3GPP TSG-RAN WG2 Meeting #113 Electronic R2-21xxxxx

Elbonia, 25 January – 05 February 2021

**Agenda item: 8.5.4**

**Source: Nokia, Nokia Shanghai Bell**

**Title: [DRAFT] Summary of Email Discussion [AT113-e][506][IIoT] QoS RAN enhancements (Nokia)**

**WID/SID: NR\_IIOT\_URLLC\_enh - Release 17**

**Document for: Discussion and Decision**

# 1 Introduction

This document is prepared for the following email discussion in RAN2#113e:

* [AT113-e][506][IIoT] QoS RAN enhancements (Nokia)

Scope:

* + - Identify set of open issues for QoS RAN enhancements that need to be addressed based on company contributions and identify any agreeable aspects to be discussed in the first week session
    - Get company inputs on opens issues (to be kicked off after first session)

Intended outcome:

* + - Set of issues that should be discussed in the first session and any proposals that could be agreeable
    - Set of additional issues that should be addressed but with lower priority

Deadline for providing comments:

* + - Companies comments on the summary: January 25th

Based on the proposals in the summary [20], we have made the following agreements in the online sessions:

**Agreements**

- Assumption: communication service availability is not needed on top of survival time [confirm over email and clarify what CSA is]

*-* RAN2 confirms that specification enhancement for survival time support may only needed for uplink. Downlink is addressed by implementation and no specification impacts.

*-* Support for survival time in UCE is up to network configuration.

Nevertheless, there are many other open issues that have to be discussed and confirmed, including:

* Need of additional new QoS parameters other than survival time
* Traffic Patterns for survival time
* Methods of survival time state monitoring
* Methods of survival time violation avoidance
* UE knowledge of survival time requirement
* TSCAI from UE

Please provide your contact information when responding:

|  |  |  |
| --- | --- | --- |
| **Company** | **Contact Name** | **Email** |
| Nokia | Ping-Heng Wallace Kuo | Ping-Heng.Kuo@nokia.com |
| Ericsson | Zhenhua Zou | Zhenhua.Zou@ericsson.com |
| CATT | Pierre Bertrand | pierrebertrand@catt.cn |

# 2 Main Issues

## 2.1 Need of Additional New QoS Parameters

Previously SA2 has confirmed that Survival Time (ST) will be included as a new QoS parameter in the normative specifications to support requirements of TSC applications, which is optionally provided to RAN as a new TSCAI element. On top of survival time, some companies have proposed that RAN2 should also ask SA2 to provide some other information, including:

1. Communication Service Availability (CSA) [4]
2. Burst Ending Time (BET) [3][6]
3. Burst Spread [5][18][19]
4. Service Reliability [1][3]

For (1) Communication Service Availability, the papers including [1], [5], [8], [9], and [15] have explicitly expressed that CSA is not needed, while only [4] think the need of CSA at RAN is worth considering. During the online sessions, we have the following discussions:

|  |
| --- |
| *Proposal 1a: RAN2 confirms communication service availability is not needed on top of survival time.*  Qualcomm doesn’t agree. Nokia thinks that no matter what CSA is the network should do its best from RAN perspective to not violate survival time. Ericsson also doesn’t think we need it and survival time is optional. CMCC also doesn’t think we need this parameter and doesn’t help the RAN side to perform scheduling. Lenovo, Samsung also thinks like Nokia. Huawei has some sympathy with Qualcomm so the network should know how hard it should it try.  - Qualcomm thinks that it is impossible for the network to meet survival time but how hard should the network try. If we want to replace wireline we need to be very reliable and we need all the possible mechanisms.  - Intel thinks that as long as there is no new requirements from SA1/SA2 we don’t need to define anything new. |

According to the latest agreements that we have reached, it is now assumed in RAN2 that CSA is not needed.

**Question 1: Do you agree RAN2 can confirm the information of communication service availability is NOT needed at RAN on top of survival time ?**

|  |  |  |
| --- | --- | --- |
| **Company** | **YES/NO** | **Comments** |
| Nokia | YES | We must note that CSA is an application layer end-to-end requirement, and communication service interruption could occur in a path segment external to 5G RAN – in this case there is nothing RAN could do and therefore CSA is not useful for RAN to know whether it could relax itself for survival time support. In our view RAN should always make the best effort to avoid communication service interruption in the 5G RAN domain, regardless what the required CSA is - because how many times the applications has failed already (due to e.g. failure in other segments of the end-to-end path) is transparent to 5G RAN anyway. |
| CATT | Yes | We have the same understanding as Nokia and this is SA2 scope anyways. |

For (2) Burst Ending Time and (3) Burst Spread, somehow both of them are relating to the timing that a data burst is (fully) exposed to RAN. Since SA2 is still working on Burst Spread, and the definition of which is not entirely clear, the rapporteur tends to think RAN2 can wait until further clarification from SA2 is available.

**Question 2: Do you agree that RAN2 should wait for SA2 to provide more clear instruction/information relating to Burst Spread, before initiating any discussion on Burst Spread and Burst Ending Time ?**

|  |  |  |
| --- | --- | --- |
| **Company** | **YES/NO** | **Comments** |
| Nokia | YES | We think there is no need for RAN2 to discuss burst spread or burst ending time at this stage, as SA2 is still working on the related topics. |
| CATT | Yes | We also think Burst Spread and Burst Ending Time are overlapping/redundant information so should be discussed in SA2. |

For (4) Service Reliability, which essentially represents the mean time between failures, has also been proposed as a possible new QoS parameter that RAN could utilize. A more formal definition of communication service reliability is:

|  |
| --- |
| **TS 22.104**  **communication service reliability:** ability of the communication service to perform as required for a given time interval, under given conditions. |

RAN2 could discuss if this is needed.

**Question 3: Do you agree that RAN may need the QoS information relating to Service Reliability?**

|  |  |  |
| --- | --- | --- |
| **Company** | **YES/NO** | **Comments** |
| Nokia | NO | Similar to our responses in Q1, this is an end-to-end application requirement, so it doesn’t really mean much to RAN as RAN cannot cope with any failure occur in path segments external to 5G system anyway. Essentially the knowledge of PER and survival time should be sufficient for RAN to do its job properly. |
| CATT | No | We have the same understanding as Nokia and this is SA2 scope anyways. |

## 2.2 Traffic Pattern for Survival Time

According to the LS from SA2 (R2-2010692), cyclic traffic for deterministic applications has been assumed for the work of survival time, which means only periodic traffics should be considered:

|  |
| --- |
| **R2-2010692 / S2-2007880 :**  3GPP SA WG2 has studied a key issue for introducing “*Use of Survival Time for Deterministic Applications in 5GS”* and selected the solutions as the basis for normative work. One of the basic principles of the agreed conclusions is that SMF determines TSCAI Survival Time and sends it to the NG-RAN. The Survival Time in TSCAI may be expressed, assuming cyclic traffic of the deterministic application:  …… |

Among the papers submitted to RAN2 #113e, [6] would like to confirm if this is the common understanding in RAN2. As pointed by [2] and [3], it is indicated in the SA2 TR that survival time is conveyed together with TSCAI periodicity parameter, and hence it is quite clear RAN2 should only consider periodic traffics in Rel-17. During the online session, one company mentioned that it may depends on RAN3’s outcome on how survival time in TSCAI is expressed. However, the rapporteur cannot comprehend why the traffic patterns targeted by SA2 could be changed by RAN3’s conclusion on signalling format.

**Question 4: Do you agree that RAN2 can confirm we should only consider periodic traffics for survival time, at least in Rel-17 ?**

|  |  |  |
| --- | --- | --- |
| **Company** | **YES/NO** | **Comments** |
| Nokia | Yes | It cannot be more clear that SA2 only considers survival time for periodic traffics, so we don’t see any reason why RAN2 should consider non-periodic traffics. |
| CATT | Yes | Same view as Rapporteur |

## 2.3 Assumptions about Message Segmentation

It seems companies have different understanding about whether segmentation is possible in the higher layer. More specifically, how many PDCP SDUs does an application message comprise? The paper [15] has proposed that RAN2 should assume that one message corresponds to one SDU, and thus no segmentation is applied to the message at the higher layer. On the other hand, [3], [13], and [19] have assumed that a burst/message could be consisted of more than one SDUs, and the message is considered as lost when at least one of its segments is not successfully transmitted before the required delivery time. The assumption about segmentation may influence the methods that RAN2 should adopt to monitor survival time state and prevent consecutive error. Hence, it is suggested that RAN2 should discuss the assumption about the number of PDCP SDUs per application message. From the rapporteur point of view, there are three options:

* **Option 1:** RAN2 assumes one application message is conveyed by one PDCP SDU
* **Option 2:** RAN2 assumes one application message is conveyed by multiple PDCP SDUs, and the number of PDCP SDUs per application message is fixed.
* **Option 3:** RAN2 assumes one application message is conveyed by multiple PDCP SDUs, and the number of PDCP SDUs per application message can be varied.

**Question 5: What assumption RAN2 should make about the number of PDCP SDUs per application message?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Option(s)** | **Comments** |
| Nokia | 1 or 2 | In the LS from SA2 (R2-2010692), it has already indicated the following:  … where a data burst corresponds to a single application message.  And since burst size is typically fixed for deterministic applications, we think either Option 1 or Option 2 should be assumed. |
| CATT | 1 | At least for the most stringent usecases which require very fast reaction time (those on top rows of Table 5-2.1 below) considering the deterministic and periodic nature of the traffic and the small payloads (20-50 bytes), it is a very safe assumption to consider that each message is carried in a single PDCP SDU. Note though that it does not make a big difference, if the trigger for increasing the reliability is a transmission failure, whether the transmission carries the complete or a fraction of the message, in any case the safest is to consider that the message failed even if only a fraction failed. |

## 2.4 Monitoring of Survival Time State

To avoid consecutive transmission failure, the transmitter needs to identify **when** it should enter the survival time state and take appropriate measures (i.e. boosting the reliability of a packet), in order to prevent survival time violation. By looking at the LS from SA2 (R2-2010692) [21], it is expected that RAN enhancement should be able to support all performance requirements listed in Table 5.2-1 in TS 22.104 [22]:

|  |
| --- |
| **R2-2010692 / S2-2007880 :**  ……  SA2 also discussed whether receiving Survival time over NGAP is sufficient assistance information for NG-RAN to address the performance targets laid out by SA1 in Table 5.2-1 in TS 22.104.  …… |

Some of the more stringent requirements listed in Table 5-2.1 of TS 22.104 (V17.4.0) are shown below:



There are a few things that are worth noting from these requirements listed in TS 22.104:

* RAN enhancements should be able to address use cases where the maximum end-to-end latency, transfer interval, and survival time are all extremely short (e.g. 500 us), which implies RAN only has extremely limited time to monitor and respond to survival time state. The maximum end-to-end latency also include the time required to transfer data from CN to RAN, which means the Uu interface delay budget is even smaller.
* For use cases such as *Wired-2-wireless link replacement*, the maximum transfer interval covers two wireless links (in accordance to Table 5-2.1 in TS 22.104: *NOTE 5: Communication includes two wireless links (UE to UE).*). Therefore, there could be a cross-dependency between monitoring/responding to survival time state in one link and status of another link.

RAN2 should take these into account when discussing the potential enhancement to support survival time requirement, in order to identify viable solutions that can really fulfil the requirements laid out by SA1. Based on the papers submitted to RAN2 #113e, the possible mechanisms for survival time state monitoring are:

* **Option 1: Monitoring based on PDCP SN** [9]

Rather than monitoring whether a transmission failure has occurred, the transmitter proactively boost the reliability and/or increase transmit diversity of PDCP SDUs corresponding to every N-th incoming message (wherein the value of N is the maximum number of consecutive message error that the application can tolerate) based on the PDCP sequence number (SN), to ensure at least one message in every N messages can be successfully transmitted. So, the transmitter autonomously enters survival time state for every N-th message.

* **Option 2: TX-side Timer** [5][8][13][15][16]

Introduce a new survival timer at the transmitter side. If the message cannot be successfully transmitted before the timer expiration (e.g. the timer may be expired at the point BAT + 5G-AN PDB), the transmitter may enter survival time state and boost the reliability of subsequent messages. The timer could be associated to certain events such as HARQ/ARQ feedback.

* **Option 3: RX-side Timer + New Feedback**[1]

Introduce a new survival timer at the receiver side. If the message cannot be successfully received before the timer expiration, the receiver should provide a fast feedback and trigger the transmitter to enter the survival time state.

* **Option 4: PDCP****Discard Timer**[19]

It is proposed to reuse PDCP discard timer. When the timer is expired and the PDCP SDU is discarded, it is deemed as transmission failure and the transmitter may enter the survival time state.

* **Option 5: HARQ****ACK/NACK**[2][12][13][14][15][18]

The transmitter may rely on HARQ feedback to determine if it should enter survival time state. For instance, if a NACK or a re-TX grant is received at MAC layer, the transmitter may enter the survival time state and boost the reliability of later messages. The HARQ ACK/NACK may be used as a trigger for the TX-side timer mentioned above.

* **Option 6: ARQ Feedback**[5][15][18]

By utilizing the RLC-layer ARQ feedback such as status report, the transmitter may identify a gap between the latest successfully delivered message and the previous successfully delivered message, which implies some messages may have lost. Hence the transmitter may enter survival time state in such situations. The ARQ feedback may be used as a trigger for the TX-side timer mentioned above.

* **Option 7: Grant/PUSCH deprioritization [**9]

When an uplink grant for a LCH with survival time requirement is deprioritized or if the associated PUSCH is cancelled due to intra/inter-UE prioritization, the UE enters survival time state to boost reliability of subsequent data.

* **Option 8: Expiration****of cg-retransmission timer** [9]

This is applicable to unlicensed band only. When the cg-retransmission timer is expired, the UE may deem it as NACK and enter survival time state.

* **Option 9: Ratio****of packet loss in a message**[13][19]

The transmitter may consider a message as lost and enter the survival time state when a certain ratio of packets within a message are not delivered successfully, e.g when any PDCP SDU in a burst is lost [19]. This assumes a message/burst may contain multiple packets.

* **Option 10: No need at UE and observation by gNB** [3]

For UL transmission, gNB can expect when a packet should arrive at the gNB (from the TSC AI and the assumption that the traffic is periodic) and observe that the packet is not delivered, while for UE does not need to monitor the survival time state.

From the rapporteur perspective, some of the methods listed above could be jointly applied or complement to each other. It is suggested that RAN2 should discuss which of these options (or combination of these options) should be considered to support all performance requirements listed in Table 5-2.1 in TS 22.104.

**Question 6: Which options (or combination of options) can fulfil performance requirements listed in Table 5-2.1 in TS 22.104 and should be considered? Please tick the options that should be considered and provide your rationale.**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Company** | **Option(s) that should be considered** | | | | | | | | | | **Rationale** |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| Nokia | V | ffs |  |  |  |  | V |  |  |  | Given that the maximum end-to-end delay is as small as 500us, basically RAN should transmit the message as soon as it arrives, in order to minimize the latency. But the time interval between messages (transfer interval) is also so short, it means RAN only has an extremely short time (if not none) to detect survival time state, if it is based on the status of the previous message. Thus, Option 5 is basically infeasible for more stringent cases as a HARQ RTT is typically longer than 500us (as observed in [2]), not to mention Option 6 where the ARQ feedback delay is even much longer! Essentially, we should avoid the situations where a message has arrived and is ready to be transmitted, but we do not process it because we are still waiting for feedback relating to the previous message to decide the survival time state– this is totally unacceptable for stringent TSC/URLLC services.  Option 1 is the only viable solution for **ALL** (including the more stringent cases) use cases, as the survival state is **NOT** dependent on the status of previous message(s), and the survival state can be determined immediately even before the burst arrive. Besides, it is much simpler in terms of both specification impacts and implementation – as e.g. PDCP doesn’t have to check HARQ status like in Option 5. Additionally, the use cases involving 2 wireless links can be easily supported without cumbersome signalling for cross-link coordination. Option 7 could also be feasible as survival state can be detected quite immediately.  Option 2 - we think it depends on how the timer is used – e.g. what is the triggering condition of this timer. This can be FFS.  Option 3 – developing new feedback may involve RAN1 as well, the impact is too much.  Option 4 – it doesn’t work for survival time state triggered by message loss in Uu interface.  Option 8 – we’ve agreed UCE will not be pursed.  Option 9 – similar issue to Option 5&6, how do we detect the message loss and react in such a short time ? |
| CATT |  |  |  |  | V |  |  |  |  |  | Unlike stated by Nokia above, [2] precisely shows that survival time triggered by NACK reception on PDCCH is part of the only practical methods, latency-wise, to address the most stringent usecases from SA1. We also don’t get the point regarding “*a message has arrived and is ready to be transmitted, but we do not process it because we are still waiting for feedback relating to the previous message*”. The latency analysis in [2] includes the preparation time for the next message and, in case of CGs, as we know since R15, the UE must wait anyways until the last minute, according to R1 timeline, before processing the PDU in case it is deprioritized by a higher priority DG.  Regarding Option 1, as we understand it, for such traffic types, the UE would proactively boost every other packet transmission, even if the link is in a reliable steady state. That sounds overkill efficiency-wise. |
|  |  |  |  |  |  |  |  |  |  |  |  |

## 2.5 Methods of Avoiding Intolerable Consecutive Errors

Once the transmitter has entered the survival time state (e.g. based on one or more of the methods listed in Section 2.4), the transmitter may adapt the way to process the subsequent data packets (i.e. strive to transmit the upcoming packet with higher reliability or lower latency), in order to avoid consecutive failure that eventually leads to survival time violation. Based on the papers submitted to RAN2 #113e, the options can be categorized as following:

* **Category 1: PDCP Duplication** [1][2][5][6][7][13][14][15][16][18]

PDCP duplication is activated when the transmitter enters survival time state, in order to improve reliability and latency performance of subsequent data packets. In particular, for UL the UE could either activate duplication autonomously, or controlled by the network.

* **Category 2: Adaptive L2/L1 configuration/parameters** [1][2][6][9][10][12][13][15][16][18]

Layer-2 configurations such as LCH settings (e.g. priority [2] and mapping restriction rules [9]) could be changed when the transmitter enters survival time state. For instance, PDCP may switch between RLC entities with different LCH configurations to process a PDCP PDU [9]. Moreover, Layer-1 parameters such as CG/SPS configurations [12], MCS, number of repetitions, and TX power could be changed when the transmitter enters survival time state.

* **Category 3: Others** 
  + Allowable transmission in measurement gaps [14]:A new rule that allows the UE to transmit data with survival time requirement even if the resource overlaps with the measurement gap.
  + UE reporting of excessive consecutive data burst loss [19]: The UE reports the consecutive data burst loss when the number of consecutively lost data bursts exceeds a configured threshold, which allows the gNB to schedule more reliable uplink resources.
* **Category 4: gNB scheduling [3]**

At this stage RAN2 should first discuss which direction in the categories above should be considered.

**Question 7: Which category(ies) of solutions listed above should RAN2 consider for avoiding intolerable consecutive errors? Please provide your rationale.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Category(ies) should be considered** | **Rationale** |
| Nokia | 1 and/or 2  (Slightly prefer Option 2) | We think both options can be considered further. For PDCP duplication (Option 1), it requires CA or DC deployment, which may not be the case especially when TSC private networks are considered where available spectrum could be rather sparse.  Option 2 is generally more versatile as it is applicable to any deployment. We think altering LCH mapping restriction (e.g. allowed CG list) is a simple way to boost reliability when needed. In particular, by having 2 (or more) RLC entities for a PDCP with different LCH configurations, the PDCP can simply switch between these RLCs in accordance to whether it is in survival time state or not. Such switching behaviour is very similar to split bearer operation. |
| CATT | 1 and/or 2 | These 2 options seem to be the simplest and also most achievable within the available reaction time. |

# 3 Other Issues

## 3.1 UE Knowledge of Survival Time Requirement

In [5], it is proposed that the UE should also obtain information relating to survival time requirement, so it is able to measure/count message loss. In light of this, a new NAS-PDU in NAS signalling is proposed, such that the gNB may notify the UE about survival time requirement.

**Question 8: Do you agree RAN2 should introduce a new NAS-PDU in NAS signalling for gNB to notify the UE about survival time requirement?**

|  |  |  |
| --- | --- | --- |
| **Company** | **YES/NO** | **Comments** |
| Nokia | NO | We do not see any clear benefits of introducing this. Resource allocation and scheduling is anyway conducted by gNB. |
| CATT | No | Same view as Nokia |

## 3.2 TSCAI from UE

In [11], it is proposed that the UE may provide TSCAI to the gNB, especially when the core network is not able to provide such information. However, from rapporteur point of view this is not in the WI scope, RAN2 should focus on potential RAN enhancement for new QoS parameters provided by the core (e.g. survival time). Nonetheless, we may discuss to confirm whether we should still consider it.

**Question 9: Do you agree that RAN2 can confirm acquisition of TSCAI from UE is beyond the WI scope?**

|  |  |  |
| --- | --- | --- |
| **Company** | **YES/NO** | **Comments** |
| Nokia | YES | This is clearly out of scope. |
| CATT | Yes | Same view as Nokia. SA2 has not considered this so far. |

# 4 Conclusion

TBD

# References

[1] R2-2100216 RAN enhancements based on new QoS related parameters, Huawei, HiSilicon

[2] R2-2100223 Discussion on Survival Time, CATT

[3] R2-2100234 RAN enhancements based on new QoS related parameters, Ericsson

[4] R2-2100269 RAN Enhancement to support new QoS, QUALCOMM Europe Inc. - Italy

[5] R2-2100328 Further considerations on new QoS, ZTE Corporation, Sanechips, China Southern Power Grid Co., Ltd

[6] R2-2100418 Topics on new QoS handling, Fujitsu

[7] R2-2100449 Discussion on RAN enhancements based on Survival Time, III

[8] R2-2100614 Support for Survival Time and Burst Spread, Intel Corporation

[9] R2-2100718 Views on RAN Enhancement for New QoS Parameters, Nokia, Nokia Shanghai Bell

[10] R2-2100831 Disucussion on RAN enhancement to support survival time, vivo

[11] R2-2100856 Scheduling Assistance Information for support of new QoS, Apple

[12] R2-2100857 Reliability enhancements for CG/SPS, Apple

[13] R2-2102229 Consideration on RAN enhancement based on new QoS , OPPO (Revision of R2-2100892)

[14] R2-2100922 Discussion on the support of survival time, Lenovo, Motorola Mobility

[15] R2-2101066 Open issues with survival time and proposal for way forward, Samsung Electronics GmbH

[16] R2-2101509 Enhancements based on new QoS requirements, InterDigital

[17] R2-2101521 Implication of survival time, LG Electronics UK

[18] R2-2101615 Discussion on the support of new QoS parameters in RAN, CMCC

[19] R2-2101673 RAN impacts of the survival time, Beijing Xiaomi Mobile Software

[20] R2-2102254 Summary of Agenda Item 8.5.4: RAN enhancements based on new QoS Nokia

[21] R2-2102254 LS on Use of Survival Time for Deterministic Applications in 5GS, SA2

[22] TS 22.104 V17.4.0, Service requirements for cyber-physical control applications in vertical domains, Oct. 2020.