PARAAEGIS: 并行保护用于灵活的隐私保护联邦学习

Zihou Wu*, Yuecheng Li*, Tianchi Liao†, Jian Lou†, Chuan Chen*,*

*School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China
†School of Software Engineering, Sun Yat-sen University, Zhuhai, China

ABSTRACT

联邦学习（FL）面临一个关键的两难境地：现有的保护机制如差分隐私（DP）和同态加密（HE）强制执行严格的权衡，迫使在模型效用和计算效率之间做出选择。这种缺乏灵活性阻碍了实际实施。为了解决这个问题，我们引入了 ParaAegis，这是一个并行保护框架，旨在让从业者灵活控制隐私-效用-效率的平衡。我们的核心创新是一个战略性的模型分区方案。通过将轻量级 DP 应用于不那么关键、低范数的部分模型，同时使用 HE 保护其余部分，我们创建了一个可调系统。一个分布式投票机制确保了对这种分区的一致性。理论分析证实了在相同隐私水平下的效率和效用之间的调整。至关重要的是，实验结果表明，通过调整超参数，我们的方法能够灵活地优先考虑模型准确性或训练时间。

Index Terms— 联邦学习，差分隐私，同态加密

1. 介绍

联邦学习（FL）是一种分布式机器学习范式 [1]，因其能够促进数据孤岛之间的有限数据流动而引起了越来越多的关注。然而，尽管 FL 具有明显的优势，但也面临着潜在的挑战，其中隐私和安全方面的挑战是目前一个重要的研究方向。由于这个过程涉及在复杂的网络环境中传输信息，恶意设备更容易窃听链路并拦截传输的模型。虽然本地数据并未直接包含在模型参数中，但有研究表明可以从传输的模型 [2,3] 中重构出本地训练数据。

差分隐私（DP）和同态加密（HE）是最广泛用于确保隐私和安全的技术。DP 提供了一个正式的数学框架来衡量系统的隐私性。在 FL 的背景下，它通常涉及梯度裁剪和噪声添加等技术以满足 DP 要求。另一方面，HE 允许对加密数据执行加法或乘法等计算，而不会泄露底层信息。DP 倾向于降低模型准确性，随着隐私保护级别的增加，影响也越来越大 [4]。HE 在保持数据安全的同时，显著增加了训练时间和通信开销 [5]。因此，在 FL 中平衡隐私、实用性和效率仍然是一个重大挑战，当前的研究正在探索结合不同保护方法的混合方法以解决这些权衡问题。

一些研究探讨了平衡隐私、准确性和效率的潜力 [6–8] 。然而，这些工作并没有进一步探究隐私、效用和效率之间隐含权衡关系的优化问题。为了解决这个问题，我们提出了 ParaAegis，一个 DP-HE 并行利用的 FL 框架，在该框架中模型参数被分为两部分，并分别通过 DP 和 HE 进行保护。为了处理客户端间模型划分一致性的难题，我们提出了一种投票机制。在这一机制下，客户端将本地分区上传至服务器，然后服务器统计每个索引出现的次数，并选择最频繁出现的部分作为全局分区。我们的理论分析和实证结果表明，涉及 DP 和 HE 并行保护的模型分区提供了隐私-效用-效率权衡的重要灵活性。此外，在固定分区比例下，基于最大范数选择 HE 部分的分区策略进一步提高了模型准确性，同时没有损害隐私和效率。代码可在 https://anonymous.4open.science/r/ParaAegis/ 获得。

2. 初步的

在 FL 的背景下，我们考虑 n 个客户端，每个客户端都有一个不同的本地数据集。第 i 个客户端的本

地目标函数表示为 $f_i := \sum_{(\boldsymbol{x},y)\in\mathcal{D}_i} \ell_i(\boldsymbol{w};\boldsymbol{x},y)$，其中 $\boldsymbol{w}$ 和 $(\boldsymbol{x},y)$ 分别是参数和样本。全局目标是各本地目标加权平均的结果，权重由每个客户端的数据样本数量决定。在学习过程中，客户端和服务端交替交换本地更新 $\boldsymbol{u}_i$ 和全局更新 $\boldsymbol{u}$。

为了保护传输的更新免受反转攻击，一些方法建议应用 θ-范数裁剪和高斯噪声注入以提供客户端级差分隐私的理论保证（请参阅 [4] 的形式定义）：

$$\hat{\boldsymbol{u}}_i := \frac{\boldsymbol{u}_i}{\max\{1, \|\boldsymbol{u}_i\|^2\,\theta\}} + \boldsymbol{z}_i \sim \mathcal{N}(0, \sigma_z \boldsymbol{I}). \quad (1)$$

动态规划对效率的影响可以忽略不计。然而，其对效用的负面影响主要来自于剪枝和噪声添加，一些研究表明这些操作的幅度与模型准确性之间存在负相关关系。

前述方法通常会牺牲模型准确性。另一种优先考虑实用性而牺牲效率的方法是全同态加密 HE [9]。使用这种密码学方法，客户端通过公钥将本地更新加密为密文 $[\![\boldsymbol{u}_i]\!]$。然后服务器利用同态加法原语在加密域中聚合这些密文，从而生成一个加密的全局更新 $[\![\boldsymbol{u}]\!]$。最后，客户端可以使用私钥解密此结果。在整个过程中，HE 的密码学特性保证所有信息保持计算安全。HE 对实用性的影响可忽略不计，而其效率开销通常与明文字长呈线性增长。在各种 HE 方案中，CKKS 密码系统 [10] 特别适用于联邦学习，因为它支持实值向量的明文空间，并具有内在的批量处理能力。因此，本文中的 HE 组件基于 CKKS 方案。

Algorithm 1 提议的 ParaAegis

初始化全局模型 $\boldsymbol{w}_0$ 并将其广播给客户端
for $t = 1$ to T do
  随机抽样客户端 $\mathcal{C}_t \subseteq [M]$
  for all $i \in \mathcal{C}_t$ in parallel do
    在本地训练模型 K 个周期；
    计算局部更新 $\boldsymbol{u}_i^t := \boldsymbol{w}_i^{t,K} - \boldsymbol{w}_i^t$；
    根据最高 $r\%$ 范数选择 $\boldsymbol{u}_i^t$ 的索引以构建分区向量 $\boldsymbol{v}_i^t$；
    将 $\boldsymbol{v}_i^t$ 上载到服务器；
  end for
  聚合分区 $\boldsymbol{v}^t := \mathrm{Voting}(\boldsymbol{v}_i^t)$；
  发送 $\boldsymbol{v}^t$ 给客户端；
  for all $i \in \mathcal{C}_t$ in parallel do
    根据 $\boldsymbol{v}^t$ 将 $\boldsymbol{u}_i^t$ 分为 $\boldsymbol{u}_{i,\mathrm{DP}}^t$ 和 $\boldsymbol{u}_{i,\mathrm{HE}}^t$；
    通过等式 1 扰动 $\boldsymbol{u}_{i,\mathrm{DP}}^t$ 到 $\hat{\boldsymbol{u}}_{i,\mathrm{DP}}^t$；
    使用公钥加密 $\boldsymbol{u}_{i,\mathrm{HE}}^t$ 到 $[\![\boldsymbol{u}_{i,\mathrm{HE}}^t]\!]$；
    将 $\hat{\boldsymbol{u}}_{i,\mathrm{DP}}^t$ 和 $[\![\boldsymbol{u}_{i,\mathrm{HE}}^t]\!]$ 上载到服务器；
  end for
  分别将更新聚合为 $\hat{\boldsymbol{u}}_{\mathrm{DP}}^t$ 和 $[\![\boldsymbol{u}_{\mathrm{HE}}^t]\!]$，然后发送给所有客户端；
  for all $i = 0$ to $M-1$ in parallel do
    使用密钥解密 $[\![\boldsymbol{u}_{\mathrm{HE}}^t]\!]$ 到 $\boldsymbol{u}_{\mathrm{HE}}^t$；
    根据 $\boldsymbol{v}^t$ 整理 $\hat{\boldsymbol{u}}_{\mathrm{DP}}^t$ 和 $\boldsymbol{u}_{\mathrm{HE}}^t$ 成为 $\boldsymbol{u}^t$；
    $\boldsymbol{w}^{t+1} := \boldsymbol{w}^t + \boldsymbol{u}^t$ ；
  end for
end for

3. 方法论

在本节中，我们详细阐述了所提出的 DP-HE 并行保护方法 ParaAegis。我们的方法详见算法 1，并受到以下观察的启发：深度学习模型中的所有参数对学习过程的贡献并不相同。具有较大范数的更新倾向于更有力地指导模型的收敛。然而，正如方程 1 所示，这些大范数的更新也最容易受到 DP 固有的裁剪和噪声注入的影响，这可能会损害模型的效用。

我们框架的核心理念是将本地更新向量 $\boldsymbol{u}_i^t$ 分成两个不相交的子集。一个小型高范数子集，记为 $\boldsymbol{u}_{i,\mathrm{HE}}^t$，通过同态加密（HE）进行保护，以保持其精度。剩余的大多数参数 $\boldsymbol{u}_{i,\mathrm{DP}}^t$ 则通过差分隐私（DP）机制进行保护。这种划分使得采用 DP 保护的组可以使用一个较小的裁剪阈值，从而减少噪声并保持效用。同时，将同态加密的应用限制在一个小子集上能够最小化其计算和通信开销。

CKKS 方案的基于向量的性质需要所有客户端有一个共同的划分，这要求在划分索引上达成共识。我们通过受 FL 模型聚合启发的服务器聚合投票机制实现这一共识（图 1）。在这个过程中，每个客户端对其最高范数参数的索引进行“投票”作为 $\boldsymbol{v}_i$。然而，以明文形式传输这些索引会构成重大隐私风险，因为它

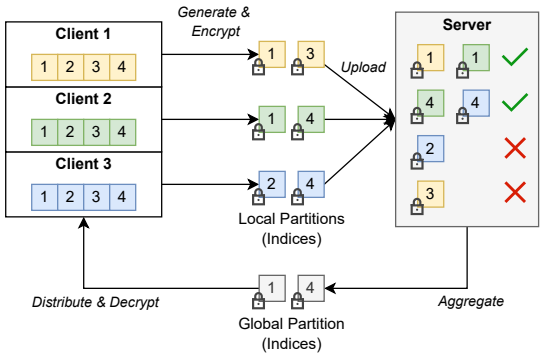

Fig. 1. 投票机制的说明。每个客户端为 HE 保护提议索引。服务器聚合这些提议并选择被提议次数最多的前 k 个索引（例如，索引 1 和 4）来形成所有客户端的全局分区。

会泄露有关本地更新的信息。为减轻这一问题，索引被加密。由于服务器的任务仅限于计票，标准非同态加密就足够了。服务器接收所有加密的局部划分，通过选择出现频率最高的索引来确定全局划分（可识别为唯一的密文），然后将该划分分发给所有客户端以应用于各自的更新。

ParaAegis 框架中固有的权衡可以通过分区比例进行调整。为简化起见，我们将 r 定义为分配给 HE 保护分区的总参数的比例，剩余的 $(1-r)$ 部分则由 DP 进行保护。该比率充当关键杠杆：当 r 减小时，保护更多依赖于 DP，这在可能牺牲效用的情况下提高了效率；而较大的 r 则将平衡转向 HE，以计算开销为代价提高精度。

关键的是，在 ParaAegis 中，这一划分比例无需固定，允许采取一种更灵活的方法以适应训练动态。这是基于观察到在训练的早期阶段，反向传播梯度通常具有较大的范数，并建立了收敛的一般方向；通过噪声干扰它们可能会减慢进程。相反，在后期阶段的梯度倾向于具有较小的范数，并且可以从随机性的正则化效果中获益，使它们更能容忍噪声。受此启发，我们为划分比例设计了一个动态衰减机制，其中加密保护参数的数量随着训练进展逐渐减少。我们设定了一个初始 HE 比率 r_0 和一个衰减率 $\lambda \in (0,1)$，并在每一轮之后，根据 $r_t := \lambda r_{t-1}$ 更新受 HE 保护的参数的比例。这一机制允许我们在早期训练阶段通过最小化收敛路径中的偏差来专注于准确性，然后将重点转向后期阶段的效率，通过减少 HE 的计算成本来缩短训练时间同时保持模型效用。我们将在随后的实验分析中验证此设计的理由。

4. 分析

我们建立了 ParaAegis 的隐私和收敛性保证。由于篇幅限制，我们仅展示主要定理。完整的详细证明见附录 A。

假设 1. *我们假设 FL 收敛分析的标准条件 [11]：一个平滑的损失函数 f，其梯度范数和方差在客户端和数据样本之间是有界的。*

定理 1 (ParaAegis 的收敛性). *在假设 1 成立的情况下，且如果注入的噪声满足 (ε,δ)-DP，则 ParaAegis 的收敛性由以下公式限制*

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla f(\boldsymbol{w}^t)\|^2] \leq O\left(\frac{1}{T}\right) + \underbrace{C_1\cdot(1-r)}_{\text{Clipping Effect}} + \underbrace{\frac{C_2\cdot(1-r)\ln(1/\delta)}{N^2\epsilon^2}}_{\text{Privacy Noise}}, \tag{2}$$

其中 C_1 和 C_2 是依赖于问题参数（如梯度边界和 Lipschitz 常数）的常数，而 N 是客户端的数量。

通过调整分区比率 r 和隐私预算 (ε,δ)，可以清楚地看到，随着隐私和效率的直接变化，由定理 1 所示的收敛界表示的效用相应受到影响。

5. 实验

实验设置。我们在 Imagenette [17] 数据集和 ResNet-18 模型 [18] 上进行实验，其参数规模足以突出我们的方法在平衡隐私、效率和效用方面的优势。对于联邦学习设置，客户端的总数是 $N=10$。训练总共进行了 $T=50$ 全局轮次，每个客户端每轮执行 $K=3$ 局部周期。关于训练超参数，批量大小 $|\mathcal{B}|$ 设为 32，学习率 η 为 0.01，梯度裁剪阈值 θ 为 1。

基准。为了评估 Paraaegis 在隐私-效用-效率权衡方面的灵活性，我们将其与 CKKS-FedAvg [13] 和 DP-FedAvg [12] 进行基准测试。我们还将差分隐私和同态

类别	算法	准确率（%）			时间（秒）			效率比		
		IID	Dir(1)	Dir(0.5)	IID	Dir(1)	Dir(0.5)	IID	Dir(1)	Dir(0.5)
No Privacy	FedAvg [1]	80.93	79.38	77.98	3571	3665	3699	2.27	2.17	2.11
基线	DP-FedAvg [12]	20.28	17.12	12.87	3007	3084	3112	0.67	0.56	0.41
	CKKS-FedAvg [13]	81.14	79.28	77.63	18527	14971	14568	0.44	0.53	0.53
	Serial [14]	10.54	11.29	12.35	19956	16718	16370	0.05	0.07	0.08
	Varying DP [15]	24.21	21.76	19.85	2494	2534	2789	0.97	0.86	0.71
	Stevens et al. [16]	18.78	17.07	13.75	3635	3708	3543	0.52	0.46	0.39
ParaAegis-静态	1%	67.85	63.48	57.16	6149	6006	5577	1.1	1.06	1.02
	5%	73.03	70.07	66.19	6888	6962	6099	1.06	1.01	1.09
	10%	75.63	73.63	71.1	7776	7706	6379	0.97	0.96	1.11
	20%	77.53	75.37	73.39	9530	9565	7153	0.81	0.79	1.03
动态 PARAEGIS	$5\% \times 0.99$	72.09	68.91	64.4	5561	5536	5299	1.3	1.24	1.22
	$10\% \times 0.99$	74.5	71.75	68.5	6238	6210	5679	1.19	<u>1.16</u>	<u>1.21</u>
	$10\% \times 0.95$	68.82	65.13	58.97	5619	5628	5400	<u>1.22</u>	<u>1.16</u>	1.09
	$10\% \times 0.9$	70.87	67.01	60.33	8762	8732	7073	0.81	0.77	0.85

Table 1. 不同算法在 ResNet-18 上的性能对比。效率比计算公式为（准确率 / 时间）×100。对于 ParaAegis-Static，参数是固定比例 r。对于 ParaAegis-Dynamic，它是 $r_0 \times \lambda$。

加密的串行组合 [14] 作为有竞争力的替代方案进行考虑。此外，还包括两种在保证隐私的情况下权衡效用-效率的方法，其中前者利用幅度变化的差分隐私 [15]，其缩放因子设置为 0.9，后者采用差分隐私、秘密共享和带错误学习方案 [16]。为了监控具有相对严格隐私约束的环境，差分隐私的隐私预算设置为 $(\varepsilon, \delta) = (1, 10^{-5})$。表 1 记录了在不同参数设置下 Paraaegis 与基准方法之间的分类准确性和运行时间比较。为了更直观地展示性能和效率之间的权衡，我们定义准确率与训练时间的比例为效率比率，这反映了这种权衡的质量。以下是表 1 中观察到的结果：(1) 随着 HE 比例 r 的增加，模型准确性相应提高，但时间开销也增加了。(2) DP-FedAvg 和 CKKS-FedAvg 可以被视为 ParaAegis 的极端情况。在它们之间的效用与效率权衡中，每种方法都牺牲了一个方面以在另一个方面获得优势。这也表明了 ParaAegis 通过并行结合差分隐私和同态加密为模型训练提供了更多调整效率-效用权衡的灵活性。(3) 两种保护方法的串行组合导致准确性和效率同时下降，使其不适合需要平衡效率和效用的情景。(4) 额外的两种方法（变化 DP 和 Steven 等）由于完全依赖差分隐私，表现相对较差，而 ParaAegis 可以避免这一缺点。(5) 在动态变体中，选择适当的参数 r 和 λ，可以在更短的时间内实现更好的准确性，与静态版本相比，实现了改进的隐私-效率-效用权衡。

Strategy	Max (Proposed)	Min	Rand
Accuracy (%)	63.64	12.40	15.78

Table 2. 不同分区策略的准确性比较

θ	0.01	0.1	1	10
Acc.	68.99	69.11	71.88	65.16
N	5	10	25	50
Acc.	72.49	65.16	56.40	48.14

Table 3. 不同裁剪阈值或客户端数量的精度比较

分区策略的消融研究。在本实验中，我们固定了分区比例，并比较不同分区策略以展示我们提出的分区策略的优势。实验中涉及的分区策略如下。最大值，作为 ParaAegis 采用的策略，选择每个客户端局部更新中范数最高的参数集作为 HE 部分；最小值，相比之下最大值，选择每个客户端局部更新中范数最低的参数集作为 HE 部分；随机，客户端随机选择一定数量的参数作为 DP 部分。为了便于观察差异，我们将分区比例固定在 $r = 0.1$，并将全局轮次的数量设置为

$T = 20$。从表 2 中可以看出，最小值和随机策略表现出相似的不良且不可接受的收敛行为，而只有最大值达到了良好的精度。这一现象表明，最大值策略有效地选择了对 HE 保护最重要的参数，使它们免受噪声引入的扰动。

超参数敏感性。我们评估了 ParaAegis 对两个关键超参数的敏感性：差分隐私剪切阈值 θ 和客户端数量 N。首先，通过测试 θ 确定了最优剪切阈值。表 3 的结果表明 $\theta = 1$ 能够达到最佳准确性。这个最优值显著小于典型的 DP-FL 设置中的数值 [19]，这证明了我们的投票机制有效地减少了受差分隐私噪声影响的更新组件的范数。其次，我们通过改变 N 来考察客户端规模的影响。观察到随着 N 增大（表 3）准确率下降，表明达成共识随着客户端多样性的增加变得更加困难。因此，ParaAegis 最适合用于跨数据孤岛的联邦学习环境，这一发现与其他基于同态加密的联邦学习框架 [20] 一致。

6. 结论

在本文中，我们提出了一种新颖的并行混合保护方法用于 FL。我们的方法将模型参数分为两部分，这两部分分别由 DP 和 HE 进行保护，并通过调整划分比例 r 和衰减率 λ（在动态变体中）来有效地灵活权衡隐私、效用和效率。

7. REFERENCES

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Artificial intelligence and statistics. PMLR, 2017, pp. 1273–1282.

[2] J. Geiping, H. Bauermeister, H. Dröge, and M. Moeller, "Inverting gradients-how easy is it to break privacy in federated learning?," Advances in neural information processing systems, vol. 33, pp. 16937–16947, 2020.

[3] A. Athalye, N. Carlini, and D. Wagner, "Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples," in International conference on machine learning. PMLR, 2018, pp. 274–283.

[4] H. B. McMahan, D. Ramage, K. Talwar, and L. Zhang, "Learning differentially private recurrent language models," in International Conference on Learning Representations, 2018.

[5] J. Ma, S. Naas, S. Sigg, and X. Lyu, "Privacy-preserving federated learning based on multi-key homomorphic encryption," International Journal of Intelligent Systems, vol. 37, no. 9, pp. 5880–5901, 2022.

[6] A. G. Sébert, M. Checri, O. Stan, R. Sirdey, and C. Gouy-Pailler, "Combining homomorphic encryption and differential privacy in federated learning," in 2023 20th Annual International Conference on Privacy, Security and Trust (PST), 2023, pp. 1–7.

[7] C. Hu and B. Li, "Maskcrypt: Federated learning with selective homomorphic encryption," IEEE Transactions on Dependable and Secure Computing, vol. 22, no. 1, pp. 221–233, 2024.

[8] Yuecheng Li, Lele Fu, Tong Wang, Jian Lou, Bin Chen, Lei Yang, Jian Shen, Zibin Zheng, and Chuan Chen, "Clients collaborate: Flexible differentially private federated learning with guaranteed improvement of utility-privacy trade-off," in Forty-second International Conference on Machine Learning, 2025.

[9] R. L. Rivest, "Cryptography," in Algorithms and complexity, pp. 717–755. Elsevier, 1990.

[10] Jung Hee Cheon, Andrey Kim, Miran Kim, and Yongsoo Song, "Homomorphic encryption for arithmetic of approximate numbers," in International conference on the theory and application of cryptology and information security. Springer, 2017, pp. 409–437.

[11] X. Zhang, X. Chen, M. Hong, Z. S. Wu, and J. Yi, "Understanding clipping for federated learning: Convergence and client-level differential privacy," in International Conference on Machine Learning, ICML 2022, 2022.

[12] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. Quek, and H. V. Poor, "Federated learning with differential privacy: Algorithms and performance analysis," IEEE transactions on information forensics and security, vol. 15, pp. 3454–3469, 2020.

[13] F. Qiu, H. Yang, L. Zhou, C. Ma, and L. Fang, "Privacy preserving federated learning using ckks homomorphic encryption," in International Conference on Wireless Algorithms, Systems, and Applications. Springer, 2022, pp. 427–440.

[14] X. Zhang, D. Huang, and Y. Tang, "Secure federated learning scheme based on differential privacy and homomorphic encryption," in International Conference on Intelligent Computing. Springer, 2024, pp. 435–446.

[15] X. Yuan, W. Ni, M. Ding, K. Wei, J. Li, and H. V. Poor, "Amplitude-varying perturbation for balancing privacy and utility in federated learning," IEEE Transactions on Information Forensics and Security, vol. 18, pp. 1884–1897, 2023.

[16] T. Stevens, C. Skalka, C. Vincent, J. Ring, S. Clark, and J. Near, "Efficient differentially private secure aggregation for federated learning via hardness of learning with errors," in 31st USENIX Security Symposium (USENIX Security 22), 2022, pp. 1379–1395.

[17] J. Howard, "Imagenette: A smaller subset of 10 easily classified classes from imagenet," https://github.com/fastai/imagenette, 2019, Accessed: 2024-09-04.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[19] N. Ponomareva, S. Vassilvitskii, Z. Xu, B. McMahan, A. Kurakin, and C. Zhang, "How to dp-fy ml: A practical tutorial to machine learning with differential privacy," in Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023, pp. 5823–5824.

[20] A. Blanco-Justicia, J. Domingo-Ferrer, S. Martínez, D. Sánchez, A. Flanagan, and K. E. Tan, "Achieving security and privacy in federated learning systems: Survey, research challenges and future directions," Engineering Applications of Artificial Intelligence, vol. 106, pp. 104468, 2021.

[21] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in Proceedings of the 2016 ACM SIGSAC conference on

computer and communications security, 2016, pp. 308–318.

[22] C. Dwork, A. Roth, et al., "The algorithmic foundations of differential privacy," Foundations and Trends® in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211–407, 2014.

[23] S. Reddi, Z. Charles, M. Zaheer, Z. Garrett, K. Rush, J. Konečný, S. Kumar, and H. B. McMahan, "Adaptive federated optimization," 2021.

A. 详细定理分析

本节提供了隐私-效用-效率权衡的详细数学表达。

A.1. 隐私分析

由于所提出方法中高斯噪声机制的考虑，隐私预算可以通过矩量计算理论 [21] 计算得出，主要是引理 1。

引理 1. *存在常数 c_1, c_2，使得对于任意的 $\varepsilon \leq c_1 q^2 T$，其中 $q = \frac{n}{N}$，如果我们选择*

$$\sigma_z \geq c_2 \frac{n^2 T \ln(1/\delta)}{N^2 \varepsilon^2}.$$

，则 FedAvg 是 (ε, δ)-DP 对于任何 $\delta > 0$。

备注 1. *此外，噪声振幅可以通过隐私预算 [12, 15] 表示，即*

$$\sigma_z^2 = (\frac{\Delta f}{\varepsilon})^2 \cdot 2qT \ln(1/\delta),$$

其中 Δf 是目标函数的敏感度。本文中，敏感度指客户端上传的连续两次更新之间差分隐私部分范数的最大差异。根据相关研究，敏感度可定义为

$$\Delta f = \min_i \frac{2\theta}{|\mathcal{D}_i|}.$$

前述的隐私预算仅指本地更新中的差分隐私部分，而对于同态加密部分，我们可以利用密码算法的特性。由于当前公钥 cryptosystems [9] 的计算安全性，攻击者在没有密钥的情况下从加密更新中获得任何信息（包括与差分隐私定义相关的成员信息）的概率可以忽略不计，只要密钥长度足够长。因此，ParaAegis 的隐私完全依赖于差分隐私部分，这类似于木桶效应。

我们进行定理 1 的证明。

证明. 由于同态加密的密码学特性，可以从 HE 部分提取的信息量在计算上是可以忽略不计的，因此由 HE 保护消耗的隐私预算也几乎为零。同时，根据引理 1，我们建立了 DP 部分噪声幅度与相应隐私预算之间的关系。此外，通过 DP [22] 的并行组合定理，梯度暴露的整体隐私预算是仅由分配给 DP 部分的隐私预算决定的。 □

A.2. 详细假设

在所提出的混合保护下的 FL 问题的收敛性中利用了以下假设，这些是假设 1 的详细版本。

假设 2 (利普希茨光滑). *存在 L，使得对于任意的 $\boldsymbol{x}, \boldsymbol{y}, i$，都有 $\|\nabla f_i(\boldsymbol{x}) - \nabla f_i(\boldsymbol{y})\| \leq L \|\boldsymbol{x} - \boldsymbol{y}\|$。*

假设 3 (有界局部方差). *存在 σ_l，使得对于任意的 t, k, i，均有 $\left\|\boldsymbol{g}_i^{t,k} - \nabla f_i(w_i^{t,k})\right\|^2 \leq \sigma_l^2$。*

假设 4 (有界全局方差). *存在 σ_g，使得对于任意的 i，均有 $\nabla f_i(\boldsymbol{w}) - \nabla f(\boldsymbol{w}) \leq \sigma_g^2$。*

假设 5 (有界梯度). *存在一个 G，使得对于任意的 t, k, i 都有 $\left\|\boldsymbol{g}_i^{t,k}\right\|^2 \leq G^2$ 成立。*

A.3. 定理 1 的证明

在本小节中，我们概述了在深入研究收敛性详细证明之前所需的准备工作。

首先，我们澄清 HE 部分和 DP 部分的定义，即对于任意向量 $\boldsymbol{x}$ 和任何 $l \in [\dim \boldsymbol{x}]$，其 HE 部分 $\boldsymbol{x}_{\mathrm{HE}}$ 和 DP 部分 $\boldsymbol{x}_{\mathrm{DP}}$ 定义为

$$\begin{aligned}\boldsymbol{x}_{\mathrm{HE}}[l] &= \boldsymbol{x}[l] \cdot \mathbb{1}\{l \in \boldsymbol{v}\} \\ \boldsymbol{x}_{\mathrm{DP}}[l] &= \boldsymbol{x}[l] \cdot \mathbb{1}\{l \notin \boldsymbol{v}\}\end{aligned}$$

我们列出了剪辑系数、其均值以及差值，这些是由张等人发现的。[11].

$$\begin{aligned}\alpha_i^t &:= \min\left(1, \frac{\theta}{\sum_{k=0}^{K-1}\left\|\breve{\boldsymbol{g}}_{i,\mathrm{DP}}^{t,k}\right\|}\right), \\ \bar{\alpha}^t &:= \frac{1}{N}\sum_{i=1}^{N}\alpha_i^t,\ \tilde{\alpha}^t := \frac{1}{N}\sum_{i=1}^{N}\left|\alpha_i^t - \bar{\alpha}^t\right|.\end{aligned}$$

局部更新可以根据上述定义很容易地表示成以下形式。

$$\boldsymbol{u}_i^t := \boldsymbol{w}^{t+1} - \boldsymbol{w}^t = -\eta\sum_{k=0}^{K-1}\left(\alpha_i^t\breve{\boldsymbol{g}}_{i,\mathrm{DP}}^{t,k} + \breve{\boldsymbol{g}}_{i,\mathrm{HE}}^{t,k}\right)$$

由利普希茨光滑性，我们有

$$\begin{aligned}&\mathbb{E}\left[f(\boldsymbol{w}^{t+1})\right] - \mathbb{E}\left[f(\boldsymbol{w}^t)\right] \\ \leq&\mathbb{E}\left[\langle\nabla f(\boldsymbol{w}^t), \boldsymbol{w}^{t+1} - \boldsymbol{w}^t\rangle\right] + \frac{L}{2}\mathbb{E}\left[\left\|\boldsymbol{w}^{t+1} - \boldsymbol{w}^t\right\|^2\right] \\ =&\mathbb{E}\left[\left\langle\nabla f(\boldsymbol{w}^t), \frac{1}{n}\sum_{i\in\mathcal{C}}\boldsymbol{u}_i^t\right\rangle\right] + \frac{L}{2}\mathbb{E}\left[\left\|\frac{1}{n}\sum_{i\in\mathcal{C}}\left(\boldsymbol{u}_i^t + \boldsymbol{z}_i\right)\right\|^2\right] \\ =&\underbrace{\left\langle\nabla f(\boldsymbol{w}^t), \mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{C}}\boldsymbol{u}_i^t\right]\right\rangle}_{A_1} + \underbrace{\frac{L}{2}\mathbb{E}\left[\left\|\frac{1}{n}\sum_{i\in\mathcal{C}}\boldsymbol{u}_i^t\right\|^2\right]}_{A_2} + \underbrace{\frac{L\sigma_z^2\theta^2 rd}{2n^2}}_{\text{caused by noise}}\end{aligned} \tag{*}$$

其中 $\boldsymbol{z}_i = [\mathbb{1}\{l \in \boldsymbol{v}_i\} \cdot \mathcal{N}(0, \sigma_z)]_l$ 和最后一项是由噪声引起的。

对于 A_1，我们将采样客户端上的期望转换为所有客户端上的期望。

$$\begin{aligned}A_1 =&\left\langle\nabla f(\boldsymbol{w}^t), \mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{C}}\boldsymbol{u}_i^t\right]\right\rangle \\ =&\left\langle\nabla f(\boldsymbol{w}^t), \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{u}_i^t\right]\right\rangle \\ =&\left\langle\nabla f(\boldsymbol{w}^t), \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{u}_i^t - \bar{\boldsymbol{u}}_i^t)\right]\right\rangle \\ &+\left\langle\nabla f(\boldsymbol{w}^t), \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\bar{\boldsymbol{u}}_i^t\right]\right\rangle\end{aligned} \tag{**}$$

然后，我们关注 (**) 的第一项并得到

$$\begin{aligned}&\left\langle\nabla f(\boldsymbol{w}^t), \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{u}_i^t - \bar{\boldsymbol{u}}_i^t)\right]\right\rangle \\ \overset{i}{\leq}&\left\langle\nabla f(\boldsymbol{w}^t), \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\eta|\alpha_i^t - \bar{\alpha}^t|\boldsymbol{g}_{i,\mathrm{DP}}^{t,k}\right]\right\rangle \\ =&\frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\eta\mathbb{E}\left[|\alpha_i^t - \bar{\alpha}^t|\left\langle f(\boldsymbol{w}^t), \boldsymbol{g}_{i,\mathrm{DP}}^{t,k}\right\rangle\right] \\ \overset{ii}{\leq}&\eta\mathbb{E}\left[\tilde{\alpha}^t\right]KG^2\bar{\rho}_t^2\end{aligned}$$

其中 (i) 是从 $\boldsymbol{u}_i^t$ 和 $\bar{\boldsymbol{u}}_i^t$ 的定义得出的，而 (ii) 是通过有界梯度假设进行的。

限定 (**) 的第二项，我们有

$$
\begin{aligned}
&\left\langle \nabla f(\boldsymbol{w}^t), \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N \bar{\boldsymbol{u}}_i^t\right]\right\rangle \\
=&\mathbb{E}\left[\left\langle \nabla f(\boldsymbol{w}^t), \frac{1}{N}\sum_{i=1}^N \breve{\boldsymbol{u}}_i^t\right\rangle\right] \\
=&\frac{-\eta\beta_t K}{2}\left\|\nabla f(\boldsymbol{w}^t)\right\|^2 - \frac{\eta\beta_t}{2K}\mathbb{E}\left[\left\|\frac{1}{\eta\beta_t N}\sum_{i=1}^N \breve{\boldsymbol{u}}_i^t\right\|^2\right] \\
&+\frac{\eta\beta_t}{2}\underbrace{\mathbb{E}\left[\left\|\sqrt{K}\nabla f(\boldsymbol{w}^t) - \frac{1}{\eta\beta_t N\sqrt{K}}\sum_{i=1}^N \breve{\boldsymbol{u}}_i^t\right\|^2\right]}_{A_3}
\end{aligned}
$$

第二个方程来自于对于任何向量 $\boldsymbol{x},\boldsymbol{y}$，$\langle \boldsymbol{x},\boldsymbol{y}\rangle = -\frac{1}{2}\|\boldsymbol{x}\|^2 - \frac{1}{2}\|\boldsymbol{y}\|^2 + \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2$ 都成立。注意到 $\breve{\boldsymbol{g}}_i^t := \nabla f_i(\boldsymbol{w}_i^t)$ 和 $\breve{\boldsymbol{g}}_i^{t,k} := \nabla f_i(\boldsymbol{w}_i^{t,k})$，我们依次分解 A_3：

$$
\begin{aligned}
A_3 =&K\mathbb{E}\left[\left\|\nabla f(\boldsymbol{w}^t) - \frac{1}{KN}\sum_{i=1}^N\sum_{k=0}^{K-1}(\alpha_i^t\breve{\boldsymbol{g}}_{i,\mathrm{DP}}^{t,k} + \breve{\boldsymbol{g}}_{i,\mathrm{HE}}^{t,k})\right\|^2\right] \\
=&K\mathbb{E}\left[\left\|\frac{1}{KN}\sum_{i=1}^N\sum_{k=0}^{K-1}(\breve{\boldsymbol{g}}_i^t - \alpha_i^t\breve{\boldsymbol{g}}_{i,\mathrm{DP}}^{t,k} - \breve{\boldsymbol{g}}_{i,\mathrm{HE}}^{t,k})\right\|^2\right] \\
\le&\frac{1}{N}\sum_{i=1}^N\sum_{k=0}^{K-1}\mathbb{E}\left[\left\|\breve{\boldsymbol{g}}_i^t - \alpha_i^t\breve{\boldsymbol{g}}_{i,\mathrm{DP}}^{t,k} - \breve{\boldsymbol{g}}_{i,\mathrm{HE}}^{t,k}\right\|^2\right] \\
\le&\frac{1}{N}\sum_{i=1}^N\sum_{k=0}^{K-1}\mathbb{E}\left[\left\|\breve{\boldsymbol{g}}_i^t - \breve{\boldsymbol{g}}_i^{t,k}\right\|^2\right] \\
&+\frac{1}{N}\sum_{i=1}^N\sum_{k=0}^{K-1}\mathbb{E}\left[\underbrace{\left\|\breve{\boldsymbol{g}}_i^{t,k} - \alpha_i^t\breve{\boldsymbol{g}}_{i,\mathrm{DP}}^{t,k} - \breve{\boldsymbol{g}}_{i,\mathrm{HE}}^{t,k}\right\|^2}_{A_4}\right] \\
\le&\frac{1}{N}\sum_{i=1}^N\sum_{k=0}^{K-1}\left(L^2\mathbb{E}\left[\left\|\boldsymbol{w}^t - \boldsymbol{w}_i^{t,k}\right\|^2\right] + G^2\right) \\
\le&L^2 5K^2\eta^2(\sigma_l^2 + 6K\sigma_g^2) + L^2 30Q^3\eta^2\left\|\nabla f(\boldsymbol{w}^t)\right\|^2 + KG
\end{aligned}
$$

其中第一个不等式来自于詹森不等式，第三个来自于 L-光滑性条件下的 f，最后一个来自于 [23] 的引理 3，即对于任意的 k 都有 $\frac{1}{N}\mathbb{E}\left[\left\|\boldsymbol{w}^t - \boldsymbol{w}_i^{t,k}\right\|^2\right] \le 5K^2\eta^2(\sigma_l^2 + 6K\sigma_g^2) + 30Q^3\eta^2\left\|\nabla f(\boldsymbol{w}^t)\right\|^2$。

此外，寻找 A_4 的上界的具体细节可以描述为

$$
\begin{aligned}
A_4 =&\left\|\breve{\boldsymbol{g}}_{i,\mathrm{DP}}^{t,k} - \alpha_i^t\breve{\boldsymbol{g}}_{i,\mathrm{DP}}^{t,k}\right\|^2 = (1-\alpha_i^t)^2\left\|\breve{\boldsymbol{g}}_{i,\mathrm{DP}}^{t,k}\right\|^2 \\
\le&(1-\alpha_i^t)^2(1-\rho_t^2)G^2 \le G^2
\end{aligned}
$$

其中第一个等式来自定义 $\breve{\boldsymbol{g}}_{i,\mathrm{DP}}^{t,k} + \breve{\boldsymbol{g}}_{i,\mathrm{HE}}^{t,k} = \breve{\boldsymbol{g}}_i^{t,k}$ 而第二个不等式则来源于事实 $\alpha_i^t \le 1, \rho_t \le 1$。

现在我们将注意力转向 A_2 在*中的情况。

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\frac{1}{n}\sum_{i\in\mathcal{C}}\boldsymbol{u}_i^t\right\|^2\right] \\
=&\mathbb{E}\left[\left\|\frac{1}{n}\sum_{i\in\mathcal{C}}\sum_{k=0}^{K-1}\eta(\alpha_i^t\boldsymbol{g}_{i,\mathrm{DP}}^{t,k} + \boldsymbol{g}_{i,\mathrm{HE}}^{t,k})\right\|^2\right] \\
=&\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{C}}\sum_{k=0}^{K-1}\eta^2\beta^2\left\|\boldsymbol{g}_i^{t,k}\right\|^2\right] \\
\le&\frac{\eta^2G^2}{n^2}\mathbb{E}\left[\sum_{i\in\mathcal{C}}\sum_{k=0}^{K-1}(\beta_i^{t,k})^2\right] = \frac{\eta^2G^2\mathbb{E}\left[\bar{\beta}_t^2\right]}{n}
\end{aligned}
$$

将方程相加，我们得到

$$
\begin{aligned}
\mathbb{E}\left[f(\boldsymbol{w}^{t+1})\right] \le& f(\boldsymbol{w}^t) + \eta\tilde{\alpha}^t(1-\tilde{\rho}_t^2)KG^2 \\
&-\frac{\eta\bar{\beta}_t K}{2}\left\|\nabla f(\boldsymbol{w}^t)\right\|^2 - \frac{\eta\bar{\beta}_t}{2K}\mathbb{E}\left[\left\|\frac{1}{\eta N\bar{\beta}_t}\sum_{i=1}^N\breve{\boldsymbol{u}}_i^t\right\|^2\right] \\
&+\frac{\eta\bar{\beta}_t}{2}\left(5L^2K^2\eta^2(\sigma_l^2 + 6K\sigma_g^2) + 30L^2K^3\eta^2\left\|\nabla f(\boldsymbol{w}^t)\right\|^2 -\right. \\
&+\frac{\eta^2LG^2K}{2n} + \frac{L\sigma_z^2\theta^2 d(1-r)}{2n^2}
\end{aligned}
$$

通过设置 $\eta \le \frac{1}{\sqrt{60}KL}$ 进行简化：

$$
\begin{aligned}
f(\boldsymbol{w}^{t+1}) \le& f(\boldsymbol{w}^t) - \frac{\eta\bar{\beta}_t K}{4}\left\|\nabla f(\boldsymbol{w}^t)\right\|^2 + \tilde{\alpha}^t(1-\tilde{\rho}_t^2)KG^2 \\
&+\frac{5}{2}\eta^3\bar{\beta}_t L^2K^2(\sigma_l^2 + 6K\sigma_g^2) + \frac{1}{2}\eta_l^2\bar{\beta}_t KG^2 \\
&+\frac{\eta^2LG^2K}{2n} + \frac{L\sigma_z^2\theta^2 d(1-r)}{2n^2}
\end{aligned}
$$

在从 $t=0$ 到 $T-1$ 求和，两边除以 $\frac{\eta KT}{4}$，用 ε 替换 σ_z 并根据引理 1 用 δ 替换后，重新整理得到

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\bar{\beta}_t\left\|\nabla f(\boldsymbol{w}^t)\right\|^2\right] \leq \frac{4}{\eta KT}\mathbb{E}\left[f(\boldsymbol{w}^0)-f(\boldsymbol{w}^T)\right]$$
$$+4G^2\mathbb{E}_t\left[\tilde{\alpha}^t(1-\bar{r}^t)\right]+10\eta^2L^2K(\sigma_l^2+6k\sigma_g^2)\mathbb{E}_t\left[\bar{\beta}_t\right]$$
$$+2G^2\mathbb{E}_t\left[\bar{\beta}_t\right]+\frac{\eta^2LG^2K}{2n}+\frac{2TL\theta^2d(1-r)\ln(1/\delta)}{N^2D_{\min}^2\varepsilon^2}.$$

通过省略与 r 和 (ε,δ) 无关的项，我们得到定理 1。