3GPP TSG-RAN WG3 Meeting #123bis R3-242160

Changsha, China, from April 15 to April 19 2024

Agenda Item: 11.2

Source: ZTE, Ericsson

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

Document for: other

# 1 Introduction

This TP follows discussions in R3-24xxxx.

# 2 Text Proposal

<<<<<<<<<<<<<<<<<<<< First Change >>>>>>>>>>>>>>>>>>>>

## 4.1 AI/ML based Network Slicing

### 4.1.1 Use case description

*Editor Note: Capture the description of use case*

Network slicing is defined as a technology that logically separates network functions and resources into multiple network slices within a common physical infrastructure. Each NS represents an independent virtualized end-to-end network, providing tailored service for a specific communication scenario (e.g., eMBB, mMTC, or URLLC).

Network slicing can provide service differentiation and ensure SLAs (Service Level Agreements) for each service type by allowing the establishment of custom slices for various types of services with different QoS requirements.  
The NG-RAN plays a key role in taking mobility and Radio Resources Management decisions for the purpose of meeting target requirements mapping to the SLA of each supported network slice. In this context AI/ML comes as a powerful enabler for coordination of actions across NG-RAN nodes with the final goal of fulfilling the NG-RAN level requirements derived from the SLAs of the supported slices.

AI/ML algorithms can analyze network performance, UE level performance, UE level service requirements and network level LSA requirements to perform optimal resource management and selection of actions for network slicing.

### 4.1.2 Solutions and standard impacts

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

#### 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.

Note: gNB is also allowed to continue model training based on AI/ML model trained in the OAM

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.

#### 4.1.2.4 Input data of AI/ML based Network Slicing:

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

From local node:

- Current/Predicted resource status per slice

- Current/Predicted slice available capacity

- Legacy predicted UE trajectory

From neighbouring NG-RAN nodes:

- Current/Predicted resource status per slice

- Current/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

#### 4.1.2.5 Output data of AI/ML based Network Slicing:

AI/ML-based network slicing model in NG-RAN node can generate following information as output:

* Predicted resource status per slice
* Predicted slice available capacity
* Resource management decisions for resources within RRM policies (Internal output)
* Slice aware mobility decisions (Internal output)

#### 4.1.2.6 Feedback of AI/ML based Network Slicing:

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

* Resource Status per slice level updates from target NG-RAN node
* Slice Available Capacity updates from target NG-RAN node
* Legacy UE performance feedback for those UEs handed over from the source NG-RAN node

#### 4.1.2.7 Potential standard impacts:

Following standard impacts is 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 NG-RAN nodes:
  + Predicted Resource Status per slice level between neighbouring NG-RAN nodes and source NG-RAN node
  + Predicted Slice Available Capacity between neighbouring NG-RAN nodes and source NG-RAN node

<<<<<<<<<<<<<<<<<<<< End of Changes >>>>>>>>>>>>>>>>>>>>