3GPP TSG-RAN WG3 #114-e R3-216100

**E-meeting, 1st – 11th November 2021**

Source: CATT

Title: (TP for 37.817) AI/ML on Mobility

Agenda Item: 18.4.3

Document for: Discussion and decision

# Introduction

This contribution captures the agreement on AI for mobility based on R3-215911.

# TP for TR 37.817 (on the basis of v0.3.0)

//////////////////////////////////////////////////////////////////////skip unrelated text//////////////////////////////////////////////////////////////////////

### 5.3.1 Use case description

Mobility management is the scheme to guarantee the service-continuity during the mobility by minimizing the call drops, RLFs, unnecessary handovers, and ping-pong. For the future high-frequency network, as the coverage of a single node decreases, the frequency for UE to handover between nodes becomes high, especially for high-mobility UE. In addition, for the applications characterized with the stringent QoS requirements such as reliability, latency etc., the QoE is sensitive to the handover performance, so that mobility management should avoid unsuccessful handover and reduce the latency during handover procedure. However, for the conventional method, it is challengeable for trial-and-error-based scheme to achieve nearly zero-failure handover. The unsuccessful handover cases are the main reason for packet dropping or extra delay during the mobility period, which is unexpected for the packet-drop-intolerant and low-latency applications. In addition, the effectiveness of adjustment based on feedback may be weak due to randomness and inconstancy of transmission environment. Besides the baseline case of mobility, areas of optimization for mobility include dual connectivity, CHO, and DAPS, which each have additional aspects to handle in the optimization of mobiltity.

Mobility aspects of SON that can be enhanced by the use of AI/ML include

* Reduction of the probability of unintended events
* UE Location/Mobility/Performance prediction
* Traffic Steering

**Reduction of the probability of unintended events associated with mobility.**

Examples of such unintended events are:

* Intra-system Too Late Handover: A radio link failure (RLF) occurs after the UE has stayed for a long period of time in the cell; the UE attempts to re-establish the radio link connection in a different cell.
* Intra-system Too Early Handover: An RLF occurs shortly after a successful handover from a source cell to a target cell or a handover failure occurs during the handover procedure; the UE attempts to re-establish the radio link connection in the source cell.
* Intra-system Handover to Wrong Cell: An RLF occurs shortly after a successful handover from a source cell to a target cell or a handover failure occurs during the handover procedure; the UE attempts to re-establish the radio link connection in a cell other than the source cell and the target cell.
* Successful Handover: During a successful handover, there is underlying issue.

### 5.3.2 Solutions and standard impacts

*Editor Note: Capture the solutions for the use case, including potential standard impacts on existing Nodes, functions, and interfaces*

## 5.3.2.1 Locations for AI/ML Model Training and AI/ML Model Inference

Considering the locations of AI/ML Model Training and AI/ML Model Inference for mobility solution, the following two options are considered:

* The AI/ML Model Training function is deployed in OAM, while the Model Inference function resides within the RAN node
* Both the AI/ML Model Training function and the AI/ML Model Inference function reside within the RAN node

Furthermore, for CU-DU split scenario, following option is possible:

* AI/ML Model Training is located in CU-CP or OAM, and AI/ML Model Inference function is located in CU-CP

## 5.3.2.2 AI/ML Model Training in OAM and AI/ML Model Inference in NG-RAN node



 Figure 5.3-1 AI/ML Model Training in OAM and AI/ML Model Inference in NG-RAN node

Step 1. The NG-RAN node configures the MDT measurement information on the UE side and sends configuration message to UE including configuration information.

Step 2. The UE collects the indicated measurement, e.g., UE measurements related to RSRP, RSRQ, SINR of serving cell and neighbouring cells.

Step 3. The UE sends MDT measurement report message to OAM via NG-RAN node including the required measurement.

Step 4. Model Training. Required measurements are leveraged to training ML model for trajectory prediction.

Step 5. OAM sends ML Model Deployment Message to deploy the trained/updated ML model into the NG-RAN node(s). The NG-RAN node can also continue model online training based on the received AI/ML model from OAM.

Note:The definition of online training is FFS.

Step 6. The NG-RAN node obtains the measurement report as inference data for UE mobility optimization.

Step 7. Model Inference. Required measurements are leveraged into Model Inference to output the prediction, e.g. UE trajectory prediction, target cell prediction, target NG-RAN node prediction, etc.

Step 8. The NG-RAN sends the AI/ML model performance feedback to OAM.

Step 9. According to the prediction, recommended actions or configuration are executed for Mobility Optimization.

## 5.3.2.3 AI/ML Model Training and AI/ML Model Inference in a NG-RAN node



Figure 5.3-2: Model Training and Model Inference both located in RAN node

Step 1. The NG-RAN node1 configures the measurement information on the UE side and sends the configuration message to the UE including configuration information.

Step 2. The UE collects the indicated measurement, e.g., UE measurements related to RSRP, RSRQ, SINR of serving cell and neighbouring cells.

Step 3. The UE sends measurement report message to NG-RAN node1 including the required measurement.

Step 4. Model training. Required measurements are leveraged to training the ML model for mobility optimization.Step 5. The NG-RAN node1 obtains the measurement report as inference data for real-time UE mobility optimization.

Step 6. Model Inference. Required measurements are leveraged into Model Inference to output the prediction, including e.g., UE trajectory prediction, target cell prediction, target NG-RAN node prediction, etc.

Step 7. According to the prediction, recommended actions are executed for Mobility Optimization.

Step 8. The NG-RAN node 1 send handover request message to the NG-RAN node 2.

## 5.3.2.4 Input data

The following data is required as input data for mobility optimization.

**Input Information from the UE:**

* FFS UE historical location information from MDT, e.g., Latitude, longitude, altitude, cell ID
* Radio measurements related to serving cell and neighbouring cells associated with UE location information, e.g., RSRP, RSRQ, SINR
* UE historical serving cells and their locations
* Moving velocity
* FFS predicted traffic

**Input Information from the neighbouring RAN nodes:**

* UE’s successful handover information in the past and received from neighbouring RAN nodes
* UE’s history information from neighbour
* Position, resource status, FFS QoS parameters of historical HO-ed UE (e.g., loss rate, delay, etc.)
* Resource status and utilization prediction/estimation
* SON Reports of handovers that are successful, too-early, too-late, or handover to wrong (sub-optimal) cell
* Information about the performance of handed over UEs
* Resource status prediction

**Input Information from the local node:**

* UE trajectory prediction output (will be used by the RAN node internally)
* Local Resource status prediction

If existing UE measurements are needed by a gNB for AI/ML-based mobility optimization, RAN3 shall reuse the existing framework (including MDT and RRM measurements). FFS on whether new UE measurements are needed.

## 5.3.2.5 Output data

* UE trajectory prediction (Latitude, longitude, altitude of UE over a future period of time)
* FFS whether the UE trajectory prediction is an internal output to the node hosting the Model Inference function
* Estimated arrival probability in CHO and relevant confidence interval
* Predicted handover target node, candidate cells in CHO, may together with the confidence of the predication

##  5.3.2.6 Standard impact

To improve the mobility decisions at a gNB (gNB-CU), a gNB can request mobility feedback from a neighbouring node. Details of the procedure are FFS.

* **Potential Xn interface impact:**
	+ Delivery of the UE mobility prediction from the source NG-RAN node to the target NG-RAN node;
	+ Predicted resource status info from candidate target NG-RAN node to source NG-RAN node
	+ Performance Feedback of the received UE mobility/performance prediction from the target NG-RAN node to the source NG-RAN node(FFS).

//////////////////////////////////////////////////////////////////////end//////////////////////////////////////////////////////////////////////