**3GPP TSG-RAN2 Meeting 119-e** **R2-220xxxx**

**Online, 17th – 29th August, 2022**

**Agenda item: 8.15**

**Source: Huawei, HiSilicon**

**Title: Summary of** **[AT119-e][310][R18 Others - Low Latency] LS on Low latency (Huawei)**

**Document for: Discussion and Decision**

1. Introduction
* **This is the summary of below offline discussion. Reply LS to SA2 will be prepared based on the results of this discussion.**
* [AT119-e][310][R18 Others - Low Latency] LS on Low latency (Huawei)

Discuss LS response on Low latency

Deadline: To be set by rapporteur

* **Reference contributions:**
1. R2-2208007 Proposed response to SA2 LS R2-2203930 on low latency Nokia, Nokia Shanghai Bell discussion Rel-18 FS\_5TRS\_URLLC
2. R2-2207043 Draft reply LS on RAN feedback for low latency Qualcomm Incorporated LS out Rel-18 FS\_5TRS\_URLLC To:SA2 Cc: RAN1, RAN3
	1. *(moved from 8.5.1)*
3. R2-2207768 Consideration on meeting very low latency requirement in TDD ZTE Corporation, Sanechips, China Southern Power Grid Co., Ltd discussion Rel-17 NR\_IIOT\_URLLC\_enh-Core R2-2205732
4. R2-2207775 [DRAFT] Reply LS on RAN feedback for low latency ZTE Corporation, Sanechips LS out Rel-17 NR\_IIOT\_URLLC\_enh-Core R2-2205734 To:SA2 Cc:RAN3
5. R2-2206963 LS on RAN feedback for low latency (S2-2201767; contact: Huawei) SA2 LS in Rel-18 FS\_5TRS\_URLLC To:RAN2 Cc:RAN1, RAN3
6. R2-2208134 Discussion on RAN feedback for low latency Ericsson discussion Rel-18
7. R2-2208687 Discussion on RAN feedback for low latency enquired by SA2 Huawei discussion Late
8. R2-2208688 Draft reply LS on RAN feedback for low latency Huawei LS out Rel-18 FS\_5TRS\_URLLC To:SA2 Cc:RAN1, RAN3 Late
* **Contact list**

|  |  |  |
| --- | --- | --- |
| Name | Company | Email |
| Tao Cai | Huawei, HiSilicon | tao.cai@huawei.com |
| Lu Ting | ZTE | lu.ting@zte.com.cn |
| Mattias Bergström | Ericsson | Mattias.a.bergstrom@ericsson.com |
| Ralf Rossbach | Apple | rrossbach@apple.com |
| Yunsong Yang | Futurewei | yyang1@futurewei.com |
| Chunli Wu | Nokia | Chunli.wu@nokia-sbell.com |
| Sherif ElAzzouni | Qualcomm | selazzou@qti.qualcomm.com |

1. Discussion

## 2.1 Regarding Q1

In Q1, SA2 asks:

1. What are the possible values for the periodicity of the TDD cycle that RAN can support? This question is related to Problem 1.

The problem 1 is described as

"One potential problem considering low latency applications is that the arrive time of the packets may not fit well with the TDD cycle used in the network. RAN just receives the traffic flow periodicity and burst arrival times but cannot influence them. For example, if a downlink packet arrives at an uplink slot, then it has to wait for the first downlink slot to be transferred and vice versa (please see Figure 1). This creates additional delay (e.g. more buffering time) to the traffic flows. This can be an issue for QoS Flows requiring PDB 5 ms or lower."



According to the contributions of this meeting, there are several categories of answers for SA2 Q1, besides all suggest to answer " the possible values for the periodicity of the TDD cycle that RAN can support ":

**Option 1: Besides answer supported TDD cycle, e.g., {0.5ms, 0.625ms, 1ms, 1.25ms, 2ms, 2.5ms, 3m, 4m, 5ms, 10ms}, also answer to SA2 that, due to various reasons, for example, packets arriving to the gNB will experience buffering (i.e. resulting in increased delay) if their time of arrival is not aligned with the transmission opportunities of the TDD subframe, there is need to adjust burst arrival time in TDD scenario.**

**Option 2: Besides answer supported TDD cycle, e.g., {0.5ms, 0.625ms, 1ms, 1.25ms, 2ms, 2.5ms, 3m, 4m, 5ms, 10ms}, also answer to SA2 that, no issue of scheduling delay considering TDD UL-DL cycle/pattern thus there is no need to adjust burst arrival time in TDD scenario.**

**Option 3: Only answer supported TDD cycle, e.g., {0.5ms, 0.625ms, 1ms, 1.25ms, 2ms, 2.5ms, 3m, 4m, 5ms, 10ms}.**

**Question 1, which option(s) above would your company support?**

|  |  |  |
| --- | --- | --- |
| Company | Option(s) | Further comments |
| ZTE | **Option 1** | We suggest to include the following points in the response LS mainly based on the detailed analysis in R2-2208134 and also R2-2208007, R2-2207768:* The NR TDD configuration framework has a hierarchical structure consisting of:
	+ Cell-specific TDD pattern: taking the union of the possible periodicity values under only *pattern1* (with periodicity P1) and both *pattern1* (with periodicity P1) and *pattern2* (with periodicity P2), there are a total of 11 possible periodicities that the gNB can configure via *tdd-UL-DL-ConfigurationCommon*: {0.5 ms, 0.625 ms, 1 ms, 1.25 ms, 2 ms, 2.5 ms, 3 ms, 4 ms, 5 ms, 10 ms, 20 ms};
	+ UE specific TDD pattern: the gNB can also configure a UE-specific UL/DL TDD configuration via RRC signalling in *tdd-UL-DL-ConfigDedicated.*
	+ Dynamic TDD uplink/downlink pattern: the gNB can also use DCI to additionally specify some (or all) of the flexible symbols in the semi-static TDD pattern to be downlink or uplink symbols.
* With the hierarchical configuration, the minimal TDD cycle (e.g. TDD-UL-DL symbols pattern cycle) can even be one slot, which depends on the SCS configuration, e.g., the minimum value for SCS SCS 120kHz is 0.125ms or the minimum value for SCS 960kHz is 0.015625ms.
* Even though 3GPP specs allow a very flexible TDD pattern, there are reasons for a limited selection in practice, for example, due to interferences including coexistence issues with neighbouring cells and other networks deployed on adjacent spectrum. One example is that, for macro network deployments using FR1 TDD spectrum with 30 kHz subcarrier spacing, TDD cycles with 5 milliseconds or 2.5 milliseconds periodicity are typically used to facilitate coexistence with legacy LTE TDD deployments.
* In the practical deployment case mentioned above, if a DL packet arrival mismatches with DL transmission occasions, there may be buffering for a period of time waiting for the subsequent DL slots/symbols for transmission. Similar issue exists for UL traffic.
 |
| Ericsson | 2 | To give the full picture to SA2 it is good to clarify that even though the (static) TDD patterns, it is possible to dynamically adjust the TDD pattern and hence no need to adjust burst arrival time for TDD. |
| Apple | Slight preference for Option 1. Option 2 may suffice in certain scenarios only. | In general, the gNB can figure out burst arrival time and periodicity based on TSCAI from the CN. TDD patterns can be made very flexible and controlled dynamically, so gNB implementation might be able to adjust the configuration accordingly. However, dynamic adjustment of TDD pattern can only be used if the UE supports it. Secondly, requirements exist to align TDD pattern for the purpose of interference reduction, e.g., between different networks. Thus, it seems reasonable to assume the gNB may experience a limited amount of buffering if the arrival time of incoming packets on the N3 i/f and the air interface transmission time is not aligned. Practical TDD cycle configurations for need to be established in a real network deployment as well. |
| Huawei, HiSilicon | Option 1 | Though the SA2 LS Q1 is not very clear on whether or not RAN2 should provide answer on there is need to adjust burst arrival time in TDD scenario, we think it will helpful to indicate, while the TDD cycle can be small, there is limitation in real-world deployment that TDD cycle cannot be changed with all the possibility in the specification. If the practical limitation exists, e.g. due to inter-cell interference and the burst arrival time can be adjusted in the application layer, it would certainly help to fulfill the requirement of low latency services.  |
| Futurewei | Option 3 | At this point, we are unsure whether adjusting BAT will result in reduction of the delay budget available for the Uu link, given a same E2E delay. It would be good to find out from SA2 whether adjusting BAT will result in reduction of the delay budget available for the Uu link, given a same E2E delay. |
| Nokia | Option 1 | Coexistence with legacy LTE TDD can be mentioned as well, thus it is not always possible in the deployment to apply those values even though supported by signalling. |
| Qualcomm | 1 | For factory network, 30kHz sub-carrier spacing is commonly used in FR1 as we assumed in the simulation evaluation for 5G-ACIA (RP-210490). In this case, TDD UL-DL pattern must be aligned with 0.5ms slot length, and it can be configured as one of the periodicity values of {0.5ms, 1ms, 2ms, 2.5ms, 3ms, 4ms, 5ms, 10ms}.Note that the results mentioned here are a best-case scenario when the Burst Arrival time is 1) Arriving at an aligned time to undergo the required buffering and other RAN related procedure in time for transmission in the next UL (or DL) TDD occasion. 2. Aligned with Burst Arrival Times of all conteding flows to ensure that the next available TDD resources are sufficient to carry the arriving burst. In case the bursts are not aligned then this number can go up to account for the latency to waiting for an **available TDD** resource. |

## 2.2 Regarding Q2

In Q2, SA2 asks:

1. SA2 could not conclude whether a similar issue existing in FDD scenario (i.e. Problem 2) as Problem 1. Please RAN2 confirm whether it exists or not.

According to the contributions, most companies think there is no similar issue for FDD, and one possible issue is indicated. There are below options for answer to SA2 Q2.

**Option 1: There is no need to adjust burst arrival time in FDD scenario as in FDD there is no the UL-DL symbols pattern issue.**

**Option 2: There are possible issues, e.g. the granularity of periodicity for SPS and ConfiguredGrant is same for FDD and TDD which may have impact on the service delay.**

**Question 2, which option above would your company support?**

|  |  |  |
| --- | --- | --- |
| Company | Option | Further comments |
| ZTE | **Option 2**Also acceptable to **Option 1** | In FDD, as there is no the UL-DL symbols pattern issue, the SPS and ConfiguredGrant can be configured in any symbol at which the packet arrives.However, the granularity of periodicity for SPS and ConfiguredGrant is same for FDD and TDD which may have impact on the service delay. |
| Ericsson | 1 | See reasingin in the Ericsson paper above. |
| Apple | 1 | The gNB can assign a suitable SPS/CG config – there are many different options. In addition, dynamic scheduling can be used if needed.  |
| Huawei, HiSilicon | 1 |  |
| Futurewei | Option 1 |  |
| Nokia | 1 |  |
| Qualcomm | 1 | From RAN2 perspective, there is no issue of scheduling delay in FDD scenario. For DL FDD, the network can always use a dynamic assignment to schedule a packet. For UL FDD, the network can configure a CG with very short periodicity for URLLC traffic. So, the scheduling delay is not a concern in FDD scenario |

## 2.3 Regarding Q3

In Q3, SA2 asks:

1. Does RAN see any additional aspects that SA2 should consider for the study?

According to the contributions, there are following additional aspects raised by companies:

**Aspect 1) If the arrival time does not match the radio resource pattern (e.g. TDD-UL-DL symbols pattern), i.e. the arrival time of DL QoS flow is UL symbol, or the arrival time of UL QoS flow is DL symbol, the transmission will be delayed.**

**Aspect 2) If the service is with very low latency requirement and the arrival time jitter is large, it will bring large challenge to RAN node. SA2 can consider further enhancement, e.g., de-jitter mechanism in CN to try to make sure that the packets arrive at RAN node just before it can be sent over Uu interface.**

**Aspect 3) the appropriate staggering of BAT can provide significant benefits for system capacity as well as power consumption with XR applications, so the relevant mechanisms for RAN to adjust BAT is highly recommended.**

**Question 3, Which aspect(s) would your company agree/disagree to suggest SA2 for further consideration?**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Company | Aspect 1 | Aspect 2 | Aspect 3 | Further comments |
| e.g. company-name | Agree/disagree | Agree/disagree | Agree/Disagree |  |
| ZTE | Agree | Agree | Neural | The Aspect 1 and Aspect 2 mainly elaborate the issue mentioned for Q1. For Aspect 3, we think the main requirement of XR is to avoid collision that may occur when multiple large size I frames arrive at the same time, which is problematic even in FDD. We feel this is not same type issue as the delay issue we discussed above that is caused by the misalignment between service direction and resource mode in TDD. But yes, the solution for such different scenarios may be similar, e.g., more traffic information may need to be exchanged between CN and RAN. |
| Ericsson | Disagree. SA2 understands that one cannot send UL in a DL-slot and send DL in an UL-slot. And as per Q1, this is not an issue since TDD pattern can be adjusted to avoid issues. | Disagree.If there is jitter in the data arriving in the RAN (from CN, i.e. DL data), the data will be buffered if it cannot be transmitted immediately. RAN (of course) has buffering capabilities. Hence there is no “large challenge” if the data has jitter.Note, de-jittering increases latency and if CN would have a de-jittering function, that function would **increase** latency. If de-jittering should happen (at all) it should happen close to the consumer of the data. If CN would de-jitter DL data it does not help since there can arise new jitter in the data in the RAN and in the end between CN and the application. One can de-jitter once again before delivering to application, but that increase latency. In UL it’s the same, if CN would de-jitter the data before it is sent to the consumer (e.g. via the internet) new jitter can again arise after the CN has de-jittered it, so also that is not beneficial. | Disagree | So to summarize: No additional issues. |
| Apple | Agree | Neutral | Neutral |  |
| Huawei, HiSilicon | Agree  | Neutral | Neutral |  |
| Futurewei | Agree | Disagree. Given a same E2E delay, de-jitter mechanism in CN takes delay budget away from the Uu link, making things worse. | Cautiously disagree. Unsure whether adjusting BAT will result in reduction of the delay budget available for the Uu link, given a same E2E delay. | However, we don’t mind that RAN2 ask SA2 what means the CN has for adjusting BATs and whether adjusting BAT will result in reduction of the delay budget available for the Uu link, given a same E2E delay. |
| Nokia | Agree |  |  | Agree with aspect 1. Aspect 2 is in our view already covered by Aspect 1, so it is sufficient to communicate whether the problem exists or not without suggesting enhancements to SA2. Aspect 3 is a bit speculative, considering the topic has not been discussed in RAN2 before and to be discussed in Rel-18 XR SI. |
| Qualcomm | Agree | Disagree | Agree | On Aspect 1, as explained in Q1, from a practical standpoint, the latencies mentioned are in an ideal scenarios when the bursts arrive exactly in time to undergo the additional RAN latency just in time for a TDD transmission opportunity to be available. Realistically, of course, the immediate transmission opportunity may not be ready to carry the burst if the BAT is not appropriately configured. On Aspect 2, we disagree the mechanism that the packet is buffered for de-jittering at CN node, because it will simply increase the latency of most packets and does not help the RAN node to relax the tight delay budget. Instead of CN node, there is a de-jitter buffer at application client which stores the burst data until the periodic deadline. So, Qualcomm recommends the mechanism to define this deadline of de-jitter buffer as a new delay metric instead of the legacy PDB. For more details, we have proposed the delivery deadline in R2-2207044. On Aspect 3, as the proponent company we reiterare our views in R2-2207043 (where more details and illustrations can be found) that If the BATs are aligned with each other, the XR traffic of the UEs contend for resources, decreasing the capacity. On the contrary, if the RATs of UEs are staggered relative to each other, the cell can serve each UE in series to fully utilize the resources, and it is beneficial to meet the PDB, improving the capacity with XR applications. Note that without staggering appropriately, it would be very challenging to meet the latency for a limited number of flows even, as the resources that can be used for XR flows are a very limited bottleneck fraction of the available resources over time. From power perspective, the appropriate staggering of BATs can reduce the active time of each UE and improve the power consumption by staying longer in sleep mode |

 **[Summary]**

## Conclusion