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 |
| LG | SunYoung LEE | ssunyoung.lee@lge.com |
| Samsung | Milos Tesanovic | m.tesanovic@samsung.com |
| Fujitsu | Ohta, Yoshiaki | [ohta.yoshiaki@fujitsu.com](mailto:ohta.yoshiaki@fujitsu.com) |
| Apple | Ralf Rossbach | rrossbach@apple.com |
| MediaTek | Pradeep Jose | pradeep[dot]jose[at]mediatek[dot]com |
| Qualcomm | Sherif ElAzzouni | selazzou@qti.qualcomm.com |
| China Telecom | Jincan Xin | xinjc@chinatelecom.cn |
| Huawei | Tao Cai | [Tao.cai@huawei.com](mailto:Tao.cai@huawei.com) |
| TCL | Hejun Wang | Hejun.wang@tcl.com |

# 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. |
| Ericsson | Yes |  |
| LG | Yes |  |
| Samsung | Yes | In addition to Nokia’s views about the superfluity of CSA in some scenarios (which we agree with), we wanted to add that sharing CSA with RAN even in cases where interruption occurs within RAN seems redundant to us as it is determined by ST and PER. Additionally, meeting PER is in any case a target and not an absolute requirement. For purposes of this release, we see no need to share CSA with RAN as the appropriate configuration of PER and ST achieves the same thing. |
| Fujitsu | Yes | As discussed online on Tuesday, gNB should be responsible for data scheduling to meet QoS/QoE (e.g. survival time) without CSA. |
| MediaTek | Yes |  |
| Qualcomm | No | CSA is crucial to Survival Time Operation. We raise two main points to clarify our understanding of survival time and CSA:   * When a flow enters “survival mode”. i.e., survival time is started, the network would deploy some of the methods in Section 2.5 to avoid expiry. In this case, gNB needs to know the CSA, as more repetitions/higher Tx power/PDCP duplication are needed for a flow with higher CSA requirement, since this flow’s expiry would be very costly from an application standpoint. However, targeting very high reliability for ALL flows in survival mode would be quite inefficient since another flow’s CSA requirement may not be that strict (the application can withstand occasional survival time expiries). * Following a blockage/shadowing (common to cluttered industrial environment) that causes multiple flows to enter survival and contend for resources, it is useful for RAN to know their respective importance using CSA metrics to prioritize available resources optimally and correctly (which flows can be left to expire, if necessary). We view entry of survival time as a tail-event that rarely happens, however, due to the importance of some safety application flows with high CSA requirements ~99.999999% reliability, we think it is very important to give the gNB a formula of how to preserve CSA when those rare blockages happen * Note that Packet Error Rate (PER) target is not useful here since successive errors are typically correlated, and the CSA cannot be simply estimated as PER^survival\_time. gNB will override PER in survival mode to a higher reliability target to avoid survival expiry. That higher reliability target would be naturally inferred from CSA.   We can breifly address some of the counter-points raised during the meeting   * “Nokia thinks that no matter what CSA is the network should do its best from RAN perspective to not violate survival time”: Reiterating that configuring the maximum number of repetitions/highest Tx power/maximum duplication/LCH capability (or any number of avoidance methods to be agreed on in RAN2) in survival mode to target maximum reliability irrespective of CSA would be quite inefficient and drastic. Some flows only have a CSA of 99.999% and do not require the same survival expiry avoidance methods as 99.999999% flows, especially that those measures would mean more resources that can disrupt other flows or severely limit network capacity by heavily under-provisioning the resources. * “Intel thinks that as long as there is no new requirements from SA1/SA2 we don’t need to define anything new.” SA2 is considering making this a requirement so this warrants serious consideration. * “Ericsson also doesn’t think we need it and survival time is optional” CN signalling survival time means that indeed the gNB must take some measures to avoid survival time expiry if it’s signalled. Again, we strongly think there should be some reliability measure that guides what the gNB does since not all flows in survival mode are equal. * “CMCC also doesn’t think we need this parameter and doesn’t help the RAN side to perform scheduling” Again we provide the example of two flows in survival mode, one with CSA 99.999999% and the other with 99.999%, it is expected that knowing this information is useful for gNB to rescue the more urgent one with more drastic expiry consequences. Alternatively, gNB should not interrupt all flows to rescue a flow with low CSA requirement from expiring. |
| China Telecom | Yes | No matter what the CSA is, the network should try its best to avoid communication service interruption. |
| Apple | Yes | Depending on the solution to enhance reliability, both CSR and CSA can help reduce the probability of exceeding the survival time (ST). Given the current solutions in RAN, we think that PER and ST are enough. |
| Huawei | Depends on how CSA is used | We don’t think gNB can directly use CSA. On the other hand, as indicted by SA2, there are services with the same ST but different CSA. Normally gNB schedules radio resource to satisfy PER and when the transmission fails and the system enters “survival time state”, gNB needs to guarantee the second (if survival time equals to transfer interval) transmission to succeed which would be resource consuming. In this state, a “second PER” would be useful for gNB to schedule “survival time state” resource and we think CSA could be the basis to determine the “second PER”. |
| TCL | Yes | To some extent, the information of communication service availability can be expressed as survival time. If the survival time is expired, the service enters downstate, means that the communication service is unavailable. In RAN perspective, it will do the best to meet the survival time, make sure the service will not enter downstate, so that the communication service is available. |

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. |
| Ericsson | No | As Rapporteur indicates, the issue is that RAN is not aware of the timing of a data burst. In the current SA2 spec 23.501 v16.6.0, it only indicates the latest possible time when the **first** packet of the data burst arrives, but in Ericsson’s view it is also important to know the latest possible time when the **last** packet of the data burst arrives. RAN2 can discuss whether it is beneficial to have this knowledge at RAN, which is independent from the SA2 discussion. The SA2 discussion is more along the line on if and how this can be provided.  Ericsson agrees to postpone the discussion on any RAN2 spec enhancements. |
| LG | Yes |  |
| Samsung | Yes |  |
| Fujitsu | Yes | As in [6]. |
| MediaTek | No | We tend to agree with the rationale from Ericsson, that it is important to know when the last packet of a burst arrives. RAN2 can discuss whether this information is useful at a RAN level. |
| Qualcomm | Yes | We think this issue is currently an SA2 discussion, so it would make sense to wait for their conclusion.   From a technical standpoint, we think that the TSN-like traffic is periodic and tight (often tied to a CG/SPS instance). If some jitter would push the BET later than that CG/SPS occasion, then that burst would automatically belong to the next occasion and the survival time would be monitored accordingly. There is no need for explicit BET signalling. |
| China Telecom | Yes | Before SA2 provides a clear definition for Burst Spread and Burst Ending time, there is no need for RAN2 to analyse potential enhancements. |
| Apple | No | RAN2/SA2 should consider either Burst End Time or Burst Spread, both parameters are useful in our view. We can postpone the discussion until SA2 provides further information, however, this bears the risk that SA2 has already concluded on a direction. We tend to agree with MediaTek and think that RAN2 can discuss whether it makes sense to indicate a preference upfront. |
| Huawei | Yes | We believe Burst Spread (or BET) won’t be crucial for network performance at this stage and we can ask further clarification by SA2. |
| TCL | Partly Yes | As is mentioned above, SA2 is still working on the (3) Burst Spread, we can wait further clarification from SA2.  However, discussion of (2) Burst Ending Time can be started in RAN2, for it may be useful in the survival time discussion, e.g., from when to start the survival time duration. |

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. |
| Ericsson | Depends on Q1 | If Question 1 can be confirmed, we are fine not to indicate CSR. Otherwise, RAN2 needs to further discuss the need for additional parameters, since in our view CSA alone is not sufficient. |
| LG | No |  |
| Samsung | No |  |
| Fujitsu | No | The definition in TS22.104 is unclear with respect to what the given conditions are. It is difficult for the gNB to take the parameter into account. |
| MediaTek | No | Agree with Nokia that this end-to-end service requirement does not translate to anything meaningful in RAN. A combination of PER and survival time should be correctly configured by the CN for a flow, for RAN to meet the end-to-end requirement of the service carried on the flow. |
| Qualcomm | No | CSA+Survival time captures the link-level reliability requirement so adding an explicit Service Reliability QoS metric is redundant. |
| China Telecom | No | Similar to Q1, the network should focus on how to avoid communication service interruption, rather than the mean time between failures. |
| Apple | No | Depending on the solution to enhance reliability, both CSR and CSA can help reduce the probability of exceeding the survival time (ST). Given the current solutions in RAN, we think that PER and ST are enough. |
| Huawei | Related to Q1 | We believe further clarification on the relation between CSA and CSR is needed and gNB cannot use CSR directly as well. |
| TCL | Partly Yes | For survival time that can tolerate several packet loss, more QoS information is needed to indicate the detail reliability requirement of the upcoming packet. For example, one packet loss and two or more consecutive packet loss would lead to different degree stringent reliability requirement of the subsequent transmission.  With such information, the UE or RAN can decide whether to boost the reliability of the subsequent transmission to the highest level or to gradually improve subsequent transmission reliability. |

## 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 |
| Ericsson | Yes | This is clearly indicated in the SA2 LS. |
| LG | Yes |  |
| Samsung | Yes |  |
| Fujitsu | Yes | It is our understanding. |
| MediaTek | Yes |  |
| Qualcomm | Yes |  |
| China Telecom | Yes | The assumption from SA2 LS is the start point for RAN2 work. |
| Apple | Most likely | The information currently available in TR 23.700-20 and SA2 draft CRs seems to indicate periodic deterministic traffic as a main target for the use of survival time, however, the SA2 SID and WID also mentions “deterministic” applications as a whole. Since survival time can be used with non-periodic deterministic traffic and the SI scope does not preclude these cases, we may as well confirm our understanding with SA2. |
| Huawei | Yes | We agree to consider only periodic service for Rel-17. |
| TCL | Not sure | To some extent, the aperiodic traffics are taken into account while defining the survival time as “time” instead of “counting”, plus the SA1 has started to think of aperiodic service, the aperiodic traffics should be considered.  However, as there is no conclusion about aperiodic traffics on survival time discussion yet, we should wait for SA2 work, or we can send a LS to SA2 on whether the aperiodic traffic should be considered.  For now, we should consider periodic traffics for survival time, whether the aperiodic traffics be considered should depend on the SA2 reply and RAN2 schedule. |

## 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. |
| Ericsson | Option 3 | In the Table 5.2-1 of TS 22.104, it is written that the message size can be 15 k to 250 k, which does not fit into one PDCP SDU. In my understanding, for video frames, the message size can vary (due to compression algorithms), and so the number of PDCP SDUs may vary.   | Communica­tion service availability: target value | Communication service reliability: mean time between failures | Message size [byte] | Remarks | | --- | --- | --- | --- | | > 99,9999 % | ~ 1 year | 15 k to 250 k | Mobile robots – video-operated remote control (A.2.2.3) | |
| LG | 1 | Option 1 provides a conservative method for fast reaction or precautions against survival time expiry. Although the message may not fit into one PDCP SDU and the size may vary, assuming one PDCP SDU per application message has no harm on guaranteeing service continuity and rather be a safe and simple criterion. |
| Samsung | Option 1 | Just wanted to clarify what we meant by our proposal (in our submission) to avoid segmentation of higher-layer messages. We were referring to e.g. IP packets. We did not mean to imply that there will be no bursts – of course there will be bursts but they will consist (in our understanding) of multiple IP packets, rather than segmented (fragmented) IP packets.  Regarding the IP fragmentation for an application message, URLLC traffic is very time sensitive and packets are relatively small. If one IP packet is fragmented and conveyed by multiple PDCP SDUs, on-time delivery will not be guaranteed by lower layers. So, it is reasonable to assume that each time-sensitive IP packet mapped to one PDCP PDU.  We are further concerned about IP packet inspection, which is not only time/processing consuming but also a complicated interaction between AS and IP. Therefore, it should be avoided as much as possible. |
| Fujitsu | Pending | RAN2 should ask SA2 on the application message segmentation. |
| MediaTek | 1 | Option 1 is a conservative choice that can be used for the most stringent usecases with fast reaction times. |
| Qualcomm | 1 | Survival time is typically used for small payload packets. We do not see a reason not to assume that those packets should be mapped to one PDCP SDU. We think mapping survival time requirements to variable size application layer bursts is not the correct approach. |
| China Telecom | Option 3 | According to 22.104, the message size may be very large and variable. Therefore, an application message should be segmented into multiple packets. |
| Apple | Option 3+1 | TR 22.832 mentions that segmentation is possible, and we assume that SA2 is going to take this into account.  The vast majority of industrial use-cases shown in TS 22.104 should fit into a single PDCP SDU. Given the low latency required, it will be good if higher layer segmentation/fragmentation can be avoided for as much as possible. However, we do not think this is actually feasible in all cases. We have the same understanding as Ericsson that application layer message sizes larger than the max PDCP SDU size can be expected in some of the use-cases.  According to TS 22.104, some examples with variable or larger message sizes are mobile robots (15 K to 250 K) and industrial wireless sensors / asset monitoring (25 K, 250 K). For other cases, the application layer message size may be well within the PDCP SDU size limit yet exceed the typical MTU size on IP level (e.g., 2000 bytes for robotic aided surgery). |
| Huawei |  | Agree with Fujitsu, we shall ask SA2. |
| TCL | Option 1 or 3 | For the uncertainty of the application message size, the assumption of one application message conveyed by fix number of multiple PDCP SDUs would lead to resource waste and unnecessary restrictions to upper layer. |

## 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] OK for licensed

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.

* **Option 11: LBT failures [4]**

In unlicensed band, UE may enter survival time state after LBT failure and not wait for timer expiry.

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** | **11** |
| 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. |
| Ericsson |  |  |  |  |  |  |  |  |  | V |  | We want to clarify first that supporting these requirements from Table 5-2.1 of TS 22.104 (V17.4.0) does not mandate enhancements when the survival time is known (or not). Since the requirements can be achieved by gNB pro-actively providing robust allocations, e.g. by resource over-provisioning. This pro-active allocation may anyway be needed when the survival time is very short, since reactive mechanisms require feedback of transmission success receivable within the survival time which may be infeasible.  The only case to consider here is UL periodic traffic (see Q4). In this case, gNB knows when packets should arrive (from TSC AI), checks if they are successfully received, and triggers follow-up actions (e.g., allocating additional resources) if entering survival time mode. We don’t see any benefits to additionally indicate these feedbacks to UE. Moreover, there is always a delay to acknowledge the successful reception, which does not work with a short survival time.  More importantly, we don’t see any reasonable ways for UE to know the reception status in a fast and reliable way:  If the feedback is on the MAC layer: Network can provision multiple HARQs re-tx to meet the PER/PDB target and so these mechanisms do not work. It is not clear how UE knows from which consecutive unsuccessful transmissions the survival time starts. Additionally, there is no HARQ feedback for the UL transmission and so UE does not know explicitly if a MAC PDU is successfully delivered or not.   * If the feedback is on the RLC/PDCP layer: typically, it has a long delay on these layers. Within this time, the gNB can react itself and schedule the UE more robustly for subsequent packets. |
| LG | (V) |  |  |  |  |  |  |  |  | V |  | We basically share the view from Ericsson. We think reactive way wouldn’t be efficient/fast and not sure whether reactive way can keep the survival time well.  The best is the network provides proactive way by providing more resources, activating PDCP duplication, or assigning proper priority of logical channel.  Doing some actions by UE itself, which may have impact on resource management and scheduling of the network, may not be desired. Thus, discussion first starts with why network control is insufficient. |
| Samsung |  | X | X |  | X | X |  |  |  |  |  | These options require smallest standards effort in our view, while providing clearly defined triggers. |
| Fujitsu | V | V |  | V | V | V | V |  |  | V |  | Just because of down-selection:  3 needs new PUCCH format, which should be avoided.  8 is only to UCE and not common to license band.  9 is slower than 5 and 6. |
| MediaTek |  | X |  |  | X |  |  |  |  | X |  | Agree with Ericsson that our baseline assumption should be that the NW can detect whether packets have arrived or not at the expected time (derived from the TSC AI) and can take corrective follow-up action on entering survival time.  Only if there are issues found with this assumption should UE based methods be introduced. If the gNB cannot react fast enough to meet survival time restrictions, faster UE based recovery actions (based on HARQ which is the earliest indication of success/failure) can be considered. |
| Qualcomm |  |  |  | ✓  4  (T) | ✓  5 |  |  |  |  | ✓  10 | ✓  11 | Agree with Ericsson that gNB (knowing the periodic schedule) can infer missing packets and activate survival time (Option 10).  Option 5 NACK can be a good candidate for licensed spectrum. We introduce Option 11 for unlicensed, where a survival state can be triggered by LBT failure.  We do not support introducing new timers at this stage, and thus we tentatively support exploring option 4 if a timer is needed (assuming PDCP discard timer can be used withing the required ST constraints). |
| China Telecom | V | V |  |  |  |  | V |  |  |  |  | Option 1 can prevent survival time violation for all stringent cases. But it will sacrifice resource efficiency.  Option 2 can perform fast packet loss detection. But the details of the timer configuration should be further discussed.  Option 7 can enter the survival time to boost reliability quite immediately. |
| Apple | (✓)  1 | ✓  2 |  |  | ✓  5 |  | (✓)  7 |  | ✓  9 | ✓  10 |  | In our understanding, a combination of a pre-defined configuration in conjunction with the UE autonomously applying extra redundancy [12] is the most suitable approach. The UE applies a pre-configuration provided by the network or prescribed by the specification upon entering survival state and acts accordingly.  The trigger point to enter survival state needs to be clearly defined and a TX-side timer-based approach or a certain number (0,1…N) of unsuccessful transmissions attempts are possible initial options, however, more analysis is required. The methods of option 5 are also suitable.  1/ will reduce spectrum efficiency, therefore, conditions to enable this option would need to be carefully selected.  3/ relies on an additional feedback mechanism that can cause additional delay and waste of radio resources.  4/ and 6/ might be too slow.  7/ could complement autonomous adaptation at the UE side, however, impact on UE implementation may not be negligible and needs to be evaluated.  8/ and 11/ may not be required (it was agreed that support for survival time in UCE is up to network configuration)  10/ is not very clear how the network would trigger additional protection. The method may need to be complemented with autonomous actions at the UE side. |
| Huawei |  | 2 | 3 |  | 5 | 6 |  |  |  |  |  | For option 1, we believe there is issue with SN gap based solution need to be solved. For option 2 and 5, if the mapping between message and PDCP SDU is clarified, it is possible to use HARQ feedback, otherwise it is difficult to tell from TB failure. We think ARQ feedback might work. |
| TCL |  | 2 |  |  | 5 |  |  |  |  |  |  | For the tight delay requirement, it is better for the UE to perform autonomous ST detection which means option 2 and option 5 are suitable. For option 11, depends on the discussion of IIoT on UCE. |

## 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. |
| Ericsson | Category 4 | For any approaches related with UE autonomously switch between resources, they do not make any sense since they require gNB to have reserved UL resources just in the case the survival time mode is entered. The probability of this happening is very small around 10^-4 to 10^-5, corresponding the PER.  Since gNB is the first entity to know that the survival time mode is entered, it can schedule a more reliable transmission of the subsequent packets, e.g., by a dynamic grant. |
| LG | Category 4  (Category 2) | As answered in Q6, UE autonomous action may have impact on resource management and gNB scheduling. Furthermore, as long as there is already ongoing data transmission, it may not sufficient to activate some UE autonomous behaviour when a problem is soon to be detected. Thus, we basically think it should be network to provide proper scheduling/resource to keep the survival time.  However, in case a UE autonomous action is deemed necessary, we think there should be a mechanism to transmit such concerned data (approaching survival time) prior to other on-going data. |
| Samsung | Category 1 and Category 2 | Categories 1 and 2 are straightforward uses of existing robustness techniques and require minimal standardisation work. |
| Fujitsu | Categories 1 and 2 | For TSC, redundancy and quickness are inevitable although it may cause inefficiency of radio resource usage and signalling overhead. |
| MediaTek | 4 or 1 | Cat 1: A recognized mechanism since Rel-15 to improve data reliability while making best use of resources through time and frequency diversity.  Whereas the following is beyond this WI:  Cat 2: Rerouting PDCP data to different logical channels is quite complicated to specify, implement and test.  Cat 3.1: Failures are experienced at cell edge where the probability of a handover is high. Delays to handovers at cell edge by overriding measurement gaps would end up deteriorating performance, instead of helping improve the situation.  Cat 3.2: Burst loss reporting is not fast enough. The gNB would be able to react faster than this. |
| Qualcomm | 1,2,4 | There is a strong case for autonomous UE PDCP activation in survival mode specially for Unlicensed band to increase reliability.  The benefits of Category 2 are a little vaguer for now since TSN flows are expected to have highest LCH priority, so there is very little room (if any) for LCH reliability improvement. The same goes for L1 improvements, but RAN2 should do a little bit more assessment. Category 4 can be supported as an implementation choice when possible. |
| China Telecom | Category 1 and Category 2 | Option 1 is a straightforward scheme to improve reliability, but it will sacrifice resource efficiency.  Option 2 improves reliability by flexible L2/L1 configuration, which is more efficient. |
| Apple | Categories 1 and 2 (slightly prefer 2) | Due to stringent latency requirements, methods to enable additional protection on lower layers are most suitable. Categories 1 and 2 are our preferred ways to introduce additional protection during survival time.  We think that enhancements for extended or different CG/PHY configs can allow HARQ InitialTx and HARQ ReTx to become more adaptive. Unacknowledged MAC PDUs sitting in the HARQ buffer can be retransmitted with more reliability, more repetitions, differently spaced in time and frequency, etc.  Enhancements may apply during survival time only (reactive mode). As an alternative, similar enhancements may be used for any Tx/Retx, regardless of survival time (proactive mode).  PDCP duplication, alternative CG configs, HARQ InitialTx and/or ReTx enhancements, proactive transmission through alternative or modified CG configs, LCH mapping restrictions, or a separate allowedCG-list for dormant CGs are all possible methods, and a combination of different options is possible.  In general, the protection method applied during survival time can be configurable by the network based on UE capabilities. |
| Huawei | 1 or 2 | The solution shall be focused on uplink and the critical criterion shall be the latency of the action shall be as short as possible. |
| TCL | 1, 2 and partly 3 | The solutions mentioned in category 1 and category 2 are sufficient to meet the reliability and low latency requirement of the upcoming transmission in survival time. Also the solutions can be obtained with enhancements based on the legacy mechanisms with less specification changes.  However, it is not necessary for UE to boost reliability or increase transmission diversity as soon as it enters survival time, especially for the service that can tolerate more than one transmission failure. This might lead radio resource waste.  The UE can calculate the consecutive data burst loss and boost transmission reliability or/and increase transmission diversity when the number of consecutive lost data burst exceeds a configured threshold. Or the UE reports the consecutive data burst loss when the number of consecutive lost data burst exceeds a configured threshold which allows the gNB to schedule more reliable uplink resource, as is mentioned in the solution category 3. |

# 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 |
| Ericsson | No | UE does not need to actively track the survival time state, see answers above, and thus there is no need for this NAS-PDU. Furthermore, the UE should follow the configuration parameters from gNB which has considered survival time. If provided to UE, the handling of the survival time would not align between UE and gNB. |
| LG | No |  |
| Samsung | No |  |
| Fujitsu | Pending | Will see discussion result of Question 6. Having said that, this seems to be up to NW scheduling. |
| MediaTek | No | Signalling survival time to the UE is quite pointless. The NW needs to configure the UE to perform an action (increase reliability) in response to a trigger (entering survival time mode). The ‘trigger’ and ‘action’ need to be discussed rather than the signalling of ‘survival time’ to the UE. |
| Qualcomm | No | Same as Nokia. Also, SA2 scope. |
| China Telecom | No | Considering the gNB could measure/count the message loss, we don’t see the need to define a new NAS-PDU in NAS signalling. |
| Apple | TBD | We tend to think that a configuration mechanism on AS level would allow for faster turnaround in terms of enablement / reconfiguration, and such mechanism might likely be DRB specific. However, it seems a bit early to conclude on this. We would prefer to keep both NAS based and AS based signalling methods on the table and revisit once more details are known about the trigger conditions to enter/leave the survival time as well as the extra protection itself. |
| Huawei | TBD | We are open to discuss further |
| TCL | NO | There is no need to introduce a new NAS-PDU for the gNB forwards the parameter to the UE is sufficient. |

## 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. |
| Ericsson | Yes | If UE can provide TSC AI to gNB, then it should be possible to forward this to the core network and then forward to the gNB. In any case, this seems to fit better in the SA group. |
| LG | Yes |  |
| Samsung | Yes |  |
| Fujitsu | Yes |  |
| Apple | No | TSCAI has been put in place by SA2 after a recommendation from RAN2 because companies thought that the CN is a better place to provide assistance information than the UE. If there is no TSCAI available in some deployment (which seems likely, see [11]), then the assumption that the CN provides TSC assistance information to the gNB is no longer valid. Part of the RAN2 design for Rel-17 IIoT/URLLC depends on the availability of assistance information at the gNB and we are currently discussing potential additional parameters and RAN enhancements. Hence, in our view, this problem is related to the scope of the work item.  We would like to highlight the benefit of the availability of assistance information at the gNB and think it is useful to consider TSCAI from UE as an enhancement in the RAN2 design.  In case this is not agreeable, considering that RAN2 is anyway about to send an LS to SA2, we could as well mention the issue to SA2 and ask in which scenarios TSCAI is not going to be available, whether it includes periodic deterministic traffic etc. |
| MediaTek | Yes |  |
| Qualcomm | Yes |  |
| China Telecom | Yes | Out of WI scope. In principle all the assistance information comes from CN side. So, if all the companies agree with this scenario, a LS to SA2 is needed. |
| Huawei |  | We may ask SA2. |
| TCL | Yes | For gNBs supporting the new QoS parameters to get the QoS parameters from the core is sufficient. |

# 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.