3GPP TSG-RAN WG3 Meeting #125bis R3-245736

Hefei, China, 14th – 18th Oct 2024

Agenda Item: 20.2

Source: Ericsson (moderator)

Title: Summary of Offline Discussion on CB: # AIPHY

Document for: Approval

# Introduction

**CB: # AIPHY**

**- Discuss the open issues above**

**- Capture agreements and open issues**

(moderator - E///)

# For the Chairman’s Notes

# Discussion-First round

Below are the agreements taken during the online session:

|  |
| --- |
| **Consider aggregated gNB case first, then split architecture.**  Case 3b:   * **Re-using the Measurement procedure defined in NRPPa to support requesting and reporting the related measurements to LMF for Case 3b. The details of the related measurements are still pending and subject to further discussion in RAN1.**   Case 3a:   * **The AI/ML model (inference and training functions) is at the gNB.**   Open issues:   1. **Model Training: Label information as training data from LMF to gNB? Data collection from UE or TRP or from LMF?** 2. **Model performance monitoring: OptionA or OptionB?**  * **Option A. gNB performs monitoring metric calculation for its own model.** * **Option B. LMF performs monitoring metric calculation for the model located at the gNB.** |

## Model training for case 3a

The following was captured as “To be continued”:

**Model Training: Label information as training data from LMF to gNB? Data collection from UE or TRP or from LMF?**

Agreements from RAN1 on this topic are as follows:

|  |
| --- |
| Agreement  For training data generation of AI/ML based positioning Case 3a and 3b, the measurement and its related data (e.g., timestamp) are generated by TRP/gNB.  Agreement  For training data generation of AI/ML based positioning Case 3a, the label and its related data (e.g., time stamp) can be generated by at least:   * LMF   Note: transfer of label and its related data is out of RAN1 scope.  Note: whether other network entities can generate label for Case 3a is out of RAN1 scope. |

|  |  |  |
| --- | --- | --- |
| **Case** | **Entities for training data generation** | |
| **Part A** | **Part B** |
| Case 1 | PRU  Non-PRU UE | * PRU * Non-PRU UE with estimated location * LMF |
| Case 2a |
| Case 2b |
| Case 3a | TRP/gNB | At least LMF |
| Case 3b | * PRU * Non-PRU UE with estimated location * LMF |

In light of the agreements above, it is proposed to agree to the following:

Proposals:

**Proposal 1: LMF provides training data to NG-RAN (part B information). FFS on the content of the label (e.g., UE location of PRU). FFS on the signalling design.**

**Proposal 2: How does the LMF get the label info is outside of RAN3 scope**

**Proposal 3: NG-RAN gets training data from TRP (part A information). The content of measurements (e.g., existing or new measurements) is pending RAN1 decisions.**

## Model performance monitoring: OptionA or OptionB?

RAN1´s agreements that are relevant in this respect are listed below:

|  |
| --- |
| Agreement  For AI/ML positioning Case 3a, for performance monitoring metric calculation in label-based monitoring, from RAN1 perspective, Option A and Option B are feasible,   * Option A. NG-RAN node performs monitoring metric calculation for its own model. * Option B. LMF performs monitoring metric calculation for the model located at the NG-RAN node.   Note: Final selection of Option A and Option B is out of RAN1 scope. Potential support of Option A and/or Option B is pending RAN3 confirmation.  Note: Exact method to perform monitoring metric calculation is up to implementation.  Note: For Option A, RAN1 assumes that user data privacy needs to be preserved. |

From the online discussion, the results below represent companies’ positions:

|  |
| --- |
| **Option A. gNB performs monitoring metric calculation for its own model.**  **Option B. LMF performs monitoring metric calculation for the model located at the gNB.**  E///, ZTE, HW, SS, CATT, Xiaomi, Lenovo: Option A  Nok, CMCC: Both options are feasible. Not ready to make decision.  NEC: Slightly prefer OptB, fine with OptA.  Lenovo: Both options are feasible.  The node owns the model should perform the model performance monitoring. |

It is worth noting that “model monitoring”, in the context of the RAN1 agreements, consists of monitoring the model performance so to trigger actions like model re-training/updating.

With this respect, actions like model training/retraining/updating can be only carried out by the node that hosts the model, as that node is aware of the model`s implementation and of how such implementation depending actions can be carried out.

Therefore, we believe that Option A is the most appropriate option as it keeps model monitoring at the gNB (which is the node hosting the model).

We also would like to point out that the gNB needs to receive from the LMF Part B of the training data (i.e. the ground truth for the gNB model predictions). The very same information can be used by the gNB to monitor the model performance. Therefore, there is no impact on the standard in supporting Option A.

Additionally, we observe that 8 companies prefer option A, and 2 are fine with both options.

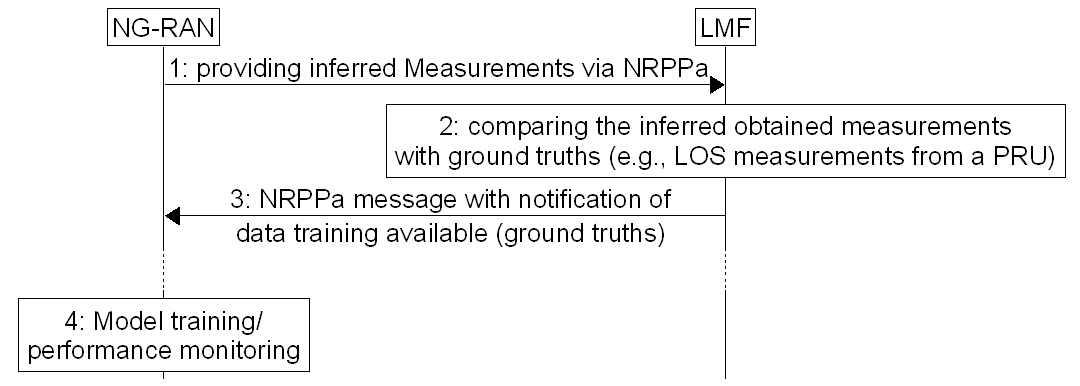
For the above reasons it is proposed to take a WA to move forward:

**Proposal 4: Take WA: RAN3 will proceed with Option A: NG-RAN node performs monitoring metric calculation for its own model.**

Based on RAN1´s agreements, it should be possible for the NG-RAN to receive feedback information (Part B) from the LMF on the inferred AI/ML measurements, which could be helpful to assist the NG-RAN for model retraining (if needed).

|  |  |  |
| --- | --- | --- |
| **Case** | **Entities for training data generation** | |
| **Part A** | **Part B** |
| Case 1 | PRU  Non-PRU UE | * PRU * Non-PRU UE with estimated location * LMF |
| Case 2a |
| Case 2b |
| Case 3a | TRP/gNB | At least LMF |
| Case 3b | * PRU * Non-PRU UE with estimated location * LMF |

Such feedback information can be based on, e.g., the LMF comparing the derived UE location (based on the received inferred measurements) with the ground truth (e.g., UE location coming from PRU/LoS measurements). The NG-RAN could use the feedback information from LMF to identify if the predictions made are of bad quality and if the model should be retrained.



**Proposal 5: The NG-RAN hosting the AI/ML model can receive feedback information from the LMF on the reported inferred measurements. The feedback information can indicate possibility of retraining the model based on providing ground truths.**

Therefore, based on NG-RAN own’s model monitoring, a gNB can request or not for training data from the LMF.

One further point concerns how an LMF requests for inferred measurements from the NG-RAN.

It seems plausible to assume that the LMF shall be able to trigger a request for inferred positioning measurements towards a gNB or to request a gNB to stop reporting inferring measurements (and instead to, e.g., report legacy measurements).

**Proposal 6: the LMF is able to trigger reporting of inferred positioning measurements from a gNB and to stop such reporting**

## Impacted RAN3 specifications

The following RAN3 specifications are foreseen to be impacted:

* NRPPa: obviously 😊
* F1AP: to support split case (later) in both case 3a and 3b
* TS 38.305: stage 2 positioning including new flow-charts
* TS 38.413: NRPPa is transported over NGAP, if new procedures are defined, they need to be mentioned in NGAP
* TS 38.470: for new procedures, if applicable
* TS 38.300: for new AI/ML model LMF use case
* TS 38.401: if applicable

Proposed allocation of BL CRs:

|  |  |
| --- | --- |
| **BL CR** | **Company** |
| NRPPa | E/// |
| F1AP | Xiaomi |
| 38.413 | Huawei |
| 38.305 | CATT |
| 38.300 | Nokia |
| 38.401 | ZTE |
| 38.470 | CMCC |
| ??? | QC |

# Conclusion, Recommendations

# References