**SA WG2 Meeting #160-Ad-Hoc-eS2-2400560**

**January 22th – 29th, 2024; Elbonia (revision of S2-24xxxxx)**

**Source: vivo (rapporteur), MediaTek Inc. (rapporteur), China Mobile, KDDI, NTT DoCoMo, Futurewei?, Samsung? Orange, ZTE?, Ericsson?**

**Title: New key issue for Vertical Federated Learning between 5GC and AF for cross-domain AI/ML coordination.**

**Document for: Approval**

**Agenda Item: 19.15**

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

*Abstract of the contribution: The intent of this proposed KI is to study enabling collaborative AI/ML operation with Vertical Federated Learning involving NWDAF and AF.*

# 1 Discussion

**Introduction**

The proposed KI is applicable to WT#2 i.e.,

*-WT2: Study whether and what potential enhancements are needed to enable 5G system to assist in collaborative AI/ML operation involving 5GC/NWDAF and/or AF for “Vertical Federated Learning (VFL)”. The work will be based only on and limited to the scope of justified use cases.*

*NOTE 7: RAN and UE aspects are out of scope. Solutions based on interactions between the application client and 5GS are out of scope. The necessary communication between AF and UE application client to support the collaborative AI/ML operation is understood as no normative procedure impact. Horizontal FL procedure defined in R18 should be taken into account and reused whenever possible.*

*NOTE 8: coordination with SA6 is required.*

As SP-231800, Vertical Federated Leaning (VFL) is an essential technique for cross-domain AI/ML framework. For example, VFL will address the following challenges.

-Blocker of data usage: Mobile networks consisting of multiple domains, such as CN and AF, generate vast and beneficial data. Instead, some of them have yet to be fully utilized.

* The first blockage is private implementations; each domain has private implementation but is unwilling to report its data.
* Second is the privacy-sensitive data, such as UE location. These raw data cannot be exposed due to the operator's or government's privacy protection policy.

- Needs diversity: Defining analytics for each application is impractical due to the diversity of application needs. Now, AF has to put up with the uniform standardized Analytics output.

**Background for VFL (Vertical Federated Learning):**

Federated learning is a distributed machine learning technique that trains machine learning models collaboratively among multiple independent training entities, each using its own dataset. There are two general categories of federated learning: horizontal federated learning (which is supported among network data forwarding (NWDAF) nodes in Rel-18) and vertical federated learning (which is being studied for Rel-19).

The most fundamental difference between these two categories is the **training data partitioning**. As shown in Figure X, in horizontal federated learning, all entities have the same features in their training data, meaning that there are different samples of the data in each entity, but all samples have the same set of features. However, in vertical federated learning, the training data samples in different entities have different features, meaning that for a given sample, some features may not be available in some of the training entities.

****

Figure X Training data partitioning in Horizontal vs. Vertical Federated Learning

[source: https://arxiv.org/pdf/2107.03428.pdf]



Figure 1 Exampe of Architecture for a vertical federated learning system

1. Suppose that companies A and B would like to jointly train a machine learning model, and their business systems each have their own data. In addition, Company B also has label data that the model needs to predict.
2. For data privacy and security reasons, A and B cannot directly exchange data. In order to ensure the confidentiality of the data during the training process, a third-party collaborator C is involved.
3. We assume the collaborator C is honest and does not collude with A or B, but party A and B are honest-but-curious to each other.

See details via Federated Machine Learning: Concept and Applications <https://arxiv.org/pdf/1902.04885.pdf>

Another possibility for vertical federate learning is shown below, where no collaborator is needed as in Figure 1.

 

Figure 2 Example of Architecture for a (split) vertical federated learning system (training and inference)

1. Suppose that companies A, B and C, that are represented by the different locations in Figure 2, would like to jointly train a machine learning model, and their business systems each have their own data.
2. For data privacy and security reasons, A, B and C cannot directly exchange data. In order to ensure the confidentiality of the data, inter-layer information (weight, gradients) are exchanged between different parts of a single model.
3. Training cycles are repeated until the model converges.

In Rel-18, ML model sharing between NWDAFs has been studied as a part of Horizontal FL. However, the ML model sharing between NWDAF and AF has not been studied. Regarding analytics exposure and feedback, there is no Analytics up to Rel-18 for VFL training.

# 2. Proposal

It is proposed to agree the following key issue into TR 23.700-84

Start of Change(all new changes)

# 5 Key Issues

## 5.X Key Issue #X: 5GC Assistance to Vertical Federated Learning with 5GC and AF

### 5.X.1 Description

This key issue aims to provide solutions for enabling “vertical federated learning” involving 5GC/NWDAF and/or AF for cross-domain AI/ML coordination.

In Rel-18, ML model sharing between NWDAFs has been studied as a part of Horizontal Federated Learning. However, the ML model sharing between NWDAF and/or AF has not been studied,( e.g. when the NWDAFs and/or AFs are in different domains, locations, regions etc).

To enable cross-domain AI/ML coordination, Vertical Federated Learning (VFL), can be considered as an alternative mechanism for distributed functionalities of an ML model. Note that, as scoped in Rel-19, 5GC/NWDAF and/or AF are involved for VFL.

This Key Issue aims to study architecture enhancement to support VFL, which allows the cooperative AI/ML training and inference of NWDAF and AF with the following aspects:

- Identify use cases for supporting VFL (e.g., between AF and NWDAF/5GC) and under which conditions, and for which entities is VLF justified as a technique to train ML models.

- Whether and how to support architecture enhancement for the existing procedures on ML model training and/or inference that will occur after the VFL procedure for ML training when using VFL techniques in particular:

* Which network entities (e.g. NWDAF, other 5GC NF, AF) may be involved in VFL and whether and how the existing NF discovery and selection needs to be enhanced.
* Whether and how ML Model training and/or inference procedures needs to be enhanced
* Whether and which new performance monitoring capabilities are required to support VFL in the 5GC
* Whether and how to enhance ML model sharing between the entities involved in VFL
* How to support sample and feature alignment among the participating network entities when performing VFL

NOTE X: Application layer-based VFL requiring communication between AF and UE application client, is out of scope.

NOTE Y: During the study on this KI, consultation with SA3 is required for handling security aspects.

NOTE Z: RAN and UE aspects are out of scope.

NOTE W: Procedure will re-use as much as possible the existing procedures defined for HFL in 3GPP TS 23.288 [x].

End of Changes