Intelligent Health Systems – From Technology to Data and Knowledge
E. Andrikopoulou et al. (Eds.)

doi:10.3233/SHTI250406

Enhancing Trust by a Keycloak-Flower Integration for Federated Machine Learning

Matthaeus MORHART[a], Johanna SCHWINN[a],
Seyedmostafa SHEIKHALISHAHI [a], Michael WELLNHOFER[a],
Ludwig Christian HINSKE[a], and Mathias KASPAR[a,1]
[a]*Digital Medicine, University Hospital of Augsburg, Augsburg, Germany*
ORCiD ID: Matthaeus Morhart https://orcid.org/0000-0003-0379-5702

Abstract. Since its introduction, federated learning (FL) has attracted a lot of attention in the medical field, but its actual application in healthcare organisations remains limited. Flower is a leading FL framework known for its good documentation and wide application. To close security gaps, we propose to integrate Keycloak with gRPC and Flower to improve identity and access management. We have developed a lightweight Python module that integrates both and also validates the client's code with the server before execution. The system has been tested in a simple prototype, but further work and security testing is required for a complex evaluation.

Keywords. Keycloak, Flower, Trust

1. Introduction and Background

Since its first mention on Pubmed in 2017, federated learning (FL) has received a lot of attention in the medical field. FL is a machine learning approach to train a model on separated datasets without sharing that data, ensuring privacy and security. Many research groups are evaluating the topic, resulting in many reviews and new FL frameworks [1,2]. Their actual practical application between multiple healthcare locations, however, is still rarely done. Of the existing frameworks, Flower appears to be one of the best documented and most widely used, leading to a collaboration with NVIDIA [3]. Flower uses gRPC (gRPC Remote Procedure Calls) as an open source, high performance and widely supported framework to efficiently connect services within single or across multiple data centres.The issue of security in all its forms is a constant question that needs to be solved or improved, also in FL. In contrast to FL, the concept of federated analysis, as for example provided by DataShield [4], is somewhat older. This is accompanied by the integration and use of more comprehensive authorization tools such as Keycloak [5], which is still lacking in FL.

In order to close this gap, we are proposing a deeper integration of Keycloak into gRPC and, thus, Flower for code verification, identity and access management.

[1] Corresponding Author: Dr. Mathias Kaspar; E-mail: mathias.kaspar@uni-a.de.

2. Methods and System design

With gRPC [6] Flower already provides encrypted secure communication between all Flower clients and the server, but no user validation and authentication. By integrating Keycloak into Flower, we added the possibility to validate clients and servers via an established authentication and authorisation tool. We do this by using tokens in a realm (i.e., for a single project). The communication between clients, servers and Keycloak is always encrypted and provides OAuth2 as authorization protocol.

The actual integration we propose is based on Keycloak user accounts that are created by the Keycloak administrator on a server that might be hosted at the same site as the Flower server. The user credentials must be changed after receiving.

We developed an own Python module that combines the Keycloak and gRPC Python modules in following process: (1) the Keycloak module is used to create and get a token using the user and the realm credentials at the Keycloak server via an application programming interface (API). (2) The tokens are integrated into the encrypted gRPC communication. Server and client tokens are then mutually validated. (3) In addition, a hash of the client.py file (this is the project-specific FL code that is executed by Flower on the client-site) is calculated on the client side and transferred to the server. The server only establishes a connection if client.py has not been changed.

3. Discussion and Conclusions

We developed a simple python module that combines the Keycloak and gRPC APIs to improve trust in establishing FL projects with multiple healthcare providers and their data. We have tested it in a first prototype, but we still need to improve the development to be able to make a more complex evaluation. In particular, it is important to evaluate further ways to improve the protocol and analyze its overall security.

References

[1] Sharma S, Guleria K. A comprehensive review on federated learning based models for healthcare applications. Artif Intell Med. 2023 Dec;146:102691.

[2] Crowson MG, Moukheiber D, Arévalo AR, Lam BD, Mantena S, Rana A, Goss D, Bates DW, Celi LA. A systematic review of federated learning applications for biomedical data. PLOS Digit Health. 2022 May 19;1(5):e0000033.

[3] Announcing NVIDIA and Flower Collaboration. https://flower.ai/blog/2024-03-15-announcing-nvidia-and-flower-collaboration/ (last accessed: 2024-10-07)

[4] Wolfson M, Wallace SE, Masca N, Rowe G, Sheehan NA, Ferretti V, LaFlamme P, Tobin MD, Macleod J, Little J, Fortier I, Knoppers BM, Burton PR. DataSHIELD: resolving a conflict in contemporary bioscience--performing a pooled analysis of individual-level data without sharing the data. Int J Epidemiol. 2010 Oct;39(5):1372-82.

[5] Keycloak Open Source Identity and Access Management. https://www.keycloak.org/ (last accessed: 2024-10-07)

[6] gRPC high performance RPC framework. https://grpc.io/ (last accessed: 2024-10-07)