**3GPP TSG-RAN WG2 Meeting #112 electronic R2-200xxxx**

**Elbonia, Nov 2nd – 13th 2020**

**Agenda item:** 8.11.2

**Source:** Intel Corporation

**Title:** Report of [Post111-e][625][POS] End-to-end latency analysis (Intel)

**Document for:**  Discussion and decision

# Introduction

This contribution provides report for RAN WG2 email discussion:

* [Post111-e][625][POS] End-to-end latency analysis (Intel)

Scope: Discuss which nodes and which procedures are involved in a positioning latency analysis, and capture expected latency values where possible.

Intended outcome: Report to next meeting

Deadline: Oct 15th , 2020

Rapporteur proposes to divide the discussion in two phases:

**Phase 1**: Based on the contributions in last meeting, and the LS/agreements from RAN1, discuss which nodes, procedures, solution (UE based/UE assisted) and RRC state should be considered in our analysis, the basic procedure for positioning methods and latency assumption for different nodes, etc. The goal of this phase is to provide analysis for the Rel.16 RAT dependent NR positioning solutions.

Deadline: Sep 30th

**Phase 2**: Finalize the E2E latency value range for different positioning methods based on discussion in phase 1, and collect potential enhancements/directions to reduce the latency (companies should show the latency analysis when propose the enhancements).

Deadline: Oct 15th

To make it easier to find the correct contact delegate in each company for potential follow-up questions, the rapporteur encourages the delegates who provide input to provide their contact information in this table:

|  |  |
| --- | --- |
| Company | Delegate contact |
| COMPANY\_NAME | NAME ([email@address.com](mailto:email@address.com)) |
| Intel | yi.guo@intel.com |
| Ericsson | [ritesh.shreevastav@ericsson.com](mailto:ritesh.shreevastav@ericsson.com), fredrik.gunnarsson@ericsson.com |
| Qualcomm | Sven Fischer (sfischer@qti.qualcomm.com) |
| OPPO | qianxi.lu@oppo.com |
| CATT | Jianxiang Li (lijianxiang@datangmobile.cn) |
| Samsung | June Hwang (june77.hwang@samsung.com) |
| Xiaomi | lixiaolong1@xiaomi.com |
| vivo | yuanyuanwang@vivo.com |
| InterDigital | Jaya Rao ([jaya.rao@interdigital.com](mailto:jaya.rao@interdigital.com)), Fumihiro Hasegawa (fumihiro.hasegawa@interdigital.com) |
| Huawei,HiSilicon | [yinghaoguo@huawei.com](mailto:yinghaoguo@huawei.com) |
| ZTE | Liu Yansheng(liu.yansheng@zte.com.cn) |
| Spreadtrum | [Huifang.fan@unisoc.com](mailto:Huifang.fan@unisoc.com) |
| Nokia | Mani.Thyagarajan@nokia.com |

# Discussion

## Phase 1 (Analysis for Rel.16 NR Positioning)

As indicated in [2],

|  |
| --- |
| *RAN WG1 evaluates physical layer latency and its potential reduction for NR Rel-17 positioning solutions. In order to evaluate End-To-End latency of NR positioning solutions the input from RAN WG2 is needed on latency components of NR/NG-RAN/5GC higher layer positioning protocols.*  *RAN WG1 respectfully asks if RAN WG2 can provide a list of latency components with corresponding range of values for the existing and any potential enhanced NR positioning solutions* |

Based on RAN1 LS, RAN2 should focus on the latency analysis of NR/NG-RAN/5GC higher layer positioning protocols, and RAN1 is responsible for physical layer latency analysis.

Below MO-LR/MT-LR procedures are cited from [8], TS23.273.



Figure 6.1.2-1: 5GC-MT-LR Procedure for the commercial location services



Figure 6.2-1: 5GC-MO-LR Procedure

Based on RAN1 agreements below, seems RAN1 will do the E2E latency analysis based on positioning procedure.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Agreement:   * Physical Layer Latency Start and End times are defined as follows:  |  |  |  | | --- | --- | --- | | **Method** | **Start** | **End** | | UE assisted DL-only & DL-ECID & Multi-RTT | Transmission of the PDSCH from the gNB carrying the LPP Request Location Information message | Successful decoding of the PUSCH carrying the LPP Provide Location Information message | | UL-only method & UL ECID & Multi-RTT | Reception by the gNB of the NRPPa measurement request message | The transmission by the gNB of the NRPPa measurement response message | | UE-based | Transmission of the PDSCH from the gNB carrying the LPP Request Location Information if applicable, otherwise,   * Alt. 1: transmission of the PUSCH carrying the MG Request from the UE. * Alt. 2: Transmission of the PDSCH from the gNB carrying the LPP message containing the assistance data * Alt. 3: Start of the Reception of DL PRS   Note: Suggest to downselect this at the next meeting.  Note: The high layers latency components may be subject to adjustment for different alternatives. | Successful decoding of the PUSCH at gNB carrying the LPP Provide Location Information message if applicable, otherwise Calculation of Location Estimate at the UE | |

In addition, based on SID, the E2E latency should be analysed based on rel-16 positioning solutions, i.e. positioning delay, step 5 in MO-LR and step 12 in MT-LR. Location service delay is out of RAN scope. But it would be good to check companies’ view on this.

* 1. Evaluate the achievable positioning accuracy and latency with the Rel-16 positioning solutions in (I)IoT scenarios and identify any performance gaps. [RAN1]

**Question 1-1: For RAN2 latency analysis of Rel.16 solutions, should we only consider the latency of positioning procedure, i.e. step 5 in MO-LR/step 12 in MT-LR or the latency of location service, i.e. whole procedure in MO-LR/MT-LR?**

**Alt 1: the latency of positioning, i.e. step 5 in MO-LR/step 12 in MT-LR**

**Alt 2: the latency of location service, i.e. whole procedure in MO-LR/MT-LR**

|  |  |  |
| --- | --- | --- |
| **Company** | **Alt 1 or Alt 2** | **Remark** |
| Intel | Alt 1 | The latency of positioning, i.e. step 5 in MO-LR and step 12 in MT-LR is under RAN control. The location service is out of RAN scope, and therefore should not be used for E2E latency analysis. |
| Ericsson | Alt 1 | Yes, Nodes involving RAN procedures and protocols (LPP, RRC, MAC) should only be studied |
| Qualcomm | Modification of Alt-2 | Deferred MT-LR for periodic/triggered events as specified in TS 23.273 section 6.3 and 6.7 should be added to the analysis. We think most applications (e.g., (I)IoT scenarios) require regular (e.g., periodic) location reports to an client.  An end-to-end latency analysis needs to consider the time between the client request for location (for MR-LR or MO-LR) or the triggering of location (for deferred MT-LR) and the availability of the location at the client.  However, only the basic steps in the procedures (e.g., MT-LR, MO-LR) need to be considered:   * No inclusion of roaming. * For MT-LR, no inclusion of privacy check based on current location (i.e., steps 16-23). * For MO-LR, no transfer to 3rd party (i.e., steps 8-11).   The rational for an end-to-end analysis is that user case requirements from SA1 are end-to-end and detailed definitions are available in Rel-16 to enable such analysis. However, for any enhancements that could affect the 5GCN, SA2 and CT WGs could be asked to comment. |
| OPPO | Alt-1 | Even if the other steps are needed for the latency analysis from an end-to-end perspective, it is out of RAN2 scope – i.e., sufficient for RAN2 to focus on the RAN-centric step. |
| CATT | Modification of Alt-2 | End-to-end latency analysis includes the client request and the response to client according to the definition from SA1. So we prefer to analyze end to end latency in phase 1.  Potential solutions on reduction of latency can focus on RAN’s nodes in phase2. |
| Samsung | Alt 1 | Obviously whole procedure in MO-LR/MT-LR described by the rapporteur includes not only RAN and UE but also the higher layer entities. The latter cannot be the scope of RAN2. |
| Xiaomi | Alt-1 | From RAN2 perspective, positioning procedures related RAN should be studied. And then RAN2 can send response LS to RAN1 about the higher layer latency analysis. According to the response LS from RAN2, RAN1 could trigger other WGs to study the latency reduction if necessary. |
| vivo | Alt 1 | Same view with Intel. The location service is out of RAN scope. |
| InterDigital | Alt 1 | The analysis should focus on the latency components of the MO-LR and MT-LR procedures that involve RAN procedures/signaling (i.e. LPP, NRPPs, RRC) and the latency (e.g. processing) at following nodes: LMF, AMF, NG-RAN (gNBs/TRPs) and UE.  While we agree with Qualcomm that end-to-end latency should be within the SA1 latency target, for the analysis the main focus should be latency components which are in the scope of RAN2. |
| Huawei/HiSilicon | Alt 1 | We are OK to only consider step 12 in MT-LR and step 5 in MO-LR for the sake of simplicity, yet this differs in the definition of end-to-end latency. We may call it NG-RAN positioning latency. |
| ZTE | Alt 1 | We share the same view with Intel. |
| Spreadtrum | Alt 1 | The positioning service latency is defined in TS 22.261 as the time elapsed between the event that triggers the determination of the position-related data and the availability of the position-related data at the system interface.  However, we can only focus on the part of RAN2 scope. |
| Nokia | Alt 1 | True E2E latency analysis requires efforts from multiple RAN WGs and CN WGs. The study item does not allocate objectives and time units for all WGs involved. RAN1 chose to focus on PHY layer latency analysis. Likewise, RAN2 should focus on latency analysis of higher layer RAN WGs owned protocols (LPP and NRPPa). One can argue RAN3 is responsible for NRPPs protocol but RAN2 as a stage 2 specification owner can take up the role of doing latency analysis involving both LPP and NRPPa protocols. However, any enhancement solutions to address gaps identified by such higher layer RAN protocols latency analysis but involving NRPPa protocol must also be evaluated and jointly decided by RAN2 and RAN3. |

The positioning procedure involves multiple nodes, e.g. UE, NG-RAN, AMF, LMF, VGMLC, HGMLC, UDM, external Client, NEF, AF, etc.

[4] mentioned, “*procedures external to the PLMN (between LCS client and GMLC), procedures related to roaming case (between V-PLMN and H-PLMN), and privacy verification procedures (e.g. between GMLC and UDM) are not included in the latency evaluation*”.

[5] also mentioned “*Considering the latency of the procedures between LCS client, GMLC, AMF for triggering the positioning is tightly related to deployment and out of RAN scope. The delay analysis in this contribution does not consider these procedures, and only count the procedures between UE, LMF, AMF and gNB*”.

The question is whether all involved nodes should be considered in latency analysis.

**Question 1-2: For RAN2 latency analysis of Rel.16 solutions, which nodes shall be considered? UE, NG-RAN, AMF, LMF, VGMLC, HGMLC, UDM, external Client, NEF, AF.**

|  |  |  |
| --- | --- | --- |
| **Company** | **UE, NG-RAN, AMF, LMF, VGMLC, HGMLC, UDM, external Client, NEF, AF** | **Remark** |
| Intel | UE, NG-RAN, AMF and LMF | The latency for other nodes, e.g. GMLC, etc are out of RAN WG2 scope. |
| Ericsson | Protocols and Procedures involving these entities:  UE, NG-RAN, AMF, LMF | The latency for other nodes is out of RAN2 scope. |
| Qualcomm | UE, gNB, AMF, LMF, GMLC, UDM, client/NEF/AF | For the latency of nodes considered out of RAN2 scope above, we can assume the same numbers as for other CN interfaces (e.g., AMF-LMF signalling). As previously commented, we should align the analysis with SA1 requirements in order to determine whether they can be supported. |
| OPPO | UE, NG-RAN, AMF, LMF | Even within the entities, there could be component that RAN3 view has to consulted, i.e., out of RAN2 expertise, e.g., the communication between NG-RAN/AMF/LMF. |
| CATT | UE, gNB, AMF, LMF, GMLC, UDM, client/NEF/AF | End-to-end latency analysis includes the client request and the response to client according to the definition from SA1. So we prefer to consider all modes for end to end latency analysis in phase 1.  Potential solutions on reduction of latency can focus on RAN’s nodes (e.g. UE, NG-RAN, AMF and LMF) in phase2. |
| Samsung | UE, NG-RAN, AMF, LMF | Same view with Intel. |
| Xiaomi | UE,NG-RAN,AMF and LMF | We should only study the nodes involving NG-RAN procedures. |
| vivo | UE, NG-RAN, AMF and LMF | Same view with Intel. |
| InterDigital | UE, NG-RAN, AMF and LMF | We agree with Intel and Ericsson to exclude the nodes that are out of RAN2 scope in the latency analysis. |
| Huawei/HiSilicon | UE, NG-RAN, AMF, and LMF |  |
| ZTE | UE, NG-RAN, AMF, LMF | We share the same view with OPPO. |
| Spreadtrum | UE, NG-RAN, AMF, LMF | The latency for other nodes is out of RAN2 scope. |
| Nokia | UE, gNB, AMF and LMF | RAN2 focus should be on nodes involved in the transport of LPP and NRPPa PDUs only. We should focus on gNB as a whole and not consider CU-DU split and the delays across F1 interface. If such gNB internal delay details are required, then it can be done by RAN3 at a future time. |

**Question 1-3: If the answer of Question 1-1 is alt 2, for RAN2 latency analysis of Rel.16 solutions, should MO-LR, NI-LR, MT-LR all be considered? Or only some of them.**

|  |  |  |
| --- | --- | --- |
| **Company** | **MO-LR, NI-LR, MT-LR** | **Remark** |
| Intel | MO-LR and MT-LR | If companies would like to consider Alt 2, MO-LR and MT-LR should be considered since MO-LR and MT-LR are typical cases for commercial use cases. |
| Qualcomm | Deferred MT-LR for periodic/triggered events, basic MT-LR and MO-LR | We consider Deferred MT-LR more typical for low-latency use cases. In fact, latency for a single location MT-LR or MO-LR may be high or very high in practice due to overheads in interfacing with an external client and paging the UE for an MT-LR. |
| CATT | All |  |
| Samsung | MO-LR, MT-LR | Have the same view with Intel |
| Huawei/HiSilicon | MT-LR and MO-LR | We think that NI-LR is more related to regulatory requirements, and thus should be excluded. |
| ZTE | All cases | We prefer to select alt1 in Q1-1.  If the final decision is alt2, we think all scenarios should be considered. |
| Nokia |  | This question 1-3 is unclear. Since the question is a conditional question depending on whether companies chose Alt 2 under question 1-1, we skip answering this question 1-3. However we would like to comment that irrespective of MO-LR or MT-LR, the focus for RAN2 should still be on the UE procedures box only i.e. step 5 in MO-LR and step 12 in MT-LR in the call flows shown above. |

As indicated in [1], RAN1 agreed:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Agreement:   * At least the following information is provided for positioning physical layer latency analysis:   + Source initiating request for positioning measurements/location for a given UE (UE, Network)   + Destination awaiting for positioning measurements/location for a given UE (UE, Network)   + Start and end triggers/events for physical layer latency evaluation     - For Rel.16 solutions, it is based on specification for each solution   + Initial and final RRC State of positioned UE (RRC IDLE, INACTIVE, CONNECTED) at the start and end time for the physical layer latency evaluation   + Positioning     - technique (enumeration): (1) DL-TDOA, (2) DL AoD, (3) UL-TDoA, (4) UL-AoA, (5) Multi-RTT, (6) E-CID     - type: DL, UL, DL+UL     - mode: UE-based, UE-assisted   + Latency component w/ value range and description, including information on any parallel (simultaneous) components   + Total latency value * Latency components are recommended to be captured in table and ordered consequently in time starting from the earliest one:  |  |  |  | | --- | --- | --- | | **Source [UE, NW]/Destination [UE, NW]**  **Positioning technique [DL-TDOA, E-CID, …], type [DL, UL, DL+UL], mode [UE-A, UE-B],**  **Initial and Final RRC States [IDLE, INACTIVE, CONNECTED]** | | | | **Latency Component** | **Value Range** | **Description of Latency Component** | | Start trigger |  |  | | Name of component 1 |  |  | | Name of component 2 |  |  | |  |  |  | | Name of last component |  |  | | End trigger |  |  | | Total values |  |  | |

In summary, for RAN1 physical latency analysis,

* All RRC state shall be considered (IDLE, INACTIVE, CONNECTED);
* All Rel-16 RAT dependent positioning methods shall be considered ((1) DL-TDOA, (2) DL AoD, (3) UL-TDoA, (4) UL-AoA, (5) Multi-RTT, (6) NR E-CID);
* All positioning types shall be considered (type: DL, UL, DL+UL);
* Both UE-based, UE-assisted shall be considered;
* List latency component w/ value range and description, including information on any parallel (simultaneous) components

Rapporteur’s comments, from RAN2 perspective:

* The positioning type (UL, DL, DL+UL) has been reflected by positioning techniques, therefore we do not need to list them separately;

**Question 1-4: Based on RAN1 agreements, shall both UE based and UE assisted be considered in RAN2 latency analysis? Or one of them ? If one of them, pls indicate which positioning methods.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Both , UE based only, UE assisted only** | **Remark** |
| Intel | Both |  |
| Ericsson | Both |  |
| Qualcomm | Both | For the Rel-16 methods, the "positioning signalling" is the same for UE-based and UE-assisted (e.g., Step 10 in Figure 1 below may be performed by the UE or by the LMF). Differences occur based on the location of the consumer of the position estimate (i.e., UE or external client). |
| OPPO | Both |  |
| CATT | Both |  |
| Samsung | Both |  |
| Xiaomi | Both |  |
| vivo | Both |  |
| InterDigital | Both |  |
| Huawei/HiSilicon | Both | Although we are fine with both UE based and UE-assisted, we suggest that MT-LR should not consider UE-based. MO-LR can consider both. |
| ZTE | Both |  |
| Spreadtrum | Both |  |
| Nokia | Both | The RAN1 format for capturing latency values for different latency components does not seem very intuitive. From RAN2 perspective, since our focus is on step 5 in MO-LR and step 12 in MT-LR in the call flows shown above, we should focus on capturing the latency values from different positioning techniques perspective. However, we should consider providing latency values for different scenarios involving UE-based vs UE-assisted. |

**Question 1-5: For RRC state, should we consider both scenarios for RAN2 latency analysis? Or only one of them?**

* + Scenario 1: To support MO-LR, MT-LR, NI-LR procedure, if the UE was in IDLE/INACTIVE, the state transmission to CONNECTED is needed before the positioning procedure (step 5 for MO-LR and step 12 for MT-LR) in order to exchange the message between the UE and the network.
  + Note: It is only applicable for alt 2 in question 1-1. For alt1, the transition time is not counted in positioning procedure itself although it is part of location service procedure.
  + Scenario 2: For the scenario that the UE obtains the assistance data via system information and provides the results to application inside the UE. The UE does not need to exchange the message with network, and the state transmission is not needed, i.e. initiate state could be IDLE, INACTIVE or CONNECTED, and the UE does not change the state when performing the positioning.
  + Note 1: this is based on rel-16, i.e. the UE cannot send/receive the data, reference signals in IDLE or INACTIVE mode.

|  |  |  |
| --- | --- | --- |
| **Company** | **Both, scenario 1 only or scenario 2 only** | **Remark** |
| Intel | Scenario 2 only | Scenario 1 and 2 are already supported in Rel-16. But procedure of scenario 1 is not part of step 5 for MO-LR and step 12 for MT-LR.  The scenario 1 could be considered only if the whole procedure for location service is considered, i.e. alt2 in question 1-1. |
| Ericsson | Scenario 2 including both UE-A/UE-B cases | For positioning latency there are 2 definitions:  a) TTFF: For every positioning method there is TTFF. Scenario 1 includes that and SA2, RAN3 are also part of this.  b) After TTFF is achieved; what is the requirement in terms of interval/frequency of providing the positioning to end user/client. (for example, once every 20ms).  Hence, from RAN2 perspective mainly the b) should be taken into consideration; i.e once the UE is in the ready state for performing positioning procedure; then the latency involved as below   * Obtaining necessary AD * Performing measurement * Reporting the measurement (UE assisted) * Positioning computation * Providing the position to UE (if client is in UE)   For GNSS based positioning method, TTFF depends upon whether UE has recent almanac/ephemeris data available or not. If it is then GNSS positioning procedure runs fast.  Similarly, for RAT dependent to run DL-TDOA; NR-ECID procedure is a prerequisite procedure hence this procedure is part of TTFF.  The latency aspect of TTFF that RAN2 deals with for example NR-ECID/Capability exchange can also be evaluated. However, the real aim should be to check what is the delay involved in b) i.e, After TTFF |
| Qualcomm |  | The scenarios are unclear; i.e., it is not clear why Scenario 2 considers assistance data delivery only. In any case, the UE needs to have a NAS signalling connection (i.e., CM connected state) and will be in RRC connected state after Step 6 and step 1 for MT-LR and MO-LR, respectively. |
| OPPO | Scenario-2 | We understand the intention of scenario-2 is to say that there is no need to consider the RRC state transition here. I.e., although scenario-1 is valid, the latency due to state transition is not needed to be considered for the evaluation here.  For the latency after TTFF has been achieved, in case AD is not changed, the latency would be mostly dependent on either UE internal implementation (UEB) or how frequent the network request the location information report (UEA), so the focus should be more on the TTFF part. |
| CATT | Neither | UE always is in RRC-CONNECTED mode in Rel-16 so the RRC state transition may not be considered in the latency analysis. No need to discuss Scenario 1.  As for Scenario 2, it is not accurate because UE still gets the location response from AMF by step 13 in Figure 6.2-1: 5GC-MO-LR Procedure. As Qualcomm mentioned that in any cases, UE should have a NAS signalling connection and in RRC\_CONNECTED MODE.  So prefer not to discuss the state transmission. |
| Samsung | Both | We think both scenario are supported in Rel-16 and are normal procedures. So it is necessary to analyze both scenario. Intel’s concern on that Scenario 1 is only applicable for location service procedure is also valid. Then the remaining part (state transition part) can be added to the positioning procedure as special component. |
| Xiaomi | Scenario 2 | We think the current scenario 2 is the lowest latency scenario for R16 positioning. The key point of scenario 2 is that the RRC state transmission latency will not be considered. So we think the other positioning procedures beside the RRC state transmission should not be limited and following aspects also should be considered:  Network initiating the positioning;  Acquiring assistance by LPP;  Reporting the measurement to network; |
| vivo |  | In general, we agree the intention of scenario 2. But the description of it is unclear.  Based on the LS “RAN1 respectfully asks if RAN2 can provide a list of latency components with corresponding range of values for the existing and any potential enhanced NR positioning solutions”, RAN1 is expecting us to provide the latency of various NR enhancement techniques. Idle/Inactive positioning is one of the major R17 techniques of enhancement, hence evaluating idle/inactive positioning latency is our scope. The meaning of the sentence” The UE does not need to exchange the message with network” described for Scenario 2 is ambiguous. In this case how can UE provide the evaluation results in idle/inactive state? We suggest the sentence can be changed to” state transmission is not calculated.” instead of “not need”. |
| InterDigital | Scenario 2 (with comments) | Scenario 2 should include the assistance information delivery both via SIB and dedicated NAS signaling, since the latency for delivering the assistance information via SIB or dedicated signaling may be different.  In addition, the differences in the procedures for the acquisition of assistance information for UE-based and UE-assisted should also be considered in the latency analysis. |
| Huawei/HiSilicon | Neither | We agree that if we look into NG-RAN positioning in the whole LCS service procedure, UE should be in RRC\_CONNECTED prior to step 12 for MT-LR or step 5 for MO-LR, which should be the baseline.  It is unclear why there are two scenarios on the table, with unclear description for both. In addition,   * For scenario 1, why is it only applicable to alt 2 in question 1-1? Even if we go with alt 1 in question 1-1, we may still assume that the UE is already in RRC CONNECTED state at least for Rel-16. * For scenario 2, why do we limit the case to posSIB?   In our understanding, we may assume either   * Alt 1. The initial RRC state can only be RRC\_CONNECTED.   Alt 2. The initial RRC state can be any RRC states, and in the time span for latency analysis, additional state transition if needed should be modelled. |
| ZTE | Scenario 2 | Although RRC state transfer is part of the E2E latency, RAN2 do not need to discuss the RRC state transfer in the positioning clause.  Based on our understanding, we should focus on the latency reduction in the positioning related procedures like step 5 for MO-LR and step 12 for MT-LR. The latency about RRC state transfer can be discussed in other clause if necessary. |
| Spreadtrum | Scenario 2 | We agree that RRC state transition is not needed to be considered for the evaluation. |
| Nokia |  | This question 1-5 is unclear, especially the description of Scenario 2. If we agree to focus RAN2 latency evaluation efforts to step 5 for MO-LR and step 12 for MT-LR i.e. UE positioning procedures as defined in 38.305 then Scenario 1 is out of scope of the latency analysis and we support this (i.e. we support Alt 1 for question 1-1). As for Scenario 2, we just need to decide whether we will consider a scenario where the UE has already obtained the assistance data already by using the broadcast assistance data (periodic or on-demand SI). This can be described with caveat note as to whether the latency analysis includes the latency for obtaining the assistance data or not. Our suggestion it to focus on the scenario that involves obtaining assistance data as part of the UE positioning procedures. The case of UE having already obtained the assistance data through broadcast SI can be a special case that can be left undefined. It is easy to deduct the delays involved in the assistance data steps once we know the overall latency of UE procedures that do involve the assistance data latency also. |

Refer to TS38.305, the procedures for different positioning methods are listed as below (**the details of step 5 for MO-LR and step 12 for MT-LR**).

**1 DL-TDOA/DL-AoD**



**Figure 1 procedure for DL-TDOA/DL-AoD**

**Question 1-6: Any comments on DL-TDOA/DL-AoD procedure? Can we use it for E2E latency analysis for DL-TDOA/DL-AoD?**

**Note 1: The procedure involves other nodes, e.g. GMLC is not shown since the procedure is focusing on the details of step 5 for MO-LR and step 12 for MT-LR;**

**Note 2: The transition from IDLE/INACTIVE to CONNECTED mode is not shown in the figure;**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Remark** |
| Intel | Yes |  |
| Ericsson | Another alternative also possible | When Client is in UE  For UE based, it is possible that UE can compute positioning using these methods without having to have LPP transaction. This will minimize the latency.  UE can obtain ciphering keys. NW may broadcast the relevant AD. UE can perform measurement and consume the positioning location without having to send any LPP message. This is already supported by standard, but we can discuss if any companies think it is unsupported. |
| Qualcomm | With modification | For baseline, Step 4 should not be needed. We can assume that the LMF obeys the UE capabilities and provides the required assistance data (i.e., assistanceAvailability=FALSE). In addition, step 5 would then typically occur before step 3. For a deferred MT-LR, steps 1 and 2 are not needed as an LMF only needs to obtain the UE capabilities once. Additional optimization is also possible for a deferred MT-LR (e.g. as defined in TS 23.273 clause 6.7.1), where step 3 does not occur. This is one reason to include deferred MT-LR in the evaluation – since it has the potential for lower latency than a normal MT-LR, MO-LR or NI-LR. |
| OPPO | Yes with comment | Agree with the comments above on:   * UEB case needs to be added; * Step-4 should not be considered in the baseline |
| CATT | With modification | Some steps (e.g. step1, step2, and step4) don’t always happen. We prefer not to include step1, step2 according to the procedures in TS38.305. |
| Samsung | Yes |  |
| Xiaomi | Yes | we think the following case should be considered:   * the steps 1, 2 and 4 are not always needed; * UE based positioning; |
| vivo | Yes with comment | Both TTFF and period positioning after TTFF(without step1 and 2) need consider. |
| InterDigital | With modification | We agree with Qualcomm that the sequence of steps should be modified for UE-assisted positioning as follows:  1. LPP: Request Capabilities, 2. LPP: Provide Capabilities, 3. LPP: Provide Assistance data, 4. LPP: Request Location information, 5. RRC: Location Measurement Indication, 6. RRC: Configure Measurement Gap, 7. UE Measurement, 8. LPP: Provide location information  Similarly, for UE-based positioning, the sequence of steps can be modified as follows:  1. LPP: Request Assistance data, 2. LPP: Provide Assistance Data, 3. RRC: Location Measurement Indication, 4. RRC: Configure Measurement Gap, 5. UE Measurement, 6. Location Calculation |
| Huawei/HiSilicon | Conditional | We do not think 3-1 and (4-1/5-1) are the correct sequential procedure.  In addition, steps 2-3, 4-3, and 9-3 should be Namf\_Communication\_N1MessageNotify. |
| ZTE | Yes with comment | We also perfer to remove step 4 from the baseline procedure. |
| Spreadtrum | Yes with comments | As the baseline procedure, step 4 should not be needed. Step 4 and 5 are optional for UE-based positioning because UE may have obtained LPP assistance data in advance. Step 9 is not needed for UE-based positioning while the client is UE self. |
| Nokia | Yes | We suggest not to use the term “E2E latency” but instead use “UE positioning RAN-protocol latency” which of course can be clarified with a description stating that it is the higher layer RAN protocol latency involving LPP and NRPPa only. We can also add a Note 3 saying this considers the worst-case scenario involving obtaining assistance data and use of measurement gaps. |

**2 UL-TDOA/UL-AoA**



**Figure 2 procedure for UL-TDOA/UL-AoA**

**Question 1-7: Any comments on UL-TDOA/UL-AoA procedure? Can we use it for E2E latency analysis for UL-TDOA/UL-AoA?**

**Note 1: The procedure involves other nodes, e.g. GMLC is not shown since the procedure is focusing on the details of step 5 for MO-LR and step 12 for MT-LR;**

**Note 2: The transition from IDLE/INACTIVE to CONNECTED mode is not shown in the figure;**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Remark** |
| Intel | Yes |  |
| Ericsson | Yes |  |
| Qualcomm | With modification | NRPPa Positioning Activation has also a response message.  NRPPa Positioning Deactivation could be added to the end of the procedure. |
| OPPO | Yes |  |
| CATT | Yes but | We prefer not to include step1, step2 according to the procedures in Figure 8.13.3.4-1: UL-TDOA positioning procedure in TS38.305.  Support to consider add “NRPPa Positioning Activation has also a response message.” in the latency analysis, but not include “NRPPa Positioning Deactivation could be added to the end of the procedure.” which is not related to positioning latency. |
| Samsung | Yes |  |
| Xiaomi | Yes | Steps 1 and 2 are not always needed. |
| vivo | Yes, with clarification | Step7 should be optional. For period SRS, we don’t need activation. |
| InterDigital | Yes | All steps in the procedure should be applicable in the analysis |
| Huawei/HiSilicon | Conditional | To be more precise:   * 2-3 should be Namf\_Communication\_N1MessageNotify * 5-3 should be Namf\_Communication\_N2InfoNotify * 8-3 should be Namf\_Communication\_NonUeN2MessageTransfer * 10-3 (which is currently 8-3) should be Namf\_Communication\_NonUeN2InfoNotify   In addition, an NRPPa Positioning activation response message should be required before step 8 if step 6 and 7 are presented. Nevertheless, it is better to add a Note saying that some steps may be omitted. |
| ZTE | Yes |  |
| Spreadtrum | Yes |  |
| Nokia | Yes | We suggest not to use the term “E2E latency” but instead use “UE positioning RAN-protocol latency” which of course can be clarified with a description stating that it is the higher layer RAN protocol latency involving LPP and NRPPa only. We can also add a Note 3 saying this considers the best case scenario where all neighbor gNBs/TRPs get the measurement request in parallel provide the measurement response in parallel but in practice there could be some additional latency added due to difference in signaling interaction timing involving different gNBs/TRPs. |

**3 Multi-RTT**



**Figure 3 procedure for Multi-RTT**

**Question 1-8: Any comments on Multi-RTT procedure? Can we use it for E2E latency analysis for Multi-RTT?**

**Note 1: The procedure involves other nodes, e.g. GMLC is not shown since the procedure is focusing on the details of step 5 for MO-LR and step 12 for MT-LR;**

**Note 2: The transition from IDLE/INACTIVE to CONNECTED mode is not shown in the figure;**

**Note 3: in the figure, we did not distinguish serving gNB and measured gNB and did not show multiple measured gNBs although they could be different.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Remark** |
| Intel | Yes |  |
| Ericsson | yes |  |
| Qualcomm | With modification | NRPPa Positioning Activation has also a response message.  NRPPa Positioning Deactivation could be added to the end of the procedure.  Step 9 and 15 need to happen concurrently.  For baseline, Step 11 (Request Assistance Data should not be needed (see our response to Question 1-6). For a deferred MT-LR, steps 1 and 2 can also be avoided. |
| OPPO | Yes | We wonder if the procedure for DL w.r.t UE and for UL w.r.t RAN can happen in parallel instead of in sequence when evaluating the whole latency, but surely PRS measurement and SRS transmission cannot happen simultaneously due to the usage of measurement gap for PRS reception. |
| CATT | Yes but | We prefer not to include step1, step2 according to the procedures in Figure 8.10.4-1: Multi-RTT positioning procedure in TS38.305.  Support to consider add “NRPPa Positioning Activation has also a response message.” in the latency analysis, but not include “NRPPa Positioning Deactivation could be added to the end of the procedure.” |
| Samsung | Yes |  |
| Xiaomi | Yes | Steps 1 and 2 are not always needed. |
| vivo | Yes, with clarification | Step7 should be optional. For period SRS, we don’t need activation. |
| InterDigital | Yes, with modification | Step 11 may be skipped since the assistance information can be provided by LMF after LPP capability transfer (Steps 1 and 2). In addition, in reference to TS 38.305 Fig 8.10.4-1 (Multi-RTT positioning procedure) certain modifications in the sequence of steps would be necessary such that the DL-PRS measurements and the UL SRS measurements occur concurrently |
| Huawei/HiSilicon | Condition | Same comments combining DL-only and UL-only. |
| ZTE | Yes |  |
| Spreadtrum | Yes but | We have same concerns as OPPO. |
| Nokia | Yes | We suggest not to use the term “E2E latency” but instead use “UE positioning RAN-protocol latency” which of course can be clarified with a description stating that it is the higher layer RAN protocol latency involving LPP and NRPPa only. |

**4 NR E-CID**



**Figure 4-1 procedure for Downlink NR E-CID**

**Question 1-9: Any comments on Downlink NR E-CID procedure? Can we use it for E2E latency analysis for Downlink NR E-CID?**

**Note 1: The procedure involves other nodes, e.g. GMLC is not shown since the procedure is focusing on the details of step 5 for MO-LR and step 12 for MT-LR;**

**Note 2: The transition from IDLE/INACTIVE to CONNECTED mode is not shown in the figure;**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Remark** |
| Intel | Yes |  |
| Ericsson | Yes | Step 4 could be seen as optional, since DL E-CID is based on measurements that the UE has available. |
| Qualcomm |  | The assumptions for Step 4 require some clarification. According to TS 38.305:  "Although NR E-CID positioning may utilise some of the same measurements as the measurement control system in the RRC protocol, the UE generally is not expected to make additional measurements for the sole purpose of positioning; i.e., the positioning procedures do not supply a measurement configuration or measurement control message, and the UE reports the measurements that it has available rather than being required to take additional measurement actions." |
| OPPO | Yes |  |
| CATT | Yes but | We prefer not to include step1, step2 according to the procedures in TS38.305.  Step 4 can be ignored since there is no measurement action in UE after step3. |
| Samsung | Yes |  |
| Xiaomi | Yes | Steps 1 and 2 are not always needed. |
| vivo | Yes | Step 4 is RAN1 scope and we can leave it to RAN1 to decide. |
| InterDigital | Yes | We agree with Qualcomm that Step 4 need further clarification. In particular, there appears to be a range for the latency corresponding to the UE measurement (Step 4): Lower bound is the processing delay for the UE to include the available measurements in the measurement report and the Upper bound is the sum of the duration for making new measurements (when not available) and processing delay |
| Huawei/HiSilicon | Conditional | Steps 2-3 and 5-3 should be Namf\_Communication\_N1MessageNotify. |
| ZTE | Yes, but | We share the same view with Qualcomm and Ericsson. Based on the description in TS38305, step 4 may be optional. |
| Spreadtrum | Yes | Agree with Ericssion |
| Nokia | Yes | We suggest not to use the term “E2E latency” but instead use “UE positioning RAN-protocol latency” which of course can be clarified with a description stating that it is the higher layer RAN protocol latency involving LPP and NRPPa only. Latency in step 4 for NR E-CID could be slightly different if only available measurements are provided back to LMF. Should clarify this in a Note 3. |



**Figure 4-2 procedure for Uplink NR E-CID**

**Question 1-10: Any comments on Uplink NR E-CID procedure? Can we use it for E2E latency analysis for Uplink NR E-CID?**

**Note 1: The procedure involves other nodes, e.g. GMLC is not shown since the procedure is focusing on the details of step 5 for MO-LR and step 12 for MT-LR;**

**Note 2: The transition from IDLE/INACTIVE to CONNECTED mode is not shown in the figure;**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Remark** |
| Intel | Yes |  |
| Ericsson | Yes | Step 3 could be seen as optional, since DL E-CID is based on measurements that the UE has available. |
| Qualcomm |  | Step 3 and/or Step 2 are unclear; UE seems not transmitting to enable gNB measurements. We assume the RRC Measurement Configuration (Step 2) is for RRM measurements, therefore, some SRS configuration/activation seems needed as well. Also, some UE measurement report seems missing. |
| OPPO | Yes | As commented by Qualcomm, the UE measurement in Step-2 needs some clarification. |
| CATT |  | We are wondering why step 2 is required in NR Uplink E-CID. And step 3 is optional. |
| Samsung | Yes |  |
| Xiaomi | Yes | In TS 38.305:  “If the LMF in step (1) requested UE measurements (e.g., E-UTRA RSRP, E-UTRA RSRQ measurements, etc.), the NG-RAN node may configure the UE to report the measurement information requested as specified in TS 36.331 [13], TS 38.331 [14].”  So the UE measurement reporting procedure is needed and the step 3 is optional. |
| vivo |  | Agree with Qualcomm. |
| InterDigital | Yes | Steps 2 and 3 needs further clarification. Similar to our comment in Question 1-9, there could be a latency range for the steps 2 and 3 as a result of reusing the existing signaling and measurements |
| Huawei/HiSilicon | Conditional | To be precise, step 4-3 should be Namf\_Communication\_N2InfoNotify |
| ZTE | Yes | Same view with Qualcomm. |
| Spreadtrum | Yes | As commented by Qualcomm, Step 2 is unclear. |
| Nokia | Yes | We suggest not to use the term “E2E latency” but instead use “UE positioning RAN-protocol latency” which of course can be clarified with a description stating that it is the higher layer RAN protocol latency involving LPP and NRPPa only. Latency in step 3 for NR E-CID could be slightly different if only available measurements are provided back to LMF. Should clarify this in a Note 3. |

**Question 1-11: Based on RAN1 agreements, shall all RAT dependent positioning methods be considered in RAN2 latency analysis, ((1) DL-TDOA, (2) DL AoD, (3) UL-TDoA, (4) UL-AoA, (5) Multi-RTT, (6) NR E-CID including UL NR E-CID)? or only some of them? If some, pls indicate which positioning methods.**

|  |  |  |
| --- | --- | --- |
| **Company** | **All or some** | **Remark** |
| Intel | All |  |
| Ericsson | 1, 3, 5 | Angle based and ECID could be to some extend covered by 1,3, 5 |
| Qualcomm |  | We see no need for E-CID, since low-accuracy positioning should not require low-latency, but O.K. to include. |
| OPPO | 1-5 | E-CID should be of low interest due to its low accuracy. |
| CATT | All |  |
| Samsung | All | We think there is no clear preference or popularity on positioning method in real use case. So it is necessary to explore the all of them. |
| Xiaomi | All |  |
| vivo | 1,3,5 |  |
| InterDigital | 1, 3, 5, 6 | The latency analysis should be performed only for positioning methods that have differences in the procedures and sequence of steps. Clearly, the latency analysis for positioning methods with common procedure steps (e.g. DL-TDoA and DL-AoD) can be reused |
| Huawei/HiSilicon | All |  |
| ZTE | All | We prefer to consider all positioning methods in latency analysis. The positioning SID does not indicate that RAN2 should only study the latency reduction for the high accuracy positioning methods. |
| Spreadtrum | All |  |
| Nokia | All | Since there are many procedural commonalities, across LPP and across NRPPa, for multiple RAT-dependent positioning techniques it makes sense to consider all RAT-dependent techniques. |

Based on the procedures shown above, the procedure for DL-TDOA and DL-AoD is same, and forUL-AoA and UL-TDOA is same.

**Question 1-12: In RAN2 latency analysis, can DL-TDOA and DL AoD be handled together? And UL-TDoA and UL-AoA be handled together, i.e. in the same table?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Remark** |
| Intel | Y | From RAN2 perspective, there is no different although from physical layer perspective, the measurement delay could be different. |
| Ericsson | Y |  |
| Qualcomm | Yes |  |
| OPPO | Yes |  |
| CATT | Yes | The measurement delay of DL-TDOA and DL-AoD may be different, so the value range could be different in DL-TDOA and DL AoD. |
| Samsung | Yes | RAN2 point of view, those i.e., DL-TDOA and DL AoD has the same procedure except the actual measurement latency in PHY. So does UL. Therefore we think RAN2 can do our job and the box of physical layer can be added later. |
| Xiaomi | Yes |  |
| vivo | Yes |  |
| InterDigital | Yes |  |
| Huawei/HiSilicon | Yes |  |
| ZTE | Yes |  |
| Spreadtrum | Yes |  |
| Nokia | N | We think it is good to have per positioning technique evaluation results even if there are commonalities between two different techniques. This way, in the future, we can add additional assumptions or additional details for any positioning technique to compute a new latency value for a particular positioning technique. We consider these evaluations as a baseline set of results which could later be re-evaluated if a different set of assumptions needs to be made or additional signaling details (e.g. gNB internal signaling between CU and DU or differences in signaling between LMF and neighbor gNBs/TRPs) needs to be taken in to account. |

Regarding the table used for collecting the latency analysis, RAN1 agreed below template:

|  |  |  |
| --- | --- | --- |
| **Source [UE, NW]/Destination [UE, NW]**  **Positioning technique [DL-TDOA, E-CID, …], type [DL, UL, DL+UL], mode [UE-A, UE-B],**  **Initial and Final RRC States [IDLE, INACTIVE, CONNECTED]** | | |
| **Latency Component** | **Value Range** | **Description of Latency Component** |
| Start trigger |  |  |
| Name of component 1 |  |  |
| Name of component 2 |  |  |
|  |  |  |
| Name of last component |  |  |
| End trigger |  |  |
| Total values |  |  |

Rapporteur’s comments:

* Source [UE, NW]/Destination [UE, NW] are not needed, it can be reflected in procedure steps if we show the procedure figure together with the table;
* Initial and Final RRC States [IDLE, INACTIVE, CONNECTED] is not needed for each procedure since it is common for all positioning methods, and could be captured in separate table, or just as separate steps in the table;

If we follow RAN1 table, and consider MO-LR, MT-LR, 5 positioning methods (same table for DL-TDOA/DL-AoD and same table for UL-TDOA/UL-AoD), 3 UE states, 2 modes, we need 60 tables.

If we have positioning methods specific table and common table, the total number will be 26 (20+6); In addition, if we only consider positioning procedure instead of location service procedure, then only table 1 below is needed, the total number will be 20;

Therefore Rapporteur would suggest to use below table to capture latency analysis in phase 2:

**Table 1: positioning method specific table**

|  |  |  |
| --- | --- | --- |
| **Positioning technique [DL-TDOA/DL-AoD, Downlink NR E-CID, …], mode [UE-A, UE-B (IDLE, INACTIVE, CONNECTED)],**  **Focusing on the latency caused by steps in figure 1-figure 4** | | |
| **Latency Component** | **Value Range** | **Description of Latency Component** |
| Step 1 |  |  |
| Step 2 |  |  |
| Step 3 |  |  |
| .. |  |  |
|  |  |  |
|  |  |  |
| Total values |  |  |

**Table 2: common procedure** **(rely on the answer of question 1-1, i.e. whether Alt2 is used for E2E latency analysis)**

|  |  |  |
| --- | --- | --- |
| **e.g. location request for MO-LR, MT-LR when the UE is in IDLE, INACTIVE or CONNECTED, and location response, etc.**  **Refer to the figure 6.2-1 for MO-LR (except step 5) and 6.1.2-1 for MT-LR (except step 12);** | | |
| **Latency Component** | **Value Range** | **Description of Latency Component** |
| Step 1 |  |  |
| Step 2 |  |  |
| Step 3 |  |  |
| .. |  |  |
|  |  |  |
|  |  |  |
| Total values |  |  |

**Question 1-13: Do companies agree to use table1 (positioning methods specific) and table 2 (common procedure) described above to capture the latency analysis results? Or any other suggestions?**

**Note: The need of table 2 relies on the answer of question 1-1, i.e. whether Alt2 is used for E2E latency analysis)**

|  |  |  |
| --- | --- | --- |
| **Company** | **Table 1 and Table 2, Table 1 only** | **Remark** |
| Intel | Table 1 | See above, we only need to consider step 5 for MO-LR and step 12 for MT-LR, and therefore table 1 is enough for RAN2 analysis. |
| Ericsson | Table 1 |  |
| Qualcomm |  | A simpler and more flexible approach is as follows.   * Construct a table T0 containing entries for atomic operations (e.g. LPP processing time, transmission + propagation time over a given link, location computation time, measurement time) and include labels and agree latency values. Labels might, for example, include those used for Q 1-14 below or could be more generic (e.g. O1, O2, O3 etc.). * Construct a single table T1 with entries for composite operations (e.g. LPP or NRPPa end to end message transfer time) which can belong to a positioning or location related procedure. Map each composite operation to constituent atomic operations and/or to other (simpler) composite operations and determine latency values and include labels. * Define and agree a message flow for each type of positioning procedure (DL, UL, UL-DL) and map each procedure to constituent atomic and composite operations used in each procedure. Determine the overall latency using tables T0 and T1. It is possible that variants of the positioning procedures will be needed for deferred MT-LR and certain enhancements of location solutions. Enter the overall latency for each positioning procedure in a table T2. * Define and agree a message flow for each location solution (MT-LR, MO-LR, NI-LR, deferred MT-LR) and map each message flow to a positioning procedure and atomic and composite operations. Determine the overall latency using tables T0, T1 and T2. Enter the overall latencies in a table T3 which can provide end to end latency. * Attempt to account for differences in UE initial/final state and UE based versus UE assisted mode and other such differences via differences in atomic and/or composite operations referenced from tables T0 and T1. This should lead to fixed differences in Table T3 – which can then assume some default configuration and include deltas for the differences in a Note.   The value of this approach is to better identify common operations and common positioning procedures used in different location solutions and ensure that the same latency assumptions are used. This may also reduce documentation (e.g. number and size of tables). As an example, a location solution S1 might list the constituent operations being used as O1, O2, O3, O4 etc. with the end to end latency being the sum of entries in tables T0 and T1 for these operations. An enhanced version of S1 can then list (e.g.) a subset of these operations (e.g. O1, O3, O4 etc. with O2 omitted). It would then be clear that O2 was omitted and that O1, O3 and O4 were common. |
| OPPO | Table-1 |  |
| CATT | Table1 + Table2 | Initial and Final RRC States [IDLE, INACTIVE, CONNECTED] can be ignored based on Rel-16 solutions because UE always steps/stays in connected mode.  Both table1 and table2 should be included. |
| Samsung | Table 1 | RAN2 analysis needs the table 1 and it’s enough. |
| Xiaomi | Table 1 |  |
| vivo | Table-1 |  |
| InterDigital | Table 1 (with comments) | For the analysis, the values for latency components for different positioning methods should be determined, hence the structure corresponding to Table 1 would be applicable. However, since different positioning methods share common procedure steps, the analysis can be performed for the positioning methods identified in Question 1-11 and then reused for the others |
| Huawei/HiSilicon | Table 1 only | For Table 1, we suggest that only the following 18 cases are evaluated   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Non-UE-based for MT-LR or  MO-LR | DL-only | UL-only | Multi-RTT | DL E-CID | UL E-CID | | CONNECTED | Y | Y | Y | Y | Y | | INACTIVE | Y | Y | Y | Y | Y | | IDLE | Y | Y | Y | Y | Y |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | UE-based for MO-LR | DL-only | UL-only | Multi-RTT | DL E-CID | UL E-CID | | CONNECTED | Y | N | N | N | N | | INACTIVE | Y | N | N | N | N | | IDLE | Y | N | N | N | N | |
| ZTE | Table 1 |  |
| Spreadtrum | Table 1 | Table 1 is enough. But we have one concern that why the total number of table is 20. We think is 10. Because there are 5 positioning methods and 2 mode. |
| Nokia |  | I think we can just capture the RAN-level latency for UE positioning procedures in RAN (step 5 for MO-LR and step 12 for MT-LR) for each positioning technique in separate sub-sections in the TR. We can also extract as observations the latency values for the following and document it separately: what is the latency for Capabilities exchange, assistance data exchange, time to indicate and obtain the measurement gap configuration, NRPPa position information request/response, NRPPa measurement request/response. The Table 1 and Table 2 with per step granularity is not required in our opinion. |

In addition, it would be good to have common understanding on the latency assumption for:

* State transition: IDLE to CONNECTED, INACTIVE to CONNECTED;
* Processing delay:
  + UE RRC processing delay, UE LPP processing delay (capability transfer, assistance data transfer, location request), UE MAC processing delay;
  + gNB RRC processing delay, gNB NRPPa processing delay;
  + AMF processing delay;
  + LMF processing delay;
* transmission delay:
  + UE from/to gNB;
  + gNB from/to AMF;
  + AMF from/to LMF;
  + AMF from/to GMLC;

**Question 1-14: Companies are invited to provide the input on latency assumptions (Unit= ms)**

**Table 3: Idle/INACTIVE to CONNECTED, and UE processing time on RRC, LPP and MAC (Unit= ms);**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Company** | **TIdle-conn** | **TInac-conn** | **TUE-RRCReconf** | **TUE-RRCDLInfo** | **TUE-RRCULInfo** | **TUE-RRCLocationMeas** | **TUE-LPPCapab** | **TUE-LPPAssi** | **TUE-LPPLocationRe** | **TUE-MAC-SRSAct** |
| **Intel** | **36.3-62.5** | **11.3-18.5 [9]** | **10 [10]** | **5** | **2** | **2** | **20-80 [10]** | **10** | **5** | **1** |
| **Ericsson** | **48** |  |  |  |  |  |  |  |  |  |
| **CATT** | **ignore** | **ignore** |  |  |  |  |  |  |  |  |

**Table 4: gNB processing time on RRC and NRPPa, AMF/LMF processing time, transmission delay between nodes, gNB measurement delay and LMF calculation delay (Unit= ms);**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Company** | **TgNB-RRC** | **TgNB-NRPPa** | **TgNB-NAS/LPP** | **TAMF** | **TLMF** | **TUE-gNB** | **TgNB-AMF** | **TAMF-LMF** | **TAMF-GMLC** | **TgNB-Meas** | **TLMF-Calc** |
| **Intel** | **3 [9]** | **3** | **3** | **3** | **3** | **0** | **3-10** | **3-10** | **3-10** | **RAN1** | **5** |
| **Ericsson** |  |  |  |  |  |  |  | **1-2** |  |  | **2** |
| **CATT** |  |  |  |  |  |  |  |  |  |  | **30** |
|  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |
| --- | --- |
| **Company** | **Remark** |
| Intel | 1 **TIdle-conn** (INACTIVE to CONNECTED), referred to TS37.910 [9];  2 **TIdle-conn** (IDLE to CONNECTED), on top of **TInac-conn,** added   |  | | --- | | *RRCReconfig+SMC = T UE processing (10-15ms)*  *2 Initiate NAS = 2\*T gNB processing (2\*3ms) + TAMF processing + 2\*T-AMF-gNB* |   3 **TUE-LPPCapab** (LPP capability processing time) referred to RRC capability processing time in TS38.331 [10];  4 **TUE-LPPAssi** (LPP assistance data) referred to RRCReconfiguration processing time in TS38.331 [10];  5 **TgNB-RRC** referred to TS37.910 [9], *the processing delay in gNB (L2 and RRC) has been reduced to 3 ms*  6 **TgNB-NRPPa** and **TgNB-NAS/LPP** referred to TgNB-RRC  7 **TgNB-Meas** should be similar to **TUE-Meas,** and wait for RAN1**.**  8 The transmission delay is tightly related to backhaul situation; |
| Ericsson | An optimized AMF/LMF for latency can have 1-2ms delay/latency between AMF to LMF. For Non-Public Network where 5GS can be deployed locally (in factory premises) this delay/latency could further be lower. Location can be computed in location server every 2ms. |
| Qualcomm | |  |  |  |  | | --- | --- | --- | --- | | Delay Component | | Assumption  [ms] | Comment | | MAC-CE Processing | TMAC-CE | 3 | For SRS activation.  TS 38.214; the command goes into effect 3 ms after the UE sends Ack for the PDSCH carrying the transport block that included the MAC-CE. | | RRC Message Procesing | TUL-RRC  TDL-RRC | 5 | NAS UL/DL Transfer and Location Measurement Indication. Since these procedures are not listed in TS 38.331, section 12, the minimum value for the RRC procedures in Table 12.1-1 of 38.331 is assumed. | | Transition from RRC Inactive to Connected state | TRRCconnect | 15 | 11.3 ms – 18.5 ms dependent on configuration/deployment (TR 37.910). Assume an "average" value. | | RRC Reconfiguration | TRRC-ReConf | 10 | TS 38.331, section 12. | | gNB processing for determining measurement gaps | TgNB-Proc | 10 | Determining/scheduling appropriate measurement gaps as requested by the UE. | | LPP Message Processing | TLPP | 10 | Decoding/encoding and processing of the encapsulated LPP PDU (e.g., providing/ obtaining relevant parameter to/from physical layer).  The same is assumed for all LPP Procedures. | | NRPPa Message Processing | TNRPPa | 10 | Decoding/encoding and processing of the encapsulated NRPPa PDU (e.g., providing/ obtaining relevant parameter to/from physical layer).  The same is assumed for all NRPPa Procedures. | | DL-PRS/UL-SRS Measurement Time | TDL-Meas |  | Depends on RAN1 | | TUL-Meas | | Position Calculation | TPosCalc | 30 | Value could also be determined by RAN1 | | Core Network Signalling | TUL-NG-C  TDL-NG-C  TUL-NL1  TDL-NL1  TNlmf  TLCS-Resp | 3 | Transmission, propagation and processing time. Depends on network configuration; e.g. signaling link bandwidth values and distances between nodes.  On average, we assume the same value on all CN interfaces. | |
| CATT | **T**Idle-conn and**TInac-conn** can be ignored based on Rel-16 solution. |
| Samsung | We have the same understanding as Intel. |
| Xiaomi | We share the same view with Intel. |
| Huawei/HiSilicon | We are generally fine with the numbers, except the following two:  **TUE-MAC-SRSAct 🡪 2**  **TUE-gNB 🡪 0.5 (allowing one-slot scheduling)**  Also based on our understanding, RAN1 already defined the time span for physical layer latency, and we only need to consider the operations not overlapping with the RAN1 share. |
| Nokia | This question may need more time to discuss but at a glance it looks like you are using 37.910 and 38.331 as a basis to form the assumptions. The RRC\_INACTIVE to RRC-CONNECTED state transition latency looks fine. I presume the RRC\_IDLE to RRC\_CONNECTED is adjusted using the inactive to connected as a baseline, which is also fine.  The UE processing delay for RRC reconfiguration, UL information transfer and Location measurement indication RRC procedures are OK but I wonder why the delay for DL information transfer is more than for UL information transfer. As these are passed through to upper layers shouldn’t the delay assumption for these two be the same.  The UE processing delay for LPP capabilities seems too high. Unlike RRC capability, the LPP capabilities exchanged should be less and so the delay for this should also be less.  The UE processing delay for LPP assistance data is OK but for Location request, shouldn’t it take in to account the measurement delay or position estimation delay?  The rest of the gNB processing delay and signaling transmission delays seems OK. For transmission delay, it is good to provide some rationale as to which lower layer configurations are assumed like SCS and if it is based on the CP latency in 37.910.  Finally, we suggest renaming the symbols used for processing delays to have “proc” in the name to make it clear. Examples: TUE-RRCReconf to TUEProc-RRCReconf, TUE-LPPCapab to TUEProc-LPPCapab. TgNB-NRPPa to TgNBProc-NRPPa. |

## Phase 2 discussion

### Latency analysis on Rel-16 RAT dependent positioning methods

### Potential solutions/directions on latency reduction

# Summary

To be added:

# Reference

[1] Chairman's Notes RAN1#102-e v022

[2] R1-2007264 LS on Latency of NR Positioning Protocols, RAN1

[3] R2-2006672 Discussion on ehancements for commercial use cases, CATT

[4] R2-2006578 Discussion on R17 positioning enhancement, Huawei, HiSilicon

[5] R2-2006750 Consideration on the support of low latency requirement, Intel Corporation

[6] R2-2007587 End-to-end latency reduction for DL/UL positioning, InterDigital, Inc.

[7] R2-2008261 [AT111-e][612][POS] Assumptions for analysis of commercial use cases (Ericsson) Ericsson

[8] TS 23.273, 5G System (5GS) Location Services (LCS); Stage 2.

[9] TS37.910

[10]TS38.331