Review on distributed learning for IoT-Edge-Cloud continuum

Audris Arzovs[1*] and Krisjanis Nesenbergs[1]
[1]*Institute of Electronics and Computer Science, Riga, Latvia*

*Contact: {Audris.Arzovs}@edi.lv

***Abstract*—In this abstract we present our ongoing work in literature review of distributed learning in Internet of Things - Edge - Cloud continuum, with the goal of defining the niches where we can develop tools beyond state of art, that would provide more seamless and efficient development of systems on this continuum, that natively provide benefits such as privacy, security, energy efficiency and ease of deployment.**

***Keywords*—IoT, Edge, Cloud, distributed learning, federated learning, split learning, transfer learning, differential privacy, homomorphic encryption.**

I. Introduction

The power of Internet of Things (IoT) devices has increased to the level of being able to run machine learning (ML) models both, for inference and training [1], [2], [3], [4], [5], which previously was only the domain of powerful Cloud and Edge devices. Due to this, opportunities have arisen for integration and orchestration of artificial intelligence (AI) based reasoning across the whole IoT-Edge-Cloud Continuum (IECC). The main benefits of IECC are the opportunities for local data processing and actuation when possible in order to achieve better response times and energy efficiency while offloading heavier tasks to Edge and Cloud devices if required due to power constraints or needs to integrate multiple data streams in the models. Unfortunately IECC systems also have multiple unsolved problems, related to efficient distribution of the learning algorithms and data across the whole continuum, while preserving as much privacy and security as possible.

This problem can be mitigated with distributed learning approaches, such as differential privacy and federated learning, which have already been thoroughly reviewed for the more powerful Edge-Cloud continuum [6], [7], [8], but there is still not a full understanding of best practices in integrating distributed learning across the whole IECC.

This ongoing work aims to review and understand the potential approaches and unsolved problems in applying distributed learning approaches to IECC, and specifically with emphasis on IoT devices, as those have only recently become capable of running powerful ML algorithms.

Below we introduce our current findings on distributed learning for IECC.

This research is funded by the Latvian Council of Science, project "Smart Materials, Photonics, Technologies and Engineering Ecosystem" project No VPP-EM-FOTONIKA-2022/1-0001.

II. Literature review

In order to develop new approaches for sustaining the privacy and security of ML models and the inter-communication of the model results between more powerful devices and IoT devices we start by looking at benefits of distributed learning, that can allow improved security and privacy of ML models and their results for example by using Differential Privacy and/or Homomorphic Encryption. There are numerous Distributed Learning approaches and implementation techniques [9].

Federated Learning (FL) is the approach in which many ML models are trained simultaneously and the results from these models are aggregated at a central node or distributed between the network nodes in a decentralized setting. This decreases the necessary communication between the IoT devices and a server because full raw data is not being sent to edge or cloud servers, and model results are sent instead. This already increases the privacy of the approach by default although there still are methods with which an adversary can retrieve the information from the model weights that is similar to the raw input data [10], [11], [12]. A Federated learning network consists of edge devices e.g. smart phones on which the local ML model training will take place. The results of these local models will be sent to a central aggregator which combines all received results into one ML model and distributes this model to all network nodes for another training round. The training process can be synchronous and asynchronous because some nodes may become a bottleneck and the training needs to be continued nevertheless. The same process may take place without a central aggregator [11], [13].

In Split learning (SL) the ML models are split in two or more parts and distributed to many devices. This network of devices train one model by communicating the results or rather the smashed data to the next device in line to continue the inference or for example backpropagation. In its essence this approach is completely sequential because each part is dependent on the previous one and the training cannot take place in parallel as it was the case in the FL setting. Which is why in some cases FL shows better results than SL. Although in this setting we have increased the security of our model by splitting the model in many parts because an adversary has less chance to do a model inversion attack when the model parts aren't capable of predicting features. The downside of

this approach is that the necessary communication bandwidth may be increased by each model part sending the results to the next one [10], [14], [15].

Transfer learning is a popular research direction in low powered devices. Where we take out parts from already pre-trained networks and implement them in smaller networks e.g. taking a feature extractor from a CNN, keeping it static and adding some smaller layers on top of it for possible training and put this network on an IoT device. This approach is suitable for less powerfull devices because the main training operations for the main feature extraction have already been done beforehand and only smaller training operations are being done afterwards. Taking granted that Edge devices usually are powerfull enough to be part of FL and SL networks this approach becomes too specific for the low power devices setting [2], [4], [5], [9].

Some of the mentioned drawbacks of these methods can be improved by combining them together. For example if we combine Federated with Split learning, we get an approach which has a split model but with parallel operation. On top of that, if we combine Transfer with Federated learning we allow Transfer learning to work in a federated setting [10], [14].

Extra security and privacy measures need to be incorporated because none of the mentioned approaches provide complete protection against privacy leakage and a rigid security. Differential privacy is one of the most popular privacy protection mechanisms. This method adds noise to the transferable dataset in a way that it hides the influence of one specific record to make it almost impossible for the adversary to retrieve information about certain records by keeping the characteristics of the whole dataset intact. When an adversary queries differentially private datasets the retrieved answers will be close to identical when the specific record is or isn't part of the dataset. Other privacy methods such as k-anonymity have been proven to be too prone to correlation attacks because it doesn't protect against correlations between existing datasets but only takes out directly identifying information [11], [16], [17], [18].

By default each approach works in plaintext fashion. An adversary has no problem to obtain the information that the networks are exchanging. Secure multi party computation and more specifically Homomorphic encryption helps to aleviate this problem by allowing us to work with encrypted information just as it was plaintext [11].

Just as in a blockchain network also here there is a problem of malicious nodes infecting the network. Robust aggregation schemes have been developed to analyse the node activity and their data to predict whether the node is trustworthy and should be allowed to influence the training results and should be allowed to take part in the training rounds.

The computation time costs that these security methods bring makes us worry about implementing them in real-time system scenarious [19].

Balancing between privacy and ML model accuracy while having small communication latency and fast data processing is still a problem to be researched further. Each implementation requires consideration of these parameters either to reach the target specification or search for a compromise [18] [20] [19] [11].

REFERENCES

[1] N. Llisterri Giménez, M. Monfort Grau, R. Pueyo Centelles, and F. Freitag, "On-device training of machine learning models on microcontrollers with federated learning," *Electronics*, vol. 11, no. 4, p. 573, 2022.

[2] K. Kopparapu and E. Lin, "Tinyfedtl: Federated transfer learning on tiny devices," *arXiv preprint arXiv:2110.01107*, 2021.

[3] S. S. Saha, S. S. Sandha, and M. Srivastava, "Machine learning for microcontroller-class hardware-a review," *IEEE Sensors Journal*, 2022.

[4] J. Lin, L. Zhu, W.-M. Chen, W.-C. Wang, C. Gan, and S. Han, "On-device training under 256kb memory," *arXiv preprint arXiv:2206.15472*, 2022.

[5] H. Cai, C. Gan, L. Zhu, and S. Han, "Tinytl: Reduce memory, not parameters for efficient on-device learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 11285–11297, 2020.

[6] H. Wu, Z. Zhang, C. Guan, K. Wolter, and M. Xu, "Collaborate edge and cloud computing with distributed deep learning for smart city internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8099–8110, 2020.

[7] R. Kozik, M. Choraś, M. Ficco, and F. Palmieri, "A scalable distributed machine learning approach for attack detection in edge computing environments," *Journal of Parallel and Distributed Computing*, vol. 119, pp. 18–26, 2018.

[8] T. L. Duc, R. G. Leiva, P. Casari, and P.-O. Östberg, "Machine learning methods for reliable resource provisioning in edge-cloud computing: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 5, pp. 1–39, 2019.

[9] W. Li, H. Hacid, E. Almazrouei, and M. Debbah, "A review and a taxonomy of edge machine learning: Requirements, paradigms, and techniques," *arXiv preprint arXiv:2302.08571*, 2023.

[10] Q. Duan, S. Hu, R. Deng, and Z. Lu, "Combined federated and split learning in edge computing for ubiquitous intelligence in internet of things: State-of-the-art and future directions," *Sensors*, vol. 22, no. 16, p. 5983, 2022.

[11] C. Briggs, Z. Fan, and P. Andras, "A review of privacy-preserving federated learning for the internet-of-things," *Federated Learning Systems: Towards Next-Generation AI*, pp. 21–50, 2021.

[12] D. Li, J. Lai, R. Wang, X. Li, P. Vijayakumar, B. B. Gupta, and W. Alhalabi, "Ubiquitous intelligent federated learning privacy-preserving scheme under edge computing," *Future Generation Computer Systems*, vol. 144, pp. 205–218, 2023.

[13] D. J. Beutel, T. Topal, A. Mathur, X. Qiu, J. Fernandez-Marques, Y. Gao, L. Sani, K. H. Li, T. Parcollet, P. P. B. de Gusmão, *et al.*, "Flower: A friendly federated learning research framework," *arXiv preprint arXiv:2007.14390*, 2020.

[14] C. Thapa, P. C. M. Arachchige, S. Camtepe, and L. Sun, "Splitfed: When federated learning meets split learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 8485–8493, 2022.

[15] J. Liu and X. Lyu, "Clustering label inference attack against practical split learning," *arXiv preprint arXiv:2203.05222*, 2022.

[16] N. Li, W. H. Qardaji, and D. Su, "Provably private data anonymization: Or, k-anonymity meets differential privacy," *CoRR, abs/1101.2604*, vol. 49, p. 55, 2011.

[17] C. Dwork, "Differential privacy," in *Automata, Languages and Programming* (M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, eds.), (Berlin, Heidelberg), pp. 1–12, Springer Berlin Heidelberg, 2006.

[18] E. Bagdasaryan, O. Poursaeed, and V. Shmatikov, "Differential privacy has disparate impact on model accuracy," *Advances in neural information processing systems*, vol. 32, 2019.

[19] P. Ruzafa-Alcázar, P. Fernández-Saura, E. Mármol-Campos, A. González-Vidal, J. L. Hernández-Ramos, J. Bernal-Bernabe, and A. F. Skarmeta, "Intrusion detection based on privacy-preserving federated learning for the industrial iot," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 1145–1154, 2021.

[20] X. Shen, Y. Liu, and Z. Zhang, "Performance-enhanced federated learning with differential privacy for internet of things," *IEEE Internet of Things Journal*, vol. 9, no. 23, pp. 24079–24094, 2022.