**3GPP TSG-RAN WG2 Meeting #113-eDRAFT R2-2xxxxxx**

**Electronic meeting, Jan 25th – Feb 5th, 2021**

**Agenda item:** 8.11.2

**Source:** CATT

**Title:** [Post112-e][617][POS] Evaluation of latency enhancement solutions (CATT)‎

**Document for:** Discussion and Agreement

# 1 Introduction

This is to continue discussion of the solutions considered in [AT112-e][607], and evaluate for performance the solutions identified. Related RAN1 and RP agreements can be taken into account and evaluated for RAN2 impact. The latency enhancements in this discussion is limited in [AT112-e][607] [1].

* [Post112-e][617][POS] Evaluation of latency enhancement solutions (CATT)

 Scope: Continue discussion of the solutions considered in [AT112-e][607], and evaluate for performance the solutions identified. Related RAN1 and RP agreements can be taken into account and evaluated for RAN2 impact.

 Intended outcome: Report to next meeting

 Deadline: Long

Rapporteur would like to have the following schedule for this email discussion to have enough time for preparing the summary report.

* Phase 1 (2021-01-06): Companies are invited to provide inputs and comments to questions.
* Phase 2 (2021-01-12): Rapporteur will provide draft summary with proposals, companies are invited to provide comments to the summary proposals.

The remainder of this document is organized as the following. Section 2 provides related RAN1 and RP agreements. Section 3 contains the questionnaire on performance evaluation of the latency solutions. The purpose is to collect the views and identify the commonalties and differences in order to provide proposals for suitable TP.

# 2 RAN1 and RP agreements

The agreement on latency in potential positioning enhancements after RAN1 #103-e meeting is below:

|  |
| --- |
| Agreement:Capture the following in the TR:* The enhancements of signaling & procedures for reducing NR positioning latency are recommended for normative work, including DL and DL+UL positioning methods
	+ The details of the solutions are left for further discussion in normative work, which may include the following aspects:
		- Latency reduction related to the measurement gap
		- Latency reduction related to the reporting and request of the measurements (e.g., via RRC signaling, MAC-CE and/or physical layer procedure, and/or priority rules)
		- Latency reduction related to measurement time
* The following enhancements of signaling & procedures for reducing NR positioning latency can be studied and specified, if needed
	+ Latency reduction related to the request and response of positioning assistance data (e.g., via RRC signaling, MAC-CE and/or physical layer procedure)
	+ Latency reduction related to the reception of DL PRS (e.g., priority rules for the reception of DL PRS)
* No assumptions are made on whether the LCS architecture specified in TS 23.273 is enhanced or not.
 |

There is no any agreement on latency in RAN #90-e and no relative objective is added in new WID of ePOS [2]. All participants to this discussion are encouraged to leave their name/contact in section 6.

# 3 Latency enhancement

In section 3.1-3.5, several aspects about latency enhancements will be discussed.

Latency enhancement solutions on measurement gap and the reporting and request of the measurements were discussed in RAN2#112-e meeting, which are parts of agreement in RAN1#103-e meeting on latency. Meanwhile there is no any latency reduction aspects mentioned in WID of ePOS in RAN #90-e because it depends on the agreement both from RAN1 and RAN2.

So we are going to discuss here how to align with the agreement from RAN1 from RAN2’s perspective in section 3.1 and 3.2.

|  |
| --- |
| * + The details of the solutions are left for further discussion in normative work, which may include the following aspects:
		- Latency reduction related to the measurement gap
		- Latency reduction related to the reporting and request of the measurements (e.g., via RRC signaling, MAC-CE and/or physical layer procedure, and/or priority rules)
 |

## 3.1 Measurement gaps optimizations (recommended by RAN1)

Measurement Gap is about 18-22ms (step 13-14) based on the analysis in R2-2009023. The request and configuration of measurement gap results in additional latency due to the transmission and reception of RRC signaling.

|  |  |  |
| --- | --- | --- |
| Step 13 RRC Measurement Gap configuration | 13-13.5 | TgNBProc-RRC+ TUE-gNB+ TUEProc-RRCReconfProcessing delays: 13ms- UE:  TUEProc-RRCReconf = 10ms- gNB: TgNBProc-RRC= 3msSignalling delay:0-0.5ms- UE-gNB: TUE-gNB= 0-0.5ms |
| Step 14 a DL PRS measurement | TUE-measc  | RAN1 inputs |
| Step 14 b UL SRS measurement | TgNB-measc  | RAN1 inputs |

So measurement gaps (MG) optimizations can reduce the latency caused by measurement gap request procedure. Solutions are summarized below according to R2-2009023 and R2-2008886:

* Option1: MG-less operation e.g. UE may operate w/o measurement gaps to process DL PRS.

Measurement Gap is about 18-22ms (step 13-14). The latency caused by measurement gap request procedure could be reduced.

* Option2: Support of semi-persistent a-periodic MGs, their pre-configuration and association with MG configuration ID
* Option 3: Avoiding or minimizing the latency due to measurement gap configuration.

As an example, the UE may be triggered to perform measurement of DL PRS based on lower layer signaling (e.g. in MAC CE) from gNB without configuration of measurement gap. The configuration of certain criteria/rules in the UE for determining whether to perform measurement of PRS based on a configured timer or priority indication can be considered for eliminating measurement gap configuration.

* Option4: Fast activation of measurement gap configuration:

UE sends indication to gNB using lower layer signaling to either skip or request a measurement gap configuration. The gNB may then activate/deactivate a preconfigured measurement gap (e.g. in MAC CE) based on the indication sent by the UE.

**Rapporteur’s comments:**

7/11 companies thought it was RAN1/4 business during the email discussion in [AT112-e][607].

Since there is agreement on measurement gaps optimizations from RAN1, so RAN2 will review this aspect in this email discussion.

**Q1-1: Do you agree with RAN1 agreement to study the following aspect for latency reduction from RAN2’s perspective?**

* **Latency reduction related to the measurement gap (MG) optimizations.**

|  |  |  |
| --- | --- | --- |
| **Company name** | **Agree/Disagree** | **Comments** |
| vivo | Agree | We agree with the RAN1 conclusion to specify MG optimizations for latency reduction, not limited to option 1,2,3,4 listed above.RAN1 conclusion: The enhancements of signalling & procedures for reducing NR positioning latency, including DL and DL+UL positioning methods. The details of the solutions are left for further discussion in normative work, which may include the following aspects:* + the measurement gap
	+ the measurement request and reporting (e.g., via RRC signalling, MAC-CE and/or physical layer procedure, and/or priority rules)
	+ the measurement time
 |
| InterDigital | Agree | Considering RAN1 conclusion, the signalling and procedures that allow for skipping MG configuration or using preconfigured MGs can be considered for normative work. |
| Qualcomm | Disagree (as RAN2-centric objective) | We agree with the "RAN1 Agreement". However, this seems a RAN1 centric objective only. The feasibility and evaluation of the various options should be further studied in RAN1 (and probably RAN4). Any RAN2 impacts will follow from the RAN1 conclusions and should not require a RAN2 study. |
| Intel | Agree | As discussed in last RAN2 meeting, companies believed the measurement gap should be discussed in RAN1 and RAN4. RAN2 can follow RAN1 on this considering RAN1 have concluded the need on this, and recommended to continue the work in work phase.  |
| Nokia | Agree to study | Since RAN1 has recommended to consider measurement gap configuration related enhancements for normative work, RAN2 can also work in this area but the RAN2 impacts for the solutions on the table need to be identified and highlighted first by the solution proponent before we can study from RAN2 perspective. Then there is the issue of time availability for such a study in RAN2. One meeting to complete this study is not enough. The best RAN2 can do in the one meeting available is to agree what solutions we need to focus for further study in the WID phase. |

**Q1-2: If you agree above aspect which can be further discussed, please provide your views: e.g. performance evaluation of the solutions above, or specifying the solutions above in detail.**

|  |  |
| --- | --- |
| **Company name** | **Comments** |
| vivo | We suggest following RAN1 conclusion and not further limiting the scope of the solutions before WI. |
| InterDigital | Given that Options 1/3 and 2/4 are aligned with the conclusions made by RAN1, we think the corresponding solutions should be captured in the TR as examples.  |
| Intel | Capture the option 1-4 in the TR, and do the further down selection in WI phase.  |
| Nokia | Performance evaluation or solution details or RAN2 impacts for a particular solution needs to come from the solution proponents. This approach to get company views in this one email discussion or in the one meeting left for the study item is not enough to make decisions about the solutions on table.  |

## 3.2 Measurement report optimization (recommended by RAN1)

According to R2-2009897, it is proposed to introduce a positioning measurement report, with configured grant (CG) mechanism. The gNB can allocate resources for the UE that has been requested to perform positioning procedure. The configured grant resources are used by the UE to transmit the positioning measurement results. Note: configured grant mechanism has been part of the NR URLLC. Hence, the specification impact can be kept the minimum.

Additionally, in R2-2009039, it is also stated that Grant Free UL Transmission enables reduce UL transmission delays and achieve URLLC Reliability targets.

Grant Free UL Transmission enables reduce UL transmission delays and achieve URLLC Reliability targets. For low latency and reliability requirements, it is required to support UL GF transmission with multiple repetitions (i.e. UL data transmission without scheduling request). Additionally, this pre-allocated grant should adapt to the PRS period, so the best latency result is performed.



Figure3-1: configured grant resource adapt PRS repetition period

This configured grant can be defined as positioning use only uplink resources. CG need adopt the positioning window, but gNB doesn't know the offset or the timing of the completed positioning measurement/calculation.

The following proposed solutions are from companies, according to the comments in [AT112-e][607].

* Option 1(summarized from companies’ comments): Using the existed CG-based transmission for a certain logical channel.

It is already supported by the configuration of logical channel in NR Rel-16 which is up to the network implementation to configure CG. However there is NRPPa impact. gNB may get the PRS period from LMF via NRPPa.

* Option2: New type or separate CG for positioning which is used to adapt the PRS period and positioning specific configured grant may be introduced in Rel-17.

This can be used as positioning use only uplink resources, so that periodic positioning measurement report could be sent without waiting any L1 signals.

**Rapporteur’s comments:**

Five companies’ comments in [AT112-e][607] are summarized as option1 i.e. using the existed CG-based transmission for a certain logical channel. Companies are invited to review the two options above in this email discussion.

**Q2-1: Do you agree with RAN1 agreement to study the following aspect for latency reduction from RAN2’s perspective?**

* **Latency reduction related to the reporting and request of the measurements**

|  |  |  |
| --- | --- | --- |
| **Company name** | **Agree/Disagree** | **Comments** |
| vivo | Agree | We agree with the RAN1 conclusion to specify reporting and request of the measurements for latency reduction, not limited to option 1,2 listed above.RAN1 conclusion: The enhancements of signalling & procedures for reducing NR positioning latency, including DL and DL+UL positioning methods. The details of the solutions are left for further discussion in normative work, which may include the following aspects:* + the measurement gap
	+ the measurement request and reporting (e.g., via RRC signalling, MAC-CE and/or physical layer procedure, and/or priority rules)
	+ the measurement time
 |
| InterDigital | Agree | We agree for studying solutions using CG for reducing latency associated with measurement reporting  |
| Qualcomm | Depends on the context | The introduction text to this question refers to the configured grant (CG) mechanism. However, the specific question Q2-1 is related to a generic RAN1 agreement.(1) We assume the CG based mechanism is proposed for periodic and triggered location reports (i.e., deferred MT-LR for periodic and triggered events, as specified in TS 23.273). If the proposal is to use CG resources for measurement reporting in RRC\_INACTIVE state, this should generally be supported. However, there seems no additional impact; i.e., if the Rel-17 "Small data Transmission (SDT)" framework is used. For RRC\_CONNECTED state, this seems not needed.(2) Assuming the question is related to the following general RAN1 agreement:"*Latency reduction related to the reporting and request of the measurements (e.g., via RRC signaling, MAC-CE and/or physical layer procedure, and/or priority rules).*"This should be studied together with "Location Server Functionality in the RAN" (Q5-1 below), since it seems one signalling end-point is a gNB/NG-RAN Node. Any "priority rules" seems a RAN1-centric objective. |
| Intel | Agree, but | RAN2 can follow RAN1 on this considering RAN1 have concluded the need on this, and recommended to continue the work in work phase.But the solution details shall be left to RAN1, and not limit to CG based solution. |
| Nokia | Agree to study | Since RAN1 has recommended to consider measurement reporting related enhancements for normative work, RAN2 can also work in this area but the RAN2 impacts for the solutions on the table need to be identified and highlighted first by the solution proponent before we can study from RAN2 perspective. Then there is the issue of time availability for such a study in RAN2. One meeting to complete this study is not enough. The best RAN2 can do in the one meeting available is to agree what solutions we need to focus for further study in the WID phase. |

**Q2-2: If you agree above aspect which can be further discussed, please provide your views: e.g. performance evaluation of the solutions above, or specifying the solutions above in detail.**

|  |  |
| --- | --- |
| **Company name** | **Comments** |
| vivo | For option2 we think the CG period can be configured same or integer multiples of PRS period. |
| InterDigital | We think Option 1 and Option 2 can be applicable for latency reduction and should be captured in TR. The details of how the CG can be properly aligned and triggered with suitable offset upon completion of PRS measurement can be discussed during WI stage. |
| Qualcomm | Measurement request and reporting (e.g., via RRC signalling, MAC-CE and/or physical layer procedure) can be studied for a Location Server in the RAN. The RAN Location Server can then provide the location result directly to the LCS Client (e.g., via user plane; section 3.2.3 below).If the NG-RAN node needs to transfer any location report to an LMF in the 5GC anyhow, this seems not needed. |
| Intel | We shall leave the details to RAN1, and not limit to CG based solution.  |
| Nokia | Performance evaluation or solution details or RAN2 impacts for a particular solution needs to come from the solution proponents. This approach to get company views in this one email discussion or in the one meeting left for the study item is not enough to make decisions about the solutions on table. |

## 3.3 Other potential aspects

RAN1 agreed in RAN1#103-e meeting:

|  |
| --- |
| * The following enhancements of signaling & procedures for reducing NR positioning latency can be studied and specified, if needed
	+ Latency reduction related to the request and response of positioning assistance data (e.g., via RRC signaling, MAC-CE and/or physical layer procedure)
	+ Latency reduction related to the reception of DL PRS (e.g., priority rules for the reception of DL PRS)
 |

### 3.3.1 Request and response of positioning assistance data optimization

According to R2-2009023, SRS configuration+activation (step 3-8) is 66- 133ms and LPP assistance data is 28-44.5ms, if the latency consumption of these two parts can be reduced, the total E2E latency can be further optimized.

According to R[2-2010096](file:///E%3A%5CWORK%5C1%203GPP%5CMeeting%5CRAN2%20112-e%5C2%20During%5CDocs%5CR2-2010096.zip), Latencies for Deferred MT-LR Event Reporting is provided as below:

**Table 19: Latencies for Deferred MT-LR Event Reporting.**

|  |  |
| --- | --- |
|  | **End-to-End Latency [ms]** |
| **LMF only** | **LMF and LSS** |
| **Baseline** | **Configuration Signalling in Advance** | **Configuration Signalling in Advance for DL-only Positioning** | **Configuration Signalling in Advance and LSS** | **Configuration Signalling in Advance and LSS for DL-only Positioning** | **LSS with Positioning and Event Reporting in RRC\_INACTIVE state** |
| UL+DL Positioning | 284-535.5 | 164-320 | NA | 100-150 | NA | 61-98.5 |
| UL-only Positioning | 221-448 | 139-287.5 | NA | 76-120.5 | NA | 55-91 |
| DL-only Positioning | 218-402.5 | 124-229.5 | 72-135.5 | 92-137.5 | 54-89.5 | 53-86.5 |

So SRS configuration and PRS configuration optimizations can reduce the latency caused by SRS/PRS confi[guration](file:///E%3A%5CWORK%5C1%203GPP%5CMeeting%5CRAN2%20112-e%5C1%20Before%5C%E6%96%87%E7%A8%BF%E8%A7%84%E5%88%92%5CPOS%5CCR%5Cbackup%5CR2-200xxxx%20Minor%20corrections%20on%20description%20of%20sfn0-Offset%20in%20SSB-Configuration.docx). Here are the solutions proposed in R2-2009023 and R2-2010096:

* Option 1: DL PRS assistance information can be pre-configured to UE which can reduce the LPP assistance data (step 11 LPP Request Location Information): 28-44.5 ms. Multiple DL PRS configurations can be associated with DL PRS configuration ID and activated when necessary;
* Option 2: SRS for positioning configuration information can be pre-configured to UE which can reduce SRS configuration+activation (step 3-8): 66- 133ms. Multiple configurations of SRS for positioning can be associated with SRS for positioning configuration ID and activated when necessary;
* Option 3: Specify signalling and procedures for Deferred MT-LR (as proposed in R2-2010096) to support positioning configuration signalling in advance;

**Rapporteur’s comments:**

8/11 companies supported request and response of positioning assistance data aspect to be further studied, and 6/10 companies support option 3 in [AT112-e][607]. Some companies’ comments show that they were not sure what preconfiguration means. Preconfiguration means PRS configuration before LPP session, in order to reduce the latency of step 11 in option 1.

**Q3-1: Do you agree with RAN1 agreement to study the following aspect for latency reduction from RAN2’s perspective?**

* **Latency reduction related to the request and response of positioning assistance data (SRS and PRS configuration optimizations).**

|  |  |  |
| --- | --- | --- |
| **Company name** | **Agree/Disagree** | **Comments** |
| vivo | Agree | We tend to follow the RAN1 agreement to study the request and response of positioning assistance data not limited to the 3 options above.RAN1 agreement:Enhancements of signalling & procedures for reducing NR positioning latency related to* + the request and response of positioning assistance data (e.g., via RRC signalling, MAC-CE and/or physical layer procedure)
	+ the reception of DL PRS (e.g., priority rules for the reception of DL PRS)
 |
| InterDigital | Agree | We agree for studying enhancements related to request/response of positioning assistance data for latency reduction |
| Qualcomm | Depends on the context | Similar to our response to Q2-1, the request/response of positioning assistance data via RRC signalling, MAC-CE and/or physical layer procedure can be considered if there is Location Server Functionality in the RAN (since the source of the Assistance Data or one signalling endpoint seems to be a NG-RAN node). In general, a LMF can provide assistance data and location requests (Q2‑1) a time T in advance of when the location estimate is needed. This can be applied to both, immediate and deferred location requests. An LMF can also provide assistance data at the start of a deferred location request which may be valid for some or all of the later location event reports. |
| Intel | Agree |  |
| Nokia | Agree to study | Since the enhancement related to reducing latency involved in providing assistance data is recommended by RAN1 for further study in normative phase, RAN2 should take the outcome of the RAN1 study also. However, signalling enhancement solutions are in RAN2 domain and hence should be driven by RAN2 rather than RAN1 proposing to use lower layer signalling instead of current LPP signalling. Once again, the RAN2 impacts for the solutions on the table need to be identified and highlighted first by the solution proponent before we can study from RAN2 perspective. |

**Q3-2: If you agree above aspect which can be further discussed, please provide your views: e.g. performance evaluation of the solutions above, or specifying the solutions above in detail.**

|  |  |
| --- | --- |
| **Company name** | **Comments** |
| vivo | RRC signalling, MAC-CE and/or physical layer procedure should be studied by RAN2. |
| InterDigital | We think the solutions discussed in Option 1 and 2 associated with preconfiguring PRS and/or SRSp in UE/RAN and activating the pre-configurations when triggered should be captured in the TR. Further details related to signalling and procedures can be discussed during WI stage.  In addition, in the case of mobility, similar enhancements for latency reduction may be considered where the UE need not request and receive assistance data when the (primary) serving cell is changed (as in Rel-16 LPP procedure). In this case, the procedure where the UE can receive suitable assistance data in advance via the serving cell (e.g. prior to HO) may be considered for latency reduction.    |
| Qualcomm | Assistance Data request and response (e.g., via RRC signalling, MAC-CE and/or physical layer procedure) can be studied for a Location Server in the RAN, since the source of the Assistance Data or one signalling endpoint would be a NG-RAN Node. For an LMF in the 5GC, assistance data (and location requests (Q2-1)) can be provided a time T in advance of when the location estimate is needed for both, immediate and deferred location requests.In addition, Assistance Data can be provided via positioning SI and stored in the UE (i.e., no additional impact). |
| Intel | The solutions details can be further discussed/down selected in WI phase.  |
| Nokia | Performance evaluation or solution details or RAN2 impacts for a particular solution needs to come from the solution proponents. This approach to get company views in this one email discussion or in the one meeting left for the study item is not enough to make decisions about the solutions on table. |

### 3.2.2 Enhancements for prioritized transmission of PRS/SRS

According to R2-2008886, in Rel-16, both PRS and SRSp are assigned with low priorities. As a result, PRS is not received or SRSp is not transmitted/dropped when either transmission of data in DL/UL or other reference signals are scheduled.

In Rel-17, it can be envisioned that supporting prioritized positioning based on the assignment and indication of higher priority for the reception/transmission of PRS/SRSp may enable satisfying the low latency positioning requirements. For DL-based positioning, the priority indication for PRS may be either indicated by LMF in assistance information or indicated by RAN in lower layer/RRC signalling. The UE may trigger the reception and measurement of PRS based on the received priority indication.

For UL-based positioning, the UE may trigger the transmission of SRSp based on the reception of the priority indication in lower layer/RRC signalling. The priority of the positioning reference signal can be associated with the type of the positioning reference signal (e.g. periodic vs. aperiodic positioning reference signals).

So some company proposed to support prioritization of PRS and/or SRSp.

**Rapporteur’s comments:**

10/11 companies thought it would be better handled by RAN1at first in [AT112-e][607]. It seems RAN2 can study the signalling procedure based on RAN1 input.

According to the latest agreement from RAN1 in #103-e meeting, only DL PRS is included:

* + Latency reduction related to the reception of DL PRS (e.g., priority rules for the reception of DL PRS)

Rapporteur invites companies to discuss if RAN2 agrees to align with RAN1.

**Q4-1: Do you agree with RAN1 agreement to study the following aspect for latency reduction from RAN2’s perspective?**

* **Latency reduction related to the reception of DL PRS (e.g., priority rules for the reception of DL PRS).**

|  |  |  |
| --- | --- | --- |
| **Company name** | **Agree/Disagree** | **Comments** |
| vivo | Agree, but | This should be studied by RAN1 and out of RAN2 scope. |
| InterDigital | Agree | Given RAN1 agreement on the topic, we think the enhancements related to signalling and procedures for supporting priority rules for receiving the DL PRS can be studied in RAN2  |
| Qualcomm | Disagree (as RAN2-centric objective) | This seems a RAN1 centric objective only. The feasibility and evaluation of the various options should be further studied in RAN1 (and probably RAN4). Any RAN2 impacts will follow from the RAN1 conclusions. |
| Intel | Agree, but | This can be contained in the WI scope as recommended by RAN1, but should be studied in RAN1.  |
| Nokia | Disagree | The study of prioritization of PRS reception by UE relative to other reference signals or data reception should be left to RAN1. Only after RAN1 decision, and if agreed by RAN1, RAN2 should discuss signalling support for such PRS reception prioritization like priority indication etc. |

**Q4-2: If you agree above aspect which can be further discussed, please provide your views: e.g. performance evaluation of the solutions above, or specifying the solutions above in detail.**

|  |  |
| --- | --- |
| **Company name** | **Comments** |
| InterDigital | We think applying priority rules for PRS reception provides more flexibility at UE/RAN for satisfying the low latency requirements associated with positioning service and other data transmission. Without priority rules, transmission of data and other RS may take precedence over DL PRS, hence resulting in higher latency for positioning. Further details on signalling and procedures associated with prioritized handling of PRS can be discussed during WI stage.  |
|  |  |
|  |  |

### 3.2.3 Architecture enhancement aspect

The functionality of Support for Location Server functionality in the RAN is mentioned in some contributions and according to the analysis of R2-2010096.

Location Server functionality in the RAN (e.g., LMC) could reduce the positioning procedure latency significantly analysed by R2-2010096.

One architecture to support Location Server functionality in the RAN (e.g., LMC) is shown in following figure:



A possible Rel-16 procedure for splitting between LMF (for "component A") and LSS (for "component B") defined in R2-2010096 would be the Deferred MT-LR for periodic and triggered events specified for commercial location services. The procedures for "component A" can be performed in advance of when it is needed. A location request for an MT-LR or MO-LR can include a time T of when the location is required. The LMF would then perform the procedures required for "component A" before the time T. A very small latency for "component B" would allow a client to treat a location estimate as current as there would be little time for location degradation due to movement of the target UE.

Additionally, in R[2-2009023](file:///E%3A%5CWORK%5C1%203GPP%5CMeeting%5CRAN2%20112-e%5C2%20During%5CDocs%5CR2-2009023.zip), it is also proposed to reduce the number of Hops between gNB, AMF and LMF as far as possible, so as to achieve the positioning requirement of greatly reducing end-to-end delay.

Both two companies mentioned local positioning functionality as below:

R2-2010096 Proposal 2: Specify support for location server functionality in the RAN (referred to as "Location Server Surrogate" (LSS)). The LSS should support at least the following functions:

- Processing of LCS Event Reports;

- cordinating UE and TRP measurement reports;

- performing position calculation (in case of UE-assisted mode);

- reporting UE location estimates to (external) clients.

R2-2009023 Proposal: To reduce the latency, following enhancement directions are considered in WI phase:

* Reduce the number of hops between gNB, AMF and LMF, e.g. Local NR positioning in NG-RAN (To reduce the latency caused by the transmission/processing from AMF/LMF, i.e. only gNB is shown in the positioning);

It is not within RAN1 scope to analyse positioning architecture enhancements to enable such more efficient signaling & procedures.

**Rapporteur’s comments:**

4/8 companies in [AT112-e][607] supported the architecture enhancement can be studied in this SI because the positioning procedure latency significantly is reduced (at least 23% - 41% ) proposed by R2-2010096 which are based on RAN2 assumptions.

4/8 companies didn’t support it because of the following reasons in [AT112-e][607]:

1. The evaluation for local LMF-based positioning has extensively discussed in past by SA2 and RAN3, which does not make any conclusion on the latency gain of local LMF compared with LMF being deployed physically adjacent to gNB.

2. The privacy issue of user data may be occurred if we introduce the location server.

3. The way to compare the latency of LSS/local LMF-based positioning with that of LMF-based positioning in R2-2010096 and R2-2009023 is questionable.

4. Other options such as deploy 5GS within factory premises may be looked for if there is latency caused by transport. For example, with Non Public Network one can deploy not only local LMF but also local AMF; as such the whole 5G Core can be within factory premises.

Companies are invited to discuss this aspect here to find a way forward.

**Q5-1: Do you agree to study the architecture enhancement aspect for latency reduction in this SI?**

* **E.g. functionality of support for Location Server functionality in the RAN.**

|  |  |  |
| --- | --- | --- |
| **Company name** | **Agree/Disagree** | **Comments** |
| InterDigital | Agree | We agree for studying the support for location server functionality in RAN |
| Qualcomm | Agree | We think some architecture enhancements are required for significantly reducing the latency; e.g., to achieve latency targets of 100 ms and below as shown in R2-2010096. In addition, the request/response of location measurements or assistance data via lower layer signalling (e.g., MAC-CE, DCI) as discussed in Q2-1 and Q3-1 above could be exploited.Regarding the comments from [AT112-e][607] above:1. Latency has only been evaluated qualitatively in the past (e.g., R3-193587). However, with the RAN2 latency assumptions a more quantitative analysis is possible (R2-2010096). In addition, only a "full location server" in the RAN has been discussed previously. However, the proposal at hand is to move only some location server functionality to the RAN. 2. Even for a "full location server" in the RAN (Local LMF/LMC) , the Local LMF/LMC would not need to know the UE identities (e.g. R3-193586).3. The comparison is based on the signalling steps required and should be rather obvious.4. An LSS can achieve latency of 100ms (R2-2010096) without any special optimization of network deployment (which may be possible only in limited scenarios). In addition, a "local AMF" has not been described/evaluated by the proponent.We further note that support for Location Server functionality in the RAN is already defined for GERAN and UTRAN since R99 – the motivation for this being to reduce latency. |
| Intel |  | Considering it has been discussed in Rel-16 in RAN2, RAN3 and SA2, and finally RAN3 left the decision to SA2. It is difficult for RAN2 to consider this again without the information from other groups.  |
| Nokia | Agree with comments | The main issue with discussing local LMF type solutions to reduce latency is that it spans multiple areas requiring expertise from multiple working groups and the challenge is to come to a mutual agreement on what can RAN2 contribute towards such a study. We know this was studied in both RAN2 and RAN3 before but there was no progress due to lack of consensus. If we can agree that from RAN2 side, we can just discuss the latency analysis already done in R2-2010096 and just document RAN2 findings in the TR then we can wait for RAN3 and SA2/SA3 to address architecture and security aspects whenever that may be. That would help with progressing towards a final decision on the solution in our view. |

**Q5-2:** **If you agree above aspect which can be further discussed, please provide your views: e.g. performance evaluation of the solutions above, or specifying the solutions above in detail.**

|  |  |
| --- | --- |
| **Company name** | **Comments** |
| InterDigital | From the previous discussions in [607], it was highlighted that significant latency reduction and more efficient coordination for PRS/SRSp configuration and processing of DL and/or UL measurements are possible when supporting certain location server functionalities (e.g. LMC) in RAN. In this regard, the associated signalling enhancements and analysis related to latency reduction when supporting LMC in RAN can be captured in the TR.  |
| Qualcomm | We evaluated the performance in R2-2010096. |
| Nokia | RAN2 can just discuss the latency analysis already done in R2-2010096 and just document RAN2 findings in the TR. |

### 3.2.4 Capability procedure aspect

Based on R2-2009023, LPP capability exchange is about 33-88.5 ms. In R2-2008810, it is stated that Positioning capabilities of UE may be reported to AMF directly before there is a location request, instead of to LMF via LPP session, in order to reduce the positioning latency. This solution also works for the positioning in Idle/Inactive mode. AMF can store these capabilities before UE steps into RRC\_CONNECTED mode.

Additionally, in R2-2010072, it is stated that Time to First Fix should be considered for positioning latency studies:

As defined in 22.261: **Time to First Fix (TTFF):** time elapsed between the event triggering for the first time the determination of the position-related data and the availability of the position-related data at the positioning system interface.

Hence, before starting first positioning measurements; any activity/transactions that the device would require should be part of TTFF; for example, capability exchange, any pre-requisite procedure such as NR-ECID, retrieval of first assistance data to perform measurements.

**Observation 1**: Time to first fix should be considered in latency studies and any improvements in this area can be studied. Considering TTFF in latency may relax the other core latency requirements for performing measurements and reporting to the location server for positioning computation.

Potential improvement during TTFF can be storage of UE positioning capabilities by AMF. AMF would thus forward it to LMF as depicted in below diagram for the MT-LR procedure.

Below are the proposals from the three companies:

R2-2008810 Proposal 7: Support the process that UE location capabilities report to AMF in idle/inactive directly without entering into RRC\_CONNECTED mode in LPP session, in order to reduce the latency and support the positioning in Idle/Inactive mode.

R2-2010072 Proposal 2: RAN2 to consider solutions that would save latency during capability transfer and send an LS to SA2 to provide solution that minimizes latency in retrieving capability from UE to LMF via LPP.

R2-2009023 Proposal: To reduce the latency, following enhancement directions are considered in WI phase:

* Skip the capability procedure (can reduce the latency caused by exchange of capability as above)

**Rapporteur’s comments:**

The capability procedure aspect seems more like a requirement to SA2 to provide solution. So the requirement can be analyzed following the steps from RAN2’s perspective:

* Step1: Does capabilities of UEs stored in core network (e.g. AMF or LMF, etc) save the latency of capability procedure in LPP session?

Some company believes that Time to first fix should be considered in latency studies considering TTFF in latency may relax the other core latency requirements for performing measurements and reporting to the location server for positioning computation.

The latency of capabilities procedure in LPP session will be saved if LMF already knew the capabilities of this UE, whatever LMF how gets the capabilities, e.g. by itself or from AMF before.

* Step2: Which node of core network is supposed to store the location capabilities of UEs?

There are several options discussed in [AT112-e][607]:

* AMF: Some companies support AMF because it is already storing UL SRS for positioning capabilities. Further AMF stores other NAS capabilities, paging capabilities, UE NW capabilities.

The reason why LMF is not supported is that LMF as such should be stateless and different LMF will be chosen by AMF. Further, it is not guaranteed that UE ID (SUPI) would always be available in LMF. As providing UEID to LMF is OPTIONAL.

* LMF: Some companies believed the capabilities could also be stored at an LMF (which seems a possible implementation option already since Rel-9).
* Either: It’s up to SA2 to make the decision which node should take the responsibility.

Since 5/11 companies in [AT112-e][607] thought this requirement was unclear that why an AMF should store positioning capabilities, companies are invited to discuss the aspect.

**Q6-1: Do you agree to study the capability procedure aspect for latency reduction in this SI?**

|  |  |  |
| --- | --- | --- |
| **Company name** | **Agree/Disagree** | **Comments** |
| vivo | Agree | Optimization of capability can help extremely improve latency. |
| InterDigital | Agree | Given the benefits for latency reduction, we agree for studying optimizations related to capability transfer procedure  |
| Qualcomm | Depends on the context | For a deferred MT-LR, the LMF can receive the UE capabilities at the start of the procedure. Since the same LMF can be used for subsequent location reports (according to clause 6.3.1 in TS 23.273), there need be no extra latency to transfer the UE capabilities. This is already supported in Rel-16. In the case of time varying capabilities, a UE could provide any new capabilities along with an event report to an LMF which would not increase latency and is also possible with the Rel-16 solution.For an immediate location request, similar to our response to Q2-1 and Q3-1 capabilities (as well as assistance data and location requests) can be provided/requested a time T in advance of when the location estimate is needed.Storing UE positioning capabilities in the NW can only work for capabilities which are static and/or long-term valid. Obviously, storing capabilities which depend on the current radio configuration (e.g., some UL-PRS capabilities) is not meaningful/possible. Other capabilities, such as DL-PRS processing capabilities etc. may be time varying too (e.g., may depend on the current UE situation (e.g., for power saving, a UE may advertise lower DL-PRS processing capabilities in certain scenarios)). Or in general, if a user turns-off location services, a UE may not advertise positioning capabilities at all, etc. Therefore, the benefit of storing positioning capabilities in the NW seems limited. If it were to be supported, an LMF might still need to receive an indication from the UE that the stored capabilities are still valid and/or an indication of any changed capabilities. |
| Intel | Agree | We do see the benefit. We can send LS to SA2 in WI phase, to check which node is desirable to store the positioning capability, AMF or LMF.  |
| Nokia | Disagree | We are open to study any valid, reasonable solution to improve the latency introduced by LPP positioning capability exchange, but the proposed solution is just moving the latency from LPP layer to the 5GC side, but many details of this solution are still lacking. These solutions are also treading in to 5GC areas which means RAN2 cannot independently decide on these enhancements. Hence the capability enhancements can be deprioritized for the time being in RAN2. |

**Q6-2: If you agree above aspect which can be further discussed, please provide your views: e.g. which node is supposed to store the capabilities?**

|  |  |
| --- | --- |
| **Company name** | **Comments** |
| InterDigital | The enhancements to LPP capability transfer procedure (e.g. for MT-LR) where the UE context (related to positioning capability) can be retrieved by LMF should be studied by RAN2. On whether the AMF (e.g. at UDM) or LMF stores the UE context may be up to SA2’s discretion.  |
| Qualcomm | Question Q6-1 generally asks for studying "capability procedure aspect for latency reduction". One such aspect would be performing "configuration signalling in advance". As commented in our response to Q6-1, storing capabilities in the NW seems less useful, since they can be time-varying, and at least some capabilities would have to be requested on-demand anyhow.  |
|  |  |

### 3.2.5 Parallel handling enhancement aspect

In some contributions, it was proposed that in parallel handling some location-related messages and steps can further reduce the total end to end latency.

Here are the solutions proposed in R[2-2009577](file:///E%3A%5CWORK%5C1%203GPP%5CMeeting%5CRAN2%20112-e%5C2%20During%5CDocs%5CR2-2009577.zip) and R[2-2008886](file:///E%3A%5CWORK%5C1%203GPP%5CMeeting%5CRAN2%20112-e%5C2%20During%5CDocs%5CR2-2008886.zip):

Option 1：For UL-TDOA/UL-AOA positioning method, some NRPPa messages can be merged into one message, such that the total end to end latency can be further reduced.

The main enhancements are as following:

* NRPPa positioning information request and NRPPa measurement request are merged into one message;
* NRPPa positioning information response and NRPPa measurement response are merged into one message;
* gNB can immediately active SRS without SRS activation request from LMF and it implies LMF don’t need to send SRS activation request.

Option 2：For UL&DL-based positioning methods, RAN2 to study potential enhancements related to provisioning of PRS and SRSp, coordinated triggering of SRSp transmission and PRS reception, and measurement report transmission/forwarding

**Rapporteur’s comments:**

8/11companies didn’t support the parallel handling enhancement aspect because of the following reasons mentioned in [AT112-e][607] :

* 6 companies believed the merging of message appeared more implementation and could be possibly done by implementation.
* One company mentioned that option 1 seems generally not feasible, since TRPs can only be configured e.g., once the SRS is known.

So companies are invited to discuss this aspect here.

**Q7-1: Do you agree to study the parallel handling aspect for latency reduction in this SI?**

|  |  |  |
| --- | --- | --- |
| **Company name** | **Agree/Disagree** | **Comments** |
| vivo | Agree | But this is based on gNB implementation. |
| InterDigital | Agree | While we agree that certain aspects of parallel handling (e.g. related to NRPPa signalling) can be done through implementation, for DL+UL based positioning certain steps in LPP and RRC signalling that can be skipped, merged or done concurrently should be studied.  |
| Qualcomm | Unclear | Unclear what needs to be studied. If possible/sensible in certain scenarios, procedures can be executed in parallel by implementation.For example. messages for the first two bullets above can already be sent in parallel. There seems no gain to physically combine the messages in that case, since that also requires transfer at exactly the same time (a possible restriction) and increased implementation.  |
| Intel | Disagree | To our understanding, it is kind of implementation issue, e.g. for option 1 the network could send the messages almost in the same time, and then additional delay due to transmission can be ignored; for option 2, the network may coordinate SRS transmission and PRS reception by implementation. Therefore we should not spend efforts on such optimization.  |
| Nokia | Disagree | Option 1 should be pursued in RAN3. The coordinated triggering of PRS and SRSp solution in Option 2 is unclear as to what the standards impacts are. |

**Q7-2: If you agree above aspect which can be further discussed, please provide your views: e.g. performance evaluation of the solutions above, or specifying the solutions above in detail.**

|  |  |
| --- | --- |
| **Company name** | **Comments** |
| InterDigital | For UL+DL based positioning, we think the following aspects described in Option 2 can be considered for latency reduction enhancements:* NRPPa signaling for including DL-PRS configuration along with the request for UL positioning information (involves RAN3)
* RRC signaling for configuring the alignment and timing for UL+DL positioning such that DL PRS measurement and UL SRSp transmission can be done with low latency (e.g. contiguously)
 |
|  |  |
|  |  |

# 4 Conclusion

Based on company feedback, the following is observed and proposed:

TBD

# 5 References

1. R2-2010868 [AT112-e][607][POS]Gathering of latency enhancement solutions (CATT), CATT
2. RP-202900 New WID on NR Positioning Enhancements, CATT, Intel Corporation, Ericsson
3. 3GPP TR 38.857 V1.0.0 (2020-12)
4. R2-2010669 Summary of 8.11.2 Enhancements for commercial use cases
5. R2-2008810 Further discussion on enhancements for commercial use cases, CATT
6. R2-2008886 Discussion on End-to-End Latency Reduction for DL/UL Positioning, InterDigital, Inc.
7. R2-2009001 Report of [Post111-e][625][POS] End-to-end latency analysis (Intel), Intel Corporation
8. R2-2009023 Solution directions to reduce end-to-end latency, Intel Corporation
9. R2-2010096 NR Positioning Latency Analysis and Enhancements, Qualcomm Incorporated
10. R2-2010276 Discussion on IDLE INACTIVE pos, on-demand PRS and latency analysis, Huawei, HiSilicon
11. R2-2010277 Discussion on R17 positioning enhancement, Huawei, HiSilicon
12. R2-2010072 Enhancements for commercial use cases, Ericsson
13. R2-2009039 Discussion on positioning enhancement, vivo
14. R2-2009137 Discussion on positioning enhancements for commercial use cases, Spreadtrum Communications
15. R2-2009577 Positioning enhancements on RRC idle/inactive UE and latency reduction, Beijing Xiaomi Electronics
16. R2-2009897 Considerations on potential positioning enhancements, Sony
17. R2-2010627 Discussion on enhancement for commercial use cases, Samsung R&D Institute UK
18. R2-2008261 [AT111-e][612][POS] Assumptions for analysis of commercial use cases, Ericsson

# 6 Participants

|  |  |
| --- | --- |
| **Company Name** | **Participant name/contact** |
| InterDigital | Jaya Rao (jaya.rao@interdigital.com), Fumihiro Hasegawa (fumihiro.hasegawa@interdigital.com) |
| Qualcomm | sfischer@qti.qualcomm.com |
| Nokia | Mani.Thyagarajan@nokia.com |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |