**3GPP TSG-SA2 Meeting #159 *S2-2311407***

**Xiamen, China, 09 – 13 October (revision of S2-2310185)**

|  |
| --- |
| *CR-Form-v12.2* |
| **CHANGE REQUEST** |
|  |
|  | **23.288** | **CR** | **0916** | **rev** | **1** | **Current version:** | **18.3.0** |  |
|  |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* |
|  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Proposed change affects:*** | UICC apps |  | ME |  | Radio Access Network |  | Core Network | **X** |

|  |
| --- |
|  |
| ***Title:***  | Clarification on the ML Model Metric in Federated Learning |
|  |  |
| ***Source to WG:*** | vivo  |
| ***Source to TSG:*** | SA2 |
|  |  |
| ***Work item code:*** | eNA\_Ph3 |  | ***Date:*** | 2023-10-11 |
|  |  |  |  |  |
| ***Category:*** | F |  | ***Release:*** | R18 |
|  | *Use one of the following categories:****F*** *(correction)****A*** *(mirror corresponding to a change in an earlier release)****B*** *(addition of feature),* ***C*** *(functional modification of feature)****D*** *(editorial modification)*Detailed explanations of the above categories canbe found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | *Use one of the following releases:Rel-8 (Release 8)Rel-9 (Release 9)Rel-10 (Release 10)Rel-11 (Release 11)…Rel-16 (Release 16)Rel-17 (Release 17)Rel-18 (Release 18)Rel-19 (Release 19)* |
|  |  |
| ***Reason for change:*** | To align with clauses 6.2A.2, the FL Server NWDAF should compute and transport the ML Model Accuracy Information to the consumer instead of the global ML Model Metric. Some parameters should be aligned between Nnwdaf\_MLModelTraining\_Subscribe and Nnwdaf\_MLModelTrainingInfo\_Request service operation, such as local ML Model metric. And the modification should be reflect to the corresponding procedure clauses.It’s not clear how to use the ML Model Accuracy Check Flag during the FL model training procedure. |
|  |  |
| ***Summary of change:*** | Modification in clause 6.2C.2.2, 7.10.4 and 7.11.2.* Change the ‘global ML model metric’ to the ‘ML Model Accuracy Information’ in 6.2C.2.2;
* Change the ‘ML Model Accuracy’ in clause 7.10.4 and the ‘global ML model metric’ in clause 7.11.2 to the ‘global ML Model Accuracy’;
* Update the ‘accuracy of local model’ to the ‘local ML Model metric’ in clause 7.10.4;
* Add the description of ML Model Accuracy Check Flag in step 4 and step 7 of clause 6.2C.2.2.
 |
|  |  |
| ***Consequences if not approved:*** | Confusion between “ML Model Accuracy Information” and “ML Model metric”. |
|  |  |
| ***Clauses affected:*** | Clause 6.2C.2.2, 7.10.4 and 7.11.2. |
|  |  |
|  | **Y** | **N** |  |  |
| ***Other specs*** |  | **X** |  Other core specifications  | TS/TR ... CR ...  |
| ***affected:*** |  | **X** |  Test specifications | TS/TR ... CR ...  |
| ***(show related CRs)*** |  | **X** |  O&M Specifications | TS/TR ... CR ...  |
|  |  |
| ***Other comments:*** |  |
|  |  |
| ***This CR's revision history:*** |  |

\* \* \* Start of Changes \* \* \*

#### 6.2C.2.2 General procedure for Federated Learning among Multiple NWDAF Instances



Figure 6.2C.2.2-1: General procedure for Federated Learning among Multiple NWDAF

0. The consumer (NWDAF containing AnLF or NWDAF containing MTLF) sends a subscription request to FL server NWDAF to retrieve an ML model, using Nnwdaf\_MLModelProvision service as defined in clause 7.5 including Analytics ID, ML model metric (e.g., ML model Accuracy), Accuracy reporting interval, pre-determined status (ML model Accuracy threshold or Time when the ML model is needed).

NOTE 1: The ML model Accuracy threshold can be used to indicate the target ML Model Accuracy of the training process and the FL server NWDAF may stop the training process when the ML model Accuracy threshold is achieved during the training process.

 If the consumer (i.e. the NWDAF containing AnLF or NWDAF containing MTLF) provides the Time when the ML model is needed, the FL Server NWDAF can take this information into account to decide the maximum response time for its FL Client NWDAF(s).

1. FL Server NWDAF selects NWDAF(s) containing MTLF (FL Client NWDAF(s)) as described in clause 6.2C.2.1.

2. FL Server NWDAF sends a Nnwdaf\_MLModelTraining\_Subscribe or Nnwdaf\_MLModelTrainingInfo\_Request to the selected NWDAF containing MTLF (FL Client NWDAF(s)) that participates in the Federated learning to perform the local model training and determine the interim local ML model information based on the input parameter in the request from FL Server NWDAF, including ML model metric and initial ML model. The request also includes the maximum response time before which the FL Client NWDAF has to report the interim local ML model information to the FL Server NWDAF.

3. [Optional] Each FL Client NWDAF collects its local data by using the current mechanism in clause 6.2 of TS 23.288 [5] if the Client NWDAF has not local data available already.

4. During Federated Learning training procedure, each FL Client NWDAF further trains the ML model provided by the FL Server NWDAF based on its own data and reports the interim local ML model information to the FL Server NWDAF in Nnwdaf\_MLModelTraining\_Notify or Nnwdaf\_MLModelTrainingInfo\_Response. The Nnwdaf\_MLModelTraining\_Notify or Nnwdaf\_MLModelTrainingInfo\_Response may also include the Status report of FL training with local ML model metric computed by the FL Client NWDAF and Training Input Data Information (e.g. areas covered by the data set, sampling ratio, maximum/minimum of value of each dimension of data, etc.) in the FL Client NWDAF.

 The Nnwdaf\_MLModelTraining\_Notify or Nnwdaf\_MLModelTrainingInfo\_Response also includes the global ML Model Accuracy when the ML Model Accuracy Check Flag was included in the Nnwdaf\_MLModelTraining\_Subscribe or Nnwdaf\_MLModelTrainingInfo\_Request (as described in step 7), the global ML Model Accuracy is calculated by the FL Client NWDAF using the local training data as the testing dataset.

NOTE 2: The parameters in characteristics of local training dataset are up to the implementation.

 The ML model, which is sent from the FL Client NWDAF(s) to the FL Server NWDAF during the FL training process, is the information needed by the FL Server NWDAF to build the aggregated model based on the locally trained ML model(s).

 If the FL Client NWDAF is not able to complete the training of the interim local ML model within the maximum response time provided by the FL Server NWDAF, the FL Client NWDAF shall send the Delay Event Notification that include the delay event indication, an optional cause code (e.g. local ML model training failure, more time necessary for local ML model training) and the expected time to complete the training if available to the FL Server NWDAF before the maximum response time elapses.

4a. [Optional]If FL Server NWDAF receives notification/response that the FL Client NWDAF is not able to complete the training within the maximum response time, the FL Server NWDAF may send to the FL Client NWDAF an extended maximum response time in Nnwdaf\_MLModelTraining\_Subscribe or Nnwdaf\_MLModelTrainingInfo\_Request, before which the FL Client NWDAF has to report the interim local ML model information to the FL Server NWDAF. Otherwise, the FL Server NWDAF may indicate FL Client NWDAF to skip reporting for this iteration. FL Server NWDAF includes the current iteration round ID in the message to indicate that the request is to modify the training parameters of the current iteration round.

 Alternatively, the FL Server NWDAF may inform the FL Client NWDAF to cease the ML model training by sending termination request and to report back the current local ML model updates.

5. The FL Server NWDAF aggregates all the local ML model information retrieved at step 4, to update the global ML model. The FL Server NWDAF may also compute the ML Model Accuracy Information, e.g. based on the local ML model metric(s) , the global ML Model Accuracy(s) (which is calculated by the FL Client NWDAF using the local training data as the testing dataset) or by applying the global model on the validation dataset (if available). The FL Server NWDAF may update the global ML model each time a FL Client NWDAF provides updated local ML model information as part of FL or the FL Server NWDAF may decide to wait for local ML model information from all FL Client NWDAF before updating the global ML model.

 If the FL Server NWDAF provides the maximum response time for the FL Client NWDAF(s) to provide the interim local ML model information in step 2, or the extended maximum response time in step 4a, the FL Server NWDAF decides either to wait for the FL Client NWDAF(s) which have not yet provided their interim local ML model within the (extended) maximum response time or aggregates only the retrieved local ML model information instances to update global ML model. The FL Server NWDAF makes this decision, considering the notification/response from the FL Client NWDAF or, if the notification is not received, based on local configuration.

6a. [Optional] Based on the consumer request in step 0, the FL Server NWDAF sends a Nnwdaf\_MLModelProvision\_Notify message to update the ML Model Accuracy Information to the consumer periodically (e.g. a certain number of training rounds or every 10 min) or dynamically when some pre-determined status is achieved (e.g. the ML Model Accuracy threshold is achieved or training time expires).

6b. [Optional] The consumer decides whether the current model can fulfil the requirement, e.g. ML Model Accuracy Information is satisfactory for the consumer and determines to stop or continue the training process. The consumer re-invokes Nnwdaf\_MLModelProvision\_Subscribe service operation as used in step 0 to stop or continue the training process.

6c. [Optional] Based on the subscription request sent from the consumer in step 6b, the FL Server NWDAF updates or terminates the current FL training process.

 If the FL Server NWDAF received a request in step 6b to stop the Federated Training process, steps 7 and 8 are skipped.

7. If the FL procedure continues, FL Server NWDAF determines FL Client NWDAF as described in clause 6.2C.2.2 and sends Nnwdaf\_MLModelTraining\_Subscribe or Nnwdaf\_MLModelTrainingInfo\_Request that includes the aggregated ML model information to selected FL Client NWDAF(s) for next round of Federated Training. The request may also include the ML Model Accuracy Check Flag, that indicates the FL Client NWDAF(s) to use the local training data as the testing dataset to calculate the Model Accuracy of the global ML model provided by the FL Server NWDAF.

8. Each FL Client NWDAF updates its own ML model based on the aggregated ML model information distributed by the FL Server NWDAF at step 7.

NOTE 3: The steps 3-8 should be repeated until the training termination condition (e.g. maximum number of iterations, or the result of loss function is lower than a threshold) is reached.

When the Federated Training procedure is complete, the FL Server NWDAF requests the FL client NWDAF(s) to terminate the FL procedure by invoking Nnwdaf\_MLModelTraining\_Unsubscribe service with a cause code that the FL process has finished and optionally with the final aggregated ML model information. Then the FL client NWDAF(s) terminate the local model training and if the final aggregated ML model information is received from the FL server NWDAF, the FL client NWDAF(s) can store it for further use.

After the training process is complete, the FL Server NWDAF may send Nnwdaf\_MLModelProvision\_Notify that includes the globally optimal ML model information to the consumer.

\* \* \* Next Changes \* \* \*

### 7.10.4 Nnwdaf\_MLModelTraining\_Notify service operation

**Service operation name:** Nnwdaf\_MLModelTraining\_Notify

**Description:** NWDAF notifies the consumer instance of the trained ML model that has subscribed to the specific NWDAF service. The NWDAF can also use this service to indicate to consumer it will terminate the ML model training.

**Inputs, Required:**

- Notification Correlation Information: this parameter indicates the Notification Correlation ID that has been assigned by the consumer during ML model training.

**Inputs, Optional:**

- Set of the tuple (Analytics ID, ML model Information as defined in clause 6.2F.2;

- ML Correlation ID, when for Federated Learning;

- Corresponding Use case context;

- Termination Request: this parameter indicates that NWDAF requests to terminate the ML model training, i.e. NWDAF will not provide further notifications related to this request, with cause code (e.g. NWDAF overload, not available for the FL process anymore, etc.);

- ML Model ID: this parameter identifies the provisioned ML model;

- Global ML Model Accuracy: The model accuracy of the global ML model, which is calculate by the FL Client NWDAF using the local training data as the testing dataset;

- Status report of FL training: local ML Model metric and Training Input Data Information (e.g. areas covered by the data set, sampling ratio, maximum/minimum of value of each dimension, etc.), which are generated by the FL Client NWDAF during FL procedure;

- Delay Event Notification: as defined in clause 6.2F.2;

- Iteration round ID.

NOTE: The detail reasons in the cause code are up to stage 3.

**Outputs, Required:** Operation execution result indication.

**Outputs, Optional:** None.

\* \* \* Next Changes \* \* \*

### 7.11.2 Nnwdaf\_MLModelTrainingInfo\_Request service operation

**Service operation name:** Nnwdaf\_MLModelTrainingInfo\_Request

**Description:** Request information about NWDAF ML model training with specific parameters.

**Inputs, Required:**

- Analytics ID as defined in Table 7.1-2.

- ML Model Interoperability information.

- ML Model ID: identifies the provided ML model.

**Inputs, Optional:**

- ML Model Information (i.e. file address (e.g. URL or FQDN) of ML Model that needs to update).

- ML Training Information (i.e. data availability requirement, time availability requirement).

- ML Preparation Flag.

- ML Model Accuracy Check Flag.

- ML Correlation ID.

- Termination Request, when terminating the Federated Learning identified by the ML Correlation ID and optionally indicating the reason, e.g. FL Client NWDAF is unselected by the FL Server NWDAF for the FL process, or the FL process is suspended, etc.

- Training Filter Information.

- Use case context.

**Outputs Required:** When the request is accepted: Operation execution result indication. When the request is not accepted, an error response with cause code (e.g. NWDAF does not meet the ML training requirements, ML training is not complete, NWDAF overload, not available for the FL process anymore, etc.).

NOTE: The detail reasons in the cause code are up to stage 3.

**Outputs, Optional:**

- ML Model ID.

- Set of the tuple (Analytics ID, ML model Information (i.e., file address (e.g. URL or FQDN) of updated ML Model).

- ML Correlation ID, when for Federated Learning.

- Corresponding Use case context.

- Global ML Model Accuracy: The model accuracy of the global ML model, which is calculate by the FL Client NWDAF using the local training data as the testing dataset;

- Status report of FL training: local ML model metric and Training Input Data Information (e.g. areas covered by the data set, sampling ratio, maximum/minimum of value of each dimension of data, etc.), which are generated by the FL Client NWDAF during FL procedure.

- Delay Event Notification with the following parameters:

- delay event indication: this parameter indicates that the FL Client NWDAF is not able to complete the training of the interim local ML model within the maximum response time provided by the FL Server NWDAF.

- [OPTIONAL] cause code (e.g. local ML model training failure, more time necessary for local ML model training, etc.).

- [OPTIONAL] the expected time to complete the training.

\* \* \* End of Changes \* \* \*