高空平台站（HAPS）：架构与系统性能”，载于IEEE第93届车辆技术会议（VTC2021-Spring）论文集，芬兰赫尔辛基，2021年，第1-6页。[22] M. Elmahallawy 和 T. Luo，“FedHAP：利用协作式高空平台站面向低地球轨道星座的快速联邦学习”，载于IEEE第14届无线通信与信号处理国际会议论文集，中国南京，2022年，第1-6页。

[23] T. Zhang, L. Gao, C. He, M. Zhang, B. Krishnamachari, and A. S. Avestimehr，《物联网联邦学习：应用、挑战与机遇》，IEEE《物联网杂志》，第5卷，第1期，第24–29页，2022年3月。[24] N. Rieke 等，《联邦学习与数字健康的未来》，NPJ《数字医学》，2020年。[25] X. Chen, G. Zhu, Y. Deng, and Y. Fang，《采用网内聚合的多跳无线网络联邦学习》，IEEE《无线通信汇刊》，第21卷，第6期，第4622–4634页，2022年6月。[26] F. Sattler, S. Wiedemann, K. R. M¨uller, and W. Samek，《基于非独立同分布数据的鲁棒且通信高效的联邦学习》，IEEE《网络与学习系统汇刊》，第31卷，第9期，第3400–3413页，2020年9月。[27] V. D. Nguyen, S. K. Sharma, T. X. Vu, S. Chatzinotas, and B. Ottersten，《用于无线物联网网络资源分配的高效联邦学习算法》，IEEE《物联网杂志》，第8卷，第5期，第3394–3409页，2021年3月。[28] Y. Mao, J. Zhang, and K. B. Letaief，《面向能量采集设备的移动边缘计算动态计算卸载》，IEEE《精选通信领域期刊》，第34卷，第12期，第3590–3605页，2016年12月。[29] Y. Wei, F. R. Yu, M. Song, and Z. Han，《采用自然演员–评论家深度强化学习优化雾计算赋能物联网中的缓存、计算和无线电资源》，IEEE《物联网杂志》，第6卷，第2061–2073页，2019年4月。[30] Q. Ren, O. Abbasi, G. K. Kurt, H. Yanikomeroglu, and J. Chen，《高空平台站（HAPS）辅助智能交通系统中的缓存与计算卸载》，IEEE《无线通信汇刊》，第21卷，第11期，第9010–9024页，2022年11月。[31] A. Mohammed, A. Mehmood, F. Pavlidou, and M. Mohorcic，《高空平台（HAP）在全球无线连接中的作用》，IEEE会刊，第99卷，第11期，第1939–1953页，2011年11月。[32] F. Dovis, R. Fantini, M. Mondin, and P. Savi，《高空平台（HAP）传播通道的小尺度衰落》，IEEE《精选通信领域期刊》，第20卷，第3期，第641–647页，2002年4月。[33] X. Cao, P. Yang, M. Alzenad, X. Xi, D. Wu, and H. Yanikomeroglu，《机载通信网络：综述》，IEEE《精选通信领域期刊》，第36卷，第9期，第1907–1926页，2018年9月。[34] Xi. Zhou, C. Liu, J. Zhao，《面向元宇宙移动增强现实的联邦学习资源分配》，arXiv预印本 arXiv:2211.08705，2022年。[35] Q. Ren, O. Abbasi, G. K. Kurt, H. Yanikomeroglu, and J. Chen，《高空平台站（HAPS）辅助智能交通系统计算》，载于 IEEE 全球通信大会（GLOBECOM）会议论文集，西班牙马德里，2021年，第1–6页。[36] S. Liu, H. Dahrouj, and M. S. Alouini，《卫星–HAPS–地面一体化网络中的联合用户关联与波束成形》，arXiv预印本 arXiv:2204.13257，2022年。[37] M. Hong, M. Razaviyayn, Z. Q. Luo, and J. S. Pang，《涉及大数据的块结构优化统一算法框架：在机器学习和信号处理中的应用》，IEEE《信号处理杂志》，第33卷，第1期，第57–77页，2016年1月。[38] A. Khalili, S. Akhlaghi, H. Tabassum, and D. W. K. Ng，《异构网络上行链路中的联合用户关联与资源分配》，IEEE《无线通信快报》，第9卷，第6期，第804–808页，2020年6月。[39] S. Boyd and L. Vandenberghe，《凸优化》，英国剑桥：剑桥大学出版社，2004年。[40] W. Khawaja, I. Guvenc, D. W. Matolak, U. C. Fiebig,and N. Schneckenburger，《无人飞行器空对地传播通道建模综述》，IEEE《通信调查与教程》，第21卷，第3期，第2361–2391页，2019年第三季度。[41]Y. LeCun，《手写数字MNIST数据库》，访问日期：1月 2022 [在线]， 网址：http://yann.lecun.com/exdb/mnist/[42] X. Chen, G. Zhu, Y. Deng, and Y. Fang，《采用网内聚合的多跳无线网络联邦学习》，IEEE《无线通信汇刊》，第21卷，第6期，第4622–4634页，2022年6月。