e3GPP RAN WG2 Meeting #112e R2-2010764

November 2nd – 13th, 2020

Agenda Item: 8.10.2.1

Source: InterDigital (email discussion Rapporteur)

Title: [DRAFT] Summary of [AT112-e][103][NTN] RACH and HARQ feedback aspects

Document for: Discussion, Decision

# Introduction

This discussion document is intended to enable continuation of user plane discussions from RAN2#112e, specifically relating to RACH and a subset of HARQ feedback-related aspects:

** [AT112-e][103][NTN] RACH and HARQ feedback aspects (IDC)**

      Scope: Discuss (a revision of) p2, p3, p5, p10, p12, p9, p13 from [R2-2010455](file:///C:\Data\3GPP\Extracts\R2-2010455%20(R17%20NTN%20WI%20AI%208.10.2.1%20Summary%20of%20%5bPost111-e%5d%5b908%5d%5bNTN%5d).docx)

Intended outcome: summary of the offline discussion with e.g.:

* List of proposals for agreement (if any)
* List of proposals that require online discussions

Please note the following deadlines for company feedback have been provided by the Chair:

* Initial deadline (for companies' feedback): Monday 2020-11-09 17:00 UTC
* Initial deadline (for rapporteur's summary in R2-2010764):  Monday 2020-11-09 23:00 UTC

The following was also noted:

Proposals marked "for agreement" in R2-2010764 not challenged until Tuesday 2020-11-10 12:00 UTC will be declared as agreed by the session chair. For the rest the discussion will continue online.

# Discussion

## Offset calculation (P2/P3)

From previous discussion in [1] the following was proposed:

*Proposal 3: From RAN2 perspective, for UE with pre-compensation capability, start of the ra-ResponseWindow and msgB-ResponseWindow is offset by UE-gNB RTD in LEO/GEO.*

Companies which did not agree expressed concerns about the accuracy of the UE timing precompensation estimate, stating that the UE may start monitoring too early/late if the estimate is inaccurate or if the timing advance had changed (e.g. due to satellite movement in LEO). However, based on agreements from the previous meeting, RAN1 is discussing a TA margin to compensate for such sources of inaccuracy (however more details may be needed from RAN1) [2]:

* *FFS: The TA margin, if needed and indicated by the network (in order to account for the TA estimation uncertainty)*

An alternative solution was proposed in [3], where if the timing reference is at the gNB the offset of the start of ra-ResponseWindow (msgB-ResponseWindow) can be made in the first PDCCH occasion after the downlink symbol that has the same symbol number, slot number and system frame number as the last uplink symbol of the PRACH occasion where msg1(MsgA) was transmitted (companies can view a detailed description in the original proposing contribution).

**Question 1: For UE with pre-compensation capability, from RAN2 perspective which option(s) do you support to offset the start of the *ra-ResponseWindow* and *msgB-ResponseWindow*? (FFS RAN1 details on TA margin included in UE-gNB RTT estimate)**

**Option 1: UE-gNB RTT estimated from Msg1/MsgA transmission;**

**Option 2: Based on DL timing (e.g. downlink symbol that has the same symbol number, slot number and system frame number as the last uplink symbol of the PRACH occasion where msg1/MsgA was transmitted);**

**Option 3: Other (please describe).**

|  |  |  |
| --- | --- | --- |
| **Company** | **Preferred Option(s)** | **Additional comments** |
| MediaTek | Option 1 | We prefer Option 1, i.e. UE-gNB RTT estimate |
| LG | Option 1 | Option 1 should be commonly used in NTN, i.e., for a UE with/without precompensation capability. |
| ZTE | Option 1 | And the details of RTT is pending on RAN1’s decision. No need to duplicated the discussion in RAN2 unless there are further RAN2 impacts identified. |
| Spreadtrum | Option 2 | There are RA-RNTI collisions caused by error of UE-gNB RTD estimation for option 1. |
| OPPO | Option 1 | We think UE can determine the offset for the start of the *ra-ResponseWindow* and *msgB-ResponseWindow* based on estimated UE-gNB RTT. TA margin might be one of potential approaches to ensure the accuracy. |
| Panasonic | Option 3 | we think it could be too early to determine how to estimate RTT so we prefer to wait for RAN1 progress on this aspects. |
| Ericsson | Option 2 | When UL and DL are aligned at the gNB, using the DL timing is the same as using the UL timing in TNs. The UE will not start monitoring the PDCCH too early nor too late.  Other options risk that if the TA estimation is not accurate, the UE may monitor for RAR outside of the 10 ms where the RA-RNTI is unique for the RA occasion used by the UE, and thus UE may accept a RAR intended for a different UE and create a Msg3 collision (or interference as the UEs TA estimate is incorrect). Similar for MsgB though if successRAR is addressed to C-RNTI there will be no increase in collisions/interference.  Using DL timing will be future proof when/if UEs without pre-compensation capabilities are introduced. |
| Qualcomm | Option 1 | The UE-gNB RTT could be simply TA or TA+Koffset broadcast for feeder link.  For Msg1 following is what specified in TS 38.321.  2> start the *ra-ResponseWindow* configured in *RACH-ConfigCommon* at the first PDCCH occasion as specified in TS 38.213 [6] from the end of the Random Access Preamble transmission;  For MsgA,  1> start the *msgB-ResponseWindow* at the PDCCH occasion as specified in TS 38.213 [6], clause 8.2A;  Therefore, for 2 step RACH, RAN1 should also need to be informed on this. |
| Lenovo | Option 1 | Option 1 is straightforward. |
| Apple | Option 1 | We should wait for the RAN1 discussion to conclude on this. |
| CAICT | Option2 | Acoording to current describtion in 38.213 as follows, DL timing to start ra-ResponseWindow and msgB-ResponseWindow is used.  “The window starts at the first symbol of the earliest CORESET the UE is configured to receive PDCCH for Type1-PDCCH CSS set.” |
| Nokia | Waiting for RAN1 | We think the concern on the accuracy of the UE timing precompensation estimate is valid.  If the UE have a good understanding of the accurate propagation time and provide exact RTD as offset, Option1 is fine. However, the UE specific delay estimation accuracy as well as how to broadcast the variable common delay (e.g. due to satellite movement in LEO) are not concluded yet, thus the impact to offset is still FFS. Furthermore, in our understanding, the TA margin discussed in RAN1 is designed to solve the TA overestimation which may cause PRACH interference. It may be helpful to avoid UE start monitoring too late while TA margin is not concluded in RAN1 yet. For Option2, it may need further discussion on gNB DL-UL frame timings shift or aligned case which is not concluded yet. |
| CATT | Option 1 | We prefer Option 1 assuming UE-gNB RTT is the same as UE-gNB RTD. And we can wait for the agreement in RAN1 to take further discussion. |
| Xiaomi | Option 1 | Option 2 is actually the same as option 1, UE also needs firstly to calculate RTT then decide how much advance is needed corresponding to the DL timing. |
| CMCC | Option 1 | Option 1 is simple and could be used in many other pre-compensation scenarios in NTN. |

Another proposal addresses the ra-ContentionResolutionTimer:

*Proposal 2: For UE with pre-compensation capability, start of the ra-ContentionResolutionTimer is offset by UE-gNB RTD in LEO/GEO.*

A similar solution can be employed as above (i.e. UE-gNB estimate + TA margin), however the UE-calculated offset for the ra-ContentionResolutionTimer may additionally be refined in Msg2 by a TA. Alternatively, [3] proposes that the offset the start of the ra-ContentionResolutionTimer by starting it in the downlink symbol that has the same symbol number, slot number and system frame number as the first uplink symbol after the end of the Msg3 transmission.

**Question 2: For UE with pre-compensation capability, which option(s) do you support to offset the start of the *ra-ContentionResolutionTimer*? (FFS RAN1 details on TA margin included in UE-gNB RTT estimate)**

**Option 1: UE-gNB RTT estimate used for Msg1;**

**Option 2: UE-gNB RTT estimate used for Msg1 transmission corrected by TA in Msg2;**

**Option 3: Based on DL timing (e.g. by starting it in the downlink symbol that has the same symbol number, slot number and system frame number as the first uplink symbol after the end of the Msg3 transmission)**

**Option 4: Other (please describe).**

|  |  |  |
| --- | --- | --- |
| **Company** | **Preferred Option(s)** | **Additional comments** |
| MediaTek | Option 1/Option 2 |  |
| LG | Option 2 | Option 2 should be commonly used in NTN, i.e., for a UE with/without precompensation capability. |
| ZTE | Option 2 | After TA correction,  NW will based on this corrected TA to schedule Msg4 therefore it is preferred to use corrected TA as the offset. Also the corrected TA is more accurate which is beneficial for further power saving. |
| Spreadtrum | Option 2 |  |
| OPPO | Option 2 | We prefer Option 2. The UE-gNB RTT estimated by UE is refined in Msg2, hence the absolute UE-gNB RTT can be used to offset the start of the *ra-ContentionResolutionTimer*. |
| Panasonic | Option 3 | we think it could be too early to determine how to estimate RTT so we prefer to wait for RAN1 progress on this aspects. |
| Ericsson | Option 3 | When UL and DL are aligned at the gNB, using the DL timing is the same as using the UL timing in TNs. The UE will not starting to monitior PDCCH too early nor too late even if TA estimation is not accurate. |
| Qualcomm | Option 2 | UE-gNB RTT estimated from Msg3 transmission, for example, TA used for Msg3 transmission + common feeder link offset, if the time reference is at satellite. |
| Lenovo | Option 2 | The offset corrected by Msg2 can be more accurate. |
| Apple | Option 2 | We can retain the current RACH protocol and don’t have to introduce unnecessary additions to MSG3. |
| CAICT | Option3 | From our observation, for down link signal detection, DL timing is always used at UE side. |
| Nokia | Waiting for RAN1 | It’s better keep unified solution for ra-ResponseWindow and ra-ContentionResolutionTimer start, if possible. |
| CATT | Option 2 | The corrected TA is more accurate. |
| Xiaomi | Option 2 |  |
| CMCC | Option 2 |  |

## Extention of ra-ResponseWindow/msgB-ResponseWindow (P5)

The following proposal received very large majority support in the previous email discussion:

*Proposal 5: If the start of the ra-ResponseWindow and msgB-ResponseWindow is accurately compensated by UE-gNB RTD, ra-ResponseWindow and msgB-ResponseWindow are not extended in LEO/GEO.*

Note: it is assumed most companies still agree with this proposal (i.e. the 24/26). If a company does not reply to the below question it is assumed satisfactory.

**Question 3: For companies which *disagree* with the following proposal, please justify why an extension is necessary or an acceptable alternate wording:**

***If the start of the ra-ResponseWindow and msgB-ResponseWindow is accurately compensated by UE-gNB RTT, ra-ResponseWindow and msgB-ResponseWindow are not extended in LEO/GEO.***

|  |  |
| --- | --- |
| **Company** | **Justification for extension/alternate wording?** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

## Enabling/Disabling HARQ UL retransmission (P10/P12)

From previous discussion in [1], the following was proposed with comments focused primarily on the definition of what disabling HARQ uplink retransmission means:

*Proposal 10: From a RAN2 perspective, HARQ uplink retransmission relying on the decoding result of previous PUSCH transmission at the UE transmitter can be enabled/disabled in Rel-17 NTN (i.e. blind retransmission and slot aggregation if configured are not disabled).*

According to legacy procedure, a dynamic UL grant is associated with a HARQ process ID (PID), where upon initial PUSCH transmission the UE stores the TB in the HARQ buffer of the identified process.

If the UE receives a subsequent grant with the same HARQ PID and NDI toggled, the UE flushes the HARQ buffer and can perform LCP procedure to build a new TB.

If the TB was not successfully decoded, the gNB may provide a grant with the same HARQ PID and NDI *not* toggled, requiring the UE to retrieve the stored TB from the HARQ buffer of the identified process and perform a retansmission. The delay between initial TB transmission to reception of a retransmission grant is at least one RTT.

According to legacy behaviour, during this time new data is not assigned to this HARQ process as the NDI is not toggled. Due to large propagation delay in NTN, if enough TBs are unsuccessfully decoded (resulting in the associated HARQ PIDs not being reusable for a long time) HARQ stalling may occur.

In one interpretation of the current proposal, if HARQ uplink retransmission relying on the decoding result of previous PUSCH transmission at the UE transmitter is disabled for a HARQ process, the UE is not expected to receive a retransmission grant for that HARQ process. It is assumed this interpretation would result in the following behaviour:

* If not configured with any other retransmission mechanisms (e.g. slot aggregation, blind decoding etc.) there would be no need to store the TB in the HARQ buffer of the identified process;
* If the UE receives a grant for that HARQ PID with NDI *not* toggled, it would ignore the grant or assume a subsequent grant for that HARQ PID implicitely has NDI toggled;
* Retransmission would rely on higher layers (e.g. RLC) (or some other form of retransmission mechanism such as blind decoding).

An alternative interpretation is if the gNB did not successfully decode the TB, it may still send a grant with toggled NDI to allow new data for this HARQ PID. If the gNB intends to toggle the NDI regardless of the decoding result, there is no need to wait at least an RTT to assign a grant to the same HARQ PID.

According to RAN1 TS 38.214, as long as the transmission of the last PUSCH for the HARQ process is completed, the gNB may assign a grant to the same PID:

*The UE is not expected to be scheduled to transmit another PUSCH by DCI format 0\_0, 0\_1 or 0\_2 scrambled by C-RNTI or MCS-C-RNTI for a given HARQ process until after the end of the expected transmission of the last PUSCH for that HARQ process.*

As the duration of the PUSCH transmission is much less than an NTN RTT, the gNB can assign a HARQ PID to new data much faster and avoid HARQ stalling. Referring to P12, there was large majority support that the criteria/decision to enable/disable HARQ uplink retransmission is already under network control:

*Proposal 12: From RAN2 perspective, HARQ uplink retransmission at the UE transmitter can be enabled/disabled, but HARQ processes remain configured. The criteria to enable/disable HARQ uplink retransmission is under network control, and is signalled to UE via RRC in a semi-static manner.*

Therefore, if the network wants to disable HARQ uplink retransmission for a HARQ process it can simply provide a grant with NDI toggled without waiting for the decoding result of the previous PUSCH transmission. If the gNB wants to enable HARQ uplink retransmission, it can simply use legacy behaviour. It is noted that further discussion may be required to properly configure the value of *drx-HARQ-RTT-TimerUL*.

In summary, disabling HARQ uplink retransmission relying on the decoding result of previous PUSCH transmission at the UE transmitter can be interpreted in two ways:

1. *gNB can send grant with NDI toggled without waiting for decoding result or previous PUSCH transmission (i.e. up to gNB implementation):*

* Relies on legacy mechanisms and gNB implementation (i.e. no spec impact);
* HARQ uplink retransmission can be enabled/disabled dynamically based on NDI;
* Can send a grant with same HARQ process much faster than an NTN RTT;
* Does not place restrictions on scheduling.

1. *UE assumes it will not receive a retransmission grant based on gNB decoding result of previous PUSCH transmission:*

* If not configured with any other retransmission mechanisms (e.g. slot aggregation, blind decoding etc.) there would be no need to store the TB in the HARQ buffer of the identified process as retransmission is not expected (i.e. change of legacy behaviour);
* Retransmission relies on higher layers (e.g. RLC) or some other mechanism such as blind decoding or slot aggregation;
* If a UE receives a grant for a HARQ PID with NDI *not* toggled, it would ignore the grant or assume NDI is implicitely toggled (i.e. change of legacy behavior);
* Places restrictions on scheduling.

**Question 4: Which option do you support to enable/disable HARQ uplink retransmission relying on the decoding result of previous PUSCH transmission at the UE transmitter?:**

**Option 1: gNB can send grant with NDI toggled without waiting for decoding result of previous PUSCH transmission (i.e. up to gNB implementation);**

**Option 2: UE assumes it will not receive a retransmission grant based on gNB decoding result of previous PUSCH transmission;**

**Option 3: Other (please describe).**

|  |  |  |
| --- | --- | --- |
| **Company** | **Preferred Option** | **Additional comments** |
| MediaTek | Option 1 | We prefer Option 1 as it has no spec impact. |
| LG | Option 2 |  |
| ZTE | Option 1 (i.e., reusing legacy behavior) | We prefer to reusing the legacy behavior which has less specs impact and can guarantee both the transmission reliability and provide more flexibility while achieving the same purpose option 2 aims for. Option 2 propose a semi-static way to disabling the HARQ retransmission which might demage the resource efficiency since it is impossible to dynamically switch between the blind retransmission and normal HARQ retransmission. Also only relaying on RLC retransmission will lead to extra latency especially when radio quality becomes worse suddenly, e.g., due to blockage of obstacles. |
| Spreadtrum | Option 3 | For HARQ retransmission enabled processe based on PUSCH decoding result, TA is used as the offset to be applied to drx-HARQ-RTT-TimerUL value range. UE will only receive grant of the same HP ID with NDI toggled or not, after drx-HARQ-RTT-TimerUL expiration.  For HARQ retransmission disabled processe based on PUSCH decoding result, drx-HARQ-RTT-TimerUL is 0. UE expect to receive grant of the same HP ID with NDI toggled or not, from now on. |
| OPPO | Option 1 | We think disabling HARQ uplink retransmission is not to disable the entire UL HARQ retransmission, and retransmissions of TB in a bundle or based on blind scheduling could still be supported.  For Option 1, perhaps it should be “with NDI not being toggled”, if retransmission is the intention here. |
| Panasonic | Option 1 |  |
| Ericsson | Option 1 | This is R15 legacy, the UE never sends a retransmission except if the gNB request a retransmission by a grant on same HP ID with non-toggled NDI. No need to change the UE behaviour.  gNB is control, and if it fails to decode a transmission, the gNB may decide to adapt the link adaptation and proactively send retransmission requests for HP IDs that have not yet been reused.  Even if unlicensed is used for NTN where there is uplink HARQ feedback, the gNB can chose to not send the HARQ feedback if it does not want to.  Option 1 has low or no spec impact (depending on how we do with drx timers). |
| Qualcomm | Option 3 | gNB can send UL grant without waiting for decoding result of previous PUSCH transmission (i.e. up to gNB implementation);  However, followings are possible configurations by network for a HARQ process.  (1) The NDI in grant is always considered toggled (no retransmission expected, UE may go to sleep after PUSCH transmission if no other DRX timers are running).  (2) The NDI in grant can be toggled (new transmission) or not toggled (blind retransmission). |
| Lenovo | Option 1 | No spec impact. |
| Apple | Option 1 | No spec impact. |
| CAICT | Option1 | It seems as the simplest way. |
| Nokia | Option1 | Option1 is supported in current specification. |
| CATT | Option 1 | Reuse the legacy mechanism. |
| Xiaomi | Option 1 with changes | **gNB can send grant with NDI toggled/not toggled without waiting for decoding result of previous PUSCH transmission (i.e. up to gNB implementation);** |
| CMCC | Option 1 | Leaving it to the gNB implementation is enough. |

## Other aspects (P9/P13)

**Question 5: Which of the following aspects should be further studied in NTN?:**

1. **Report UE-calculated TA in e.g. msg3/msg5/msgA;**
2. **Enhancements to RSRP-based selection mechanism of 2-step vs. 4-step RACH;**
3. **Introduction of K\_offset in SI (pending RAN1 agreements).**
4. **LCP impact caused by disabling HARQ UL retransmission.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Supported Option(s)** | **Additional comments** |
| MediaTek | Option 1, Option 3, Option 4. | Reporting of UE-calculated TA, K\_offset and LCP impacts on disabling HARQ UL retransmissions should be studied. |
| LG | Option 2 and 4 |  |
| ZTE | Option 1 | **For option 2:**  We understand the intention of RSRP-based RA type selection is to ensure the 2-step RACH can only be selected in case the radio condition is qualified, which provide flexibility on NW side to deploy the 2-step RACH with a good resource efficiency (i.e. only take the UE with good radio condition into account and leave the UE in the edge of cell to 4-step RACH). In NTN, we don’t see new requirement on RA type selection, thus we think the current mechanism can be reused and no optimization is needed on this aspect from RAN2’s perspective.  **For Option 3:**  Since RAN1 is still discussing the details on K\_offset (e.g., cell or beam specific, explicit or implicit indication, and etc.,), what’s need to be broadcast is still uncertain, it is preferred to postpone the discussion until RAN1 make more progress.  **For Option 4:**  We understand the intention of LCP enhancement is to distinguish the grant with fast retransmission (e.g. blind retransmission) and the grant with slow retransmission (e.g. HARQ retransmission) , and ensure the LCH with strict requirement on latency can be mapped to a grant with fast retransmission. However, we think similar requirement have been discussed before in IIOT and the allowedPHY-PriorityIndex has been introduced for the similar purpose. With the PHY-priority index, UL MAC SDUs from the logical channel can only be mapped to the dynamic grants indicating PHY-priority index equal to the values configured for that LCH. In NTN, the NW can simply configure the grant with/without HARQ with different PHY-priority index. No further optimization is needed. |
| Spreadtrum | 1, 2, 3, 4 |  |
| OPPO | Option 1/2/3/4 | We think all these aspects are worth being studied in the WI. |
| Panasonic | 1,2,4 |  |
| Ericsson | 1, 2 | The features shall only be supported if they are found to be beneficial to the NTNs.  We need to wait the RAN1 discussions for Koffset  There is no need to change the LCP.  Is there any UL data that really require HARQ retransmissions? We do not think so.  As” the gNB knows what data the UE has from the BSRs and SRs and from received data thus gNB can schedule accordingly, if some data require retransmissions gNB can schedule it and not reuse the HP ID until it decoded the transmission correctly.  Changing LCP means delaying some data due to the wrong type of HP ID in a grant, and possibly we must then create a new Scheduling Request for this “SR for data requiring HARQ retransmissions” and/or “SR for data not requiring HARQ retransmissions”.  Even if the gNB schedules data on the “wrong” grant type, if gNB decodes the TB correctly, all is fine. If gNB fails to decode a TB, the gNB may send an RLC status report for critical services to trigger early RLC retx. This seems like a better way to handle “wrong type of HP ID” as most of the time gNB knows correctly what data the UE has and even when wrong HPID is used >90% of the transmissions will be successful (assuming link adaptation adjust for <10% block error rate, likely BLER target will be lower for HP IDs without HARQ feedback) so problems will be rare.  If data is lost (for example some MAC CE that is cancelled at the transmission attempt), that may happen in legacy too (but less common if gNB do retransmissions). |
| Qualcomm | 1, 2, 3, 4 | But for option 1, Msg5 should be used in 4 step RACH to report the calculated TA. |
| Lenovo | 1,2,4 | For 2-step RACH, the near-far effect may not be obvious as that in TN, i.e. there may not be a clear difference in RSRP between cell center and cell edge UEs. As a result the RSRP criterion for RA type selection may not work well. |
| Apple | 2 and 4 | RAN1 can decide on 3. 1 has a major spec impact. There is simply no need for 1 and need for unnecessary changes to current RACH procedure. There will be a lot of unnecessary complexity for UEs to send a TA that has been corrected in MSG2 by the network. |
| CAICT | 1,2,3 |  |
| Nokia | Option 2,4 | Option1 may be potential topic while the use case and benefit is not clear. Option3 can be stuied later after RAN1 reaching conclusion on open issues about K\_offset (e.g. cell or beam specific, whether/how to update, implicit or explicit signalling etc). |
| CATT | Option 2,4 | Option 1 Report UE-calculated TA is also related with Option 3. K\_offset in Option 3 is under discussion in RAN1. So we prefer RAN2 to go on discussion on option 4 and 2 at first which doesn’t depend on agreement in RAN1. Then we can come back to discuss option 1 and 3 when there is a conclusion from RAN1. |
| Xiaomi | 1,2,3,4 | For aspect 1, it should be decided whether only differential TA is indicated. |
| CMCC | Option 1, option 2 and option 4 | Option 3 should wait for RAN1 to progress. |

# Summary

<to be completed pending company input>

# Conclusions

Based on company feedback, the following are proposed:

<to be completed pending company input>

# References

1. R2-2010455 – Summary of [Post111-e][908][NTN] RACH and HARQ feedback aspects – InterDigital
2. Chairman’s Notes RAN1#102-e 8.4 v004 – RAN1 Vice Chair
3. R2-2010980 – On Random Access in NTN - Ericsson
4. TS 38.321 – Medium Access Control (MAC) protocol specification v16.1.0