Differentially Private Federated Learning in Edge Networks: The Perspective of Noise Reduction

Yiwei Li, Shuai Wang, Chong-Yung Chi, and Tony Q. S. Quek

Abstract

The proliferation of distributed sensitive data in recent years in network edge devices motivates the introduction of edge computing which moves machine learning (ML) applications from the data center to the edge of the network. On the other hand, recent demands on data privacy have called for federated learning (FL) as a new distributed paradigm. The inherent privacy protection nature of FL makes FL in edge computing a prospective framework, especially for application scenarios where privacy protection and resource utilization are critical. Nevertheless, FL also suffers from privacy leakage as the exchanged messages between edge devices and the edge server can be revealed. As such, differential privacy has drawn great attention for privacy protection in the edge FL system for its extremely low computation cost, which is readily implemented by adding well-designed noise to target data/models. However, the added noise will deteriorate learning performance, and it is challenging to get a satisfactory trade-off between privacy protection and learning performance. This article gives the first systematic study on the framework of differentially private FL in edge networks from the perspective of noise reduction. To this end, three noise reduction methods are summarized based on the intrinsic factors influencing the added noise scale, including privacy amplification, model sparsification, and sensitivity reduction. Furthermore, we discuss the ongoing challenges and propose some future directions where differential privacy can be implemented to obtain a better trade-off between privacy and learning performance.

Introduction

The past few years have witnessed a fast growth of artificial intelligence and machine learning (ML) applications deployed in edge networks, called edge intelligence or computing. In comparison with the traditional cloud computing framework relying on a powerful data center to execute these ML tasks, edge computing physically moves computation from the data center toward the edge nodes of the network where the data are usually generated. This is accomplished on the basis of the rapid evolution of the high-bandwidth mobile networks, such as 5G and 6G, and the increasing computation power of edge nodes such as resource-constrained Internet of Things (IoT) devices. The concept of edge computing has exhibited great potential in significantly reducing the traffic load resulting from data transmission, shortening the response latency for time-critical ML applications, and preserving data privacy in edge networks [1]. On the other hand, as an emerging distributed learning paradigm, federated learning (FL) [2] has gained significant interest in the last decade because of its inherent emphasis on user privacy. In particular, FL enables collaborative ML model training over massively distributed clients under the orchestration of a central server without the need to share the clients' raw private data [2].

Applying FL to the edge networks is appealing as edge computing can be undertaken within edge devices under the orchestration of an edge server (ES), which offers a prospective framework for distributed ML applications. The combination of FL and edge computing, namely edge FL (EFL), has been widely deployed in numerous edge ML scenarios where privacy protection and resource utilization are critical. The main challenges in conventional FL are privacy protection, massively distributed clients, non-i.i.d. and unbalanced data, and limited communication resources [2]. Compared with conventional FL, the EFL system confronts some new challenges due to the characteristics of edge networks, which intimately adhere to the infrastructure of edge networks, including the demand for computational efficiency and light-weight algorithm design, interference among edge devices, low data quality, the need for strong and fine-grained data security, and so on [3]. We remark that the above challenges are unique or severer in edge FL. For example, the challenge of interference among edge devices is caused by the unique characteristics of the edge communication protocols, while low data quality emerges due to the ubiquitous uncertainty in the data acquisition of noisy edge network environments. They not only bring difficulty to the system development but also put emphasis on data security and privacy. Nevertheless, FL fails to meet the security demand as the edge devices' private data may be easily revealed from periodically exchanged messages between the server and edge devices.

Numerous efforts have been devoted to the development of privacy-enhancing techniques for establishing a secure FL system. They can be categorized into cryptography-based approaches and

Yiwei Li and Chong-Yung Chi are with Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan; Shuai Wang (corresponding author) and Tony Q. S. Quek are with Information Systems Technology and Design, Singapore University of Technology and Design, Singapore.

Digital Object Identifier: 10.1109/MNET.001.2200204

0890-8044/22/$25.00 © 2022 IEEE

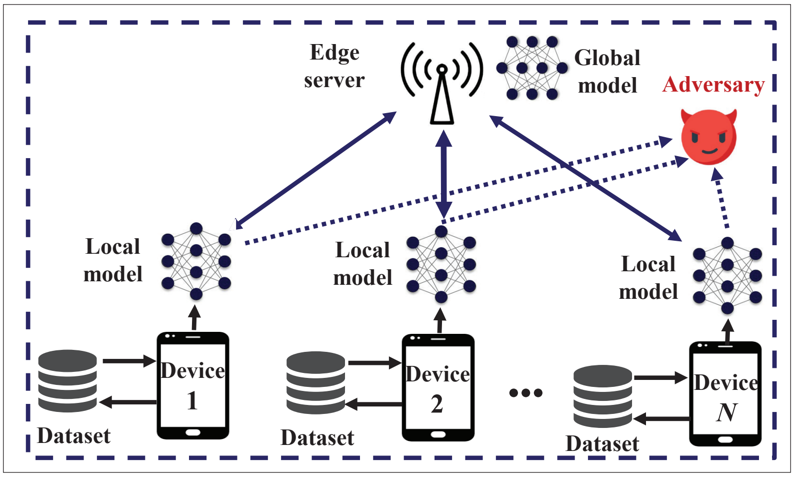

FIGURE 1. The framework of private EFL network exposed to adversaries.

differential privacy (DP) based approaches. While the cryptography-based methods [4], such as secure multiparty computation, homomorphic encryption, and secret sharing, run with a very high privacy guarantee, their success requires complicated encryption protocols, substantial extra communication and computation cost, and a strong assumption of a trustworthy server. Recently, a blockchain-based approach was also deployed in FL without the assumption of a reliable server, but the complicated protocols and high computation power are necessary. The resulting prohibitive overheads particularly for large-scale ML tasks and unrealistic assumptions make it hard for a wide deployment of cryptography-based approaches to the edge FL networks [5]. On the other hand, the use of DP in EFL is rising rapidly, thanks to its complete theoretic guarantees, algorithmic simplicity and negligible system overhead. Recent studies [6] show that specially-designed DP mechanisms can provide FL clients with strong privacy protection, and any adversary, including the dedicated server, is unable to infer sensitive information from the intermediate communications during the training process [6]. Incorporating DP into the EFL framework is also straightforward by simply perturbing the transmitted messages in each communication round with random noise. Therefore, the differentially private EFL frameworks (DP-EFL) being deployed have drawn high attention in both academic research and industry in recent days.

Nevertheless, the DP-EFL framework preserves client privacy at the cost of a training performance degradation because applying perturbation with random noise would adversely affect the convergence of learning algorithms to the desired high-accuracy ML model. Moreover, a larger noise scale means a stronger privacy guarantee but results in lower learning performance. Thus, a trade-off exists between the noise scale and the learning performance. Understanding and balancing the trade-off is a promising direction for realizing desirable DP-EFL frameworks. This naturally raises the following research question.

Question: How to reduce the noise scale in DP-EFL network without sacrificing the privacy guarantee?

Exploring the fundamental principles of the DP-EFL framework is essential to finding possible answers. Its theoretical foundation relies on the definition of DP and *privacy composition* property. DP defines a differentially private randomized mechanism which generates similar outputs for two input datasets with only one different sample. Once it is applied to the learning process, it is extremely hard to tell whether or not a specific sample is used, thus preventing data leakage. The privacy protection level is usually determined by a positive parameter ε, and a smaller ε corresponds to a stronger privacy protection. The *privacy composition* property states that sequentially applying the DP randomized mechanism would yield an accumulated privacy loss. Then, the DP-EFL framework is built by applying a DP randomized mechanism (e.g., using additive artificial noise) to the exchanged messages of each communication round. The *privacy composition* property enables modular design and privacy analysis of the complicated FL process given the total privacy budget $\bar{\varepsilon}$. Thereby, the required privacy protection level and the corresponding noise scale for each round can be readily determined [6].

The noise scale required for guaranteeing DP of the whole EFL network depends on three main factors, including privacy protection level, model dimension and sensitivity, where the term "sensitivity" refers to the magnitude of the largest change measuring the impact of a single data record on the output of a randomized mechanism [6]. Therefore, noise reduction can be progressively achieved by the analysis of these factors together with all the involved building blocks of the DP-EFL framework. Research communities have proposed several promising approaches by focusing on a specific factor. For instance, recent works [7, 8] attempted to use randomly shuffling and data subsampling to maintain privacy protection levels with a smaller noise. Additionally, the sparse vector techniques [9] or dimension selection algorithms [10] were applied to the DP-EFL framework to mitigate the noise effect on model accuracy, especially for ML applications with large models.

In this article, we aim to provide a systematic review of various existing and prospective research problems on the crucial noise-level reduction for the DP-EFL system. After an overview of the DF-EFL framework, we concentrate on the above-mentioned three intrinsic factors on which the power of the additive artificial noise relies, including a review of existing approaches, their strengths and weaknesses, followed by some new future research directions. Finally, we draw some conclusions.

Edge Federated Learning with Differential Privacy

Edge Federated Learning and Threat Model

We start with a brief introduction to the EFL framework. Figure 1 depicts the EFL system consisting of an ES and a total of *N* (which could be very large) distributed devices, and they respectively own non-overlapping and private datasets. The objective of FL is to train a high-quality ML model with the datasets from edge devices under the orchestration of ES without directly assessing the devices' private data. This can be achieved cooperatively by the ES and devices, who follow a computation and aggregation protocol to implement the designed FL algorithm. The ES receives and aggregates local ML model updates computed by the devices, and then updates the global ML model, which is then broadcast to the devices for each communication round. The process continues in a round-by-round manner until convergence to an accurate ML model.

While FL is more secure than traditional distributed ML frameworks owing to no data sharing among devices, data leakage is still possible. The exchanged messages between devices and ES may contain private information of the devices' data, and adversaries could figure out the devices' private data from these messages. In addition to the adversary, the ES and third parties can also acquire the exchanged messages during the whole FL process. Specifically, the ES may be curious about the devices' private data and recover it from the uploaded local model updates. Concurrently, the exchanged messages may be overheard by third parties who could reveal devices' private data using advanced techniques like auxiliary data or cutting-edge attacks.

The DP-EFL Framework

DP provides a strong criterion for privacy preservation together with the inherent advantage of preserving privacy without complicated computation. The purpose of DP-EFL is to build a framework that enables edge devices to preserve their privacy against the above-mentioned threats without sacrificing much accuracy of the learning model. In the EFL system, the DP is implemented by proper use of artificial noise in training process before uploading to the ES. However, we need to determine where to apply the additive noise with suitable power, for example, the local data, objective function, and local gradient/models as depicted in Fig. 2. Moreover, the noise scale and where it is applied will affect the algorithm design as well as the resulting sensitivity.

Adding noise to the local data or the objective function is impractical as it is difficult to determine the required noise scale for guaranteeing DP. In particular, owing to the complex and diverse data types such as images, videos and texts, the sensitivity cannot be calculated if the noise is injected into the local raw data. Likewise, it is hard to determine the sensitivity when the noise is injected into the non-convex objective function, especially for the ML tasks using deep neural networks. Thereby, these two schemes are seldom applied in the EFL framework because of extra performance loss for determining the sensitivity [11].

On the contrary, adding the noise to local gradients/models can be readily applied to the FL system and hence it has been widely adopted in the current DP-EFL framework. There exist quite a few schemes of privacy protection based on this approach in the FL system, such as DP-SGD [12], DP-FedAvg [13] and DP based primal-dual method (DP-PDM) [14], where additive noise is applied to the local gradient, local models and local primal variable or dual variable. More importantly, the impact of the additive noise on the learning performance can be disclosed through an algorithm convergence analysis, and the total privacy loss can be readily tracked during the training as well.

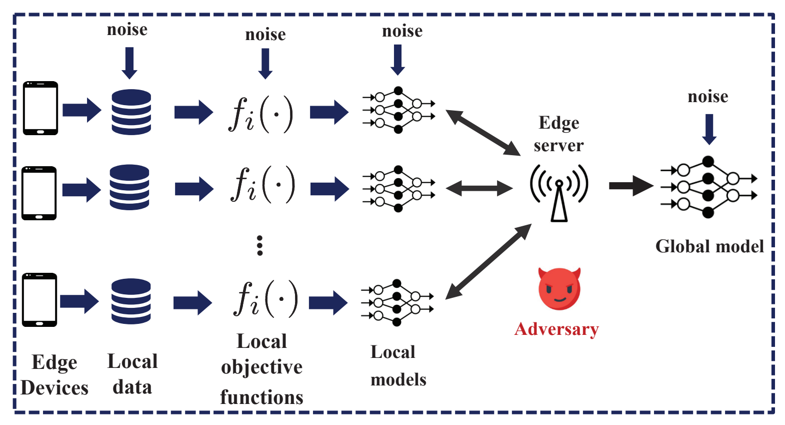

FIGURE 2. An illustration indicating where the artificial noise can be added in the DP-EFL system.

Noise Scale Calculation

Given the total privacy loss $\bar{\varepsilon}$ for a DP-EFL framework after T communication rounds of training, the scale of the additive noise σ for guaranteeing DP in each communication round is proportional to the dimension of the ML model d and sensitivity Δs while it is inversely proportional to the privacy protection level ε µ $f(\bar{\varepsilon}, T)$[6, 12], that is

$$\sigma \propto \mathcal{O}\left(\frac{d \times \Delta s}{f(\bar{\varepsilon}, T)}\right), \qquad (1)$$

where Δs is the sensitivity which relates to data distribution or data processing. The model/data dimension d indicates the number of elements perturbed with random noise. In most existing works, ε is set in the DP-EFL framework rather than the total privacy $\bar{\varepsilon}$ accumulated over T communication rounds for each client. However, this policy may lead to substantial data leakage in the sequel simply because $\bar{\varepsilon}$ (also a function of ε and T) increases with T. Instead, we consider the more practical case that $\bar{\varepsilon}$ is constrained in the DP-EFL system in our work.

Overview of Existing Works on Noise Reduction

Following the above analysis, existing works on the mitigation of the noise effect on the learning performance naturally can be categorized into three directions:

- Privacy amplification: to target the additive noise power reduction by adversely enlarging $f(\bar{\varepsilon}, T)$ under the constraint on the resulting $\bar{\varepsilon}$ after T communication rounds of training.
- Model dimension reduction: to reduce the number d of model dimensions that the additive noise is applied simply because smaller noise power is needed for smaller d, thereby mitigating the adverse effect on the learning performance.
- Sensitivity reduction: to make use of specific DP properties or find where to apply the additive noise in the DP-EFL framework such that the resulting sensitivity Δs can be reduced, thus leading to a smaller noise scale.

We first give an overview of these three kinds of approaches by comparing them with respect to three viewpoints including communication saving, computational reduction and performance loss. All aspects mean a lot to the choice of appropriate approaches in certain applications. Table 1 shows the comparison result, and Fig. 3 depicts

Noise reduction methods	Communication saving	Computation reduction	Performance loss
Data subsampling	no	no	large
Model dimension reduction	yes	yes	small
Sensitivity reduction	no	no	large

TABLE 1. Comparison of various noise reduction methods.

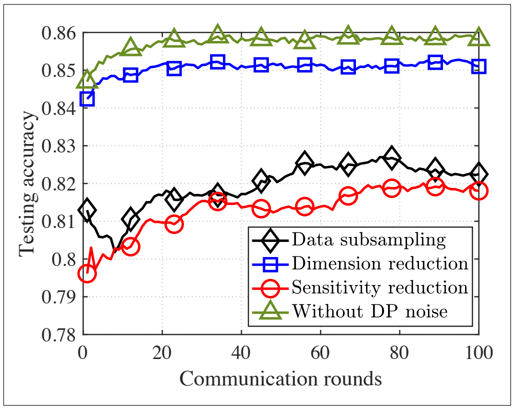

FIGURE 3. Testing accuracy (model learning performance) of FedAvg respectively adopting three typical noise reduction schemes and a baseline without DP noise.

the resulting testing accuracy (model learning performance) on the Adult dataset if they are respectively adopted in FedAvg algorithm while guaranteeing $(1, 10^{-5})$-DP in each communication round. In general, as shown in Fig. 3, the learning performance will be downgraded if each noise reduction scheme is adopted. We can observe from Table 1 and Fig. 3 that the data subsampling scheme can greatly reduce the computation load because only a small portion of local data is used in the local update. Nevertheless, this scheme also significantly downgrades the learning performance due to using the mini-batch gradient descent. The model dimension reduction method not only achieves communication saving but also may have a negligible effect on learning performance with carefully chosen parameters. It is worth mentioning that only one sensitivity reduction scheme is considered in Table 1 and Fig. 3, others may yield better performance. Next, we review these noise reduction schemes in detail.

Noise Reduction by Privacy Amplification

Many recent works on privacy amplification adopt the idea of shuffling [7] or subsampling [8] to achieve a larger $f(\bar{\varepsilon}, T)$ for noise reduction.

Data Shuffling

Data shuffling achieves privacy amplification according to the fact that the anonymity in DP will enhance privacy protection. Specifically, a sufficient number of data or updated models are collected and randomly shuffled at each local step so that any individual data can "hide in the crowd." Thus, the privacy loss for the model can be dramatically lowered as anonymity will greatly increase model uncertainty. The work [7] demonstrated that the privacy loss of the global model could be much lower than only requiring each edge device to guarantee $\bar{\varepsilon}$-DP. In the EFL system, when there exists a trustworthy server that randomly shuffle (e.g., anonymized) the local models collected from local edge devices before uploading, the global model achieves stronger privacy protection than without any shuffling operation. However, it may not be practical to perform model shuffling in real-world applications due to the concern that any trustworthy third-party seldom exists.

Data Subsampling

Applying data subsampling to a given dataset also leads to amplification of the privacy protection level of the model being trained using the subsampled data. Provided that each round guarantees (ε, δ)-DP, the work [8] demonstrated that data subsampling provides stronger privacy protection ε' (i.e., $\varepsilon' < \varepsilon$). In this way, if the total privacy loss is fixed in FL, data subsampling offers a stronger protection level $f(\bar{\varepsilon}, T)$ for each round, thereby leading to smaller noise power. It should be noted that the value of ε' is affected by different data sampling strategies, so a proper data subsampling strategy is important for different applications.

Noise Reduction by Model Sparsification

Owing to the relation of σ and d in Eq. 1, another strategy aims to reduce the noise scale by perturbing a small subset of the elements in exchanged messages. To this end, many existing related works proposed to combine model sparsification and DP, that is, the model parameter is first sparsified and the additive noise is then applied over all the dimensions of the sparsified model. One advantage of this strategy is to improve communication efficiency since the size of transmitted messages per round is significantly reduced.

To be specific, the work [9] proposed a vector sparsification approach that returns a sparse privatized vector with at most $h \leq d$ dimensional values from the original model dimension d. By this way, per-dimension privacy protection level is increased from ε to $d\varepsilon/h$, thus resulting in the sparse model with stronger protection level. To remedy this issue, recent work [10] proposed a two-stage dimension selection framework consisting of a dimension selection (DS) stage and a value perturbation stage, as illustrated in Fig. 4, where the DS stage first builds a top-k (k largest values) dimension set from the local model, from which the h important dimensions are then selected. Then, in the value perturbation stage, the value of the selected dimension is perturbed via the DP algorithms and used to construct a sparse privatized local update. Finally, a sparse privatized local model update is constructed and returned to the server.

Although the strategy of model sparsification is encouraging to reduce the noise scale, it would dramatically downgrade learning performance when the most informative dimensions set is not selected. It is acknowledged that existing works [9, 10] tried to mitigate the performance degradation by selecting the model dimensions with the largest magnitudes. Nevertheless, there may still exist much privacy deficiency because the selected dimensions may not be the most informative for the convergence of FL algorithms.

Noise Reduction by Reducing the Sensitivity

The noise reduction can also be implemented by reducing the sensitivity Δs according to Eq. 1. The sensitivity in the DP mechanism gives

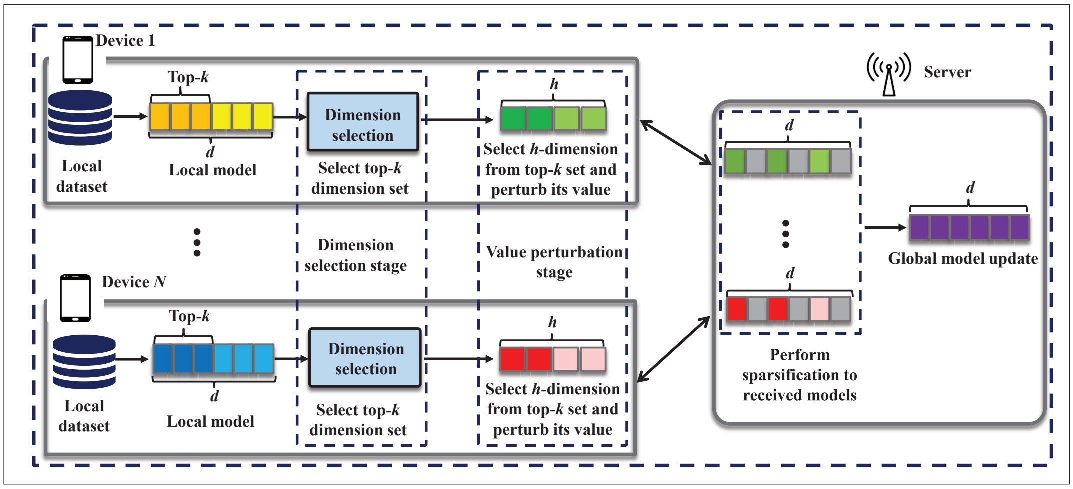
Device 1
Top-k
Local dataset
d
Local model
Dimension selection
Select top-k dimension set
h
Select h-dimension from top-k set and perturb its value
Dimension selection stage
Value perturbation stage
Device N
Top-k
Local dataset
d
Local model
Dimension selection
Select top-k dimension set
h
Select h-dimension from top-k set and perturb its value
Server
d
d
Perform sparsification to received models
d
Global model update

FIGURE 4. The framework of the two-step dimension selection in DP-EFL.

an upper bound on the amount of noise for effective privacy, but it is almost formidable to determine the sensitivity ahead of time in real-world applications. In some cases, the sensitivity is determined by data processing in order that a smaller sensitivity might be found to reduce noise. For instance, DP-SGD [12] adopts the scheme of gradient clipping on the edge devices to determine the sensitivity, hence, we can choose a smaller sensitivity during the model training to mitigate the adverse effect of noise. In addition, the sensitivity Δs may possibly be reduced by utilizing the post-processing property of DP [6]. To be specific, the post-processing property enables any nonprivate data operation on the results of DP outputs without the risk of losing their privacy guarantees, that is, maintaining the privacy protection provided by DP. Owing to the post-processing property, the noise for ensuring DP can be added somewhere with lower sensitivity before uploading to the ES. For example, if the DP outputs may meet some domain constraints, the post-processing is simply the projection of the privacy-preserving outputs onto the feasible region under such constraints, thus lowering the sensitivity.

Challenges and Future Work

As previously presented, most existing noise reduction approaches can be categorized into three classes. Despite their efforts, there are still many difficulties to overcome in order to obtain a desirable trade-off between learning performance and privacy protection. Currently, there is an increasing privacy protection demand of applying the DP-EFL to the IoT, such as online learning tasks for network advertising and positioning-based services, where it is easier for adversaries to acquire and infer private data with more available prior knowledge in such challenging scenarios. In such cases, stronger privacy protection is required but the existing DP-EFL system cannot provide it because the exaggerated privacy protection level for preventing information leakage will dramatically deteriorate the learning performance. Thus, the noise reduction methods are indispensable for DP-EFL. In view of these problems, we put forth several possible research directions to improve or supplement the aforementioned approaches.

Combination of DP and Cryptography

It may be feasible and beneficial to advisably combine DP in EFL and cryptography-based approaches because the strong privacy protection guarantee of the latter may potentially help reduce the noise scale of the former; the former can prevent the differential attack and the latter can prevent the backdoor attack or model inversion attack. Therefore, their combination may enable significant noise reduction with a stronger privacy protection level if the extra computational overhead is acceptable. For instance, combining the DP and homomorphic encryption (HE) can be applied to DP-EFL as illustrated in Fig. 5. In this case, HE is applied on the local client side to protect the local data privacy, and DP is implemented on the cloud/edge server to prevent the global model's information leakage. In this scenario, the trained global model can be used by other third-party for further data analysis, which is applicable for privacy protection of sensitive health data lying on isolated hospitals. Nevertheless, the combined privacy protection cryptography-based consumes extra computation resources of the local clients. Thus, a careful design to balance the computation and privacy protection is significant.

Mixed Privacy Protection

It allows the DP-EFL system to utilize some data available with no privacy concerns to train the model, such as synthetic datasets or datasets available for public use. In this way, less local sensitive data will be used for training, thus reducing the data privacy leakage. In other words, this scheme provides stronger privacy protection within the same communication round and constrained total privacy loss $\bar{\varepsilon}$ by using a smaller dataset in one communication round. The use of large amounts of public data for pre-training has recently enabled the learning models to achieve DP for the target task along with near state-of-the-art performance. It might be helpful to use auxiliary data to pre-train the model on the ES side followed by finely tuning the model parameters using the DP-EFL to mitigate the effect of the additive noise during the FL training. However, the associated theoretical analyses are still yet to be explored, such as convergence analysis, privacy analysis of a completed ML task and the trade-off between the privacy protection and the pre-train model.

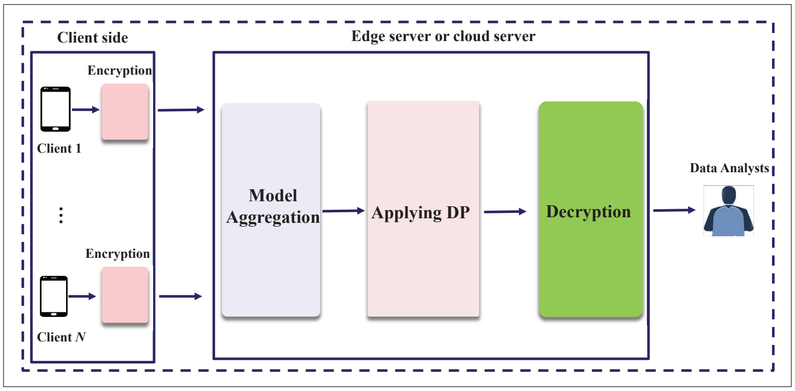

FIGURE 5. The combination of DP and HE.

Selection of Informative Models

The key to noise reduction by model sparsification is the selection of the most representative top-k dimension models, which is non-trivial in a real-world deployment. Most existing model reduction algorithms with DP are based on the Euclidean distance (absolute value) between the selected model and the target model, and thus their drawbacks include the inconsistency between the selected model and the value to the learning performance, and the learning performance backfires due to the selection of misrepresentative components of the model. Thus, the selection of the informative components of the model can be implemented by other approaches with better learning performance, such as selecting the informative components by modular scores, which have been studied for subset selection in FL recently, e.g., a utility score for each sample or client often measured by the training loss [15]. However, the selection of top-k informative components from the model of large size is also non-trivial, besides the determination of k (Fig. 4).

Conclusion

In view of various pervasive edge computing applications in 5G and 6G, where privacy protection is not only needed but also must be guaranteed, we have presented a comprehensive review from the background, central development issues and existing DP-based approaches for edge FL, that are naturally needed by the prospective DP-EFL framework that we focused on. Specifically, we concentrated on the noise reduction issue in DP-EFL and classified the existing works into three categories. Furthermore, through insightful perspective analyses and discussions on the impact of the noise scale on the learning performance of DP-EFL systems, we finally presented challenges, promising solutions and future researches yet to be investigated for the advances of the DP-EFL.

Acknowledgments

This work is supported by the Ministry of Science and Technology, Taiwan, under Grants MOST 111-2221-E-007-035-MY2 and MOST 110-2221-E-007-031. It is also supported in part by the National Research Foundation, Singapore and Infocomm Media Development Authority under its Future Communications Research & Development Programme.

References

[1] Q. Xia *et al.*, "A Survey of Federated Learning for Edge Computing: Research Problems and Solutions," *High-Confidence Computing*, vol. 1, no. 1, 2021, pp. 1–13.
[2] P. Kairouz *et al.*, "Advances and Open Problems in Federated Learning," *Foundations and Trends in Machine Learning*, vol. 14, no. 1–2, 2021, pp. 1–210.
[3] W. Y. B. Lim *et al.*, "Federated Learning in Mobile Edge Networks: A Comprehensive Survey," *IEEE Commun. Surveys & Tutorials*, vol. 22, no. 3, 2020, pp. 2031–63.
[4] C. Ma *et al.*, "On Safeguarding Privacy and Security in the Framework of Federated Learning," *IEEE Network*, vol. 34, no. 4, 2020, pp. 242–48.
[5] P. Vepakomma *et al.*, "No Peek: A Survey of Private Distributed Deep Learning," arXiv preprint arXiv:1812.03288, 2018.
[6] C. Dwork *et al.*, "The Algorithmic Foundations of Differential Privacy," Foundations and Trends in Theoretical Computer *Science*, vol. 9, no. 3-4, 2014, pp. 211–407.
[7] Ú. Erlingsson *et al.*, "Amplification by Shuffling: From Local to Central Differential Privacy via Anonymity," *Proc. Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SIAM, 2019, pp. 2468–79.
[8] B. Balle, G. Barthe, and M. Gaboardi, "Privacy Amplification by Subsampling: Tight Analyses via Couplings and Divergences," *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2018, pp. 6277–87.
[9] N. Wang *et al.*, "Collecting and Analyzing Multidimensional Data With Local Differential Privacy," *Proc. IEEE 35th Int'l. Conf. Data Engineering (ICDE)*, 2019, pp. 638–49.
[10] X. Jiang, X. Zhou, and J. Grossklags, "SignDS-FL: Local Differentially Private Federated Learning With Sign-Based Dimension Selection," *ACM Trans. Intelligent Systems and Technology (TIST)*, 2022, pp. 1–22.
[11] X. Li *et al.*, "When Deep Learning Meets Differential Privacy: Privacy, Security, and More," *IEEE Network*, vol. 35, no. 6, 2021, pp. 148–55.
[12] M. Abadi *et al.*, "Deep Learning With Differential Privacy," *Proc. ACM SIGSAC Conf. Computer and Commun. Security*, 2016, pp. 308–18.
[13] Y. Li, T.-H. Chang, and C.-Y. Chi, "Secure Federated Averaging Algorithm With Differential Privacy," *Proc. IEEE Int'l. Wksp. Machine Learning for Signal Processing (MLSP)*, 2020, pp. 1–6.
[14] Z. Huang *et al.*, "DP-ADMM: ADMM-Based Distributed Learning With Differential Privacy," *IEEE Trans. Information Forensics and Security*, vol. 15, 2019, pp. 1002–12.
[15] R. Balakrishnan *et al.*, "Diverse Client Selection for Federated Learning via Submodular Maximization," *Proc. Int'l. Conf. Learning Representations (ICLR)*, 2022, pp. 1–18.

Biographies

Yiwei Li is currently pursuing his Ph.D. degree at National Tsing Hua University, Taiwan. His research areas of interest are security and privacy protection in federated learning, optimization in machine learning with applications spanning the Internet of Things, wireless communications network and intelligent mobile edge systems.

Shuai Wang is currently a Postdoctoral Research Fellow in the ISTD Pillar of SUTD, Singapore. Before that, he received the Ph.D. degree in the School of Science and Engineering of CUHKSZ, China. His research interests include optimization algorithms for signal processing and machine learning, federated learning, security and privacy protection.

Chong-Yung Chi is a Professor, Institute of Communications Engineering, and Department of Electrical Engineering, National Tsing Hua University, Taiwan. His research interests are signal processing in wireless communications and networking, convex analysis and optimization, federated learning, hyperspectral image analysis, and graph signal processing.

Tony Q. S. Quek received the B.E. and M.E. degrees in electrical and electronics engineering from the Tokyo Institute of Technology in 1998 and 2000, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology in 2008. Currently, he is the Cheng Tsang Man Chair Professor and Head of ISTD Pillar with Singapore University of Technology and Design, as well as the Director of the Future Communications R&D Programme.