**3GPP TSG-SA5 Meeting #157 *S5-245980***

Hyderabad, India, 14 - 18 October 2024

**Source: ZTE Corporation, Huawei**

**Title: DP on data collection for UE-side Model Training**

**Document for: Endorsement**

**Agenda Item: 5.3**

# 1 Decision/action requested

***The group is asked to endorse the detailed proposal in section 4.***

# 2 References

[1] RP-242389: “LS on AIML data collection”

# 3 Rationale

## 3.1 Introduction

According to the “Table 7.2.1.3.2-1” in LS RP-242389 [1], different data collection options for UE-side model training and technics are briefly analyzed. The options of the AI/ML-specific Data Transfer Path include:

* Option 1a: UE to OTT server via either 3GPP or non-3GPP network
* Option 1b: UE ->Server for data collection for UE -side model training/OTT server
* Option 2: UE-> CN -> Server for data collection for UE-side model training/OTT server
* Option 3: UE-> gNB->OAM -> Server for data collection for UE-side model training/OTT server

The analysis of the above options and the impact on SA5 are shown as follows.

## 3.2 Observation

#### 3.2.1 Data collected for UE-side model training

Management of UE is out of the scope of SA5 and data collected for ML model training is not specified in the ML model lifecycle in TS 28.105. MnS Consumer can indicate the address(es) of the candidate training data source but the detailed training data format is vendor-specific.

In RP-242389 [1], the content of UE-side data collection is not elaborated. SA5 has defined the following measurement collection mechanism regarding UE data collection.

* MDT. UE level measurements for the 5G system are documented in TS 28.558. NG-RAN UE level measurements collection and reporting are supported by MDT mechanisms for both signalling and management-based activation, which are detailed in clause 4 of TS 32.422.
* QoE measurements. The measurements collected are DASH, MTSI, and Virtual Reality measurements, which include collecting QoE data from UEs in the specified area in case of Management Based Activation and an individual UE in case of Signalling Based Activation. The corresponding collection procedure and QoE measurement collection record content are detailed in TS 28.405 and TS 28.406, respectively.

**Observation 1: The data collected for UE-side model training is not specified in RP-242389. SA5 has defined two types of data collected from UE, which includes MDT and QoE measurements. The content of data collected for UE-side model training depends on the definition of RAN 1/2.**

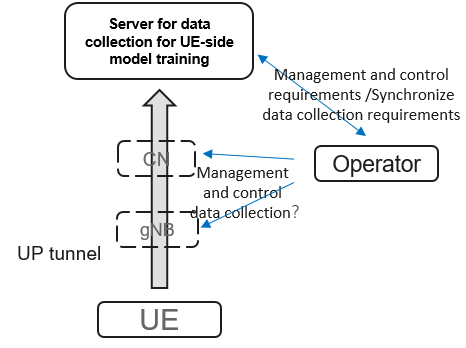
#### 3.2.2 Impact of option 1a/1b/2 on SA5

The impacts of option 1a, option 1b, and option 2 on SA5 are analyzed as follows:

**Option 1a: UE to OTT server via either 3GPP or non-3GPP network**

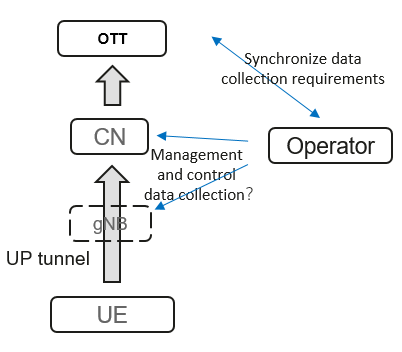
* In case of via 3GPP, considering that the data collection is transparent between UE and OTT server as described in clause 7.2.1.3.2 in TR 38.843. There is no management and control requirements for SA5.
* In case of via non-3GPP, there is no management and control requirements for SA5.

**Option 1b: UE ->Server for data collection for UE-side model training/OTT server**

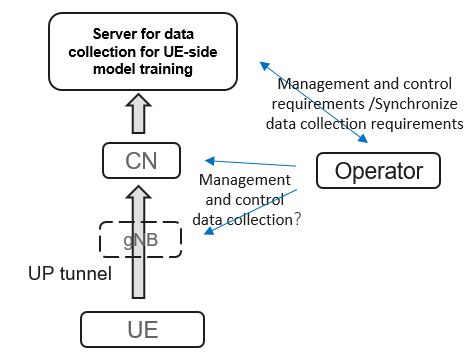


* In case of endpoint is OTT server, there is no management and control requirements for SA5.
* In case of endpoint is “Server for data collection for UE -side model training”, there is management and control requirements for SA5. Operator needs to configure the Server for data collection for UE-side model training/CN/gNB to support the data collection. If the server is not control by MNO, synchronize data collection requirements between the OTT server and the Operator may be required.

**Option 2: UE-> CN -> Server for data collection for UE-side model training/OTT server**



* In case of endpoint is OTT server for CP tunnel and/or UP tunnel, there is management and control requirements for SA5. Operator configures the CN/gNB to support the data collection. Synchronization data collection requirements between the OTT server and the Operator may be required. Data collection transfer process needs to be captured in SA5.



* In case of endpoint is “Server for data collection for UE -side model training”, if the server is fully controlled by MNO, operator configures the Server for data collection for UE -side model training/CN/gNB to support the data collection. Data collection transfer process needs to be captured in SA5. If the server is not control by MNO, operator configures the CN/gNB to support the data collection. Synchronize data collection requirements between the OTT server and the Operator may be required. Data collection transfer process needs to be captured in SA5.

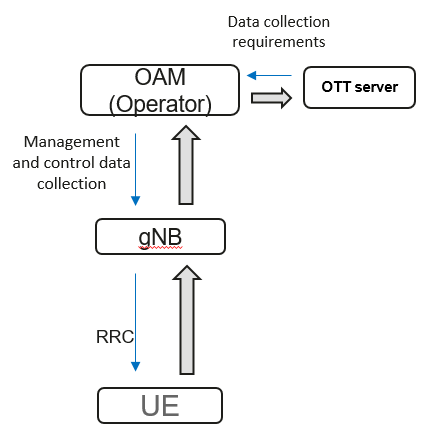
**Observation 2: For option 1b and option 2, there may be management and control requirements for SA5.**

#### 3.2.3 Analysis on Option 3 UE-> gNB->OAM -> Server for data collection for UE-side model training/OTT server

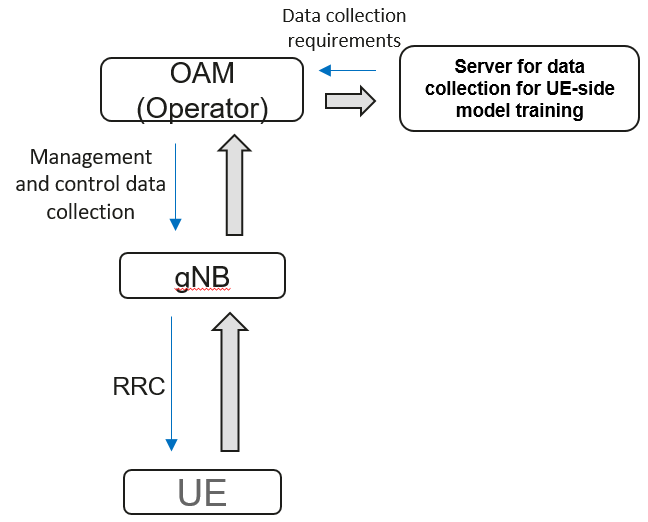
Analysis on aspects of option 3 is as follows:

* **First termination entity & AI/ML-specific Data Transfer Path**

In option3 path, the termination entity is Server for data collection for UE-side model training/OTT server. For MDT, the termination entity can be the data consumer (e.g., TCE) to collect data directly or via the management system. For QoE Measurements, the termination entity can be the Measurement Collector Entity (MCE) to collect data. These termination entity for MDT and QoE measurements can be fully controlled by OAM.



In case of endpoint is the Server controlled by MNO, operator has the control of the standardized data collection transfer process for gNB by extending existing UE data collection mechanism (e.g., MDT/QoE measurements). Data collection requirements between the Server for data collection for UE-side model training and the Operator can be internal implementation.



In case of endpoint is the OTT server or the Server not controlled by MNO, operator has the control of the data collection transfer process for gNB by extending existing UE data collection mechanism (e.g., MDT/QoE measurements). Data collection requirements between the OTT server and the Operator may be required.

* **UP/CP tunnel & Protocol layer for data transfer & Controllability of MNO on data transfer & Solution for network controllability**
* CP Tunnel. For the existing UE data collection mechanism (MDT, QMC) above, only the control plane tunnel is supported. Data collection configuration parameters and corresponding data measurements reported are within the RRC signalling.
* UP Tunnel. Data collection via user plane is not discussed in SA5.

The existing data collection mechanism only supports data transferred within the RRC signalling in the CP tunnel, which is fully controlled.

* **Possible Options for Visibility of data content in MNO and Data format**
* MDT. Only standardized data can be collected.
* QoE Measurement Collection Content. QoE measurements are contained in the measReportAppLayerContainer. QoE metrics for 3GP-DASH, VR, and MTSI are defined by SA4 in TS 26.247 and TS 26.114.

**Observation 3: The existing data collection mechanism in SA5 only supports to collect standardized data via the CP tunnel.**

## 3.3 Proposed way forward for SA5

The definition of data collected for UE-side model training in the received LS is unclear. According to the analysis above, currently, only standardized metrics can be collected from UE. If QoE metrics are in the scope of data collection for UE-side model training, SA4 should be involved in the discussion. If MDT mechanisms are reused or extended for data collection for UE-side model training, the possible negative impact on gNB needs to be evaluated holistically. For example, the transferred data volume, the number of impacted UE, data collection frequency, and the potential negative impact on the performance of gNB all should be taken into consideration.

SA5 propose to ask RAN to clarify the following aspects:

* For all options, clarify which standardized data are collected?
* For option 1b and option 2, there may be management and control requirements for SA5. If there are management support for these two options, the discussion of option1b and option2 should have SA5 involved.
* For option1b, option2 and option3, the endpoint “Server for data collection for UE -side model training” needs further clarification whether it is a “Network entity fully control by MNO” or a “Network entity not control by MNO”.

# 4 Detailed proposal

Proposal #1: Take the three obervations and ralted analysis above into consideration when drafting the LS reply.

Proposal #2: Attach this DP to the LS reply.