**3GPP TSG-RAN WG2 Meeting #112-eDRAFT R2-2010868**

**Electronic meeting, November 2nd – 13th, 2020**

**Agenda item:** 8.11.2

**Source:** CATT

**Title:** Gathering of latency enhancement solutions (CATT)‎

**Document for:** Discussion and Agreement

# 1 Introduction

This is to describe and discuss the proposed latency enhancements in a format suitable for developing into a TP. The proposed latency enhancements of the following offline discusion as per the draft summary report [1]

* [AT112-e][607][POS] Gathering of latency enhancement solutions (CATT)

Scope: Describe and discuss the proposed latency enhancements in a format suitable for developing into a TP.

Intended outcome: Text proposal in R2-2010868

Deadline: Friday 2020-11-13 0000 UTC

There are two rounds to discuss the latency enhancement solutions.

1st round is to collect companies’ view on these solutions and summarise potential agreements. Initial deadline: Monday 2020-11-09 1900 UTC

2nd round is to collect the text proposals from companies for developing into a TP. Initial deadline: Thursday 2020-11-12 0000 UTC

The remainder of this document is organized as the following. In Section 2 we provide descriptions on the proposaed latency enhancement based on company contribution [2-14]. In Section 3 the discussions are summarized with proposed TP.

# 2 Discussion

In section 2.1-2.7, these proposed latency enhancements will be discussed in several aspects.

All participants to this discussion are encouraged to leave their name/contact in section 5.

## 2.1 Support for location server functionality in the RAN

The functionality of Support for Location Server functionality in the RAN is mentioned in some contributions and according to the analysis of R2-2010096, the improvements of reducing positioning procedure latency is large.

**Observation 1:** Location Server functionality in the RAN (e.g., LMC) could reduce the positioning procedure latency significantly. With the given assumptions, the improvements can be:

- for UL+DL methods: 40% - 55%;

- for UL-only methods: 50% - 61%;

- for DL-only methods: 23% - 41%.

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")（as 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:\WORK\1%203GPP\Meeting\RAN2%20112-e\2%20During\Docs\R2-2009023.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);

**Q1: Please provide your views if location server functionality in the RAN is captured into TR as an enhancement of latency.**

|  |  |  |
| --- | --- | --- |
| **Company name** | **Agree/Disagree** | **Comments** |
| Huawei/HiSilicon | Disagree | We don’t think the location server functionality is beneficial for latency enhancement.  First, We dont think it is part of the SID  Second, the evaluation for local LMF-based positioning has already been done by SA2, which does not make any conclusion on the latency gain of local LMF compared with LMF being deployed physically adjacent to gNB and RAN3 dose not evaluate the latency gain.  C:\Users\y00397895\AppData\Roaming\eSpace_Desktop\UserData\y00397895\imagefiles\E79128FF-7765-4A23-8B57-C0C632000E50.png  Third, 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, which only counts the number of signalling. While the latency of the signaling can vary significantly with different distance of deployment between the LMF, gNB and AMF. |
| Qualcomm | Agree | The latency improvements can be seen from the analysis in R2-2010096 which is based on RAN2 assumptions. |
|  |  |  |
|  |  |  |
|  |  |  |

## 2.2 The capability procedure

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:

**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)

**Q2: Please provide your views if enhancement of capability procedure is captured into TR as an enhancement of latency.**

|  |  |  |
| --- | --- | --- |
| **Company name** | **Agree/Disagree** | **Comments** |
| Huawei/HiSilicon | Disagree | 1. We don't see much benefit for UE location capabilities reporting to AMF, since it only saves the signalling delay from UE-gNB-AMF, which is only about 3-10.5ms according to R2-2009001 but at the expense of the additional complexity of AMF.  2. This also means UE location capabilities should be stored in AMF. Serveral problems may be caused.  1) AMF needs to store all the positioning capabilities for all UEs (UE positioning capabilities are transparent to AMF) in the network, which probably incurs additional complexity and is hard to realize because the UE number in the network can be huge!.  2) There may be a lot of spec impacts including RAN2, SA2, CT1, CT4. An alternative is to store UE positioning capabilities in LMF without any spec impact. When a new LMF is selected for certain UE, the new LMF can either request context from the old LMF, or request capabilities from the UE. The LMF change can rarely happen so that there would be little  lantecy impacts. |
| Qualcomm | Unclear | Generally, we are not against capturing enhancement proposals in the TR which have been evaluated. However, why an AMF should store positioning capabilities is rather unclear, since they could also be stored at an LMF (which seems a possibible implementation option already since Rel-9). |
|  |  |  |
|  |  |  |
|  |  |  |

## 2.3 SRS configuration and PRS configuration 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:\WORK\1%203GPP\Meeting\RAN2%20112-e\2%20During\Docs\R2-2010096.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 by SRS/PRS confi[guration.](file:///E:\WORK\1%203GPP\Meeting\RAN2%20112-e\1%20Before\文稿规划\POS\CR\backup\R2-200xxxx%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. 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. 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;

**Q3: Please provide your views which option(s) of SRS configuration and PRS configuration optimizations is captured into TR as an enhancement of latency.**

|  |  |  |
| --- | --- | --- |
| **Company name** | **Option1/2/3** | **Comments** |
| Huawei/HiSilicon |  | Not sure what preconfiguration means.  Even in Rel-16, assistance data can be provided to the UE in advance, and UE will not take measurement until UE receives LPP RequestLocationInformation message. Would that be considered as pre-configuration as well?  For SRS, we assume reusing periodic MIMO-SRS can save a lot of latency as shown in our contribution R2-2010276, we do not need capability exchange, activation request/response, MAC CE activation, etc. Even for positioning SRS, we already support SP-SRS/AP-SRS for positioning that can be activated rather dynamically.  For deferred MT-LR, the procedure is already defined. Not sure what kind of enhancement we are talking about. The current event report is to AMF and then to the LMF, but we can directly enhance LPP triggered report to LMF. |
| Qualcomm | All | All options can be captured. This may require primarily procedure description changes. E.g., SRS configurations are currently deleted after cell change, state transitions, etc. or LMF does not know whether the UE has/supports assistance data via broadcast, etc. |
|  |  |  |
|  |  |  |
|  |  |  |

## 2.4 The parallel handling of positioning related messages

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:\WORK\1%203GPP\Meeting\RAN2%20112-e\2%20During\Docs\R2-2009577.zip) and R[2-2008886](file:///E:\WORK\1%203GPP\Meeting\RAN2%20112-e\2%20During\Docs\R2-2008886.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

**Q4: Please provide your views which option(s) of parallel handling of positioning related messages/steps is captured into TR as an enhancement of latency.**

|  |  |  |
| --- | --- | --- |
| **Company name** | **Option1/2** | **Comments** |
| Huawei/HiSilicon | Neither | For Option 1, we think is may only works for the cases when the psoitioning procedure only involves the serving gNB. For example, how can LMF request the measurement from neighboring gNBs together with the request of positioning information? Because at this time, LMF hasn’t received the SRS configuration from the information response.  For Option 2, similar problem occurs when the neighboring gNBs are involved in the positioning process. It’s far from easy to coordinate the timing/triggering of PRS transmission by both serving gNB and neighboring gNBs with the SRSp transmission by UE.RAN4 is discussing the spacing between SRS and PRS in R16 and this is some kind of coordination |
| Qualcomm | None | Option 1 seems generally not feasible, since TRPs can only be configured e.g., once the SRS is known. Immediate SRS activation is possible in Rel-16 e.g., with periodic SRS.  Option 2 is unclear, but looks more RAN1 centric. |
|  |  |  |
|  |  |  |
|  |  |  |

## 2.5 Measurement gaps (MG) optimizations

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

So measurement gaps (MG) optimizations can reduce the latency caused by measurement gap request procedure. Here are the solutions proposed in R2-2009023 and R2-2008886:

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

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.

**Q5: Please provide your views which option(s) of measurement gaps (MG) optimizations is captured into TR as an enhancement of latency.**

|  |  |  |
| --- | --- | --- |
| **Company name** | **Option1/2/3/4** | **Comments** |
| Huawei/HiSilicon | Option1 | We are supportive of MG-less operation.  1. This can reduce the latency caused by MG configuration.  2. The data transmission process wouldn’t be affected if there is no measurement gap for processing DL PRS. So it may bring some gain from the apect of the latency for data transmission. |
| Qualcomm | None | This seems RAN1/4 business. It seems there is already some measurement gap enhancements work ongoing in RAN4. |
|  |  |  |
|  |  |  |
|  |  |  |

## 2.6 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.

**Proposal 4:** RAN2 should study mechanisms for supporting prioritization of PRS and/or SRSp with respect to data and other reference signals for reducing positioning latency

**Q6: Please provide your view if prioritized transmission of PRS/SRS is captured into TR as an enhancement of latency.**

|  |  |  |
| --- | --- | --- |
| **Company name** | **Agree/Disagree** | **Comments** |
| Huawei/HiSilicon | Agree | We think this should be better handled in RAN1, e.g. handling between PRS/CSI-RS/SSB/PDCCH/PDSCH, and between SRS/PUSCH/PUCCH is more related to physical layer consideration.  Wait for the RAN1 input on the enhancements for prioritized transmission of PRS/SRS. |
| Qualcomm | Disagree | This seems RAN1 business. |
|  |  |  |
|  |  |  |
|  |  |  |

## 2.7 Measure report optimization

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). Additionlly, this pre-allocated grant should adapt to the PRS period, so the best latency result is performed.



Figure 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. So, a new type or separate CG for positioning need to be introduced

The following proposed solutions are from companies:

R2-2009039 Proposal 5: Grant-free UL transmission can be used to adapt the PRS period and positioning specific configured grant should be introduced in Rel-17.

R2-2009897 Proposal 8: Positioning latency reduction is facilitated by RAN2 in the context of RAN procedures/protocol on LPP signalling, RRC signalling, and MAC layer.

**Q7: Please provide your views if measure report optimization is captured into TR as an enhancement of latency.**

|  |  |  |
| --- | --- | --- |
| **Company name** | **Agree/Disagree** | **Comments** |
| Huawei/HiSilicon | Disagree | Our understanding on the configured grant for PRS measurement reporting should only be used in INACTIVE state NR positioning based on small data transmission. Allocating exclusive configured grant for positioning for CONNECTED state UE will fragment the specification, e.g. it means that physical channel – transport channel – logical channel – RLC channel - radio bearer – RRC message and up to NAS container are dedicated for an LPP message, which is breaks the standard interface between layers. UE will not be able to use CG for other RRC message, or NAS message, or even UP data.  Besides, for CG-based transmission, if you want to use apply CG transmission for a certain logical channel, this can already been enabled by the configuration of logical channel. |
| Qualcomm | Disagree | Proposal 5 seems RAN1 business, and Proposal 8 is too general/unclear. |
|  |  |  |
|  |  |  |
|  |  |  |

And, companies can input if any on the following questsion.

**Q8: Do you see any other solutions on latency that have not been covered by previous discussions?**

|  |  |
| --- | --- |
| **Company name** | **Issues and comments if any** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

## 2.8 Skeleton of text proposal

In order to reach a format suitable for developing into a TP, below please find the draft skeleton to capture companies’ text proposal in 2nd round.

# 7 Studied NR positioning enhancements

*(from objective 1c. Includes positioning techniques, DL/UL positioning reference signals, signalling and procedures for improved accuracy, reduced latency, network efficiency, and device efficiency for both RAN1 and RAN2.  
Enhancements to Rel-16 positioning techniques, if they meet the requirements, will be prioritized, and new techniques will not be considered in this case. )*

## 7.X Enhancements on latency

### 7.X.1 xx aspect

Potential solution 1: Companies’ text proposal

Potential solution 2:

### 7.X.2 xx aspect

Potential solution 1:

Potential solution 2:

Note: this skeleton is for capturing the text proposal, not the final skeleton of TR.

**Q9: Please insert your views and comments to** **the skeleton of text proposal in the table below.**

|  |  |  |
| --- | --- | --- |
| **Company name** | **Agree/Disagree** | **Comments** |
| Huawei, HiSilicon | Agree |  |
| Qualcomm | Agree |  |
|  |  |  |
|  |  |  |
|  |  |  |

# 3 Conclusion

TBD

# 4 References

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

# 5 Participants

|  |  |
| --- | --- |
| **Company Name** | **Participant name/contact** |
| Huawei, HiSilicon | Yinghao Guo  yinghaoguo@huawei.com |
| Qualcomm | Sven Fischer  sfischer@qti.qualcomm.com |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |