**3GPP TSG SA WG5 Meeting #156 S5-244694**

**Maastricht, The Netherlands 19 - 23 August 2024**

**Source: China Mobile, TIM**

**Title: Add solution of Using NDT to generate ML training data for TR 28.915**

**Document for: Approval**

**Agenda Item: 6.19.5**

# 1 Decision/action requested

***In this box give a very clear / short /concise statement of what is wanted.***

# 2 References

[1] 3GPP draft TR 28.915: “Management and orchestration; Study on management aspects of Network Digital Twin v0.3.0”.

[2] SP-231727 "New Study on management aspects of Network Digital Twin"

# 3 Rationale

This contribution proposes to add solution of using NDT to generate ML training data for TR 28.915 based on SP-231727 [2]

# 4 Detailed proposal

|  |
| --- |
| **1st Change** |

# 5 Use cases

## 5.6 Use case 6: Using NDT to generate ML training data

### 5.6.1 Description

For many use cases, ML training requires large amounts of data to guarantee good performance of the ML models. In general, the ML training data for network related use cases is obtained through historical network management data. For instance, assuming that there is a ML model supporting MDA SLS analysis described in TS 28.104 [2] clause 7.2.2, the raw feature of training data could be the enabling data, such as UL/DL throughput, uplink/downlink delay, etc., as specified in clause 8.4.2 of [2].

However, obtaining data from the network has two limitations:

* The quantity of issues happened in actual mobile network is limited.
* The variety of issues happened in actual mobile network is limited. There could be corner network issues cases that hardly happen in real life.

Sufficient ML training data plays a key role to a useful ML model. The more training data provided, the better performance of ML model. To overcome, these challenges, an NDT can be used as simulated data generation entity that simulates the network and its characteristics including problems to generate configuration and performance measurements data which can be used to enrich the ML training data set.

### 5.6.2 Potential requirements

**REQ-NDT-FUN-01** The NDT MnS producer should have the capability to allow an authorized MnS consumer to request generation of simulated network data to be used for ML training.

### 5.6.3 Potential solutions

#### 5.6.3.X solution X

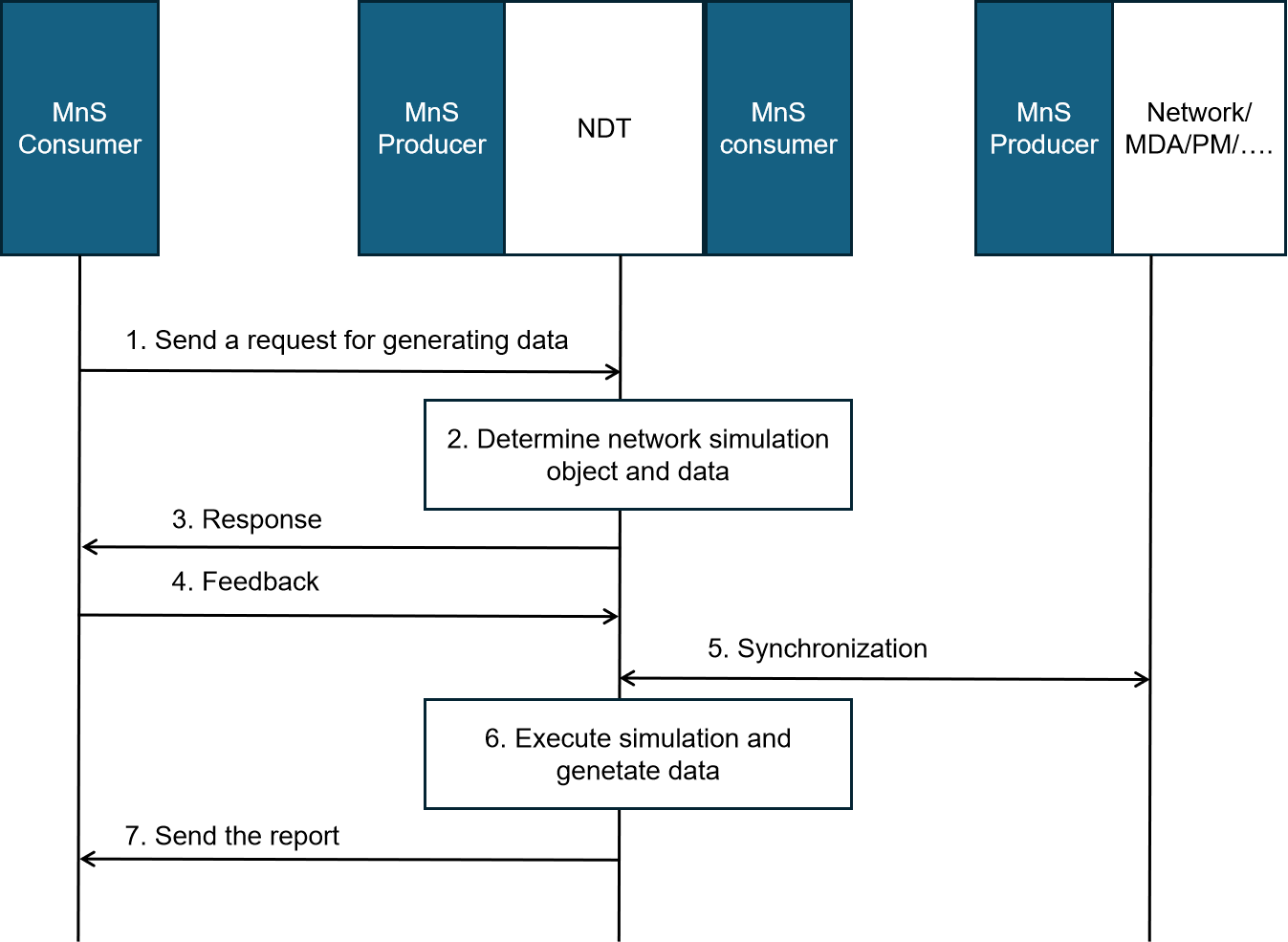


Figure 5.6.3-X: procedure of generating ML training data using NDT

1. The MnS consumer, eg, ML training Function, sends an data generating request to the MnS producer who provides NDT. The request includes the data requirements and simulation requirements for the generated ML training data.

Data requirements may include the following:

- Data source in the simulated network: indicates where the required data is generated from the, e.g., a simulated network element, slice, or subnet.

- Data type: indicates the type of data needed, e.g., PM data, log data, traffic data.

- Data subtype: if there are finer types within a certain type of data, they can be indicated by the data subtype, e.g., performance measurement type.

- Required data period: indicates the time span of the required data (e.g., one day, one week).

- Data sampling periods: indicate the sampling periods that data is collected from the simulated network in NDT, e.g., every 15 minutes, every hour, etc.

NOTE: The ML training data is generated from the simulated network in NDT.

Simulation scope: the area of the actual mobile network or the managed object that needs to be simulated in NDT. For instance, a geography area, a network slice, etc.

Simulation data: the data collected from the managed entities for NDT simulation, e.g., PM data as defined in TS 28.552/28.554, CM data as defined in TS 28.541/28.622, etc.

1. MnS producer receives the request and determines the simulation object and simulation data.
2. The MnS producer sends the response to the MnS consumer including the simulation scope and the estimated time when the ML training data can be provided
3. The MnS consumer sends the feedback to the MnS producer including whether to agree to the MnS producer's configuration of NDT.

If the MnS consumer agrees to the MnS producer's configuration of the NDT, then go to step 5; if not, the MnS producer needs to re-configure the simulation scope and simulation data.

1. MnS producer collects data from the managed entities.
2. MnS producer executes network simulation and generates the data in NDT.
3. The MnS producer sends the report to the MnS consumer. The report content may include the ML training data files or the address address from which the data may be fetched.

|  |
| --- |
| **End of Changes** |