**SA WG2 Meeting #162S2-2403965**

**April 15 – 19, 2024, Changsha, China,**

**Source: Samsung, Futurewei**

**Title: KI#2: New Solution on Support for VFL with NWDAF and AF as Participants**

**Document for: Approval**

**Agenda Item: 19.15**

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

*Abstract of the contribution: This pCR proposes a new solution for KI#2 on Support for VFL with NWDAF and AF as Participants.*

# 1 Discussion

This pCR proposes a new solution in TR 23.700-84 for KI#2 and use case #4 on Support for VFL with NWDAF and AF as Participants.

# 2. Proposal

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

\*\*\* Start of 1stchange \*\*\*

## 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 |  |  |

\*\*\* Start of 2nd change (all new text) \*\*\*

## 6.X Solution #X: Support for VFL with NWDAF and AF as Participants

### 6.X.1 Description

This solution addresses use case #4 and key issue #2. Unlike traditional centralized learning approaches, where data is pooled together in a single location, or Horizontal Federated Learning (HFL), where different entities contribute similar types of data about different samples, VFL allows for the collaborative training of machine learning models across entities that hold different types of information about the same entities or events.

The solution introduces support for VFL with NWDAF and AF as participants (active and/or passive) by means of enabling data alignment (i.e. sample and feature alignment) among entities participating in the VFL training process with a new service operation. This new service operation allows guaranteeing that training samples are the same in the participating entities even though their training features are different. The solution requires a VFL server guiding the training and distributed inference processes and VFL participants following the VFL server instructions. NWDAF is the 5GC NF acting as VFL server.

### 6.X.2 Procedures

NOTE 1: In this solution, the VFL server coordinates the VFL operation and acts as active participant with access to labels.

NOTE 2: VFL Participants in this solution can be either active participants with access to labels or passive participants without access to labels.

NOTE 3: VFL Participants may also be called VFL Clients.



Figure 6.X.2-1: Procedure for VFL with NWDAF and AF as Participants

The procedure in Figure 6.X.2-1 to support VFL operation at NWDAF is described step by step below.

1. VFL server (i.e. NWDAF) and VFL participant entities (NWDAF, AF) entities register to NRF. The registration may include their NF profiles, Analytics ID(s), Address information of NWDAF, Service Area, VFL capability type information (i.e. VFL server or VFL participant type) and Time interval supporting VFL. The latter parameter can be the same as Time interval supporting FL described in clause 5.2 of TS 23.288 [5].

NOTE 3: VFL participant type parameter can have the value of ‘active’ or ‘passive’.

2. The VFL server and participants are discovered via NRF by invoking the Nnrf\_NFDiscovery\_Request service operation.

NOTE 4: Details of the discovery mechanism are not within the scope of this solution.

3. The VFL server sends a VFL preparation request to the VFL consumers. For NWDAF participants, the existing service operations Nnwdaf\_MLModelTrainingInfo\_Request or Nnwdaf\_MLModelTraining\_Subscribe may be reused and enhanced, or a new service may be defined. For AF participants, a new AF/NEF service is required. Either way, the ML Preparation Flag is provided to check if the VFL participants can meet the ML model training requirement (e.g. Analytics ID, ML Model Interoperability information, Available data requirement, Availability time requirement, etc.). The VFL participants may respond to the VFL server indicating whether they will join the VFL operation and may include the reason in the response message if it cannot join the VFL operation.

4. The selection of VFL participants by the VFL server may happen here, fully or partially. The selection of VFL participants by the VFL server may also be performed or refined in step 8.

5. The VFL server may determine the list of required features and target samples and sends a request for data alignment (i.e. samples and features) including information required for feature and samples alignment to the VFL participants, via NEF for AF participants. This new service/service operation is required to facilitate the alignment of datasets from different sources in a VFL environment. It ensures that participating entities work with a common set of samples without revealing sensitive data. In addition, these samples are the intersection of different datasets where each entity has different features (information or attributes) for the same set of entities or individuals.

Required inputs of this service operation include dataset identifiers (i.e. unique identifiers for the datasets held by each VFL participant), alignment technique (i.e. specific methods or algorithms to be used for data alignment, such as Private Set Intersection (PSI), feature hashing, or other techniques that ensure data privacy and integrity) and Notification Target Address. Optional inputs may include an alignment correlation ID, expiry time, and additional data alignment information such as challenges or discrepancies encountered in the data alignment process.

NOTE 5: Existing service operations (e.g. Nnwdaf\_MLModelTraining\_Subscribe, Nnwdaf\_MLModelTrainingInfo\_Request) may also be enhanced to include feature and sample information for interactions between VFL participants.

6. Each VFL participant performs data alignment pre-processing. This is required for each VFL participant to ascertain its capability to meet the AI/ML model training requirements. Data alignment pre-processing may include model training demands feasibility assessment, model access and compatibility, data alignment verification, etc. At this step, VFL passive participants may determine a list of supported features and samples. A passive VFL participant may determine the list of supported features (e.g. a subset of the list of required features shared by active participant) and the list of supported samples (e.g. a subset of target samples shared by active participant) based on the information received from the active participant, its available data, corresponding NF type(s) or instances of the requested analytics ID, recent data collection operations, and ML Model Interoperability information.

7. The VFL participants notify the VFL server the result of the data alignment, via NEF for AF participants. Furthermore, the VFL participants may provide the VFL server with a decision regarding its participation in the VFL operation along the supported features and sample IDs identified (if any). The VFL participants may also include the reason in the response message if it cannot join the VFL operation.

8. If not completed at step 4, the selection of VFL participants by the VFL server may be performed or refined based on the inputs received by the VFL server from the VFL participants. At this step, VFL active participant may perform sample alignment by identifying overlap/intersection of supported samples of all VFL passive participant. A VFL passive participant selected in step 4, may be excluded from VFL training if the list of its supported samples has zero or very little overlap with supported samples of other VFL passive participants.

 Based on supported feature of each passive participant (feature dimension of each passive participant), VFL active participant may also decide how to partition data features between passive participants and assigns a subset of features to each passive participant. A passive participant selected in step 4 may be excluded from VFL training if the list of its supported feature has zero overlap with the list of required features. As the 3GPP specified analytics IDs require input data from different domains including 5GC and AF, a domain-based feature selection for cross domain VFL can be used where each domain performs VFL training based the data owned by that domain (i.e., AF performs VFL training based on AF data and 5GC performs VFL training based on 5GC data). However, within each domain also, it is important to decide how to vertically divide data features among multiple VFL participants. For example, observed service experience analytics ID, within 5GC domain, requires input data from AMF, SMF and UPF. In one approach, VFL passive participants may perform VFL training based on AMF, SMF and UPF data together. In another approach, each VFL passive participant may perform VFL training based on different NF type.

9. The VFL training process starts on this step, where intermediate training results are shared and coordinated by the VFL server, facilitating a collaborative approach to model refinement across the VFL participating entities. In the first iteration, the VFL server triggers the VFL training process by invoking a ML Model Training service subscription operation from the VFL Participant #1 (e.g. NWDAF). In the subsequent iteration, the VFL server may need to use the ML Model Training service notification operation with the Participant #2 (or last participant if more than two) (e.g. AF via NEF) to enable model refinements in each participant. Subsequent iterations may require the VFL server to subscribe or notify from/to either participant using the ML Model Training service. While the existing Nnwdaf\_MLModelTraining service can be used for NWDAF participants, a new service at the NEF/AF is required to support this functionality. In the last iteration, the VFL server informs the VFL participants that the VFL training process is completed via a suitable flag, steps 10 through 12 are skipped, and the VFL training loop is terminated.

10a.The VFL Participant #1 may perform computation to locally train its model. The computation may lead either to intermediate training results to be shared to the next participant in step 11, or to refine a previously trained model and notify the VFL server in step 13.

10b.The VFL Participant #1 may share its intermediate training results via NEF with the AF participant, using the same service operations as in step 9.

10c.Participant #2 (i.e. last VFL participant) performs computation to locally train its model. The computation may lead either to intermediate training results to be shared with the VFL server in step 11, or to refine a previously trained model and the previous VFL participant in step 10d.

10d.The VFL Participant #1 performs computation to refine its local model.

11. Intermediate training results and/or model refinement results are shared with the VFL server, via NEF if from AF, using the same ML model training service used in step 9.

12. The VFL server performs further VFL computation.

13. A consumer NF subscribes to or requests analytics from the NWDAF hosting VFL server functionality.

14. A distributed VFL inference process is triggered by the VFL server invoking a ML model inference request from Participant #1. This step can be executed with a new service or by enhancing the existing ML Model Training service since the required VFL computation is essentially the same for inference and some training cycles.

15a.Participant #1 performs local inference computation.

15b.Participant #1 shares intermediate inference results with Participant #2 via NEF when Participant #2 is an AF. In that case, the NEF/AF service used may be new or an enhanced version of the new NEF/AF service already used in step 9.

15c.Participant #2 performs local inference computation.

16. Participant #2 (i.e. the last participant) notifies the result of the distributed inference process to the VFL server using the same service as in step 14 via the notification operation.

17. The VFL performs further inference computation and derives the requested analytics.

18. The derived analytics are delivered to the NF consumer.

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

NWDAF:

- New service to support data alignment.

- Enhancements to Nnwdaf\_MLModelTraining service (or new service(s)) to support VFL preparation as well as sharing of intermediate training and inference results.

NEF/AF:

- New service to support data alignment.

- New service(s) to support VFL preparation as well as sharing of intermediate training and inference results.

\*\*\* End of 2nd change \*\*\*