**3GPP TSG-RAN WG2 Meeting #109 electronic R2-2001993**

24th Feb – 6th Mar 2020

**Agenda Item: 6.12.3 L2 measurements**

**Source: CMCC**

**Title: Report for [AT109e][805][SON/MDT]L2 open issues (CMCC)**

**Document for: Discussion and decision**

# 1 Introduction

This email discussion is based on the summary of L2 measurement in R2-2000909, mainly focus on cat b open issues.

* [AT109e][805][SON/MDT]L2 open issues (CMCC)

Intended outcome: email discussion report in R2-2001993

Deadline: CET 23:00, 2020/02/28

During RAN2#109e first week online, RAN2 achieve the following agreements:

Agreements:

1 Granularity for per UE measurement performed by UE (i.e. D1 queueing delay) is per DRB per UE for non-split case.

2 Granularity for per UE measurement performed by network (i.e D2 delay, loss rate) is per DRB per UE.

4 Capture in TS 38.314 that for PRB usage measurements that have been defined in TS 28.552, i.e. DL/UL Total PRB Usage, Distribution of DL/UL Total PRB Usage, M(T), M1(T), P(T) are measured per cell level. P(T) is the total available PRBs for this cell. M1(T) is the PRBs used for traffic transmission in this cell.

5 For EN-DC UL D1 delay measurement configuration for non-split bearer,

- D1 measurement of MN terminated bearer(including non-split bearer) can be configured by MN,

- D1 measurement of SN terminated bearer(including non-split bearer) can be configured by SN via RRC message (SRB3 or SRB1).

- For the SN terminated bearers, it is the SN to configure and calculate the UL/DL delay.

6 Capture a general definition of DL measurement in TS 38.314:

Packet delay includes RAN part of delay and CN part of delay. For RAN part, the DL delay comprises:

- D1 (the DL delay in gNB-DU), referring to 5.1.1.1.1 Average delay DL air-interface in TS 28.552

- D2 (the DL delay on F1-U), referring to 5.1.3.3.2 Average delay on F1-U in TS 28.552

- D3 (the DL delay in CU-UP), referring to 5.1.3.3.1 Average delay DL in CU-UP in TS 28.552

7 The flooring operation associated to the definition of mean number of active UEs is removed.

As guided by session chairman, the discussion will focus on only the critical issues, without which the SON/MDT feature won't work. All the others things which make SON/MDT work better should be postponed to R17.

**Therefore, for all the cat b/c issues, rapporteur suppose that if it is not a critical issue and no quick consensus reached, we would better postpone it to next release.**

# 2 Discussion on open issues

## 2.1 Per DRB granularity

In the agreement above, the original proposal 3 is not captured and can be clarified further as follows:

**Proposal 3.1: ‘Granularity for per cell measurement performed by network (only for number of active UE) is per DRB per cell.’**

**Proposal 3.2: ‘And add a clarification in 38.314 that all the per DRB per cell measurements and per DRB per UE measurements can be aggregated into per QoS level per cell by network implementation.’.**

**Q1.1: Do companies agree with above Proposal 3.1 and 3.2?**

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comments |
| CMCC | Yes | The clarification in Proposal 3.2 can be added in the front of ‘chapter 4 Layer 2 measurement’ in 38.314, to indicate it apply to all the per DRB per cell and per DRB per UE measurements. |
| QUALCOMM | yes |  |
| Ericsson | Yes | We think that the proposed clarification is good and addresses the issue raised by companies during the telco. |
| Huawei, HiSilicon | Yes |  |
| CATT | Yes |  |
| Samsung | Yes |  |
| ZTE | Yes | We support with both proposals. |

**Q1.2: Regarding to how to reflect per DRB measurement in 38.314, can we use ‘*drbid*’ in the matrix for each measurement?**

Here are 3 examples:

Number of active UE per DRB per cell:

D2.1 delay per DRB per UE:

D1 delay per DRB per UE:

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comments |
| CMCC | Yes |  |
| QUALCOMM | yes |  |
| Ericsson | Yes | I suppose, the right-hand-side of the delay equations listed above will also contain drbid in them (tSucc(i,drbid), tSched(i,drbid), tDeliv(i,drbid), tArrival(i,drbid),). |
| Huawei, HiSilicon | Yes | In addition to the matrix, we think it may be good to also mention the drbid in the relevant table. For example, for section 4.1.1.3.7 number of active UE per DRB per cell, Table 4.1.1.3.7-1 may need to be updated with adding the explanation of drbid. |
| CATT | Yes |  |
| Samsung | Yes |  |

## 2.2 PRB usage

It has been agreed that PRB usage defined by SA5 should be measured at per cell level.

Another open points is that, different from the EUTRA PRB, the NR PRB is defined in frequency but not in time. So the PRB term may not be a suitable unit for counting PRB usage.

In TS 38.211, RB is defined as “A resource block is defined as consecutive subcarriers in the frequency domain. ” So it would be better to utilize 1 RB x 1 symbol to calculate PRB usage measurement.

**Proposal 5(Cat b): Capture in TS 38.314 that the counting unit for PRB usage measurement is 1 RB x 1 symbol. (1 RB=12 sub-carrier)**

Q2: Do you agree with above proposal 5?

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comments |
| CMCC | Yes |  |
| QUALCOMM | Yes |  |
| Ericsson | Yes |  |
| Huawei, HiSilicon | Yes |  |
| CATT | Yes |  |
| Samsung | Yes |  |
| ZTE | Yes |  |

## 2.3 Delay measurement

For UL D1 delay measurement configuration in EN-DC, RAN2 has achieved following agreements

5 For EN-DC UL D1 delay measurement configuration for non-split bearer,

- D1 measurement of MN terminated bearer(including non-split bearer) can be configured by MN,

- D1 measurement of SN terminated bearer(including non-split bearer) can be configured by SN via RRC message (SRB3 or SRB1).

- For the SN terminated bearers, it is the SN to configure and calculate the UL/DL delay.

###### Issue 3.2: For split bearer, does UE perform a single D1 measurement or 2 independent D1 measurements?

Qualcomm[1] thought there is no need to report the single D1 value to both MN and SN respectively which costs unnecessary signalling overhead. While Ericsson[8] thought that two different D1 measurements needs to be performed by the UE, independently one each for MN associated D1 delay and SN associated D1 delay in the split bearer configurations. Huawei, HiSilicon[16] share same view with Ericsson that the scheduling latency are different in the two paths, but prefer the UE to report two D1s to the RAN node where it receives the measurement configuration.

**Proposal 7(Cat b): RAN2 is kindly asked to make decision among following options:**

**Option 1:**

* **For the UL PDCP packet average queuing delay measurement for split bearer, UE reports a single D1 value to the node where it receives the measurement configuration.**

**Option 2:**

* **MN and SN can independently configure the UE with D1 measurements in the split bearer configurations.**
* **The UE shall perform two independent D1 delay measurements in the split bearer configuration, one for MN associated D1 delay measurement and the other for SN associated D1 delay measurement.**
* **The UE shall report the MN configured D1 delay measurement to the MN and the SN configured D1 delay measurement to the SN.**

**Option 3:**

* **the UE should report two D1s in MR-DC to the RAN node where it receives the measurement configuration.**

**Question 3.2: For split bearer case for EN-DC, which node can configure UE with D1 measurement? And does UE report 1 single D1 or 2 separate D1?**

|  |  |  |
| --- | --- | --- |
| Company | Which node can configure D1 for split bearer? | UE report 1 single D1 or 2 separate D1? |
| CMCC | No strong view, we slightly prefer Option 2 or 3. | Option 2 or 3 |
| QUALCOMM | Option 1 | For the Split bearer, there is only one PDCP buffer, thus for the UL PDCP packet queuing delay measurement, UE can only report single D1 value for the split bearer for both MCG and SCG. |
| Ericsson | Option-2 | Option-2 |
| Huawei, HiSilicon | Postponed to R17 | We understand that D1 in TS 38.314 is referring to NR protocols, i.e. NR PDCP, NR RLC, NR MAC, NR PHY.  Based on the above understanding, the UE should perform D1 following T 38.314 with NR protocols. For split bearers for EN-DC, there is NR PDCP + either LTE RLC/MAC or NR RLC/MAC, so the UE can only get D1 measurement for NR PDCP + NR RLC/MAC.  It seems that Issue 3.2 and all options take NR-NR DC as a typical example, but in the beginning of the section it mentions “For UL D1 delay measurement configuration in EN-DC”. Since it may need more discussions on supprting D1 for protocol layers (e.g. NR PDCP + LTE RLC/MAC) and options, we suggest to postpone it to R17. |
| CATT | Option1 | The same view with QC, the average UL delay in UE is a SA2 requirement and it is calculated in PDCP layer. Since SA only needs the D1 measurement as one part of the end-to-end delay, whether the DRB is a split bearer or not should not affect the reported result. Only one result for one DRB Id is enough |
| Samsung | Option 1 | We agree with Qualcomm. Moreover, the routing of PDCP PDUs to different RLCs of a split bearer is left to UE implementation. Therefore, we don’t see an accurate way to report 2 different delay values. |
| ZTE | Option 2.  D1 measurement is configured per DRB and the D1 value is common for MN and SN in case of split bearer.  However, even D1 value is common for MN and SN, we think both MN and SN can configure the D1 measurement separately for a certain DRB, which have RLC bearers in both MN and SN accordingly. | UE should report 1 single D1 for each measurement configuration (i.e. measurement configuration received from MN or SN)  As indicated in 38.323, the PDCP SDU is randomly delivered to MN or SN based on the grant received in lower layer, thus the PDCP queuing delay shall be common to both MN and SN in case of split bearer.  However, even the D1 is measured per DRB and common for MN and SN for a split bearer, we think MN and SN can initiate the measurement for D1 separately, and if two separate measurement for D1 is configured through MN and SN RRC signaling accordingly, then the UE need to report two D1 to MN and SN accordingly. |

###### Issue 3.3: How to derive final delay for split bearer considering with or without PDCP duplication?

Huawei, HiSilicon[16] thought that for the split bearer configured with PDCP duplication, the packets of these two paths are the same. Therefore the node hosting the PDCP entity can use the min value of these two paths. For the split bearer configured without PDCP duplication, the packets of these two paths are not the same. Therefore the node hosting the PDCP entity can use the average value of these two paths.

**Proposal 8(Cat b): For the split bearer, the node hosting the PDCP entity derives the delay of the split bearers based on the delay of two paths.**

**- For split bearer with PDCP duplication, the final delay is the min value of measured results of two paths**

**- For split bearer without PDCP duplication, the final delay is the average value of measured results of two paths**

Question 3.3: Are you agree with above proposal?

|  |  |  |
| --- | --- | --- |
| Company | Support or not | Comments |
| QUALCOMM | No | The proposed split bearer with PDCP duplication UL delay measurement impacts UE a lot because it involves the PDCP duplication detection between PDCP and RLC, we prefer to reuse UL delay measurement w/o duplication method to support duplication cases for this release |
| Ericsson | No | As our preference is for MN and SN to configure and receive D1 measurements independently, the MN-CU-CP and SN-CU-CP receive these measurements independently from the UE and they report the same to the TCE directly without the need for merging the two. The OAM can perform the ‘min’ or ‘average’ on the so reported measurement results. |
| Huawei, HiSilicon | Support | For the E2E measurement from SA2, it is the node hosting the PDCP to send the RAN delay results to the CN via the GTP-U. Therefore the node hosting the PDCP need combine the measurement results of two paths. Otherwise the CN cannot get the accurate delay results. |
| CATT | Yes | As our preference is only one result for one DRB Id, for PDCP duplication the PDCP entity could consider to calculate the packet with the min delay value, if we think SA5 should know the duplication feature is used in RAN side, the network can add a duplication activation indicator together with average delay results.  For non-duplication, packets using different RLC associated to the same PDCP could be calculated together. |
| Samsung | No |  |
| ZTE | No | In our view, whether to perform combination depends on how we intent to use this measurement result.  If the result will be used to evaluate the overall quality of services for the QoS flow, then we prefer to adopt some kind of weighted average based on the throughput/data volume of each path. For example:  delay = (data volume in path 2 \*delay in path 1 + data volume in path 2 \* delay in path 2)/ (data volume in path 1 + data volume in path 2)  If the result will be used to evaluate the quality of services for the QoS flow in one cell, then simply report the result for each path separately is sufficient. |

###### Issue 3.4: Whether CA duplication has impact on the split bearer delay measurement?

Huawei, HiSilicon[16] thought that for CA, the packets are transmitted via multi-paths in the same node. Therefore there is no X2/Xn delay difference among these paths. We think the air delay of these paths are similar. Therefore the measurements performed by the UE and the measurements performed by the gNB does not need to distinguish the transmission in these paths. For the CA based duplication, the UE and gNB can measure the delay assuming the packets of these paths are different.

**Proposal 9(Cat b): For the CA duplication bearer, the UE and gNB measure the UL/DL delay assuming the packets of multi-paths are different.**

Question 3.4: Are you agree with above proposal?

|  |  |  |
| --- | --- | --- |
| Company | Support or not | Comments |
| QUALCOMM | Yes |  |
| Ericsson | Support | Agree with the argument provided by Huawei. |
| Huawei, HiSilicon | Yes |  |
| CATT | Yes | CA duplication has no difference with DC duplication if supports option1 in Issue 3.2 |
| Samsung | Yes |  |
| ZTE | Yes | We think, with this understanding, in case of CA, PSCells and SCells can perform independently per DRB per cell measurement, including packet delay, number of active UE, and so on. |

###### Issue 3.5: Unit for delay measurement

CATT[2] observed that the unit RAN2 used for UL delay in 38.314 is “millisecond” (ms). The delay defined by SA5 is in “microsecond” (us). And D2.3(F1 delay) reuses the DL F1 delay which defined in TS28.552, the units is “microsecond” (us). CATT[2] thought that the different parts of the UL delay in RAN side use the different units, and cannot be added together.

**Proposal 10(Cat b): Confirm the unit of the UL PDCP queuing delay reported by UE is “ms” and add a section of “UL F1 delay” in TS38.314 with “ms” unit.**

**Proposal 11(Cat b): RAN2 to confirm the unit of UL PDCP queueing delay is 0.1ms (Current 38.331CR), instead of 1ms (last agreement).**

Question 3.5: Are you agree with above proposals?

|  |  |  |
| --- | --- | --- |
| Company | Support or not | Comments |
| QUALCOMM | Proposal 10 is ok | For Proposal 11, we prefer to keep last RAN2 agareement |
| Ericsson | Support Proposal-11 | We are okay with proposal-11 as this introduces the possibility to exchange D1 measurement with finer granularity in higher numerologies.  Regarding the introduction of a new section of UL F1 delay, we do not think this is necessary as the OAM that receives all the split delay measurements can compute the total delay by adding them while knowing that some are in milli-second’s unit (CU-UP delay, RLC delay), some are in 100 micro second’s unit (D1 delay) and some other are in micro-second (F1 delay). As long as the OAM knows the unit of each measurement, we do not see any reason why the OAM cannot add these measurements to calculate total delay. |
| Huawei, HiSilicon | Proposal 11  + the introduction of a new section of UL F1 delay | For 0.1 ms granularity, it is not aligned with the previous RAN2 agreement and we think there are at least two reasons to have 0.1 ms granularity:  (1) RAN part of UL/DL delay feature is mainly for URLLC service, and thus there may be quite small delay. If it is referring to UL PDCP queuing delay, it may be even smaller, e.g. 0.5ms or smaller ones. If we keep “ms” unit (i.e. millisecond), the UL PDCP queuing delay may be useless  (2) As mentioned by the email rapporteur, SA5 defined delay measurements have the unit “microsecond” (us) (according to TS 28.552), which is much smaller than ours (0.1ms or 1ms). Honestly speaking, the microseond granulatiy is reasonable as 5G will be enhanced in the future so that RAN part of UL/DL delay will be smaller and smaller. Considering that we will sum all parts into a DL or UL delay from RAN2 point of view, the gap is not neglectful. So 0.1 ms for UL PDCP queuing delay is a compromise solution  For the introduction of a new section of UL F1 delay, the reasons are as below:  In the lastest 38.314 CR, there is a note:  NOTE: The total RAN part of UL packet delay measurement is the sum of D1(PDCP queuing delay, as defined in 4.2.1), D2.1(over-the-air delay, as defined in 4.1.1.2.1), D2.2(RLC delay, as defined in 4.1.1.2.2), D2.3(F1 delay, as defined in TS 28.552 [2]) and D2.4(PDCP re-ordering delay, as defined in 4.1.1.2.3)  Here it says D2.3 (F1 delay, as defined in TS 28.552 [2]), however, after internal check, we only find F1 delay for DL, i.e. 5.1.3.3.2 Average delay DL on F1-U. So the issue is that no specs have defined the UL F1-U delay, while all other measurements have clear definitions (either in RAN2 or SA5).  In order to solve the issue, we think either we try to introduce it in RAN2, or in SA5 (via a LS). If we do nothing, it will be very strange and it may lead to potential problem for implementing the UL F1-U delay (because no specs have defined it). |
| CATT | Proposal 10(ms) or use unit “us” | We just want to unify the unit of UL delay, we are both fine with microsecond or millisecond, but don’t agree to mix the unit in different UL delay part.  To use unit “ms” is the agreement of last meeting. Since all the other parts (Dx measurement) of UL delay in 38.314 use the unit “ms”, it is reasonable to also use the same unit for D1 measurement.  If “ms” is considered not enough for URLLC service, since the data in NR is scheduled per symbol, 0.1ms is also not an optimum granularity, especially for larger subcarrier space. Therefore to use the unit of “microsecond” (us) could also be considered to align with the F1 delay defined in 28.552, and the value range of the UE reported UL average delay may be extended to 0~10e6. |
| ZTE | Proposal 11 | We see the benefits to introduce finer granularity for reporting. Just for clarification, if 0.1ms granularity is only used for PDCP queuing delay? Or will it be used for all delay related measurement? |

###### Issue 3.6: Value range for *ReportInterval*

The value range for *ReportInterval* in 38.331 CR is still open. CATT[2] propose to use the following values.

Proposal 12(Cat b): Set the value range of the *reportInterval* field for UL delay measurement to:

* For PDCP queuing excess delay measurement, use “ms1024, ms2048, ms5120 or ms10240” (if the measurement is maintained in NR);
* For average PDCP queuing delay measurement, use “ms120, ms240, ms480, ms640, ms1024, ms2048, ms5120 or ms10240”.

Above proposal has already been discussed in MDT topic.

###### Issue 3.7: Introducing Min and max value for delay measurement

In addition to average delay measurement, Huawei, HiSilicon[13] see benefits to also have min and max value in the same period as the average delay.

**Proposal 13(Cat b): It is proposed to introduce min and max value for delay measurement in addition to the average delay.**

Question 3.7: Are you agree with above proposals?

|  |  |  |
| --- | --- | --- |
| Company | Support or not | Comments |
| CMCC | Support | The TP provided in R2-2001370 is fine for us.  Question for clarification, does the proposal apply to both D1 and D2 delays?  [Huawei, HiSilicon]  Our views are as below:  We think that min and max values are generally applicable for RAN part of UL/DL delay, and thus it requires to collect min/max values for each part, e.g. D1, D2, and other measurements.  Among all parts, D1 (UL PDCP queuing delay) is impacting UE and ASN1, so it is more important to consider min/max values for it. For all other parts (purely collected by network side), either we can leave network implementation to collect min/max value, or we update the specs by adding min/max values for them. Anyway, D1 is more urgent. |
| QUALCOMM | No | UL PDCP packet delay measurement has been discussed in RAN2 for a long time, we think the average PDCP queuing delay measurement is sufficient. |
| Ericsson | May be | In our understanding, this measurement is more useful at the ‘cell level’ (28.552 spec) counter and not at the ‘per UE level’ counter (38.314 spec) as different UEs in the cell might have different radio conditions and thus it will result in different min and max values compared to the average values. Therefore, we propose to send this enhancement to SA5 specification rather than in our specification. |
| Huawei, HiSilicon | Support | The average delay measurement is an average values of lots of samples. While the min and max delay measurements are to accurately identify the delay performance, e.g. the worst packet.  For example, assumed that the target URLLC delay is 0.5 ms for DL and UL RAN part delay, one may get the average value 0.4ms or smaller, and it seems ok. In this scenario, the max value (during a period) may be 10ms and it significantly impacting user experience, HOWEVER, it may be impossible to identify the “bad” packet/delay due to average delay. In other words, the intention of delay measurements is to help finding out delay problem, but unfortunately the average delay may not be enough to meet the intention.  From UE point of view, it will not introduce too much burden, because the UE doing average delay collection will anyway get delay sample of each packet. We think that min/max value will also have some impacts to 38.331 CR, but the impacts will be minor, i.e. just add min/max value in the delay reports.  Regarding CMCC’s question, our feedbacks are added above. |
| CATT | Maybe | The motivation is valid, but we should check with SA5 whether SA5 has this requirement. |
| Samsung | Not support | We think average delay is sufficient |
| ZTE | Maybe | For the measurement on the max value of delay , we understand the intention, and we also agree that it is important to identify the “bad” packet. However, compared to the measurement on the maximum value of delay, we think the more important thing is to record the number/ratio of “bad” packets, thus we prefer to prioritize the excess delay measurement. If the excess delay measurement is adopted, then we are open to the measurement on the max value of delay.  For the measurement on the max value of delay, we have similar question as CMCC on the D1/D2 issue, and we also want to clarify that whether and how to combine the D1and D2 to derive an overall max delay. We think it is not reasonable to derive the overall maximum delay by simply adding the maximum delay of D1 and D2.  In addition, considering the first packet may trigger BSR/SR/RACH, which may lead to much longer delay, we wonder whether we need to differentiate the first few packet from the following one (e.g. only consider the maximum delay in the continuously scheduling/transmission). We think this may depend on the usage of the maximum delay value measurement, and more discussion may be need to have better understanding on this aspect.  For the measurement on the min value of delay , we don’t see clear usage for this measurement, and we think it can be excluded from this release. |

###### Issue 3.8: Whether to introduce excess delay measurement in NR?

* **Support: Huawei, HiSilicon[11], ZTE[20]**
* **Not support: Qualcomm[1]**

During email discussion, companies have different understanding on the agreement 2 in last meeting, and no consensus is reached during the email discussion. Rapporteur suggests RAN2 to make final decision in the coming meeting.

Huawei, HiSilicon[11], ZTE[20] though that with the excess delay measurement, the network can know the delay distribution and thus some optimization means can be considered. Both delay measurements are useful and they are beneficial for monitoring delay status.

While, Qualcomm[1] thought that all the information provided by UL PDCP packet excess delay ratio measurement can be provided by UL PDCP packet average queuing delay. And it doesn’t bring any benefit to the network by requiring UE to report D1 as UL PDCP packet excess delay ratio meanwhile at the RAN node side, the remaining part of RAN delay(D2) is measured as average delay.

It is clearly that 2 network vendor support to introduce excess delay to know more about delay distribution. While 1 chipset vendor thinks there is no benefit. **If consensus can not be made during email discussion, we will postpone it to next release.**

**Proposal 14(Cat b): RAN2 is kindly asked to decide whether to support excess delay or not.**

**Question 3.8: In addition to the opinions in the above contributions, is there anything else companies want to comment? Otherwise, rapporteur suggest to postpone.**

|  |  |  |
| --- | --- | --- |
| Company | Support or not | Comments |
| QUALCOMM | NO | As we know, UL PDCP packet excess delay ratio measurement was introduced in LTE for the QoS verification in MDT as a compromised method of average queuing delay measurement. In NR, per SA2 requirements on E2E delay measurement, UL PDCP average queuing delay measurement is supported. All the information provided by UL PDCP packet excess delay ratio measurement can be provided by UL PDCP packet average queuing delay, there is no need to support two UL PDCP packet delay measurement methods in NR.  Secondly, QoS verification is end to end, it doesn’t bring any benefit to the network by requiring UE to report D1 as UL PDCP packet excess delay ratio meanwhile at the RAN node side, the remaining part of RAN delay(D2) is measured as average delay.  Thirdly, there is only a single UE capability introduced in NR for UL delay measurement (ulPDCP-Delay) in NR, we don’t think it can be extended implicitly to require UE to support two UL delay measurement methods.  **It is unnecessary to support UL PDCP packet excess delay ratio measurement in NR.** |
| Ericsson | Support | In our understanding, the average delay and excess delay ratio provide different insights.  In the current average delay measurements, the D1 delay is calculated for all the packets within the time duration of *reportInterval* and the average value of the same is reported to the RAN node. Currently, the smallest value that one can configure for *reportInterval* is 120ms and therefore the average delay would hide those packets which might fail to meet certain QoS criterion when most packets meet the criterion. In such a circumstance, the excess delay ratio informs exactly what percentage of the packets failed to meet this criterion. |
| Huawei, HiSilicon | Support | For excess delay measurement, it has been captured in TS 38.314 and 38.331, and the detailed standard impacts have been shown. The measurement was introduced in LTE, and we see similar use case/benefits to have it in NR.  For average delay measurement, we think it is a different story from excess delay, and it is more about how to monitor and guarantee RAN part of UL/DL delay.  Generally, we think excess delay measurement is useful from network performance and optimization point of view.  If we have both average delay and excess delay, we may have two separate UE capability bits, and we think it should be a compromise. |
| CATT | No strong view | Average delay is the requirement from SA5, whether SA5 has the requirement for excess delay should check with SA5, RAN2 should not make this decision alone. |
| ZTE | Support | We don’t consider excess delay as a compromise of average delay, excess delay provide NW additional information on distribution of packet delays, e.g. NW can understand the percentage of packets with higher delay than configured threshold.  Small clarification, if excess delay is introduced, we think the same granularity of *reportInterval* as agreed for average delay shall also be used for excess delay report configuration. The meaning is the same value range is used, not the same value. NW is flexible to configure separate report interval for excess delay and average delay reporting.  Another issue needs to be clarified is whether UE also report both average delay and excess delay if it supports both measurement. In our understanding, in this case, NW is flexible to configure UE to report both measurements or either of which taking into account UE capability, and UE just follows NW’s configuration. |

If excess delay is supported, RAN2 also needs to discuss on the capability for UL PDCP delay measurement.

Proposal 15(Cat b): If both excess delay and everage measurements are supported, RAN2 also needs to discuss on the capability for UL PDCP delay measurement.

* **Option 1: 1 capability for supporting both measurements.**
* **Option 2: 2 separate capability for average delay and excess delay.**

**Question 3.9: Only if there is consensus on excess delay, we can further share views on capability for excess delay. Otherwise, please ignore this question.**

|  |  |  |
| --- | --- | --- |
| Company | Option 1 or 2 | Comments |
| QUALCOMM | Option 2 | We don’t think it is necessary to introduce excess delay ratio measurement in NR.  And if we have to do it, separate capability for average delay and excess delay is needed. |
| Ericsson | No strong view |  |
| Huawei, HiSilicon | Option 2 | As we commented for Q3.8, we think it is acceptable to have two separate UE capability bits for two features. |
| CATT | Option 1 | More simple |
| ZTE | No strong view | If companies have strong concern on supporting both delay, it is fine to have separate capabilities. |

If Excess delay is supported, Huawei[14] thought that the M6 measurement should be a general measurement that at least covering measurements that need UE reporting, and it will be easy for RAN3 on the specification work, e.g. RAN3 needs to capture M6 measurement configuration in their specs.

Proposal 16(Cat b): M6 measurement in NR in TS 37.320 should cover at least delay ratio and average delay measurements.

Rapporteur suggest to not discuss proposal 16 until there is consensus on excess delay.

###### Issue 3.9: Some further enhancement for UL queueing delay measurement

Huawei[14] propose to introduce histogram of PDCP queueing delay.

**Proposal 17(Cat c): It is proposed RAN2 to discuss reporting of the histogram of the PDCP queuing delay. Network can configure more than one delay thresholds for the ratio reporting.**

Another proposal from Huawei[14] is that for delay for all packets, operators may want to know the x% worst delay value, e.g. x could be 99%. The reason behind is that such values can reflect the QoS on some certain levels and it can avoid too much measuerment reports. Therefore, we think the network can configure the reporting of the x% worst delay value.

**Proposal 18(Cat c): It is proposed RAN2 to discuss reporting of the x% worst delay value.**

Question 3.9: Do you support proposal 17&18 in Rel-16 or postpone it to next release?

|  |  |  |
| --- | --- | --- |
| Company | Support or not | Comments |
| CMCC | R17 | We agree with the intention to have clear view on distribution of delay. But we prefer to have the same enhancement on matrix for both D1 delay and D2 delay. Probably we can work on it in R17. |
| QUALCOMM | Rel-17 |  |
| Ericsson | R17 | Agree to postpone this to rel-17 |
| Huawei, HiSilicon | R17 | It seems not enough time for P17 and P18, so we are ok to postpone this to R17. |
| CATT | R17 |  |
| Samsung | Rel 17 |  |
| ZTE | R17 | Can be postponed to R17. |

###### ~~Issue 3.10: A general definition of DL measurement is missing~~

~~Huawei, HiSilicon [15] observed that a general definition of DL measurement is missing, and it may lead to some ambiguity on the whole DL measurement. One option is to capture a simple sentence to illustrate the definition in TS 38.314~~

~~Proposal 19(Cat b): Capture a general definition of DL measurement in TS 38.314:~~

~~Packet delay includes RAN part of delay and CN part of delay. For RAN part, the DL delay comprises:~~

~~- D1 (the DL delay in gNB-DU), referring to 5.1.1.1.1 Average delay DL air-interface in TS 28.552~~

~~- D2 (the DL delay on F1-U), referring to 5.1.3.3.2 Average delay on F1-U in TS 28.552~~

~~- D3 (the DL delay in CU-UP), referring to 5.1.3.3.1 Average delay DL in CU-UP in TS 28.552~~

Rapporteur note: The proposal above has already been agreed. No need to discuss

## 2.4 Number of UEs

###### Issue 4.1: Flooring operation may results to zeroing for low load scenario

The flooring operation is used in the mean number of active UEs definition so that the result can be defined as an integer. Ericsson and CMCC[7] observed that flooring operation for the mean number of active UEs results in ‘zeroing’ the information in low load scenario. The following agreement has been achieved.

7 The flooring operation associated to the definition of mean number of active UEs is removed.

In the annex of R2-2001112[7], a TP is provided with the following matrix for mean number of active UE.

**Question 4.1: Is the above matrix for mean number of active UE per DRB per cell acceptable for you? If not, please provide a suggestion as well.**

|  |  |  |
| --- | --- | --- |
| Company | Support or not | Comments |
| CMCC | Support |  |
| QUALCOMM | support |  |
| Ericsson | Support |  |
| Huawei, HiSilicon | Support |  |
| CATT | Support |  |
| Samsung | Support |  |
| ZTE | Support |  |

###### Issue 4.2: Introducing number of inactive context per RNA

ZTE, Sanechips[3] observes that since network is always aware of the RNA inactive UE belongs to, it is possible to count the number of inactive UE context (both mean and maximum number) stored per RNA, which can be used for the RAN node to know the resource consumption in granularity of RNA level, therefore to help for configuration of RNA at network’s side.

**Proposal 23(Cat c): The number of UE inactive context stored (both mean and maximum number) can be counted per RNA to help the network to optimize the configuration of RNA.**

**Question 4.2: Do you support proposal 23 for Rel-16 or postpone to further release?**

|  |  |  |
| --- | --- | --- |
| Company | Support or not | Comments |
| QUALCOMM | No strong view |  |
| Ericsson | Not support | An RNA can be a UE specific configuration and having different counters per RNA in a gNB could create a lot of overhead on the network side.  Additionally, the purpose of having the Inactive UE counter is to notify the OAM about the overhead of storing the UE context in the gNB. As long as the UE context is stored in the gNB, there is a memory overhead and this is independent of the RNA to which that particular UE belongs to.  Therefore, we do not support this proposal. |
| Huawei, HiSilicon | No strong view | We wonder whether it is SA5 business, because it looks like that SA5 can try to define a wider measurement (e.g. per RNA) based on what RAN2 has defined. |
| CATT | No strong view | Try to check the requirement with SA5 |
| Samsung | No strong view | We don’t see a clear benefit and need for counting inactive UEs per RNA. This measurement can always be determined from implementations. |
| ZTE | R17 | When UE in RRC Inactive state, it is required to update RNA periodically or when moving outside of RNA configured. To introduce the RNA granularity INACTIVE UE counting is not only to understand the overhead required for storing UE context, but also to help NW evaluate the paging capacity within certain area, so that NW can optimize the configuration of RNA. If the RNA is too large with many INACTIVE UE the paging capacity might be insufficient to meet the paging requirement, if the RNA is too small, the RNA update could be too frequent.  Based on above analysis, we think it is beneficial to introduce RNA granularity Inactive UE measurement, however, considering the time budget, it is fine to discuss this measurement further in R17. |

###### Issue 4.3: New measurement on Counting number of UE whose duration time in RRC\_INACTIVE is below a configured threshold

ZTE, Sanechips[3] thinks that inappropriate determination of state transition will send UE to RRC\_INACTIVE mode in vain, e.g., UE initiates RRC connection resume procedure shortly after transition to RRC\_INACTIVE mode, which delays the data transmission unnecessarily while consumes more power at UE’s side. The number of users whose duration time in RRC\_INACTIVE state is below a configured threshold can be used to diagnose if the decision on state transition is appropriate.

**Proposal 24(Cat c): To optimize the determination on state transition, measurement on the number of users whose duration time in RRC\_INACTIVE is below a configured threshold, shall be supported.**

**Proposal 25(Cat c): The number of UE, whose duration time in RRC\_INACTIVE is below a configured threshold, shall be count per cell, where the INACTIVE UE is considered “belongs to” the cell in which the UE was released from RRC\_CONNECTED state to RRC\_INACTIVE state.**

**Question 4.3: Do you support proposal 24&25 for Rel-16 or postpone to further release?**

|  |  |  |
| --- | --- | --- |
| Company | Support or not | Comments |
| QUALCOMM | Rel-17 |  |
| Ericsson | Not support | This requires the network to maintain a timer for each inactive UE which will create a large network overhead. Simillar metric can be achieved by enhancing the mobility history information (number of state transitions per cell) and this can be studied in rel-17. |
| Huawei, HiSilicon | Rel-17 |  |
| CATT | R17 |  |
| Samsung | Rel 17 |  |
| ZTE | R17 | Considering the time budget, it is fine to discuss it further in R17. |

## 2.5 Throughput

###### Issue 5.1: Whether to inform TCE that DC duplication is enabled for throughput measurement

The immediate MDT related throughput measurement is performed at the RLC level (in the DU). Ericsson[9] observes that, in the DC based DL PDCP duplication scenario, the MN RLC and SN RLC receive the same set of packets to be transmitted the UE (DL scenario). If an immediate MDT session associated to throughput measurement in both MCG and SCG is activated for a given UE by the OAM, then the MN and the SN sends the respective RLC throughput measurements to the TCE. In the DC scenario, the TCE is unaware of whether the DC based DL PDCP duplication is enabled or disabled to this UE.

Ericsson[9] thinks that, for the DC scenario, there is a benefit in knowing whether the PDCP duplication is enabled or not while including the UE specific throughput measurements as part of immediate MDT.

**Proposal 26(Cat c): If the UE is in DC scenario and if the RAN node receives the signaling based MDT request associated to UE throughput measurements, the RAN node shall notify the TCE whether the PDCP duplication is enabled or not at per DRB level. Draft LS is also provided in [6].**

**Question 5: Do you support proposal 26 for Rel-16 or postpone to further release?**

|  |  |  |
| --- | --- | --- |
| Company | Support or not | Comments |
| QUALCOMM | Support |  |
| Ericsson | Support | This is needed to calculate the overall UE throughput based on the throughput measurements received from MN DU and SN DU. |
| Huawei, HiSilicon | Postponed to R17 | At RAN2#108, we agreed:  1 Only immediate MDT is supported for EN-DC scenario in R16 MDT  2 In signaling based immediate MDT, MME provides MDT configuration for both MN and SN towards MN including multi RAT SN configuration, specifically E-UTRA and NR MDT configuration. MN then forwards the NR MDT configuration towards SN (EN-DC scenario, SN is always NR).  For EN-DC case, for LTE MN, currenlty M5 (PDCP level throughput) has been defined in TS 36.314; for NR SN, it is RLC throughput (as throughput measurements are defined for gNB as in TS 28.552). So it is impossible for RAN (LTE MN, NR SN) to report two RLC throughputs, i.e. the above description “then the MN and the SN sends the respective RLC throughput measurements to the TCE.” may not be accurate. And thus it needs more time to check P26, and maybe it is good to do it in R17. |
| CATT | Support |  |
| Samsung | Support |  |
| ZTE | No support | In our understanding, the UE throughput defines the amount of packets transmitted in each time unit, which can be used to evaluate the transmission performance, and mainly depends on the radio link quality and congestion level in the cell. The usage of the duplication indicator is not clear to us.  In addition, considering the duplication can be activated/deactivated very dynamically based on the radio quality by NW implementation, it is not clear how to deal with the duplication indicator if the duplication status changes within one time unit.  Moreover, considering the PDCP will discard the packet in one leg based on the transmission confirmation received from the other leg, it is not clear how can the NW derive the overall UE throughput based on the duplication indication. |

## 2.6 Others

###### Issue 6.1: Do we need to captue a Note on the total RAN part of DL packet delay measurement?

In the latest 38.314 CR, there is a Note on The total RAN part of UL packet delay measurement, and generally it explains how the UL packet delay forms.

NOTE: The total RAN part of UL packet delay measurement is the sum of D1(PDCP queuing delay, as defined in 4.2.1), D2.1(over-the-air delay, as defined in 4.1.1.2.1), D2.2(RLC delay, as defined in 4.1.1.2.2), D2.3(F1 delay, as defined in TS 28.552 [2]) and D2.4(PDCP re-ordering delay, as defined in 4.1.1.2.3)

However, we do not have a Note for DL delay. It is our understanding that SA2 requirement on E2E delay is to have separate DL and DL delay measurements, so it is suggested to have uniform definitions. In addition, in TR 37.816 for the SI RDCU, we had agreed on the concept of DL packet delay (shown as below).

6.2.2.2.2 DL packet delay measurement

RAN part of the DL delay is measured by gNB by DRB level. For arrival of packets the reference point is PDCP upper SAP. For successful reception the reference point is MAC lower SAP. It includes average delay in DL (e.g. average delay in CU-UP, average delay on F1-U and average delay DL in gNB-DU)\*. The delay may be converted to QoS flow level by gNB with the assumption that all QoS flows mapped to one DRB get the same QoS treatment.

Note: the measurements “average delay DL in CU-UP” and “average delay on F1-U” in the Table 6.2.2-1 could be used.

Our proposal is to add a new Note to TS 38.314 CR as below:

NOTE: The total RAN part of DL packet delay measurement is the sum of D1 (Average delay DL in CU-UP, as defined in TS 28.552 [2]), D2 (Average delay DL on F1-U, , as defined in TS 28.552 [2]) and D3 (Average delay DL air-interface, as defined in TS 28.552 [2]).

**Question X: Do we need to captue a Note on the total RAN part of DL packet delay measurement?**

|  |  |  |
| --- | --- | --- |
| Company | Support or not | Comments |
| Huawei, HiSilicon | Support | We have had a Note for RAN part of UL packet delay, so it is suggested to also add a Note for RAN part of DL packet delay and the Note does not introduce new measurements. |
| CATT | Support |  |
| ZTE | Maybe not | We think current wording is clear enough, why we shall add a note here? |
| Ericsson | Not required | We believe that the existing UL related NOTE needs to be either updated or removed.  Firstly, the existing UL related NOTE states that the ‘total RAN delay’ consists of D1+D2.1+D2.2+D2.3+D2.4. However, the D1 also includes some UE implementation specific delays. Because the definition of D1 as specified in section 4.2.1.2 states that this is the delay from packet arrival at PDCP upper SAP until the UL grant to transmit the packet is available. Based on this definition, there are two components of D1 delay.   1. D1.1: The first one being UE implementation specific i.e., the time between the packet arrival at PDCP upper SAP to the time of sending the scheduling request. 2. D1.2: The second one being the ability to send the scheduling request and receive the scheduling grant from the DU.   Of these two, the first one is completely UE implementation specific and RAN node cannot affect it. Therefore, calling the including D1 in the ‘total RAN delay’ is not correct. We propose to either remove the D1 from the existing NOTE or completely remove the NOTE.  Secondly, as stated in the objective of each of these individual measurements, the measurements are for the OAM performance observability purposes. The OAM gets to know the delay experienced by the packets being transmitted in a particular DRB in each part of the split RAN node. It is up to the OAM to use it for the total delay computation or only perform localized analysis of the delay. Additionally, the NOTE that is captured only reflects the situation in the split RAN node scenario and not in the non-split scenario. Strictly speaking one has to define this NOTE for other deployment scenarios as well. Therefore, instead of adding such additional NOTEs, we propose to remove the existing NOTE itself.  For the new proposal of adding the NOTE for DL delay, we think it is not needed based on the same arguments as listed above. |

# 3 Reference

1. R2-2000000 Remaining Issues of UL PDCP Packet Average Queuing Delay Measurement Qualcomm Incorporated discussion Rel-16 38.314 NR\_SON\_MDT-Core
2. R2-2000103 Correction and Open Issues of UL Delay Measurement CATT discussion Rel-16 NR\_SON\_MDT-Core
3. R2-2000806 Further consideration on INACTIVE UE counting ZTE Corporation, Sanechips discussion Rel-16 NR\_SON\_MDT-Core
4. R2-2000909 Summary of L2M open points and proposals CMCC discussion NR\_SON\_MDT-Core Late
5. R2-2000910 Clarification for per cell PRB usage CMCC, Ericsson, Huawei discussion NR\_SON\_MDT-Core Revised
6. R2-2001110 [DRAFT] LS on throughput measurement in DC based PDCP duplication scenario Ericsson LS out Rel-16 NR\_SON\_MDT-Core To:TSG RAN3, SA5
7. R2-2001112 Definition of number of active UEs Ericsson, CMCC discussion
8. R2-2001113 Handling PDCP queueing delay measurements in split bearer scenario Ericsson discussion
9. R2-2001119 Throughput measurement in duplication scenario Ericsson discussion
10. R2-2001147 Per DRB measurements in TS38.314 Nokia, Nokia Shanghai Bell discussion Rel-16 NR\_SON\_MDT Late
11. R2-2001368 Discussion on per DRB and excess delay measurement for L2M Huawei, HiSilicon discussion Rel-16 NR\_SON\_MDT-Core
12. R2-2001369 TP on per DRB measurements for L2M Huawei, HiSilicon pCR Rel-16 38.314 0.0.3 NR\_SON\_MDT-Core
13. R2-2001370 Discussion on min and max value for delay measurements for L2M Huawei, HiSilicon discussion Rel-16 NR\_SON\_MDT-Core
14. R2-2001371 Discussion on UL delay measurements in non MR-DC Huawei, HiSilicon discussion Rel-16 NR\_SON\_MDT-Core
15. R2-2001372 Discussion on DL delay measurements Huawei, HiSilicon discussion Rel-16 NR\_SON\_MDT-Core
16. R2-2001373 Discussion on delay measurements for MR-DC and CA Huawei, HiSilicon discussion Rel-16 NR\_SON\_MDT-Core R2-1915856
17. R2-2001419 Clarification on PRB usage ZTE Corporation, Sanechips discussion Rel-16 R2-1915418 Withdrawn
18. R2-2001440 On Mapping 5QI to DRB Mediatek Inc discussion
19. R2-2001603 Clarification for per cell PRB usage CMCC, Ericsson, Huawei, ZTE discussion NR\_SON\_MDT-Core R2-2000910
20. R2-2000907 Summary of open points in [108#91][NR/L2M] running 38.314 CR (CMCC) CMCC discussion Rel-16 NR\_SON\_MDT-Core