**3GPP TSG-SA5 Meeting #148-e S5-233501rev2**

**e-Meeting, 17th-25th April, 2023**

**Source: China Unicom**

**Title: Add New Solution for configuration of latency for URLLC in RAN over the air interface**

**Document for: Approval**

**Agenda Item: 6.8.3**

# 1 Decision/action requested

***The group is asked to approve the proposal.***

# 2 References

[1] 3GPP TR 28.832 v0.3.0: “Management Aspects of URLLC”

[2] 3GPP TS 28.541: “Management and orchestration; 5G Network Resource Model (NRM);”

# 3 Rationale

It was approved in SP-220146 to study the management aspects of URLLC and one of the objectives is to investigate configuration management related to URLLC. In order to achieve the objective mentioned above, a solution for configuration of latency for URLLC in RAN over the air interface is proposed in this contribution.

End to end latency is an important attribute used to describe the requirement for URLLC. Latency in RAN is part of end to end latency and the corresponding attribute can be used to specify the maximum allowed delay in RAN including the air interface. The definition of latency in RAN is specified in ITU-R M.2410. The description is as follows:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

*4.7.1 User plane latency*

*User plane latency is the contribution of the radio network to the time from when the source sends a packet to when the destination receives it (in ms). It is defined as the one-way time it takes to successfully deliver an application layer packet/message from the radio protocol layer 2/3 SDU ingress point to the radio protocol layer 2/3 SDU egress point of the radio interface in either uplink or downlink in the network for a given service in unloaded conditions, assuming the mobile station is in the active state. This requirement is defined for the purpose of evaluation in the eMBB and URLLC usage scenarios.*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

According to the definition above, the latency in RAN should contain the delay over the air interface which is a crucial part of end to end latency.

In TS 28.541, RANSliceSubnetProfile is used to represent the requirements for RAN slice profile. The documentation of “RANSliceSubnetProfile.dLLatency” and “RANSliceSubnetProfile.uLLatency” in TS 28.541is shown as follows.



 According to the description above, “RANSliceSubnetProfile.dLLatency” and “RANSliceSubnetProfile.uLLatency” only represent the packet processing latency within gNB exluding that of air interface. So the issue is that the existing attributes in RANSliceSubnetProfile cannot match the configuration requirement for latency in RAN as part of end to end latency.

In order to address the above issue, we propose the following solution by modifying the documentation of “RANSliceSubnetProfile.dLLatency” and “RANSliceSubnetProfile.uLLatency” to contain the latency over the air interface.

# 4 Detailed proposal

This contribution proposes to make the following changes in [1].

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

# 2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non‑specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TS 22.104: “Service requirements for cyber-physical control applications in vertical domains; Stage 1”

[3] 3GPP TS 22.261: “Service requirements for the 5G system; Stage 1”

[4] 3GPP TS 38.211: “NR; Physical channels and modulation”

[5] 3GPP TS 38.212: “NR; Multiplexing and channel coding”

[6] 3GPP TS 38.213: “NR; Physical layer procedures for control”

[7] 3GPP TS 38.214: “NR; Physical layer procedures for data”

[8] 3GPP TS 38.300: “NR; NR and NG-RAN Overall Description; Stage2”

[9] 3GPP TS 38.321: “NR; Medium Access Control (MAC) protocol specification”

[10] 3GPP TS 38.323: “NR; Packet Data Convergence Protocol (PDCP) specification”

[11] 3GPP TR 38.824: “Study on physical layer enhancements for NR ultra-reliable and low latency case (URLLC)”

[12] 3GPP TS 28.541: “Management and orchestration; 5G Network Resource Model (NRM);”

[13] 3GPP TS 28.552: “Management and orchestration; 5G performance measurements”

[14] 3GPP TS 28.554: “Management and orchestration; 5G end to end Key Performance Indicators (KPI)”

[15] 3GPP TS 38.314: “NR; Layer 2 Measurements;”

[16] ITU-R M.2410: “Minimum requirements related to technical performance for IMT-2020 radio interface(s)”

|  |
| --- |
| **2nd Change** |

5.X Issue #X: Configiration of latency for URLLC in RAN including the air interface

### 5.5.1 Description

URLLC is a service with requirement of ultra-reliability and low latency. End to end latency is an important attribute used to describe the requirement for URLLC service. Latency in RAN is part of end to end latency and the corresponding attribute can be used to specify the maximum allowed delay in RAN including the air interface. The definition of latency in RAN is specified in ITU-R M.2410. The description is as follows:

*4.7.1 User plane latency*

*User plane latency is the contribution of the radio network to the time from when the source sends a packet to when the destination receives it (in ms). It is defined as the one-way time it takes to successfully deliver an application layer packet/message from the radio protocol layer 2/3 SDU ingress point to the radio protocol layer 2/3 SDU egress point of the radio interface in either uplink or downlink in the network for a given service in unloaded conditions, assuming the mobile station is in the active state. This requirement is defined for the purpose of evaluation in the eMBB and URLLC usage scenarios.*

According to the definition above, the latency in RAN should contain the delay over the air interface which is a crucial part of end to end latency. In TS 28.541, the attribute “dlLatency” and “ulLatency” in RANSliceSubnetProfile only represent the packet processing latency in gNB exluding that of air interface. The performance of delay over the air interface can’t be evaluated simply through the existing latency configuration. Concequently, the issue is that the existing attributes in RANSliceSubnetProfile cannot match the configuration requirement for latency when taking air interface into consideration.

5.5.X Potential Solution

The requirement for configuration management of radio network providing URLLC services is:

- The OAM should have the capability of configuring the maximum allowed downlink and uplink latency in RAN including the delay over air interface.

According to the requirement above, the attributes “dlLatency” and “ulLatency” of RANSliceSubnetProfile defined in TS 28.541should be modified to contain the latency over the air interface. The specific changes are proposed.

1. The attributes “dlLatency” and “ulLatency” of RANSliceSubnetProfile<<dataType>> defined in TS 28.541 shall be modified as follows:

| Attribute Name | Documentation and Allowed Values | Properties |
| --- | --- | --- |
| RANSliceSubnetProfile.dlLatency | An attribute specifies the required DL packet transmission latency (millisecond) in RAN including the air interface of the network slice and is used to evaluate the delay between NG-RAN and UE, e.g. time between received DL packet from UPF the packet successfully received by UE. See clause 5.1.1.1.6 in TS 28.552[13]. | type: Realmultiplicity: 1isOrdered: N/AisUnique: N/AdefaultValue: NoneallowedValues: N/AisNullable: False |
| RANSliceSubnetProfile.ulLatency | An attribute specifies the required UL packet transmission latency (millisecond) in RAN including the air interface of the network slice and is used to evaluate the delay between UE and NG-RAN, e.g. time between the UL packet scheduled in UE and the packet successfully sent to UPF. See clause 5.1.1.1.7 in TS 28.552[13]. | type: Realmultiplicity: 1isOrdered: N/AisUnique: N/AdefaultValue: NoneallowedValues: N/AisNullable: False |

NOTE1: Even though this study is about URLLC, the two attributes are also applicable for latency requirement for non URLLC services.

NOTE2: The two attributes represent maximum latency allowed or target latency of RAN slice subnet.

|  |
| --- |
| **End of changes** |