**SA WG2 Meeting #162S2-2405030**

**April 15 – April 19, 2024, Changsha, China (revision of S2-2404525)**

**Source: ETRI**

**Title: Solution (KI#2): How to support Vertical Federated Learning between NWDAF and AF**

**Document for: Approval**

**Agenda Item: 19.15**

**Work Item / Release:** **FS\_AIML\_CN / Rel-19**

*Abstract of the contribution: This pCR proposes a solution for supporting VFL between NWDAF and AF to address Key Issue #2: 5GC Support for Vertical Federated Learning.*

# 1 Discussion

In order to support vertical federated learning (VFL) as a new promising AI/ML mechanism in 5GC, Key Issue #2 has been defined.

This pCR introduces a solution for supporting VFL between NWDAF and AF. As an example use case, the utilization of VFL for service experience analytics can be considered. The existing service experience analytics assumes NWDAF can collect various types of raw data, including service experience (i.e., QoE), from AF. However, due to potential privacy concerns, this may not always be feasible, necessitating alternative approaches. Utilizing VFL can help overcome this limitation.

Specifically, the solution provides customized support for VFL between NWDAF and AF, covering NF discovery and selection, ML model training and inference related procedure, initial ML model provisioning to participants, and sample and feature alignment among participants.

During the previous meeting, held in February in Athens, it was suggested that it would be beneficial for each solution to describe the relevant terminology they consider. As a result, terminology has been added accordingly to this pCR.

As experienced during the previous meeting, there are varying perspectives on terminology definitions among different companies. Considering that VFL is not a technique within the 3GPP but rather introduced in the field of AI/ML, the most rational approach is to base definitions on trusted existing literature in that domain and adapt them slightly to fit the 3GPP domain. Additionally, recognizing the constraints of the current Rel-19 release timeline, it is deemed more reasonable to only define terminologies considering the most typical system models that are regarded in the field, as opposed to exploring a wide range of diverse system models. With this approach, our proposed terminologies have been based on the following two prominent academic papers:

[REF-1] Jiang, Jiawei, et al. "Vf-ps: How to select important participants in vertical federated learning, efficiently and securely?." *Advances in Neural Information Processing Systems* 35 (2022).

- “In the Vertical Federated Learning (VFL) setting, each participant holds a disjoint set of features for the same set of entities; the coordinator also holds the ground truth labels.”

[REF-2] Liu, Yang, et al. "Vertical Federated Learning: Concepts, Advances, and Challenges." *IEEE Transactions on Knowledge and Data Engineering* (2024).

- “In a typical VFL system, passive parties communicate only with the active party, which serves as the coordinator that orchestrates the training and inference procedures. In some scenarios, a third party is involved and responsible for encryption and decryption.”

**Observation 1: In a typical VFL system, an active participant may also act as the coordinator.**

In addition to the context within the AI/ML domain, it is also crucial to consider relevant terminology and concepts already introduced in 3GPP specifications (e.g. TS 23.288), aiming to minimize unnecessary debates.

As a representative example, roles such as server and clients have been specified in the FL context in clause 5.3 of TS 23.288. While historically these roles have been defined in the context of horizontal FL, considering that both horizontal and vertical FL belong to the same FL technique, if we choose to use terms like server and client in the vertical FL context, it is essential to align their definitions and implications with those in the context of horizontal FL. Thus, it is recommended to consider the content in clause 5.3 of TS 23.288 when defining terminologies for VFL.

5.3 Federated Learning (FL) among multiple NWDAFs

Federated learning among multiple NWDAFs is a machine learning technique in core network that trains an ML Model across multiple decentralized entities holding local data set, without exchanging/sharing local data set. This approach stands in contrast to centralized machine learning techniques where all the local datasets are uploaded to one server, thus allowing to address critical issues such as data privacy, data security, data access rights.

NOTE 1: Horizontal Federated Learning is supported among multiple NWDAFs, which means the local data set in different FL client NWDAFs have the same feature space for different samples (e.g. UE IDs).

For Federated Learning supported by multiple NWDAFs containing MTLF, there is one NWDAF containing MTLF acting as FL server (called FL server NWDAF for short) and multiple NWDAFs containing MTLF acting as FL client (called FL client NWDAF for short), the main functionality includes:

**FL server NWDAF:**

- discovers and selects FL client NWDAFs to participant in an FL procedure

- requests FL client NWDAFs to do local model training and to report local model information.

- generates global ML Model by aggregating local model information from FL client NWDAFs.

- sends the global ML Model back to FL client NWDAFs to perform an additional training iteration if needed.

**FL client NWDAF:**

- locally trains ML Model as tasked by the FL server NWDAF with the available local data set, which includes the data that may not be allowed to be shared with other FL client NWDAFs due to e.g. data privacy, data security, data access rights.

- reports the trained local ML Model information to the FL server NWDAF.

- receives the global ML Model from FL server NWDAF and perform an additional training iteration if needed.

**Observation 2: The roles of server and clients in the FL context have been specified in the existing 3GPP specifications. Therefore, it is recommended to base the definitions on these specifications unless there is significant intention.**

**Observation 3: Within the above overall context, the server is the entity responsible for owning labels, thus it can be regarded as the active participant in the VFL context. On the other hand, the clients are entities that do not own labels but participate in the training process, often referred to as passive participants in the VFL context.**

**Proposal: It is proposed to only define the terms “VFL server” and “VFL client”.**

# 2. Proposal

It is proposed to adopt the following solution in TR 23.700-84.

\*\*\* Start of change (new text) \*\*\*

# 6 Solutions

## 6.0 Mapping of Solutions to Key Issues

Table 6.0-1: Mapping of Solutions to Key Issues and Use Cases

|  |  |  |
| --- | --- | --- |
|  | Key Issues | Use cases (optional) |
| Solutions | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 5 | 6 |
| #1 | X |  |  |  |  |  |  |  |  |  |
| #2 | X |  |  |  |  |  |  |  |  |  |
| #3 | X |  |  |  |  |  |  |  |  |  |
| #4 | X |  |  |  |  |  |  |  |  |  |
| #5 | X |  |  |  |  |  |  |  |  |  |
| #6 | X |  |  |  |  |  |  |  |  |  |
| X |  | X |  |  |  |  |  |  | X |  |

\*\*\* Next change (new text) \*\*\*

## 6.X Solution #X: How to support Vertical Federated Learning between NWDAF and AF

### 6.X.1 Terminology

**VFL server:**

- owns labels for a VFL task and coordinate the training and inference procedures

- distributes initial ML models to each VFL client

- Performs sample and feature alignment with VFL client

- discovers and selects VFL clients to participant in an VFL procedure

- requests VFL clients to do local model training and to report local intermediate information.

- locally trains ML Model with the available local data set

- computes the training loss and the gradients by aggregating local intermediate information from VFL clients.

- sends the gradients back to FL clients to perform an additional training iteration if needed.

**VFL client:**

- Performs sample and feature alignment based on criteria from VFL server

- locally trains ML Model as tasked by the VFL server with the available local data set

- reports the trained local intermediate information to the VFL server.

- receives the gradient information from VFL server and computes the gradient of its local model

### 6.X.1 Description

This solution addresses Key Issue #2: 5GC Support for Vertical Federated Learning.

The key operational characteristics of vertical federated learning (VFL) is to train a ML model without explicitly sharing raw data among participants involved in training and inference. However, it still require coordination between the involved participants for training and inference.

This solution targets a use case where VFL is operated between NWDAF and AF, specifically focusing on Scenario 2 as outlined in the agreed S2-2403591, which involves AF initiating the VFL training process. It addresses the following aspects as agreed in S2-2401830:

- Whether and how the existing NF discovery and selection needs to be enhanced.

- How to support sample and feature alignment among the participating network entities when performing VFL.

### 6.X.2 Procedures

#### 6.X.2.1 Procedure of NF discovery and selection

Figure 6.x.2-1 shows the procedure of NF discovery and selection.



Figure 6.x.2-1. Procedure for of NF discovery and selection

1. NWDAF or ASF (AI/ML Support Function as a new NF), with the capability to participate as VFL clients, registers its NF profile with the NRF. This profile may include VFL capability information indicating whether it can support training for VFL, VFL capability type information (i.e., “VFL server or VFL client), the NWDAF NF Type, Analytics ID(s), Address information of NWDAF, Service Area, and Time interval supporting VFL.
If NWDAF is a VFL client, it requires the NWDAF to have the capability to perform ML model training with ML models provided by the AF, serving as a VFL server, as a new functionality.

NOTE 1: A VFL server is considered to be an VFL active participant that owns the label information, while VFL clients refers to the rest of VFL participants that do not own labels but participate in the training process.

1. The NRF stores the NWDAF or ASF profile and marks it as available.
2. The NRF acknowledges that NF registration is accepted via Nnrf\_NFManagement\_NFRegister response.
3. The AF, serving as a VFL server, aims to perform VFL.
4. The AF subscribes to the VFL client NF selection assistance information by sending an initial Nnef\_ClientNFSelectionAssistance\_subscribe request, which includes one or more VFL Client NF filtering criteria. Subsequently, the AF may update the client NF filtering criteria of the subscription by invoking Nnef\_ClientNFSelectionAssistance\_subscribe and providing a Subscription Correlation ID. In this step, filtering criteria related to sample and/or feature alignment or VFL capability may be included.
5. If the AF request does not contain a Subscription Correlation ID, the NEF verifies the authorization of the AF request and executes the corresponding service operations (e.g., NF discovery) based on the VFL Client NF filtering criteria provided by the AF. If filtering criteria related to sample and/or feature alignment are included in step 5, the NEF may perform necessary operations for sample and/or feature alignment. For example, the NEF may identify a set of available NFs that have data for the same sample but different features based on the received filtering criteria.

NOTE 3: Sample and/or feature alignment may be performed by other NFs (e.g., NWDAF) as well as the NEF.

1. If the AF request contains a Subscription Correlation ID, the NEF correlates the Nnef\_ClientNFSelectionAssistance\_Subscribe request to an existing subscription based on the Subscription Correlation ID. The NEF may update the filtering criteria and subsequently trigger corresponding service operations (e.g. NF discovery) based on the updated filtering criteria.
2. The NEF interacts with different 5GC NFs to collect the required information. The interactions between the NEF and the 5GC NFs depend on the VFL Client NF filtering criteria provided by the AF. If the filtering criteria in step 5 include VFL capability, the NEF discovers candidate NFs with VFL capability through NRF. For filtering criteria that can be verified through NRF besides VFL capability (e.g., sample alignment), the NEF finds candidate NFs that satisfy them via NRF. However, for filtering criteria that cannot be directly verified through NRF (e.g., feature alignment), the NEF may check their satisfaction through interactions with 5GC NFs.
3. Based on the collected information from other 5GC NFs, the NEF consolidates all the information to derive the list(s) of candidate NFs that fulfil the VFL Client NF filtering criteria in the AF request.
4. The NEF sends an Nnef\_ClientNFSelectionAssistance\_Notify request to the AF, which includes the list(s) of candidate NFs and possibly additional information.
5. Based on the information received in step 10, the AF determines VFL client NF(s).

### 6.X.3 Impacts on services, entities and interfaces

NWDAF or ASF (as a new NF):

- Supports sample and/or feature alignment.

- Retrieves initial ML models for training from the AF.

- Performs local ML model training with the initial ML models provisioned from the AF.

- Performs ML model inference collaboration with other NWDAF(s), ASF and/or the AF.

NEF:

- Receives VFL client NF filtering criteria from the AF.

- Maps VFL client NF filtering criteria to corresponding service operations, including those required for sample and/or feature alignment.

- Derives candidate VFL client NFs and sends the information of the candidate VFL client NFs to the AF.

AF:

- Acts as the server managing the training and inference procedures for VFL.

- Provisions initial ML models for training to the NWDAF or AEF.

- Receives intermediate results from the NWDAF or AEF.

- Performs ML model inference collaboration with other NWDAF(s), ASF and/or the AF.

NRF:

- Recognizes and stores a new VFL capability for NWDAF to facilitate the discovery of VFL-capable NWDAF.

\*\*\* End of change \*\*\*