INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Federated Learning Toolkit with Voice-based User Verification Demo
Submitted to Interspeech

Prathamesh Mandke[1], *Rachel Oberst*[1], *Matthias Reisser*[1], *Avijit Chakraborty*[1], *Christos Louizos*[1], *Joseph Soriaga*[1], *Daniel Madrigal*[2], *Andre Manoel*[2], *Nalin Singal*[2], *Jeff Omhover*[2], *Robert Sim*[2]

[1]Qualcomm AI Research*, [2]Microsoft Corporation

{pmandke, roberst, mreisser, clouizos, avijitc, jsoriaga}@qti.qualcomm.com
{danielmad, amonteiroman, nasingal, jeomhove, rsim}@microsoft.com

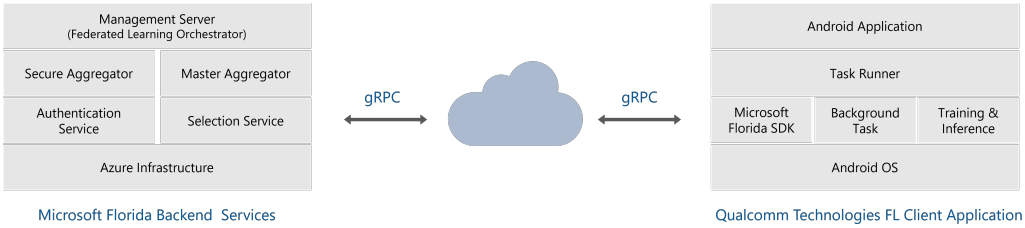

Figure 1: *System overview*

Abstract

Federated Learning (FL) enables distributed machine learning model training on edge devices, ensuring data privacy [1, 2]. However, managing such training with the devices' limited resources, heterogeneous architectures, and unpredictable availability is challenging. To address these challenges and improve on-device training for mobile devices, we present a joint effort by Qualcomm Technologies, Inc. and Microsoft. The demonstration includes a technical display of FL on Snapdragon®[1] devices, coordinated through Microsoft Florida [3], and a federated user verification example using voice samples.

1. Introduction

Federated training requires solving several challenges inherent to distributed training on edge devices. Any infrastructure and interface need to address device heterogeneity, limited resource, availability and orchestration of large federations through scalability, robustness and flexibility.

While multiple FL libraries have been developed to aid the research and development of FL algorithms [4, 5, 6, 7] or target production environments [8, 9, 10], a focus on mobile devices is often lacking. These devices represent arguably the most proliferated and interesting sources of data at the edge. Our solution is a demonstration of a combined effort from Qualcomm Technologies, Inc., a leader in mobile space innovation with the Snapdragon processors, and Microsoft, a leading cloud innovator with the Azure server offering.

Our technical demonstration is aimed at machine learning researchers, showcasing FL in action on Snapdragon devices, with training orchestrated through Microsoft Florida. A second, practical demonstration aimed at the INTERSPEECH community features federated user verification based on voice samples. This includes user enrollment, subsequent acceptance of the enrolled user, and rejection of impostors using the model trained through our FL system.

[1]Snapdragon branded products are products of Qualcomm Technologies, Inc. and/or its subsidiaries.

*Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc. and/or its subsidiaries.

2. System Design

The proposed system consists of two interacting components, as visualized in Fig. 1. The Microsoft Florida Server implements the server-side components of FL, such as the orchestration of training runs, updating of the global model and communication with the FL clients. Those clients are provided by Qualcomm AI Research, interact with the server through a dedicated SDK and implement an application for local training and inference.

2.1. FL Client Application from Qualcomm AI Research

To prepare on-device training, a ML researcher first designs the local training loop with PyTorch on a workstation. Once algorithmic components such as model and loss are determined, they are saved to TorchScript and converted to run on-device within the Qualcomm FL Client App. The App uses the Microsoft Florida SDK to communicate with the Florida server and Qualcomm's On Device Training framework to execute training and evaluation of the training progress.

During the FL training round, the Florida SDK on each device receives the global model from the Florida server and passes it on to the Kotlin-based Trainer in the Client App via a training callback. As shown in Fig. 2, the Trainer then invokes the C++ based On Device Training Framework using the Java Native Interface. The framework trains the model on the local data and sends the locally updated model back to the Florida SDK through the same call stack.

One core innovation in our offering lies in this seamless integration of the call stack from ML researcher-facing APIs for prototyping on a workstation to the on-device execution of the final FL training run. No device-specific programming knowledge is required by the ML researcher to launch an FL application, in combination with the Microsoft Florida FL Server GUI.

2.2. Microsoft Florida FL Server

Florida is a platform that provides developers and ML researchers with a suite of tools and services to create and deploy applications with FL capabilities. It features cross-platform SDKs to handle server communication and client-side training logic, backend services running on Azure Infrastructure to se-

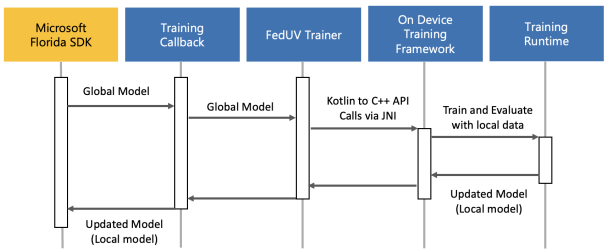

Figure 2: *On-device call stack*

curely orchestrate clients and aggregate the incoming model updates as well as user interfaces (CLI/GUI) to easily create, manage and monitor FL tasks.

The Client App via Florida SDK connects to the Florida Server to request work, download the latest model, execute the application-defined training callback and subsequently upload the model update to the server, where it is aggregated with other client updates. Optionally, the SDK can apply security mechanisms like differential privacy and secure aggregation.

3. FedUV demo

User verification (UV) is a binary decision problem of accepting or rejecting a test input based on its similarity to a user's reference data. Training UV models is a great use case for FL since the biometric data is naturally generated on device and cannot be sent to the central server due to privacy constraints.

In this demo, we replicate the VoxCeleb setup of our prior work [11] in which we propose the *FedUV* algorithm and translate it from the simulated setup to our on-device FL system. With *FedUV*, clients can participate in UV without sharing any client-specific embeddings with other participants in the federation. As such, *FedUV* provides additional levels of security compared to prior UV methods. We refer the reader to [11] for a more detailed discussion of *FedUV*.

The result of a successful *FedUV* run on the VoxCeleb dataset [12] is a trained feature-extractor which can be used without any further fine-tuning to enroll a new user. After providing enrollment voice samples, their averaged embedding can be used to accept/reject any subsequent voice-sample by comparing distances to the average embedding. A public negative dataset serves as a means to tune the distance threshold for a desired trade-off point on the ROC curve.

Technical demo For this demo, we feature an active training run of a *FedUV* model. Attendees interact with the Florida GUI and the Snapdragon devices as visualized in Fig. 3.

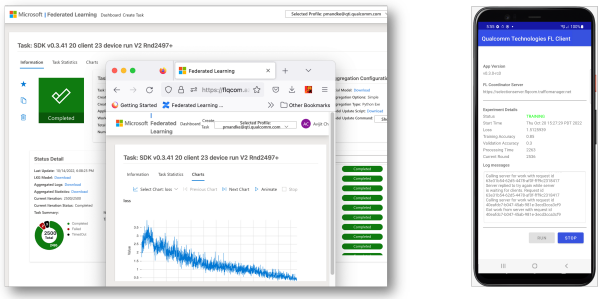

Figure 3: *Orchestration and App interface*

Interactive demo To demonstrate the result of the federated training run, we set up an interactive demo involving a Snapdragon device with a voice-sample feature extractor that has been trained to convergence using *FedUV* beforehand. A participant provides enrollment voice samples and registers as the accepted user on that device. Subsequently, other participants in the demo are shown to be denied access to the device while the initial participant continues to be given access based on their new voice samples.

4. Conclusion

Deployed ML models face the danger of concept drift or even an initial domain mismatch between training data and data observed during deployment. FL across the federation of consumers of a machine learning model serves as one potential solution to ameliorate these issues and continuously update it after deployment. FL can serve a means to avoid costly centralized dataset curation in the first place, opening up the development of new services to providers with limited resources. For spoken language in particular, FL can cover demographic diversity, accents or (microphone) hardware diversity while keeping sensitive biometric data on-device. In this Show & Tell, we demonstrate a software stack that leverages Qualcomm's and Microsoft's expertise in bringing FL from simulation to the real-world. Attendees can interact with the technology, try it for themselves and assess its implications for their own use-cases.

5. References

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *International Conference on Artificial Intelligence and Statistics*, 2017.

[2] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings *et al.*, "Advances and open problems in federated learning," *arXiv preprint arXiv:1912.04977*, 2019.

[3] D. Madrigal, A. Manoel, J. Chen, N. Singal, and R. Sim, "Project florida: Federated learning made easy," 2023.

[4] A. Ingerman and K. Ostrowski, *TensorFlow Federated*, 2019.

[5] S. Caldas, S. M. K. Duddu, P. Wu, T. Li, J. Konečnỳ, H. B. McMahan, V. Smith, and A. Talwalkar, "Leaf: A benchmark for federated settings," *arXiv preprint arXiv:1812.01097*, 2018.

[6] T. Ryffel, A. Trask, M. Dahl, B. Wagner, J. Mancuso, D. Rueckert, and J. Passerat-Palmbach, "A generic framework for privacy preserving deep learning," *arXiv preprint arXiv:1811.04017*, 2018.

[7] Y. Ma, D. Yu, T. Wu, and H. Wang, "Paddlepaddle: An open-source deep learning platform from industrial practice," *Frontiers of Data and Domputing*, vol. 1, no. 1, pp. 105–115, 2019.

[8] H. R. Roth, Y. Cheng, Y. Wen, I. Yang, Z. Xu, Y.-T. Hsieh, K. Kersten, A. Harouni, C. Zhao, K. Lu *et al.*, "Nvidia flare: Federated learning from simulation to real-world," *arXiv preprint arXiv:2210.13291*, 2022.

[9] C. He, S. Li, J. So, X. Zeng, M. Zhang, H. Wang, X. Wang, P. Vepakomma, A. Singh, H. Qiu *et al.*, "Fedml: A research library and benchmark for federated machine learning," *arXiv preprint arXiv:2007.13518*, 2020.

[10] D. J. Beutel, T. Topal, A. Mathur, X. Qiu, J. Fernandez-Marques, Y. Gao, L. Sani, H. L. Kwing, T. Parcollet, P. P. d. Gusmão, and N. D. Lane, "Flower: A friendly federated learning research framework," *arXiv preprint arXiv:2007.14390*, 2020.

[11] H. Hosseini, H. Park, S. Yun, C. Louizos, J. Soriaga, and M. Welling, "Federated learning of user verification models without sharing embeddings," in *ICML*. PMLR, 2021.

[12] A. Nagrani, J. S. Chung, and A. Zisserman, "Voxceleb: a large-scale speaker identification dataset," *arXiv preprint arXiv:1706.08612*, 2017.