SA WG2 Meeting #162 S2-2404330r03

Changsha, April 15 – April 19, 2024 (Revision of S2-2402104)

**Source: OPPO, China Telecom, ZTE, Sony, KDDI**

**Title:** **New Solution for KI#2 –** **Sample/Feature alignment and general training procedure for the VFL**

**Document for: Approval**

**Agenda Item: 19.15**

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

*Abstract of the contribution: This contribution proposes a solution for KI#2 related to the sample/feature alignment and VFL training procedure.*

# 1. Introduction/Discussion

This contribution addresses the KI#2: 5GC Support for Vertical Federated Learning. More specifically, this solution will address these bullets as following:

- Identify VFL use cases and under which conditions, and for which entities these VFL use cases show that VFL is justified to train ML models.

- Whether and how to support architecture enhancement for supporting VFL for model training and/or inference. In particular:

- Whether and how ML Model training related procedures need to be enhanced to support VFL.

- Whether and how to provide ML Models to the participants in the VFL training process.

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

5GS includes several domains and for this release, we only consider the 5GC and AF. For the 5GS assistance in collaborative AI/ML operation between 5GC and AF for Vertical Federated Learning, one of the use cases is QoE prediction which is related to the observed service experience analytics. When NWDAF provides observed service experience analytics, as in other analytics that require input data from the AF, the AF may prevent raw data to be exchanged directly between NWDAF and an external AF, as NWDAF is in the PLMN and the AF is outside the PLMN and the user data has high privacy protection needs.

Furthermore, NWDAF and AF may have different features of the same sample identity, which is a requirement of VFL. In such cases, VFL can be very helpful to break the data isolation and enable joint training between NWDAF and AF. Moreover, regardless of the entities involved in VFL, the application of VFL among different entities requires alignment of samples and features to make sure the above VFL requirement is addressed.

Therefore, this solution addresses the issue on how to support the sample/feature alignment, provide the models to the participants and VFL training procedures based on the QoE prediction use case.

# Text Proposal

It is proposed to agree the following changes to TR 23.700-84.

\* \* \* \* First change \* \* \* \*

# 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 |  |

\* \* \* \* Second change \* \* \* \*

## 6.X Solution #X: Sample/Feature alignment and general training procedure for the VFL

### 6.X.1 Description

This solution is proposed to address Key Issue #2: 5GC Support for Vertical Federated Learning, and based on the Use Case #5.

In Use Case #5, two scenarios are identified:

Scenario 1: NWDAF initiates VFL training process.

Scenario 2: AF initiates VFL training process.

This solution focuses on scenario 2 where AF is the active participant(VFL server) with label while NWDAF(s) is the passive participant(VFL client)(s).

The basic principle for this solution is that AF is the only entity which owns the label (e.g. the value for the QoE) for the training data when performing VFL model training. NWDAFs in the 5GC which has the model training capability (e.g. NWDAF with MTLF) will be other entities to train the VFL model with AF collaboratively. Since the label is in the AF, AF will act as the active participant(VFL server) and NWDAFs will be the passive participant(VFL client).

The function of the active participant(VFL server) is shown as following:

* Initiating the VFL task, performing the sample and feature alignment with the passive participant(VFL client) to make sure the training dataset in the different entity has the same sample space and different feature space.
* Sending the initial model to the passive participant(VFL client). Making sure each entity can perform the VFL model training for the same VFL task.
* Collecting the passive participant(VFL client) training result (i.e. intermediate result). Based on itself training result using its own local data, the label and the passive participant(VFL client) training result, calculating the loss information(e.g. loss function, type of loss function, loss value). Sending the loss information to the passive participant(VFL client).
* Indicating to the passive participant(VFL client) that the training progress is finished based on the comparison of the gathered training result and the label(i.e. the loss value/loss function converged) or the pre-set iteration number is reached..

The function of the passive participant(VFL client) is shown as following:

* Performing the sample and feature alignment based on the active participant(VFL server) request.
* Receiving the part of the initial VFL model from the active participant(VFL server). Using its local dataset to train the local model.
* Generating the local model training result and sending it to the active participant(VFL server).
* Storing the local well-trained model in order to support the VFL inference later.

NOTE 1: How to select the participant that has the VFL capability is not in the scope of this solution.

NOTE 2: The coordinator in the VFL is related to the security key management and the VFL model type (e.g. linear model or DNN model). Therefore, in order to make sure this solution can be used in a more general case, the coordinator role is not considered in this solution.

### 6.X.2 Procedures

##### 6.X.2.1 Sample and feature alignment procedure

In VFL, it is required active participant(VFL server) and passive participant(VFL client) have same samples and different features of the same sample identity before the model training. This solution proposed a preparation process used to have a negotiation between active participant(VFL server) (AF) and passive participant(VFL client) (NWDAF) to ensure they share the same sample space, i.e. same UEs and identify the features before the VFL model training. In this process, NEF is responsible for candidate passive participant(VFL client) discovery and willingness checking. Moreover, if any participating entity leaves or a new entity joins the VFL, this triggers the process of updating the participating entity list. Additionally, it necessitates re-aligning sample and feature profiles among all entities to ensure consistency and compatibility in the collaborative model training process.

If the samples provided by the AF involve multiple NWDAFs, e.g. the UEs served by the AF are served by multiple NWDAFs since the AF and the NWDAFs have different service area, the NEF will map the samples into multiple sub-samples and return result per sub-sample in one response to the AF. When the AF receive the response, it may decide whether to start VFL model training with following three options:

- Option1: maintain the sample space to start VFL model training with multiple NWDAFs;

- Option 2: reduce the sample space to ensure to start VFL model training with only one NWDAF;

- Option 3: stop the VFL process.



1. "VFL active participant(VFL server)" AF determines to use VFL model training because data that cannot be obtained/exposed directly from 5GC (e.g., for privacy reasons) or vendor implementation policy. The active participant(VFL server) (e.g. AF) sends VFL participant discovery request to NEF by indicating Analytics ID (e.g. Observed Service Experience), application ID, External UE IDs, filter information (e.g. Area of Interest).

2. Based on the request from AF, the NEF authorizes the request information and applies parameter mapping (e.g. external UE IDs mapping to internal UE IDs, geographical area mapping to TA(s)/ Cell-id(s)). The NEF requests for passive participant(VFL client) discovery from NRF and further queries UDM for determining the NWDAF ID and its corresponding serving UE list if the NWDAF registration in UDM is supported with the requested information provided by the AF.

3. The NRF/UDM returns available NWDAF(s) and their corresponding serving UE list to the NEF .

4. NEF send the available NWDAF(s) and their corresponding serving UE list to the AF.

1. Active participant(VFL server) (i.e. AF) sends the VFL sample and feature alignment request to the NEF including the samples (i.e. external UE IDs) the AF want to use for the VFL training, then analytics ID (e.g. Observed Service Experience), the application ID indicates which application needs to initiate the VFL operation, the feature profile indicates what feature the AF can provide and what feature(s) are needed for performing the VFL. For example, AF can indicate in the feature profile, for the observed service experience case, access speed, stall time, frame rate for AF can be used as the feature. The input data from AF as described in TS 23.288 clause 6.4.2 can also be the feature provided by the AF. The specifications regarding encryption methods. These encryption methods may encompass various techniques such as Rivest-Shamir-Adleman (RSA), ElGamal, and Lattice-based encryption, among others.
2. NEF maps the external UE IDs to the internal UE IDs based on the information exchanged with the UDM.
3. NEF sends the VFL sample and feature alignment request to the passive participant(VFL client) (i.e. NWDAF(s)) including the internal UE IDs (e.g. SUPIs), Application ID and feature profile. If multiple NWDAFs involved, the NEF map the samples into sub-samples which are corresponding to each NWDAF and sends the VFL initiation request to each NWDAF, respectively.
4. Based on the feature profile, NWDAF(s) can identify which analytics are related to the VFL request. NWDAF(s) collects the input data related to the internal UE IDs with the NFs as described in the TS 23.288 clause 6.4.2.
5. Based on the collected input data from the NFs, NWDAF(s) can determine which UE related data is available, whether the feature for the input data can be implemented/satisfied with the feature profile provided by AF. Based on the parameters received in the step 4 and the data collected in the step 5, NWDAF(s) determines the samples are same among NWDAF(s) and AF. On the other side, the feature in the same samples should be different among NWDAF(s).
6. NWDAF(s) sends the sample and feature alignment response to the NEF including the application ID, the sample and feature alignment indication, to indicate which UE and for which feature can be used in the VFL operation.
7. If multiple NWDAFs involved, the NEF aggregates the sample and feature alignment response from NWDAFs. NEF maps the internal UE IDs to the external UE IDs. The UE IDs in this step can be the subset of the UE IDs in the AF request based on the NWDAF(s) determination in order to make the sample and feature alignment.
8. NEF sends the VFL sample and feature alignment response to the AF including the sample and feature alignment indication.
9. The active participant(VFL server) verifies whether the selected samples and features meet the requirements. If the active participant(VFL server) determines that there is an insufficient quantity of samples or features that meet the specified criteria, it can halt the VFL process. Subsequently, it has the capability to initiate a new VFL process by adjusting the requirements or selecting different passive entities, where iterations are needed.
10. When introducing a new passive participant(VFL client) or when updates are required for the current one, the active participant(VFL server) evaluates whether realignment is needed for the ongoing VFL procedure or if it is necessary for initiating a new collaborative learning process. Realignment ensures that the sample and feature requirements are appropriately matched to facilitate effective collaboration. If necessary, the active participant(VFL server) initiates the sample and feature alignment procedure based on either the existing alignment requirements or an updated version.

Editor's Note: Whether to use the existing or new service operation between the AF and NEF and between the NEF and NWDAF to perform the sample and feature alignment is FFS.

##### 6.X.2.2 General training procedure for the VFL between the passive participant(VFL client) (e.g. NWDAF) and active participant(VFL server) (e.g. AF)

In the process of VFL model training, the AF and NWDAF(s) exchange intermediate data (e.g. intermediate training result, loss information). It is assumed that the active participant(VFL server) has ground truth data or labels or is able to collect ground truth data or labels. It is also assumed that the passive participant(VFL client) discovery and selection, the sample/feature alignment are done before the training process.

The VFL model training process will iterate until the AF determines the model performance satisfies requirement based on comparing with the label(i.e. the loss function/loss value converged) or the pre-set iteration number is reached.



1. After performing the sample and feature alignment procedure, NWDAF and AF has the available local datasets to perform the VFL training. Therefore, when the AF needs to initiate the VFL training, AF sends the VFL training subscribe to the NEF including the external UE IDs and feature profile received in the NEF before to indicate which sample and feature are used for the VFL training, the initial model parameters in order to configure the VFL local model to the NWDAF or the initial ML model, the VFL model correlation ID in order to let AF know the intermediate result from NWDAF is for which VFL model and VFL task. AF can then correlate it to the AF itself local result to generate the final result. The interoperability information of the model(e.g. model file format, execution environment). In this message, the VFL server may specify a maximum response time(i.e. the maximum time between receiving loss and sending back the intermediate training result) in every iteration.
2. NEF maps the external UE IDs to the internal UE IDs based on the information exchanged with the UDM.
3. NEF sends the Nnwdaf\_MLModelProvision\_Subscribe to the NWDAF including the internal UE IDs, feature profile, VFL model correlation ID and initial model parameters.
4. NWDAF(s), based on the feature profile, identifies the already collected data happened in the sample and feature alignment procedure. NWDAF uses the collected data and initial model parameters to do the local model training.
5. NWDAF(s) sends the Nnwdaf\_MLModelProvision\_Notify to the NEF including the intermediate results and the VFL model correlation ID. If multiple NWDAFs involved, the NEF aggregates the intermediate results from NWDAFs. The message may include intermediate results identification information, which identifies the intermediate results in the VFL training service, e.g., time stamp.
6. NEF sends the VFL training notify to the AF including the intermediate results and the VFL model correlation ID.
7. AF, based on the VFL model correlation ID, to correlates the intermediate result to the AF itself local result in order to generate the final result. AF calculates the loss information(e.g. loss function, loss value) based on the final result and the label stored in the AF. AF sends the VFL training update subscribe to the NEF including the loss information(e.g. loss function, type of loss function, loss value) and the VFL model correlation ID.

NOTE 1: For the QoE prediction use case, the label is QoE value which is known to the AF.

1. NEF sends the Nnwdaf\_MLModelProvision\_Subscribe to the NWDAF including the loss information(e.g. the loss value, type of loss(e.g. cross-entropy), loss function) and VFL model correlation ID.
2. NWDAF, based on the loss information, computes gradient and updates the local model. The step 4 to step 9 are repeated until the termination condition reached.
3. AF, based on the label and the test dataset, determines the VFL model is in convergence. AF sends the VFL training unsubscribe to the NEF including the training terminate indication.
4. NEF sends the Nnwdaf\_MLModelProvision\_Unsubscribe to the NWDAF to terminate the VFL training progress.

After VFL training procedure, “VFL active participant(VFL server)" NF collects trained ML models from "VFL passive participant(VFL client)" NF(s). This sequence aims to use the trained model in the inference procedure that involves the other NF.

Editor's Note: Whether to use the exist or new service operation between the AF and NEF to perform the VFL model training is FFS.

NOTE2: This procedure is needed only when the 5GC supports the VFL inference with NF that has not participated in the VFL model training.

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

NWDAF:

Enhancement to support the VFL model sample and feature alignment with the AF to make sure the sample is the same, but the feature for the sample is different between AF and NWDAF.

Based on the model and loss function provided by the AF, performing local training and update for the VFL model.

NEF:

Enhancement to support the sample and feature alignment related information and VFL model training related information exchange between AF and NWDAF.

\* \* \* \* End of changes \* \* \* \*