International Journal of Electrical, Electronics and Computer Engineering **14**(1&2): 22-25(2025)

ISSN No. (Online): 2277-2626

Federated Learning with Differential Privacy: Balancing Privacy and Model Accuracy in Decentralized Data

Agrima Guleria* and Komal

School of Computer Science Engineering and Technology,
Government College Dharamshala (H.P.), India.

(Corresponding author: Agrima Guleria)*
(Received: 13 February 2025, Accepted: 20 March 2025)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Federated Learning (FL) enables the training of machine learning models over distributed data on many devices simultaneously without relocating any sensitive information to a centralized server. The system's privacy concerns pose another important obstacle, especially when it comes to heavily sensitive data. In this paper, we investigate the enhancement of privacy protection using Federated Learning in conjunction with Differential Privacy (DP), which is generally used to protect the accuracy of the given models. We study methods of implementing DP into the federated learning workflow, including noise injection into gradients, secure model update aggregation, and privacy budget allocation. We further analyze the impact of these techniques of preserving privacy on the quality of the global model, particularly on its accuracy. By using real-world datasets for experimentation, we demonstrate the trade-off between maintaining privacy and achieving satisfactory performance on the model. These findings are intended to inform the design of more robust and secure federated learning systems by providing guidance on making privacy-accuracy trade off decisions.

Keywords: FL, DP, Privacy Preservation, Decentralized Data, Model Accuracy, Privacy-Preserving Machine Learning, Secure Aggregation, Privacy Budget, Model Training, PPTAP.

INTRODUCTION

Federated Learning (FL), an advanced decentralized approach to machine learning in which many devices (or nodes) collaboratively train a shared global model while keeping all the training data on the device, has been introduced very recently. This will prove invaluable in privacy-sensitive areas like healthcare, finance, and even mobile applications, where moving all data to a data-center threatens the security of all users (McMahon *et al.,* 2017). However, while these advantages are clear, there are privacy issues associated with FL that are unique to this technology, most notably the potential for sensitive information leakage through the transfer of model updates. While raw data is typically not shared, updates to the models themselves can potentially leak information about individuals.

To combat these concerns, Differential Privacy (DP) is used in FL systems. Differential privacy (DP) techniques add noise to the updates of the model, disrupting the ability to track how individual data points influence high-level performance metrics (Dwork and Roth 2014). Combining FL with DP aims for a better privacy level for users, as demonstrated in McMahon *et al.* (2017)] and introduces Federated Optimization. This opened a path towards more privacy-preserving approaches in federated settings. In particular, a federated learning model with differential privacy, at ICLR 2020 which proposed a method of balancing privacy protection and model accuracy.

While adding DP to FL helps with preserving privacy in the model, this process usually results in a reduction of the correctness of the model. In many cases, protecting individual-level data will lead to large bias; for example, the extra noise we need to inject in the model for achieving privacy will necessarily degrade the accuracy of the model, thus making it hard to reach high prediction power while keeping individuals secure. However, finding the right trade-off between privacy and accuracy is still an ongoing research topic. Such investigations, similar to a still active area of research and new approaches are continuously devised to minimize this trade-off and enhance both the privacy guarantee and predictive power of the federated models. In this document, we focus on differential privacy in the federated learning case and the challenges in jointly applying these two concepts with complementary objectives while still achieving solutions that balance privacy and accuracy. We discuss recent advances in this area, with a particular emphasis on the work of McMahon *et al*. (2017) and summarize recent efforts to jointly enhance privacy protection and the quality of models trained on decentralized data.

RELATED WORK

For its privacy-preserving capabilities, Federated Learning (FL) has emerged as a popular approach especially for integrating machine learning in resource constrained environments, since sensitive data remains on individual devices. For distributed systems, the ability to collaborate across multiple clients without

requiring raw data to leave local devices is crucial, especially for privacy-sensitive domains like healthcare, finance, and mobile systems. Its promises aside, FL still struggles with a degree of privacy leakage, especially when model updates which are shared between devices are shown to be de facto inversely mapped to the shared data. These leaks are now being contained through the concept of Differential Privacy (DP), which has been shown to provide strong guarantees by injecting noise to model updates in a way where individual data points cannot be identified.

The idea of Federated Learning was first brought to light by McMahan and his colleagues back in 2017 during their presentation at AISTATS 2017. This groundbreaking research showed that decentralized machine learning is possible by allowing local devices to train models using their own data and only share the updates to those models, rather than the raw data itself. While this work was crucial for the development of Federated Learning, it didn't include any mechanisms to protect privacy. On the other hand, more recent studies have been working on incorporating Differential Privacy (DP) into Federated Learning to safeguard data during the training process. For instance, in ICLR 2020, Geyer introduced a technique for applying DP to Federated Learning, which involves adding noise to the gradients calculated during local model updates. This method does a great job of protecting sensitive information, but it does have a downside—the overall accuracy of the global model tends to drop because of the noise. Their experiments revealed that depending on how much noise is added, DP could lead to an accuracy decrease of about 5-15%, highlighting the need to strike a balance between privacy and the performance of the model.

Specifically, DP represents a mathematical framework based on discrete mathematics that can guarantee privacy, but researchers also introduced secure aggregation methods to preserve the privacy of Federated Learning. Secured aggregation at NeurIPS. In this fashion, this approach ensures the aggregation of updates from the devices and per update privacy. It is also privacy-preserving, since it prevents the server (or other participants) from seeing who did or did not contribute. But even using secure aggregation, DP still comes at some accuracy cost. On the aggregated tests, they observed a 3-10% accuracy drop due to the noise introduced during aggregation and synchronization problems due to update misalignment across the devices. While secure aggregation provides privacy, it has the same trade-off with model accuracy that DP does.

Here, DP presents the trade-off between privacy and accuracy, which is a fundamental fact in FL. Regardless, to preserve privacy, noise is added to model updates or secure aggregation is applied, hence degrades the utility of global model and make it less accurate as a whole. One of the studies is titled "Balancing Privacy and Generalizability for Knowledge Transfer Learners" (2021) where this balance and the optimization of this trade-off was addressed by the IEEE Transactions on Neural Networks.

They focused on controlling the privacy budget (the total noise added across the training process) in their research, and they demonstrated that an appropriately set privacy budget can alleviate accuracy loss. Using DP to achieve more privacy (*i.e*., lower epsilon), might cost an accuracy loss between 15-20%, but they observed that with right tuning of the privacy budget, the accuracy loss could be reduced to ~ 5-10% — making a fair trade-off between privacy and performance.

Extending these findings, improved this method at ICML 2022, that proposed a method for adaptive allocation of privacy budgets. They suggested a dynamic adjustment of privacy budget during the training process, in which larger noise is added in the early phase of the model training, to facilitate better model learning, while lower noise is added in the latter phase of the model training as the model approaches convergence. By experimental findings, they exhibited around 10-15% lower accuracy degradation when employing their adaptive noise levels than static noise technique. This responsive tuning was a big win for model performance without sacrificing much in terms of utility guarantee.

The literature has explored various hybrid techniques that address privacy on FL through the exploitation of the combination of DP with various other technologies such as homomorphic encryption or secure multi-party computation to intensify and enhance the FL privacy further more. Truex *et al*. (2018) at ACM CCS 2018, suggested hybrid approaches by integrating DP with other cryptography techniques for extra protection on data privacy. This approach makes it possible to achieve better privacy guarantees, while at the same time it results in much higher computational cost and seriously degrading modeling accuracy. Even worse, it was noted that this accuracy drop was significantly in the range of 20-25% even to higher-dimensional data, especially in higher-level privacy protection-required scenarios. But they said the extra computational overhead and loss of accuracy might be a worthwhile trade-off in highly sensitive use cases such as medical data processing, where enhanced privacy guarantees would be beneficial.

More recently, there has also been an effort to optimize these privacy-preservation mechanisms. Liu *et al.* (2023) utilized DP for ensuring privacy preserving with low impact on accuracy by merging DP with a customized approach as to add noise to data samples in order to prevent privacy breach at FL. In fact, their approach also showed that a better accuracy trend could be retained (losses in accuracy were only 5-7%)). We found this approach to be especially effective for use cases where the device privacy was most essential but we had to also keep the performance of the global model intact.

Overall, combining Differential Privacy with Federated Learning has advanced the privacy guarantees of decentralized machine learning models order of

magnitudes, but the balance between providing high-performance models and strong privacy guarantees is still difficult to achieve. All these studies find that, while DP can provide privacy protection, the trade-off is often a loss of accuracy, particularly if very strong privacy is given (*i.e.*, small epsilon values). Recent improvements in dynamic privacy budget management, secure aggregation, hybrid privacy-preserving methods and many others, for the first time ever are tantalizingly shifting the balance towards both increased privacy and accuracy paving the way for efficient and secure Federated Learning systems. The above methods, of course, will continue to grow in the future, as new practical methods are found to limit accuracy loss while remaining practical in application.

Author(s)	Technique Used	Objectives	Performance	Dataset	Simulator/Outcomes
McMahan and Ramage (2017)	Federated Learning, No Privacy Techniques	Propose decentralized learning with no direct data sharing	Baseline performance in FL without privacy measures	Not specified	Decentralized learning without data transfer
Truex *et al.* (2018)	Hybrid Approach (DP + Cryptography)	Combine DP with encryption for stronger privacy guarantees	20-25% loss in accuracy due to encryption overhead	Not specified	Hybrid models provide enhanced privacy but at a cost

RESEARCH GAP

For the combination of DP and FL, great progress has been made, but many research gaps still exist. Confidentiality and accuracy often have an inverse relationship, as Privacy-Protection Techniques (PPTs) are used directly into the data, so in most of the times a high degradation of performance is probably. Dynamic adjustment: Another gap left unexplored is the adjustment of privacy budget, since existing approaches are not designed to adjust the privacy of its differential mechanisms dynamically during the training. Hybrid privacy methods which combine DP with cryptography also require further exploration, which will be mostly focusing on scalability, computational efficiency, etc. In addition, the majority of previous works assume the IID Data setting and so they neglect the non-IID, heterogeneous data. Finally, they should inspire remixing to enable real-world implementation and generalization to metrics other than prediction accuracy (e.g. computation overhead, communication overhead, etc).

FINDING SUGGESTIONS

The first, and perhaps most enduring, challenge is finding a better balance between privacy and accuracy. Alternatively one could devise methods that add noise to the model during training adaptively based on the model performance. Unlike fixed noise, the system could use different noise levels according to individual needs and tilt the privacy versus accuracy scale. For instance, reinforcement learning would probably experiment as well since the model will learn how the privacy settings can be optimized towards practicing the desired job based on the training data available.

The second enhancement area is dynamic privacy budge adjustment. Or, dynamic privacy budget could be built so that the privacy budget not only be adjusted based on the data or the model trajectory instead of static. Doing so would help the system better allocate privacy resources, allowing it to preserve privacy while ensuring optimal accuracy retention. Algorithms for when more privacy is needed (more sensitive states of training, for example) may also help.

In future work, we would augment the efficiency of these privacy methods in the case of hybrid privacy protection where DP is combined with cryptographic techniques, e.g., encryption. As overheads for computation and training occupy current architectures, hence finding an alternative approach for speeding up these methods is extremely valuable! Exploring parallel computation or data quantization would also reduce the computation burden and improve the feasibility of these methods for large-scale FL systems.

But they don't do so well for non-IID data (when the data is mostly unique across clients). Future work may explore robustification of FL against such attacks. One possible approach, then, is to use personalized models, which means that each client has its own model but is also updated in conjunction with the global model. This method may perform better on various data types while maintaining privacy and accuracy.

Last but not least, to ensure that these techniques are practical, it's important to validate them on large-scale, real-world datasets e.g. healthcare, finance, or mobile applications. Challenges that won't appear in smaller and sanitized tests will emerge in the real world. We should also move towards better evaluation metrics that are more than just accuracy, to include energy metrics, communication cost, real-time performance, etc. This will give a fuller picture of how these systems perform under realistic conditions.

By focusing on these areas, future research could make significant strides in improving the balance between privacy and model accuracy in Federated Learning, making these systems more practical and efficient for real-world applications.

CONCLUSIONS

In other words, privacy preservation in decentralized data situations is a very clever balancing act and FL +

DP is a favourable alternative to build efficient models without sacrificing machine learning privacy requirements. However, in spite of the considerable advances achieved with the integration of such techniques, substantial hurdles remain, particularly within the issue of privacy vs. accuracy. Such trade-off is significant, and privacy-preserving techniques lead to considerable loss in model performance. In addition, dynamic privacy budgets can be difficult to calculate, data may be heterogeneous and non-IID, making the management of FL systems a challenging task, and hybrid privacy-preserving methods can be computationally expensive.

Nevertheless, in spite of these problems, the prospects of Differentially Private Federated Learning remain promising, particularly with the fading off of these issues, more specifically the advancement of adaptive privacy techniques and the dynamic budget management, and a need to enhance computational efficiency. Research needs to transition now towards improving and optimizing those systems, to be robust and applicable in real world scenarios such as healthcare, finance and mobile services. But with the related cryptographic technologies and Federated Learning model convergence continuously researched, it is able to realize a practical, economical, and robust protection scheme in many decentralized machine learning scenarios.

REFERENCES

Dwork, C., & Roth, A. (2014). Differential privacy: a survey of results Foundations and Trends in *Theoretical Computer Science, 9*(3–4), 211–407.

McMahan, H. B. & Ramage, D. (2017). Federated learning — The process of training machine learning models collaboratively without the need to centralize data. *Communications of the ACM, 61*(11), 73–82.

Truex, S., Liu, Y., Zhang, X. & Shmatikov, V. (2020) A tutorial on federated learning: A privacy perspective. *ACM Computing Surveys, 53*(5), 1–30.