3GPP TSG-RAN WG3 Meeting #125 R3-244734

**Maastricht, Netherlands, 19 – 23 August, 2024**

**Agenda item: 11.2**

**Source: Nokia, Jio, Huawei, ZTE Corporation, Ericsson, Telecom Italia, InterDigital, NEC,CATT, Lenovo**

**Title:** **(TP to TR 38.743) AI/ML Network Slicing**

**Document for:** **Text Proposal**

# 1 Introduction

This paper provides a TP to TR 38.743 to capture the following agreement from the offline session CB: # AIRAN1\_Slicing:

**Slice UE performance needs to be introduced, while the granularity to support slice UE performance can be further checked in WI phase.**

# Annex TP for TR 38.743

The TP is based on TR 38.743 v1.1.0.

<<< start of changes >>>

## 4.1 AI/ML based Network Slicing

### 4.1.1 Use case description

Support of network slicing in NG-RAN is defined in TS38.300 [x].

The NG-RAN plays a key role in taking mobility, load balancing and Radio Resources Management decisions for the purpose of meeting target requirements derived from the SLA of each supported network slice.

AI/ML function can analyze metrics related to network and UE level performance related to perform optimal resource management and mobility decisions for network slicing to meet the requirements.

### 4.1.2 Solutions and standard impacts

4.1.2.1 Locations for AI/ML Model Training and AI/ML Model Inference

The following solutions can be considered for supporting AI/ML-based network slicing:

- AI/ML Model Training is located in the OAM and AI/ML Model Inference is located in the gNB.

- AI/ML Model Training and AI/ML Model Inference are both located in the gNB.

In case of CU-DU split architecture, the following solutions are possible:

- AI/ML Model Training is located in the OAM and AI/ML Model Inference is located in the gNB-CU.

- AI/ML Model Training and Model Inference are both located in the gNB-CU.

4.1.2.2 Input data of AI/ML based Network Slicing

To predict the optimized network slicing decisions, a gNB may need the following information as input data for AI/ML-based network slicing:

From local node:

- Measured/Predicted radio resource status per slice

- Measured/Predicted slice available capacity

- Legacy predicted UE trajectory

From neighbouring gNBs:

- Measured/Predicted radio resource status per slice

- Measured/Predicted slice available capacity

From the UE:

- UE measurement report (e.g., UE RSRP, RSRQ, SINR measurement, etc), including cell level and beam level UE measurements

4.1.2.3 Output data of AI/ML based Network Slicing

AI/ML-based network slicing model in a gNB can generate the following information as output:

* Predicted radio resource status per slice
* Predicted slice available capacity
* Resource management decisions for resources within RRM policies (used by gNB internally)
* Slice aware mobility decisions (used by gNB internally)

4.1.2.4 Feedback of AI/ML based Network Slicing

To optimize the performance of AI/ML-based network slicing model, the following feedback can be considered to be collected from gNBs:

* Measured Radio resource status per slice
* Measured Slice available capacity
* Legacy UE performance feedback for those UEs handed over from the source gNB
* Finer granularity UE performance feedback for those UEs handed over from the source gNB to determine UE performance for a certain slice in use by a certain UE.

 Note: Exact level of finer granularity to be defined in normative phase.

4.1.2.5 Potential standard impacts

Following standard impacts are listed for subsequent Rel-19 normative work compared with what was specified during Rel-18.

Xn interface:

- Enhanced existing procedure to collect predicted information between gNBs:

■ Predicted radio resource status per slice

■ Predicted slice available capacity

<<< end of changes >>>