EasyChair Preprint

№ 12740

DPLE: a Privacy-Enhanced and Straggler-Resilient Distributed Learning Framework for Smart Cloud

Yilei Xue, Jianhua Li and Jun Wu

April 30, 2024

DPLE: A Privacy-Enhanced and Straggler-Resilient Distributed Learning Framework for Smart Cloud

Yilei Xue
Jianhua Li
Jun Wu

***Abstract*—In the intelligent cloud setting, distributed learning encounters privacy and straggler challenges. Lagrange coded computing offers partial relief. Yet, if the number of inquisitive but honest nodes surpasses a threshold or if there are external eavesdroppers, system privacy becomes compromised. To tackle this issue, we introduce a novel approach called DPLE (Differentially Private Lagrange Encoding). Additionally, we provide theoretical analysis to determine the artificial noise variance necessary for achieving desired privacy levels within this framework. Through experimentation, we evaluate how different system parameters affect accuracy.**

I. Introduction

Cloud computing [1] offers powerful computing and storage resources, making distributed learning more feasible and practical. However, there are privacy and straggler issues in distributed learning [2], [3]. Yet, persistent challenges hinder the advancement of distributed learning within smart cloud environments. These challenges encompass issues concerning performance, fault tolerance, adaptability, and privacy considerations. Our study predominantly addresses the privacy concerns inherent in distributed learning, particularly crucial in cloud setups where physical resources are shared among multiple users, segregated by virtual machines.

Recently, the integration of coding methodologies with distributed learning has garnered increased attention as a means to uphold system privacy [4], [5]. Lagrange coded computing [6], [7] (LCC) uses interpolation polynomials for data encoding, balancing privacy, security, and resilience. Yet, it's limited to handling polynomial functions. CodedPrivateML enhances LCC with sigmoid function approximation for logistic regression within its framework. However, relying only on interpolation polynomials provides modest privacy safeguards. This paper advances privacy safeguards by integrating differential privacy (DP) [8]–[11] into the LCC framework.

II. System Model of DPLE

In Fig. 1, we depict a master-worker setup consisting of a master and N workers. The master holds the dataset $\mathbf{X}$ and label vector $\mathbf{y}$. The model weight $\mathbf{w}$ is obtained by minimizing the cross-entropy loss function. Fig. 2 depicts the system architecture of DPLE. Initially, the master normalizes dataset $\mathbf{X}$ into $\bar{\mathbf{X}}$. Then $\bar{\mathbf{X}}$ is divided into K shares $\bar{\mathbf{X}} = [\bar{\mathbf{X}}_1^T, \ldots, \bar{\mathbf{X}}_K^T]^T$, and the encoded dataset $\hat{\mathbf{X}}_i$ $(i = 1, \ldots, N)$ is denoted as $\hat{\mathbf{X}}_i = g_x(m_i)$ with

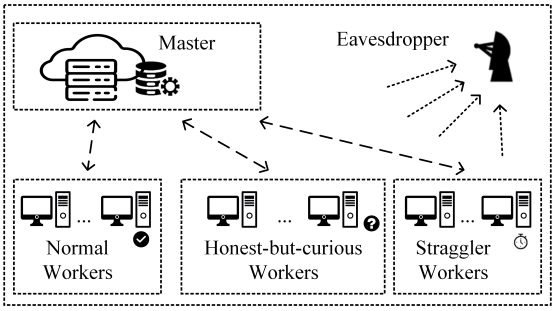

Fig. 1. Scenario.

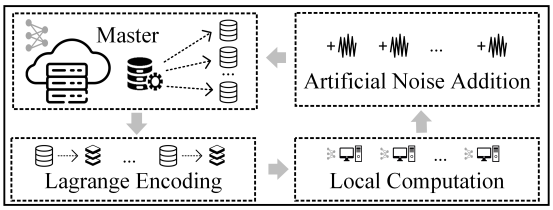

Fig. 2. System architecture.

$$g_x(\beta) = \sum_{i=1}^{K} \bar{\mathbf{X}}_i \prod_{k \neq i} \frac{\beta - u_k}{u_i - u_k} + \sum_{i=K+1}^{K+T} \mathbf{Z}_i \prod_{k \neq i} \frac{\beta - u_k}{u_i - u_k}. \quad (1)$$

In Eq.(1), $\mathbf{Z}i$ is a redundancy matrix and T signifies the maximum tolerable number of honest but curious nodes. $u_1, \ldots, u_{K+T}$ and $\{m_i\}_{i=1}^N$ are distinct numbers selected by the master. Similarly, the master uses the same set $\{m_i\}_{i=1}^N$ to encode $\mathbf{w}^{(t)}$. The i-th encoded weight vector is $\hat{\mathbf{w}}_i^{(t)} = g_w(m_i)$, where

$$g_w(\beta) = \sum_{i=1}^{K} \mathbf{w}^{(t)} \prod_{k \neq i} \frac{\beta - u_k}{u_i - u_k} + \sum_{i=K+1}^{K+T} \mathbf{v}_i \prod_{k \neq i} \frac{\beta - u_k}{u_i - u_k}. \quad (2)$$

The elements in $\mathbf{v}_i$ are randomly selected from $[-\xi, \xi]$ for some real ξ.

After encoding, the master assigns $\hat{\mathbf{X}}_i$ and $\hat{\mathbf{w}}_i^{(t)}$ to the i-th worker abd each worker computes the local function

$$f(\hat{\mathbf{X}}_i, \hat{\mathbf{w}}_i^{(t)}) = \hat{\mathbf{X}}_i^\top \cdot \hat{S}(\hat{\mathbf{X}}_i \cdot \hat{\mathbf{w}}_i^{(t)}). \quad (3)$$

$\hat{S}(z) = \sum_{i=0}^{r} a_i z^i$ is the r-th order approximation of the sigmoid function and a_i is estimated using least squares estimation.

After the i-th worker finishes computation, artificial noise $\mathbf{n}_i^{(t)}$ is added to $f(\hat{\mathbf{X}}_i, \hat{\mathbf{w}}_i^{(t)})$ to meet DP requirements. Hence, the perturbed result $\tilde{f}_i^{(t)}$ is given by

$$\tilde{f}_i^{(t)} = \hat{\mathbf{X}}_i^{\top} \cdot \hat{S}(\hat{\mathbf{X}}_i \cdot \hat{\mathbf{w}}_i^{(t)}) + \mathbf{n}_i^{(t)}. \tag{4}$$

Upon receiving $\mathcal{D} = (2r{+}1)(K{+}T{-}1){+}1$ results ($\mathcal{D}$ denotes the set of the first workers who have finished their computations), the new polynomial $\tilde{h}(\beta)$ is crafted by the master as

$$\tilde{h}(m_i) \triangleq \tilde{f}_i^{(t)} = f(\hat{\mathbf{X}}_i, \hat{\mathbf{w}}_i^{(t)}) + \mathbf{n}_i^{(t)}. \tag{5}$$

Note that $N \geq |\mathcal{D}| + S_s$, where S_s represents the maximum number of stragglers. Hence the function values at u_α ($\alpha = 1, \ldots, K$) can be expressed as

$$\tilde{h}(u_\alpha) = \sum_{i \in \mathcal{D}} [\hat{\mathbf{X}}_i^{\top} \hat{S}(\hat{\mathbf{X}}_i \cdot \hat{\mathbf{w}}_i^{(t)}) + \mathbf{n}_i^{(t)}] \cdot \prod_{j \in \mathcal{D} \setminus \{i\}} \frac{u_\alpha - m_j}{m_i - m_j} \tag{6}$$

Once the master obtains $\tilde{h}(u_\alpha)$ for $\alpha = 1, \ldots, K$, it aggregates these results

$$\sum_{\alpha=1}^{K} \tilde{h}(u_\alpha) = \sum_{\alpha=1}^{K} [f(\bar{\mathbf{X}}_\alpha, \mathbf{w}^{(t)}) + \tilde{\mathbf{n}}_\alpha^{(t)}], \tag{7}$$

and updates the gradient based on

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \frac{\eta}{m} \{\bar{\mathbf{X}}^{\top} [(\hat{S}(\bar{\mathbf{X}}\mathbf{w}^{(t)}) - \mathbf{y})] + \sum_{\alpha=1}^{K} \tilde{\mathbf{n}}_\alpha^{(t)}\}. \tag{8}$$

III. Analysis of Artificial Noise Variance

This section explores the conditions necessary for the variance of the artificial Gaussian noise $\mathbf{n}_i^{(t)} \sim \mathcal{N}(0, \sigma_{(t),i}^2)$.

We define a matrix $\mathbf{R}^{(t),i} \triangleq \hat{\mathbf{X}}_i^{\top} \odot (\mathbf{1}_d \cdot [\hat{S}(\hat{\mathbf{X}}_i \cdot \hat{\mathbf{w}}_i^{(t)})]^{\top})$ for the i-th worker. Then we find that $f(\hat{\mathbf{X}}_i, \hat{\mathbf{w}}_i^{(t)}) = \sum_{\varsigma=1}^{m/K} \mathbf{R}_\varsigma^{(t),i}$, where $\mathbf{R}_\varsigma^{(t),i}$ denotes the ς-th column of $\mathbf{R}^{(t),i}$. Assuming $\mathbf{R}_\varsigma^{(t),i}$ is bounded by $B_2^{(t),i}$ and according to Eq. (3), we obtain $\Delta\text{Gau} = 2B_2^{(t),i}$. Thus, for the i-th worker in the t-th iteration, ensuring the $(\varepsilon_i, \delta_i)$-DP is feasible if $\sigma_{(t),i}$ meets

$$\sigma_{(t),i} \geq \sqrt{2\ln(1.25/\delta_i)} \cdot 2B_2^{(t),i}/\varepsilon_i. \tag{9}$$

IV. Experiment

We validate the effectiveness of DPLE through experiments in this section. Logistic regression training is conducted on the MNIST and FashionMNIST datasets. For MNIST, 12,700 samples from classes 1 and 2 are chosen for training. FashionMNIST focuses on binary classification between 'Pullover' and 'Dress' classes. Each dataset sample comprises 785 features, with the additional feature accounting for bias. Privacy budget ε is uniformly distributed across workers, and δ is set to 0.01.

Figures 3 and 4 illustrate the training accuracy of the Fashion-MNIST and MNIST datasets under different privacy budgets. In Fig. 3, we set the parameters as $N = 50$, $K = 5$, and $T = 4$, while in Fig. 4, the parameters are $N = 50$, $T = 2$, and $K = 5$. It's observed that training accuracy increases as

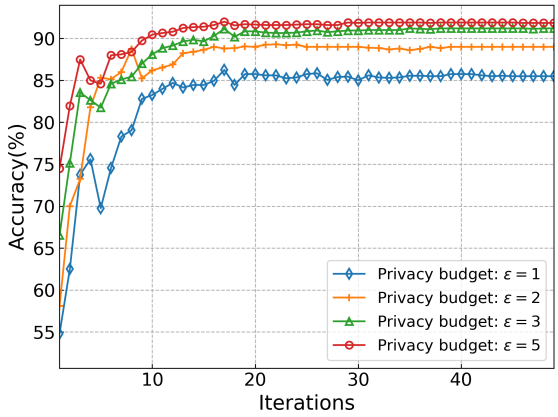

Fig. 3. The influence of varied privacy budgets.

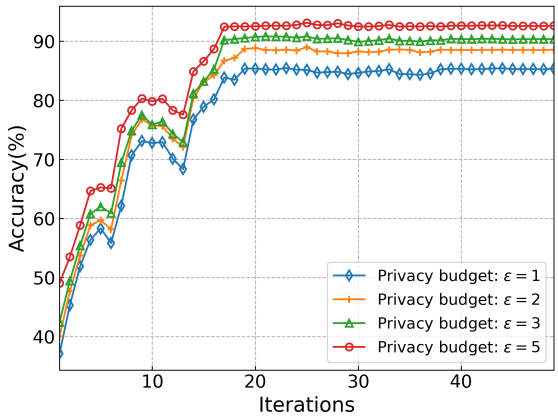

Fig. 4. The impact of varied privacy budgets.

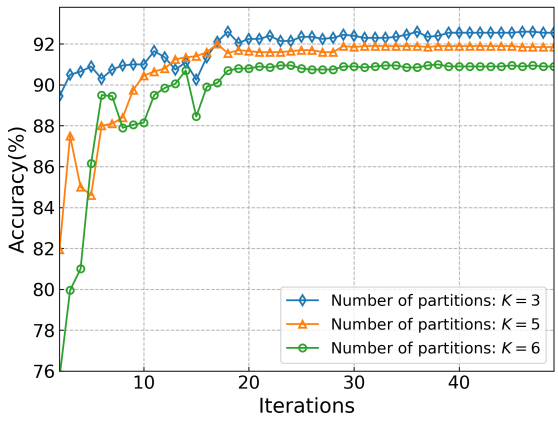

Fig. 5. The impact of varied dataset partitions with Fashion-MNIST.

the privacy budget expands, albeit with a trade-off of reduced privacy protection.

In Fig. 5, we illustrate the impact of varying dataset partitions on training accuracy, with parameters set at $N = 50$, $T = 4$, and $\varepsilon = 5$. It can be observed that increasing dataset partitions leads to a gradual decrease in training accuracy.

V. Conclusion

To counter the risks posed by numerous honest but inquisitive nodes and external eavesdroppers in the LCC framework, we integrated differential privacy into it. We then examined the necessary noise variance magnitude for attaining targeted levels of privacy protection. Experimentation investigated the impact of different system parameters on training accuracy.

References

[1] Z. Chkirbene, A. Erbad, R. Hamila, A. Gouissem, A. Mohamed, and M. Hamdi, "Machine learning based cloud computing anomalies detection," *IEEE Network*, vol. 34, no. 6, pp. 178–183, 2020.

[2] M. Klymash, M. Kyryk, I. Demydov, O. Hordiichuk-Bublivska, H. Kopets, and N. Pleskanka, "Research on distributed machine learning methods in databases," in *2021 IEEE 4th International Conference on Advanced Information and Communication Technologies (AICT)*, 2021, pp. 128–131.

[3] N. D. Vanli, M. O. Sayin, I. Delibalta, and S. S. Kozat, "Sequential nonlinear learning for distributed multiagent systems via extreme learning machines," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 546–558, 2017.

[4] E. Ozfatura, S. Ulukus, and D. Gündüz, "Coded distributed computing with partial recovery," *IEEE Transactions on Information Theory*, vol. 68, no. 3, pp. 1945–1959, 2022.

[5] S. Dhakal, S. Prakash, Y. Yona, S. Talwar, and N. Himayat, "Coded federated learning," in *2019 IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–6.

[6] S. Gupta and J. Harshan, "Analog lagrange coded computing: On the curious case of adversarial workers," in *2024 National Conference on Communications (NCC)*, 2024, pp. 1–6.

[7] J. So, B. Güler, and A. S. Avestimehr, "Codedprivateml: A fast and privacy-preserving framework for distributed machine learning," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 441–451, 2021.

[8] R. Hu, Y. Guo, and Y. Gong, "Federated learning with sparsified model perturbation: Improving accuracy under client-level differential privacy," *IEEE Transactions on Mobile Computing*, pp. 1–14, 2023.

[9] J. Soria-Comas, J. Domingo-Ferrer, D. Sánchez, and D. Megías, "Individual differential privacy: A utility-preserving formulation of differential privacy guarantees," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 6, pp. 1418–1429, 2017.

[10] J. Li, F. Zhang, Y. Guo, S. Li, G. Wu, D. Li, and H. Zhu, "A privacy-preserving online deep learning algorithm based on differential privacy," in *2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2023, pp. 559–564.

[11] M. Kim, O. Günlü, and R. F. Schaefer, "Federated learning with local differential privacy: Trade-offs between privacy, utility, and communication," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 2650–2654.