3GPP TSG RAN WG1 Meeting #102-e R1-200xxxx

eMeeting, August 17 – 28, 2020

Agenda Item: 7.2.4.2.1

Source: Moderator (Ericsson)

Title: Resource allocation for NR sidelink Mode 1 – Thread 2

Document for: Discussion, Decision

# 1 List of issues for discussion

[102-e-NR-5G\_V2X\_NRSL-Mode-1-02] Email discussion/approval covering:

* HARQ reporting to gNB
  + Details in the WA from RAN#100-e for the case of reaching the maximum number of HARQ re-transmissions for a TB.
  + Whether there are other exceptional reports to the gNB (e.g., nothing to transmit for DG, etc.) and, if so, how to address them.
  + Editorial corrections and clarifications for HARQ reporting to gNB (if any).
* Processing times
  + Processing time for SL CG type-2.
  + Whether the gNB needs to be aware of SL HARQ RTT (Z = a + b) or alternative assumptions or behaviour, if necessary.
  + Editorial corrections and clarifications for processing times (if any).

By 8/20, followed by potential TPs by 8/25 – Ricardo (Ericsson)

# Discussion

## 1.3 HARQ reporting to gNB

### Issue 1.3-1 Details in the WA from RAN#100-e for the case of reaching the maximum number of HARQ re-transmissions for a TB.

**Regarding the following working assumption made in RAN#100-e:**

1. **When the maximum number of HARQ retransmissions for a TB is reached, the UE reports ACK/NACK based on the contents of PSFCH (i.e., the same behaviour as if the maximum number of retransmissions had not been reached).**
2. **When the maximum number of HARQ retransmissions for a TB is reached, the UE reports ACK.**

**(For other answers, please explain)**

|  |  |
| --- | --- |
| **Company** | **View** |
| NTT DOCOMO | Option A. |
| Intel | Option A |
| vivo | We have no particular preference for either option. But we have some questions about option A. I understand that setting the maximum retransmission for a TB to some extent is to limit the resources used for a single TB. Since gnb is not aware of where the initial transmission takes place on a CG, it cannot differentiate the HARQ-ACK reporting with/without satisfying the maximum transmission times restriction. So if we go with option A, gnb may assign more resources for retransmission due to the reported NACK, then what is the purpose of setting the maximum retransmission times in this case? It sees this limit has virtually no impact on the number of resources used by a TB. Could the proponents of option A elaborate a bit more of the intention of setting such restriction if option A is adopted? |
| OPPO | The WA made in RAN1#100-e is as follows:  Working assumption (Q5):  In case of reaching the maximum number of HARQ re-transmissions for a TB, the UE sends one bit on the UL resources for SL HARQ-ACK reporting. The specification will specify the UE behavior (what the behavior is: FFS), and specify the contents of the report (what the content is: FFS).  This WA was partially agreed in RAN1#100bis-e for configured grant:  Agreements:   * The working assumption (as in proposal 3 in the summary) from RAN1#100-e is confirmed.   Proposal 3 (for a working assumption):   * The working assumption from RAN1#100-e is confirmed.   + In case of reaching the maximum number of HARQ re-transmissions for a TB using resources provided by a configured grant, the UE reports ACK to the gNB.     - ~~FFS whether the specification supports that the gNB configures the UE with a maximum number of transmission per TB.~~   While for dynamic grant, there is no agreement till now. @ FL, can you clarify that this issue is only for DG?  On the other hand, one remaining issue regarding the WA is how the UE knows whether/when the maximal number of transmissions is reached.  We have the following agreement in RAN1#99. For DG, the number of re-transmissions is up to gNB. Based on that agreement, how the UE knows when/whether the number of re-transmission is reached. That should be clarified, and some specification is needed. Otherwise, it cannot work.  Agreements:   * For dynamic grant, the number of retransmissions of a TB is up to the gNB. * For configured grant, the maximum number of times that a TB can be retransmitted using the resources provided by the configured grant is configured per priority per configured grant. |
| LG Electronics | First of all, the network can’t exactly know (a) which CG resource (within a period) is used for the initial TX of TB or (b) how many re-TXs of TB have been performed. In other words, only UE can know such information. With Option B, it is possible for the network to avoid allocating unnecessary re-TX resources for the TB that has been already reached to the maximum number of re-TXs. As a result, we are supportive of **Option B**. Furthermore, **when the maximum number of re-TXs for a TB is configured per priority per CG, it can be defined that this value is shared between re-TXs using CG/DG resources for the same TB**. |
| ZTE, Sanechips | Option A. |
| Sharp | Option A |
| Qualcomm | We prefer Option B  Reporting ACK provides information to the gNB to stop giving grants for this TB. In the case where the UE is not (or cannot be) aware of the maximum number of retransmission, the UE will proceed as normal for any retransmission and provide feedback based on received PSFCH. |
| CMCC | We share similar view with OPPO that this issue is only for DG and UE cannot be aware of whether the number of re-transmission is reached. So the same behaviour regardless of whether the maximum number of retransmissions has been reached or not is preferred. And from our perspective, UE does not need to know when/whether the number of re-transmission is reached, so no specification is needed. |
| Fujitsu | Option A |
| CATT | First of all, how to define the maximum number of the HARQ re-tx for a TB. The maximum number is a number X that should be configured to UE, and . If the maximum number of transmission for a TB is 10, it means this TB can be transmitted and retransmitted no more than 10.   * For **CG type-1 and Type-2**, the maximum number is configured exactly to the UE.   + For CG type-1, the configured grant can configured and indicate the first set of resource for initial transmission and re-transmission, e.g. N\_max=3, then the rest 7 re-transmission will be scheduled by DG. For DG schedule re-transmission, each time DCI can dynamically scheduled N\_max=1/2/3, and at the end of each set of scheduled resources, there will be a PUCCH resource for ACK/NACK reporting.   + For CG type-2, DCI is used to active the first transmission(s), e.g. N\_max=3 indicated in DCI. For each TB, the initial transmission and 2 re-transmissions (3 resources), the resources are indicated by DCI and configured CG period, and the rest 7 re-transmissions are scheduled by other DG if needed.   Therefore, **for CG Type-1, both gNB and UE knows the maximum number that a TB can be transmitted**. gNB should not schedule extra resources for transmission that is exceeds the maximum number (e.g. gNB will not schedule 11-th transmission resources for the TB).  **Q1: How many TX resources are configured for each TB? 10 or only 3?**  **Q2: Can the DG scheduled re-transmissions use CG resources?**  **Q3: Can a TB use resources located in more than one CG period?**   * For **DG**, there is no such a parameter that is configured to UE to indicate the maximum number of transmission. DCI is used for each set of resources for DG transmission. The first set of transmissions indicated by DCI, e.g. N\_max=3, and the rest 7 re-transmissions are scheduled by other DG if needed. **gNB knows the maximum number that a TB can be transmitted but UE does not know**. UE only perform the (re-)transmissions by using the resources scheduled by gNB and indicated.   **With the analysis above, there is no necessary for a UE to act the HARQ-ACK report in case of reaching the maximum Tx number of a TB.** |
| Huawei, HiSilicon | We do not have strong preference here, but it should be clarified how the UE know the maximum (re-)transmissions are reached.  According to the agreements in RAN1#99, the number of retransmissions of a TB for dynamic grant is up to the gNB, thus maximum number of HARQ retransmissions has not been specified for dynamic grant so far. To provide more flexibility to dynamic grant, it is also no need to define such maximum number of HARQ retransmissions, thus it seems reasonable to report ACK/NACK based on the contents of PSFCH as the general way. |
| Samsung | We prefer Option B.  In our understanding this issue is for CG, since the retx number of DG is up to gNB. For CG, if Option A is adopted, gNB has no information of how many times one TB is transmitted, thus gNB may schedule resource for retx even if the maximum retx number is reached. Therefore the scheduled resource is wasted. Option B can better reflect the situation of SL transmission.  In addition, since CG could schedule only up to 3 resources per period, we prefer to clarify the number of retx per priority per CG includes both CG-based resource and DG-based resource for retx of CG. |
| Fraunhofer | Option B.   * In our understanding, the maximum number of HARQ retransmissions in question is **relevant only to CGs**. The maximum number of retransmissions for a TB using resources of the grant is configured per priority per configured grant using the parameter *sl-CG-MaxTransNumList*. * Hence TX UEs are aware of this restriction, and can manage and maintain the number of retransmissions for a given TB. However, as LG stated, the **gNB cannot keep track of the number of retransmissions** a TX UE carries out for a single TB using the resources provided in a given CG. * In the case of option B, if the gNB receives an ACK after the maximum number of retransmissions was reached, it will carry out the same action as when it receives an ACK for a successful transmission. Hence the gNB’s response does not vary between the two events. * In the case of option A, it is important for the gNB to differentiate between a NACK when the maximum number of retransmissions has been reached and when it has not been reached.   + If NACK was sent before the maximum number of retransmissions was reached, the gNB has the option to schedule further resources for retransmission using DG.   + If NACK was sent after reaching the maximum number of retransmissions, the gNB should NOT provide further grants for retransmission. The UE has already attempted to transmit the maximum number of times, given the existing conditions, and has failed to do so successfully. Providing further resources would only result in inefficient resource utilization. * Agree with OPPO that option A would raise the **question of the functionality and purpose of the restriction on maximum number of retransmissions**.   If the value is set to the maximum, the **behaviour of the UE intended by option A can be covered by option B**, where the gNB can continue providing more resources for retransmission, but bound by the maximum upper limit. |
| Ericsson | Option A |
| Futurewei | Option A |
| Nokia, NSB | It seems to us that the fundamental problem is the meaning of “maximum number of HARQ retransmissions for a TB is reached”along with the definition of sl-CG-MaxTransNumList, which reads “maximum number of times that a TB can be transmitted using the **resources provided by the configured grant**” (emphasis added).  If we take the parameter definition as it currently stands then there does not seem to be any reason why this limit on the number of TX **using resources provided by the configured grant** should restrict the number of additional ReTX **using resources provided by dynamic grant**. Hence option A is the obvious answer.  On the other hand one could consider changing the definition of the parameter to e.g. “maximum number of times that a TB whose initial TX uses resources provided by the configured grant can be transmitted in total”. If such a change is made then option B makes sense. |

### Issue 1.3-2 Other exceptional reports to the gNB

**Are there are other exceptional reports to the gNB (e.g., nothing to transmit for DG, etc.)?**

1. **Yes (please provide details).**
2. **No further specification is necessary.**

**(For other answers, please explain)**

|  |  |
| --- | --- |
| **Company** | **View** |
| NTT DOCOMO | Yes  When a dynamic grant is provided to a UE, but if the UE skips the corresponding SL transmission, the UE should report ACK to gNB.  According to RAN2 spec., SL , is possible; hence, we think this exceptional report is needed. For example, a UE sends SR/BSR to gNB and gNB schedules SL grant based on the reported information. BSR includes destination ID/LCG ID/Buffer size. gNB predicts how many grants are necessary. However, SL grant does not include MCS indication/MIMO/DM-RS/CSI-RS/etc. They are determined by the UE itself and actual transmitted TBS is dependent on the parameters (i.e. channel condition/UE capability/etc.). The provided SL grants may be insufficient to transmit the reported buffer or may be sufficient. If sufficient, the UE could not have any transmitted data on one or more of the provided grants. This is feasible case in our understanding. gNB does not know details of actual SL communication. |
| Intel | We are not sure if the case mentioned by NTT DOCOMO needs additional handling. For example, it can be similar to the case when a configured grant is not used. If it is not handled by specification, we are supportive that a UE sends ACK in this case. |
| vivo | Option B. No.  In the case of UL, the ability to skip on CG is a mandatory capability, but skipping on DG is an optional capability. Whether the UE can skip on a UL DG depends on the presence of skipUplinkTxDynamic. Specifically, if skipUplinkTxDynamic is configured, the UE can skip on UL grants, if not, the UE cannot skip the grant and should send a MAC PDU with padding.  The gnb may provide a sidelink grant, but the UE has no data available. In this case, if SL grant skipping is allowed, then we may need to discuss whether this behavior is an optional or mandatory capability. Since skipping UL DG is an optional capability, skipping SL DG should also be optional. However, defining such a new capability may introduce new RRC parameters, which should be avoided during this maintenance phase.  Another possibility is that the UE can send a MAC PDU with padding on a predetermined resource, the UE can then report HARQ-ACK based on the PSFCH reception associated with the DG. In this case, the process for determining the HARQ information is the same as if there is SL data to be sent. Therefore, no special reports need to be defined. |
| OPPO | Yes.  In case of dynamic grant, the SL transmission resource is allocated by gNB. While if the allocated resource beyond the PDB of the packet, UE will not use the resource for SL transmission. In that case, UE should report ACK to gNB to terminate gNB’s allocation for re-transmission. |
| LG Electronics | At this moment, it is not clear that the current MAC specification supports the mechanism that a UE skips SL TX on the resources of Mode 1 DG. So, it would not be desirable for RAN1 to discuss/conclude the relevant behaviour of HARQ reporting before RAN2 makes the explicit agreement on whether to support such mechanism in Rel-16 NR V2X. So, our preference is **Option B**. |
| ZTE, Sanechips | Option B. |
| Sharp | Option A. We are supportive of specifying the case mentioned by NTT DOCOMO. |
| CMCC | We are open to this optimization so option A is slightly preferred. |
| Fujitsu | Yes  We think it is a feasible case where UE has nothing to transmit on the resources allocated by dynamic grant. For example, this case can occur when PUCCH carrying sidelink ACK is dropped due to prioritization. Since PUCCH is not transmitted, gNB will not receive any ACK/NACK and thus may treat it as DTX. For safety, gNB may continue to schedule retransmission by using dynamic grant. However, UE has nothing to transmit since it has been ACKed before. In this case, letting UE report ACK can prevent gNB from keeping scheduling retransmission. Actually, the principle is similar with what has been defined when UE has nothing to transmit on the configured grant resources. |
| CATT | Option B.  Currently, there is not exceptional cases that needs ACK/NACK reported to gNB. in terms of the case mentioned by NTT, the skiping mechanism is not supported now (also mentioned by LGE). UE will perform the transmission and use the resources scheduled by gNB. |
| Huawei, HiSilicon | Option B.  UE acquires the resource based on gNB scheduling, so gNB is aware of the SL transmission requirement, such as the latency, so the gNB could schedule a resource within the PDB. For the SL transmission which scheduled by gNB but skipped by the UE, it is due to the misunderstanding about the buffer status between UE and gNB, but this is similar with NR Uu scenario, which UL grant is skipped as well when no UL data is going to be transmitted. Therefore, we think no special issue in SL needs to be addressed. |
| Samsung | Option B.  Regarding SL skipping, we share similar view with vivo and LGE that RAN1 should not trigger this feature and introduce high layer impact. For the other issues, we consider it can be solved by current mechanism. Therefore, so further specification impact is needed. |
| Fraunhofer | Option A.  We agree with the issue being raised by NTT Docomo. If the UE does not transmit on a resource indicated by a DG, it can respond with an ACK in order to avoid further scheduling of resources by the gNB. This issue is already covered in 38.213 for CG, it can be extended for DG as well. |
| Ericsson | B  At this point, we are not convinced there are other exceptional cases. |
| Futurewei | B  While there could be exceptional cases that could potentially be investigated, we do not see any critical one needed in Rel-16 |
| Nokia, NSB | B |

## 1.4 Processing times

### Issue 1.4-1 Processing time for SL CG type-2

**The current specification only captures the PSSCH processing time for sidelink dynamic grant (TS 38.214 Clause 8.6) corresponding to the following agreement from RAN1#101-e:**

|  |
| --- |
| Agreements:   * For dynamic grant in Mode 1, a UE does not expect to be scheduled to perform a SL transmission earlier than after the end of the scheduling PDCCH.   + - is 10, 12, 23, and 36 for equal to 0, 1, 2, and 3, respectively.       * FFS other values of based on the discussion on capabilities (Q5).     - *µ* corresponds to the one of (*µDL*, *µSL*) resulting in the largest *Tproc.*     - = 1     - (parameters as defined in 38.211) |

**R1-2005339 proposes using the existing preparation time for SL CG type-2 activation. Is the following proposal agreeable?**

**Proposal:**

* **For SL configured grant type-2 activation, the UE processing time is equal to Tproc (agreed in RAN1#101-e).**

|  |  |
| --- | --- |
| **Company** | **View** |
| NTT DOCOMO | Agree |
| Intel | Agree. There seems no essential difference. |
| vivo | Agree |
| OPPO | Agree |
| LG Electronics | We are fine with FL’s proposal. |
| ZTE, Sanechips | Agree |
| Sharp | Agree. |
| Qualcomm | Agree |
| CMCC | Agree |
| Fujitsu | Agee |
| CATT | Agree. |
| Huawei, HiSilicon | Agree.  The processing time for CG type 2 should be the same as dynamic grant. |
| Samsung | Agree. |
| Ericsson | Agree |
| Futurewei | Agree |
| Nokia, NSB | Agree |

### Issue 1.4-2 Whether the gNB needs to be aware of SL HARQ RTT (Z = a + b) or alternative assumptions or behaviour, if necessary

**A few contributions (e.g., R1-2005741, R1-2005847) discuss whether it is necessary for the gNB to know the SL HARQ RTT (i.e., Z = a + b in the agreements) for being able to schedule the SL Mode 1 transmissions and to schedule the PUCCH transmissions with SL HARQ reports.**

1. **It is necessary to agree on an assumption on the HARQ RTT.**
2. **It is not necessary to agree on an assumption on the HARQ RTT. In that case,**
   * **How should the UE proceed if the time between two SL transmissions (for the same TB) is shorter than the HARQ RTT?**
   * **How should the UE proceed if the time between the last PSFCH reception and the SL HARQ report to the gNB is smaller than the processing time at the UE?**

**(For other answers, please explain)**

|  |  |
| --- | --- |
| **Company** | **View** |
| NTT DOCOMO | Option A |
| Intel | Option A.  To us it seems easier to define sidelink HARQ RTT at gNB. In this case it should be the same as in Mode-2 with the only update that ‘b’ component needs to be bounded by specification, otherwise still not possible at gNB to always respect this restriction. In our view, ‘b’ is similar to Tprep, and additional small margin could be added if necessary.  If this restriction is not introduced, then another specification effort is required to introduce a mix of blind and feedback-based retransmissions on SL depending on the size of the gap between granted resources. Or, it can lead to unnecessary resource wastage. |
| vivo | Option A.  We share the same view as intel that a mix of blind and feedback-based retransmissions should be avoided in mode-1. If the time gap between two adjacent resources provided by a DG/CG is less than ‘a+b’, UE shall map non-HARQ-feedback based MAC PDU on the resources. In this case, UE shall set the ‘SL HARQ enable/disable’ indication in SCI on these resources to ‘disable’, and UE is not expected to be provided with a PUCCH by the DG/CG as there is no associated PSFCH reception.  [vivo-2] Regarding the question: **How should the UE proceed if the time between the last PSFCH reception and the SL HARQ report to the gNB is smaller than the processing time at the UE**, I think this issue can be avoided by NW proper scheduling. For example, even for the async case where NW has no SL timing, NW can provide a conservative K1 so that the gap between PSFCH and PUCCH is always equal or larger than the minimum requirement. We don’t think additional rules are needed for UE side. |
| OPPO | Option B  For the definition of a and b, there is the following agreement. The time of b is based on UE implementation. Different UE may have different UE capability of the processing time. It is not good idea to reopen the discussion of UE capability for PSFCH processing and PSCCH/PSSCH preparing time.  For mode 1, how to promise the time between 2 allocated SL resources is larger than Z is left to implementation. The parameter a is determined by resource pool parameter, which is known by gNB. gNB can allocate SL resources with conservative assumption of b. in case the time gap between 2 resources is less than Z, UE cannot use the 2nd resource for transmission. How to report HARQ-ACK can re-use the mechanism as dropping transmission because of prioritization.  gNB can allocate the resource of PUCCH which has enough timing gap between the last PSFCH and PUCCH.  Agreements:   * In Step 2, a UE ensures a minimum time gap Z = a + b between any two selected resources of a TB where a HARQ feedback for the first of these resources is expected   + ‘a’ is a time gap between the end of the last symbol of the PSSCH transmission of the first resource and the start of the first symbol of the corresponding PSFCH reception determined by resource pool configuration and higher layer parameters of *MinTimeGapPSFCH* and *periodPSFCHresource*   + ‘b’ is a time required for PSFCH reception and processing plus sidelink retransmission preparation including multiplexing of necessary physical channels and any TX-RX/RX-TX switching time and is determined by UE implementation |
| LG Electronics | Just to be clear, we reformulated the issues to be discussed/resolved.  Issue 1) How to avoid the case that the time gap between “PSFCH RX resource” and “re-TX resource” is smaller than UE’s capability (i.e., b)?   * Considering an impact on RAN2 specification, it would not be desirable to define new UE capability signalling on “b”. **We think that this issue can be resolved by defining the maximum value of “b” that can be supported by all the UEs**.   Issue 2) How to handle the case when multiple SL HARQ bits are multiplexed in the same PUCCH resource and some of the bits are related to PSFCHs not satisfying the minimum PSFCH-to-PUCCH processing time (i.e., Tprep)?   * This problematic case can occur due to a timing misalignment between gNB and UE’s sync source. In the example of figure below, a UE doesn’t have sufficient processing time to generate SL HARQ bit for the green PSFCH when compared with the red PSFCH. **From our perspective, it can be defined that the UE sets the SL HARQ bit with insufficient UE processing time as NACK state** (i.e., for the green PSFCH). |
| ZTE, Sanechips | Option A, in light of a simple solution with a specified upper-bound of “b” (even the value of b is an UE implementation issue). From gNB implementation point of view, it is better not to force gNB to do a best-effort guess of b’s upper-bound on a per-UE basis, given the UE implementation does seem to allow a time-varying value of “b”. |
| Sharp | Option A |
| Qualcomm | Is the proposal to convey the exact value of ‘b’ to the gNB or to place an upper limit on ‘b’?  We are supportive of defining an upper bound on ‘b’, but this was discussed in Mode 2 preparation phase for this meeting, but was not prioritized. Our view is that an upper bound on ‘b’ is useful even in Mode 2 and such an upper bound should be applicable there as well.  On the other hand, the exact value of ‘b’ is already agreed to be left up to UE implementation and we don’t see the need to change the agreement.  **How should the UE proceed if the time between the last PSFCH reception and the SL HARQ report to the gNB is smaller than the processing time at the UE?**  The behvaior should be left up to UE implementation as is done when minimum PDSCH-to-PUCCH time (N1) is violated per subclause 5.3 in 38.214, where a valid HARQ-ACK is only required when processing timeline is satisfied. This is a general problem though and not specific to the value of ‘b’ |
| CMCC | Option A. We share the similar view with intel and vivo that a mix of blind and feedback-based retransmissions should be avoided in mode-1. The time of b is based on UE implementation in mode-2, for simplicity, a definition of upper-bound of “b” can be specified for mode-1 scheduling. |
| Fujitsu | Option A |
| CATT | Option A.  Similar view with Intel and vivo. With knowing the HARQ RTT, mixture transmission of blind and HARQ based will be avoided. Furthermore, resources will not be wasted. |
| Huawei, HiSilicon | Option B  As the previous agreement in mode-2, “b” is implemented by the UE and this value could be varied by different UEs, so we do not see the necessity that the gNB should know the values. For the feedback scheduling of SL to UL that does not satisfy the processing timeline, we think it would be an error case that up to UE implementation. This is also the NR Uu rule. The aim to specify the processing timeline is providing a reference for gNB scheduling, the gNB could guarantee the scheduling to meet the requirement, and the UE behaviour is unnecessary to specify if the timeline is not met. |
| Samsung | Option A |
| Fraunhofer | Option A  The gNB needs to be aware of the processing timeline of the UE, in order to schedule the PUCCH accordingly. Otherwise, it is possible that the feedback might not be ready at the UE at the scheduled PUCCH instance, and the behaviour of the UE would be unclear. |
| Ericsson | Option A. We are fine with a clarification. With such clarification, the gNB can avoid problematic cases. |
| Nokia, NSB | Option A – OK to specify an upper bound on b |

## Other comments

NOTE: I will prepare TPs or list of TPs for editorial aspects. I will share it in a later iteration.

|  |  |
| --- | --- |
| **Company** | **View** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |