Vertical Federated Learning for Privacy-Preserving ML Model Development in Partially Disaggregated Networks

Nazila Hashemi*, Pooyan Safari*, Behnam Shariati*, and Johannes Karl Fischer

Fraunhofer Institute for Telecommunications Heinrich Hertz Institute, Einsteinufer 37, 10587 Berlin, Germany, (email of the corresponding author: behnam.shariati@hhi.fraunhofer.de)

Abstract: *We present a novel framework that enables vendors and operators, with partial access to operational and monitoring features of a service, to collaboratively develop a ML-assisted solution without revealing any business-critical raw data to each other. We validate our proposal for a QoT estimation use-case.*

Introduction

Optical network automation and disaggregation are significantly transfroming the telecom ecosystem. On the one hand, network automation requires an unprecedented level of collaboration among various Network Elements (NE), from hardware to software, which in turn necessitates sharing various sets of data (e.g., telemetry metrics, device configurations, etc.) between NEs[1],[2]. On the other hand, network disaggregation decomposes the conventional black-box operational model of telecom infrastructure, which used to be proprietarily provided/operated by a single vendor, and enables multi-party ecosystems in which NEs from multiple vendors co-exist and interoperate to deliver an end-to-end network service[3].

While network operators and datacom players are pushing for network disaggregation, many NE vendors are reluctant to fully support the idea, as, among other reasons, it requires from them to share their device specific data, which are business critical, with others. Therefore, the prerequisite of data sharing for network automation and the reluctancy of vendors to share their data with each other are real showstoppers for the realization and eventually the automation of disaggregated networks. This is a challenge for which vendors, in favour of network disaggregation, are seeking accurate and reliable solutions[4].

As one of the main enablers of network automation, Machine Learning (ML) based solutions are attracting the highest attention. However, their development predimonantly relies on availability of data. It is crucial to devise a reliable solution for the development of ML models that protects the privacy of the data owners (i.e., vendors and operators) and prevent the privacy of their business-critical data to be compromised. In[5],[6], we have proposed and demonstrated a Distributed Learning Framework (DLFi) that enables the development of ML models in multi-domain scenarios in which each Domain Manager (DM) has access to the full feature set of each data instance within its domain. We implemented a category of Distributed Learning (DL) called Horizontal Federated Learning (HFL), in which the datasets of different parties have the same features but they differ in data instances[7]. HFL is perfectly compatible with multi-domain networks, as each proprietary DM has access to the entire data of the NEs that run an end-to-end intra-domain lightpath. However, this assumption is not valid for disaggregated networks as the chain of NEs creating an end-to-end lightpath may belong to multiple vendors [3],[4].

In this work, we propose a Vertical Federated Learning (VFL) solution for collaborative ML model development in disaggregated networks, in which multiple parties (e.g. vendors or an operator) hold different features of the same data instances[7]. We apply our proposal to develop a ML-assisted Quality of Transmission (QoT) classifier in partially disaggregated networks (see Fig 1) in which the transceivers (TRx) belong to a vendor different than the one of the Open Line System (OLS). We consider the scenarios in which the TRx vendor or the network operator are the provider of the VFL service, considering the ownership level of the employed features. In addition, we consider a scenario in which a third party (e.g. a software vendor different than the ones operating the data plane devices) that does not own any data develops the QoT classifier as

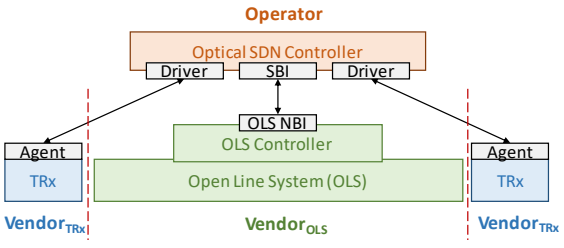

Fig 1. Multi-vendor partially disaggregated network

978-1-6654-3868-1/21/$31.00 ©2021 IEEE

* These authors have equally contributed to this work.

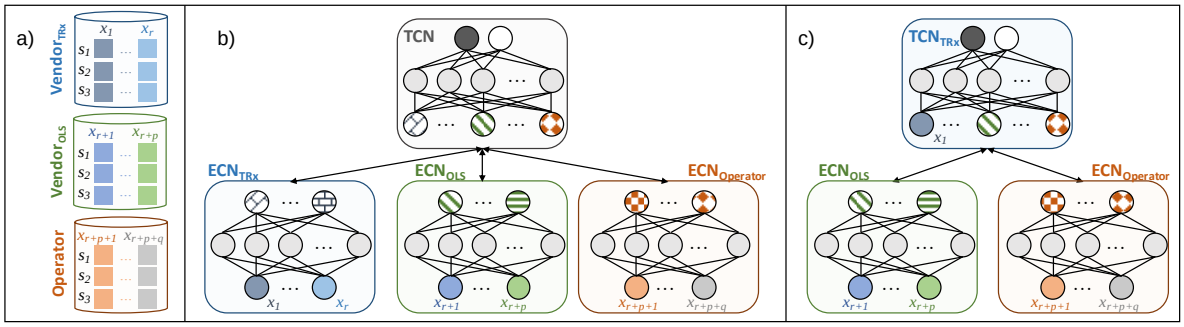

Fig 2. a) The feature set of every data instance S_i partially owned by multiple vendors or the operator, b) our VFL architecture considering a third-party VNF provider as TCN, and c) our VFL architecutre considering one of the data owners as TCN.

Algorithm 1: VFL algorithm. A set of M ECNs, each worker m maintains a feature set $x_{i:j}$ for all the data instances s_n while $n \in \{1, \dots, N\}$. The corresponding label y_n is only stored on the TCN. There are M different local models trained on ECNs with parameters $\theta_1, \dots, \theta_M$. There is a global model located at the TCN with the parameters θ_0. E is the number of training epochs with learning rates η_m, η_0 for local and global models, respectively. $h_{n,m}^k$is the embedding vector corresponding to the data instance n extracted using local model $\theta_m$on ECN m at update k.

01: **for** each **epoch** in E **while** not converged:
02: send config (e.g., batch size) to all ECNs
03: **for** each minibatch b in $\mathcal{B}$ at update k:
04: **do in parallel** on each ECN m:
05: extract $h_{n,m}^k$ using local model θ_m^k
06: send embedding vectors to TCN
07: **when** all embeddings received on TCN **do**:
08: concatenate vectors $h_n^k = \left[h_{n,1}^k, \cdots, h_{n,M}^k\right]$
09: compute global model output $\hat{y}_n = \theta_0^k(h_n^k)$
10: $\mathcal{L}(\theta_0^k) = \frac{1}{|b|}\sum_{n=1}^{|b|} l\left(\theta_0^k; \hat{y}_n, y_n\right)$
11: $g_0^k = \nabla_{\theta_0}\mathcal{L}(\theta_0^k)$
12: $\theta_0^{k+1} = \theta_0^k - \eta_0 g_0^k$
13: gradients w.r.t. embedding $g_m^k = \nabla_{h_{n,m}}\mathcal{L}(\theta_0^k)$
14: send g_m^k to each ECN m
15: **do in parallel** on each ECN m:
16: $g_{\theta_m}^k = \nabla_{\theta_m} h_{n,m}^k g_m^k$
17: $\theta_m^{k+1} = \theta_m^k - \eta_m g_{\theta_m}^k$

a Virtualized Network Function (VNF). Our solution performs similarly good as the Centralized Learning (CL) baseline, while offering a high level of privacy protection for the involved vendors and operators.

The Proposed Vertical Federated Learning

As summarized in Algorithm 1 for the VFL, we consider $M = \{1, \dots, M\}$ remote geo-distributed Edge Contributor Nodes (ECNs). Each ECN possesses the corresponding training data sets and local models. There is also a Training Coordinator Node (TCN) that manages the whole training process and unlike for HFL[5] it plays an additional role in the model training via learning a global model which outputs the class posteriors[8]. There are two different architectures considered in this paper. In the first architecture as shown in Fig 2b the provisioning of the training data is carried out through only employing ECNs. While in the second architecture a portion of the raw data (i.e., s_i data points as represented in Fig 2a) is also accessible on the TCN, see Fig 2c. The data located on each ECN m comprises N data instances s_i and a set of non-overlapping features (see Fig. 2a). It is assumed that the training labels y_n are only available on the TCN. Inspired by the work in[8], in order to preserve the data and model privacy, each ECN m learns a local model parameterized by θ_m. This local model maps the input data into a vector which is referred to as embedding h_m. Therefore for each data instance there is an embedding vector produced on each ECN. All the embeddings are sent to the TCN to be concatenated as the input of the global model. In the second architecture the raw data on the TCN is also included in this concatenation. The obtained vector is fed into the global model to output the estimated class posteriors. Given the ground truth labels y_n during the training, the loss function $\mathcal{L}$ for a mini-batch of size $|b|$ is calculated as line (10) of the Algorithm 1. In order to minimize this loss function the gradients g_0^k are calculated w.r.t. the global model parameters θ_0. The global model parameters will be updated with the learning rate η_0 according to line (12). The gradients w.r.t. the embeddings $h_{n,m}$ at update k are calculated in line (13). These gradients are sent to their corresponding ECN destinations in order to compute the gradients w.r.t the local model parameters θ_m according to line (16). The local model parameters are updated using the equation in (17) with learning rate η_m. This is continued to reach the optimum performance metrics such as validation accuracy.

Formulation of the QoT Estimation Use-case

In order to validate the performance of the proposed VFL approach, we consider a QoT classification problem in a partially disaggregated scenario in which the TRxs and the NEs of the OLS, belong to two different vendors [3],[9]. For the QoT classification use-case, we use the publicly available dataset 01[9] generated based on the networking scenario described in [10]. In this work, we perform a lightpath-based QoT classification

Table 1. The considered scenarios for our study are detailed here. The results are obtained in a virtual test-bed comprising four connected Virtual Machines (VM). Each one of the VMs, depending on the scenario, hosts a partial feature set. Each scenario is defined based on the number of contributing ECNs and the VFL architecture presented in Fig 2. Scenarios 1 and 4 are based on the architecture presented in Fig 2b, while scenarios 2 and 3 implement the one presented in Fig 2c. The second row corresponding to each scenario presents the model architecture, which is a single-layer feed-forward neural network, in terms of number of inputs, number of neurons of the hidden layer, and the size of the embedding layer. Scenario 4 excludes the features owned by $\text{vendor}_{\text{OLS}}$ and uses only the data provided by $\text{vendor}_{\text{TRx}}$ and operator. The CL option is presented as a baseline.

	VM 1	VM 2	VM 3	VM 4	Training	Validation	Test
Scenario 1	ECN_{TRx}	ECN_{OLS}	$\text{ECN}_{\text{Party3}}$	TCN	96.30 %	97.81 %	97.66 %
	3, 32, 4	20, 32, 4	8, 32, 4	12, 512, 2			
Scenario 2	×	ECN_{OLS}	$\text{ECN}_{\text{Party3}}$	TCN_{TRx}	96.08 %	97.23 %	97.03 %
	×	20, 32, 4	8, 32, 4	11, 512, 2			
Scenario 3	ECN_{TRx}	ECN_{OLS}	×	$\text{TCN}_{\text{Party3}}$	96.39 %	97.38 %	97.07 %
	3, 32, 4	20, 32, 4	×	16, 512, 2			
Scenario 4	ECN_{TRx}	×	$\text{ECN}_{\text{Party3}}$	TCN	95.97 %	97.22 %	97.31 %
	3, 32, 4	×	8, 32, 4	8, 512, 2			
CL	31, 512, 2	-	-	-	98.01 %	98.22 %	98.33 %

task, which aims at predicting the QoT metric of a single Lightpath Under Test (LUT).

Considering the structure of the lightpath based version of dataset 01, we define three feature groups: 1) TRx features, which include *freq*, *mod_order*, *lp_linerate*, 2) topology features, which include *path_len*, *avg_link_len*, *min_link_len*, *max_link_len*, *num_links*, *num_spans*, *src_degree*, *dst_degree*, and 3) network-status features, which includes the remaining 20 features of the dataset. Moreover, as illustrated in Fig 2a, we assume that feature group 1, group 2, and group 3 belong to the TRx vendor (**$\text{Vendor}_{\text{TRx}}$**), network operator, and OLS vendor (**$\text{Vendor}_{\text{OLS}}$**), respectively. According to the data ownership principles and the VFL architecture of DLFi, these three feature groups can contribute to the development of the ML-based QoT classifier by incorporating three (or two) ECNs. The considered scenarios are detailed in Table 1. Scenarios 1 and 4 assume that a 3rd party, which is neither the TRx vendor nor the operator, develops the QoT classifier as a standalone VNF, similar to the one presented in [11]. It can also represent the option where any of the vendors or the operator is the provider of the VNF. However, scenario 2 and 3 strictly assume that the TRx vendor and the operator, are the provider of the VNF, respectively.

Results and Concluding Remarks

To evaluate the performance of the scenarios in Table 1, we randomly selected a class-balance subset of 100,000 data instances from the lightpath-based version of dataset 01[9], out of which 70%, 20%, and 10%, are considered for training, validation, and test subsets, respectively. We report accuracy, defined as the number of correctly classified data instances over the total number of them. We also report the total traffic, which measures the amount of traffic exchange between the involved ECNs and the TCN of DLFi during training. As presented in Table 1, different model dimensions are adopted according to the scenario at hand. The Adam

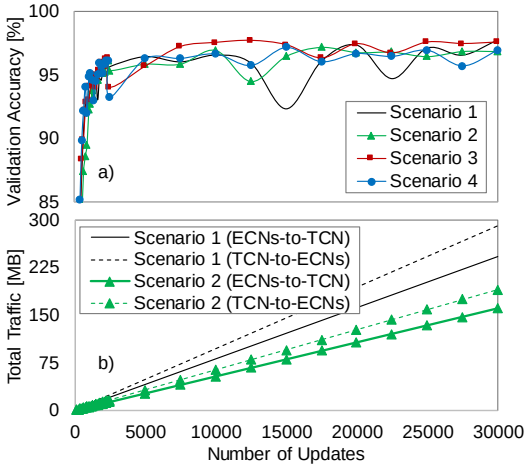

Fig 3. a) Validation accuracy and b) the evolution of the total traffic exchange between ECNs and TCN in terms of the number of updates. The traffic values for scenario 3 and 4 are identical to scenario 2, as the difference in traffic exchage is due to the different number of contributing ECNs.

optimizer with a learning rate of 0.001 is used with default hyperparameters as in[12] for the training of all the neural network models. Note that the learning rates η_0, η_m for all the local and global models are assumed to be equal. The results reported in Table 1 and Fig 3 are chosen according to 5 iterations of early stopping. Our VFL-based QoT classifier model development delivers models with similar accuracy as the CL baseline, while it protects the privacy of business-critical data of different parties. As shown in Fig 3b, the total traffic exchange among ECNs and TCN after 30,000 updates of the training is just around 500 MB.

In this paper, we proposed and verified a VFL algorithm that enables accurate ML model development, in a privacy-preserving fashion, for partially disaggregated networks. Our solution can play a key role in secure exploitation of business-critical data of vendors for automation of partially disaggregated networks.

Acknowledgement: The authors would like to thank Geronimo Bergk for providing the dataset. This work has received funding from BMBF through AI-NET PROTECT (KIS8CEL010).

References

[1] ETSI White Paper No. #40, “Autonomous networks, supporting tomorrow’s ICT business,” 1st edition, Oct 2020.

[2] D. Rafique, L. Velasco, “Machine learning for network automation: overview, architecture, and applications [Invited Tutorial],” in JOCN, vol. 10., no. 10., D126-D143, Oct 2018.

[3] E. Riccardi, P. Gunning, O. G. Dios, M. Quagliotti, V Lopez, and A. Lord, “An operator view on the introduction of white boxes into optical networks,” in JLT, vol. 36, bo. 15, pp. 3062-3027, Aug 2018.

[4] K. Kaeval, T. Fehenberger, J. Zou, S. L. Jansen, K. Grobe, H. Griesser, J.-P. Elbers, M. Tikas, and G. Jervan, “ QoT assessment of the optical spectrum as a service in disaggregated network scenarios,” JOCN, vol. 13, no. 10, pp. E1-E12, Oct 2021.

[5] B. Shariati, P. Safari, A. Mitrovska, N. Hashemi, J. K. Ficher, and R. Freund, “Demonstration of federated learning over edge-computing enabled metro optical networks,” in Proc. ECOC, Brussels, Belgium, Dec 2020.

[6] P. Safari, B. Shariati, and J. K. Fischer, “Privacy-preserving distributed learning framework for 6G telecom ecosystems,” arXiv preprint arXiv:2008.07225, 2020.

[7] Q. Yang, Y. Liu, T. Chen, and Y. Tong, ““Federated machine learning: concept and applications,” ACM Trans. Intell. Syst. Technol., Vol. 10, No. 2, Article 12. February 2019.

[8] T. Chen, X. Jin, Y. Sun, W. Yin, “VAFL: a method of vertical asynchronous federated learning”, ICML Workshop on Federated Learning for User Privacy and Data Confidentiality, July, 2020.

[9] G. Bergk, B. Shariati, P. Safari, and J. K. Fischer, “QoT Dataset Collection, ” [online – accessed May 2021] https://www.hhi.fraunhofer.de/networkdata

[10] P. Safari, B. Shariati, G. Bergk, and J. K. Fischer, ”Deep convolutional neural network for network-wide QoT estimation,” in Proc. OFC, 2021, Paper Th4J.3.

[11] B. Shariati, P. Safari, G. Bergk, F. I. Oertel, and J. K. Fischer, “Inter-operator machine learning model trading over Acumos AI federated marketplace,” in Proc. OFC, 2021, Paper M2B.7.

[12] D.P. Kingma, J. Ba, “Adam: a Method for Stochastic Optimization”, International Conference on Learning Representations (ICLR), 2015.