3GPP TSG-RAN WG2 Meeting #112-e R2-20xxxxx

E-meeting, …, 2020

Agenda Item: …

Source: Ericsson

**Title:** Summary of email discussion [Post111-e][926][SmallData] Context Fetch

Document for: Discussion and Decision

# Introduction

The discussion handles:

* [Post111-e][926][R17 Small Data] Context fetch (Ericsson) Henrik/Tuomas

Scope

• Discuss the RAN2 aspects of context fetch with and without anchor relocation and identify any issues that need further input (e.g. from RAN3 and/or SA3 etc)

• Can also discuss if there are any bottlenecks to support SDT without anchor relocation (i.e. the FFS point above)

Outcome: agreeable proposals and identified impacts to other groups

Deadline: Long

Companies are invited to respond to the questions below in time for the email discussion deadline.

# Context Fetch

The questions in this email discussion is based on company input in Tdocs submitted to RAN2-111e: R2-2007469, R2-2007197, R2-2007838, R2-2006583, R2-2007541, R2-2007195, R2-2007180, R2-2006714

## Background

A UE moved to RRC-INACTIVE state results in that the DU releases the stored UE context including corresponding tunnels established between DU and CU-UP. The UE remain CM-CONNECTED and both UE and CU-CP store the UE Context. When a UE moves within an RNA area configured by NG-RAN, the last serving gNB node keeps the UE context (and the UE-associated NG connection with the serving AMF and UPF).

Context relocation can be done by the RAN-based Notification Area (RNA) update procedure. The procedure may be triggered when the UE moves out of the configured RNA, or periodically. Figure 1 from 38.300 (therein Figure 9.2.2.5-1) shows the RNA update with context relocation.



Figure 1 RNA update with Context Relocation

The RNA update procedure is initiated by the UE sending an RRCResume with RNA update as cause value. Upon successful UE context retrieval, the receiving gNB would typically keep the UE to inactive and triggers a Path Switch Request. After the path switch procedure is performed, the receiving gNB triggers release of the context in the last serving gNB.

The RNA procedure may also be done without context relocation. In case the last serving gNB decides not to relocate the UE context (e.g. in case the UE is still within the RNA area), it sends a Retrieve UE Context Failure to the serving gNB. procedure and sends the UE back to RRC\_INACTIVE, or to RRC\_IDLE directly by an encapsulated *RRCRelease* message (i.e. RRC Release).

### Small Data

For NR SDT, when the UE temporarily access the NW for transmission of UL data, e.g by payload transmitted in a 2-step msgA or 4-step msg3 Random Access procedure, context handling in the NW may need to be extended to support small data transfer in RRC\_INACTIVE. As a result, both with and without anchor relocation needs to be considered.

**Q0: Do you agree with the general principles above (legacy)?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Response** | **Comment** |
| ZTE | Yes | Both should be supported as required by the WID (“*Context fetch and data forwarding (with and without anchor relocation) in INACTIVE state for RACH-based solutions [RAN2, RAN3]*”).   * For the case of anchor relocation, the legacy procedure can be reused without much changes. * For the case of no-anchor relocation, some discussion is needed. Specifically, the small data packet sent in the first UL message will be using the new key (assuming no changes to the existing security framework) and anchor gNB has to terminate the PDCP protocol layer for the DRB and use the new key for the data. Feasibility of this can be checked with RAN3 and SA3. |
| Huawei, HiSilicon | Yes | As captured in WID already, for RACH-based solution context fetch and data forwarding with and without anchor relocation needs to be considered. |
| Panasonic | Yes |  |
| ITRI | Yes |  |
| NEC | Yes | This has been agreed in the WID. |
| Nokia | Yes | Both options are in the scope of the WID and should be considered as per RAN2 agreement. The solution without anchor relocation may require input from RAN3: RAN3 should be consulted via LS. |
| Samsung | Yes | Agree with the legacy principles according to which context may or may not be relocated (as decided by last serving gNB). |
| Qualcomm | Yes |  |

Following the background assumptions, part of the legacy events that may initiate a context relocation request, i.e RNA update and RRC Resume request, also SDT from RRC INACTIVE contribute to the frequency and initiation for when context relocation may be considered.

### 2.2.1 General case

It can be assumed that w.r.t SDT and RA based schemes, context relocation can be realized by the following:

The receiving gNB, after resolving the gNB identity contained in the I-RNTI, sends the Retrieve UE Context Request to the Last Serving gNB where:

1. Retrieve UE Context Failure is received at Receiving gNB and UE AS context is maintained at the Last Serving gNB. SDT data is forwarded to the Last Serving gNB where it is deciphered, sent to 5GC and DL data tunneled to the Receiving gNb by the Last Serving gNb, or
2. Retrieve UE Context Response results in that UE AS context is relocated to the (new) Receiving gNB and SDT Data is either:
   1. Stored at the Receiving gNB until UE Context is relocated where it is deciphered and sent to 5GC after context relocation
   2. SDT Data is transferred to Last serving gNB where it is deciphered and sent to 5GC, and UE Context relocation is performed subsequently.
   3. Path switch to the Receiving gNB is performed.
3. The overall options for either 1 or 2 above should remain as in legacy in that it would be the Last Serving gNB that decides on UE Context relocation.

**Q1: Do companies agree with the above general assumption 1-3.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Response** | **Comment** |
| ZTE | Yes, but | We share the general understanding above with the rapporteur.  With regards to the details (e.g. whether retrieve UE context Failure is used for this or some other message etc), RAN3 can be tasked to decide this.  Then, based on the current agreement that some stored UE context will be used (at least for RLC we agreed that stored configuration will be used), the implication is that the target gNB will have to buffer the data until the response message from the serving gNB is received and then the target should:   * Either forward the stored SDT data packet (no anchor relocation case) – i.e. bullet b) in the above rapporteur text: or * Perform anchor relocation and pathswitch and establish new CN tunnel through the target (i.e. bullet b) mentioned above).   In both cases, the UE context needs to be transferred to the target gNB though and this is different from the legacy procedure (where in case of no anchor relocation, the UE context is not sent, but in the SDT case some UE context is needed – e.g. to establish the RLC bearer). This aspect should be clarified to RAN3.  Note our understanding is that in the rapporteur text bullet point C) regarding pathswitch is part of the case when anchor relocation happens (i.e. bullet a) |
| Huawei, HiSilicon | Yes in general | Bullet 2c. should not be a separate option, but is complementary to both alternatives 2a and 2b.  For both cases, i.e. with and without anchor relocation, we need to decide whether the first UL packet is buffered in the Receiving gNB for the duration of context retrieval procedure or is it always sent to the old gNB in parallel. In our opinion, user data should be buffered at the receiving/target gNB until context retrieval procedure is finalized and afterwards forwarded either to 5GC (if context is relocated) or to the last serving gNB (if context is not relocated). This is similar behavior as applied in LTE EDT. |
| Panasonic | Yes | Assumption 2a and 2b are alternative to each other, while assumption 2c is mandatory and needs to be performed after 2a or 2b. |
| NEC | Yes | Agree with HW that bullet 2c should not be a separation.  We are fine to let the last serving gNB to make decision on whether anchor relocation is to be performed. If the last serving gNB decide that anchor is not relocated, it will provide TNL information to the new gNB to forward UL data from the new gNB to the last serving gNB.  All these gNBs interaction over Xn could be concluded in RAN3. |
| Nokia | This seems to be purely RAN3 territory. | Likely an LS to RAN3 is needed. |
| Samsung | - | In our view, the detailed procedure and interaction between current gNB and last serving gNB should be discussed in RAN3 first.  Agree with Panasonic/Huawei, that 2a/2b are alternatives and 2c is performed after 2a/2b. |
| Qualcomm | Yes | For both two cases, i.e. with or without anchor relocation, the first uplink small data can be sent together with the Retrieve UE Context Request to the last serving gNB.  As the Rapporteur described above, for the without anchor relocation case, uplink small data is forwarded to the last serving gNB where it is deciphered, sent to 5GC (bullet 1). For the with anchor relocation case, the first uplink small data is transferred to Last serving gNB where it is deciphered and sent to 5GC, and UE Context relocation is performed subsequently (bullet b)  If the uplink small data has to be buffered in the serving/target gNB until the UE context retrieval response (or failure), it will cause unnecessary latency. |

### 2.2.2 Assistance Information

The decision for Context Relocation w.r.t SDT may be based on evaluating signaling load and signaling latency or other. For example, if subsequent SDT data is expected without moving UE to CONNECTED (pending agreements), context relocation may be beneficial compared to subsequent tunneling of SDT data and initiation of Context relocation procedures without path switch and context relocation. The signaling and data forwarding to the anchor CU-CP/CU-UP is expected to involve potentially increased signaling than what is the case for legacy RNAU procedure.

On the other hand; depending on if INACTIVE UEs for the most cases are not expected to connect to a Serving gNB other than the Last Serving gNB (i.e Anchor gNB), optimizations may not be as useful. For example, if a UE in most cases anyway have performed RNAU due to e.g. timer expiry and as a result, context relocation may not specifically benefit from optimizations related to SDT in particular. The RNAU timer (T380) has the possible value range 5min to 720min currently.

To aid the last serving gNb in the decision to relocate context, assistance information may be useful. However, the need for this should be balanced to the need for optimizations for this case.

Example of assistance information is traffic pattern information, BSR or other (see Q3).

**Q2: For subsequent SDT, is it beneficial for the Last Serving gNB (anchor gNB) to know if the UE needs more UL resources for subsequent data transmission after the initial small data transmission?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Response** | **Comment** |
| ZTE | Yes | In general, the serving/anchor gNB should be aware of how much data is in the UE buffer.  On Uu interface, if not all data fits in the initial UL message, the UE will include a BSR and in our view, this information is enough for the anchor gNB to make the decision. The BSR information needs to be conveyed by the target to the source in this case as the assistance information.  In general, the gNB should know based on the BSR, subscription information and the data included for DRBs whether subsequent data is to be expected. Whether any further assistance information should be included (e.g. from UE side) should be carefully considered based on the overhead this incurs and the reliability of such information etc. In our view, BSR is enough. |
| Huawei, HiSilicon | Yes | This is an important factor for the Last Serving gNB to consider when making the decision on whether to relocate the context, e.g. in case of SDT transmission with only a single UL packet, the overhead related to anchor relocation will likely exceed the overhead due to user data being forwarded. |
| Panasonic | Yes | We think BSR may not be enough for UE with frequent UL small data transmission. For instance, UE may not include BSR when provided grant from gNB is large enough to carry all UL data. In order to facilitate serving gNB to configure resources for subsequent UL data transmission and anchor gNB to make decision on whether to perform context relocation, UE can indicate traffic pattern (one possible way is to indicate periodic traffic pattern or aperiodic traffic patter) in the assistance information message. Based on the traffic pattern, anchor gNB can make the relocation decision accordingly. |
| ITRI | Yes | We share the same view as Huawei that the information about whether the UE needs more UL resources for subsequent data transmission could assist network in deciding whether to relocate the context. |
| NEC | Yes | If there is more UL data to be transmitted by subsequent SDT or entering CONNECTED state, some configuration needs to be generated and sent to the UE by the new cell by SRB1, the UE context including the PDCP and security context needs to be passed to the new gNB. Thus without anchor relocation is only applicable to one-shot SDT transmission. |
| Nokia | No, but | SDT procedure is limited by the availability of limited amount of data in the UE, hence, based on that the context retrieval is for SDT should serve as the minimum information for the Last Serving gNB. Whether additional information is needed could be discussed. |
| Samsung | - | The detailed algorithm to determine whether to relocate anchor or not should be discussed in RAN3. If RAN3 decides that certain assistance information is needed, they will inform the same to RAN2 and then RAN2 can discuss whether it is feasible to provide such information or not.  Regarding the usage of BSR as assistance info, even if BSR is included in first UL message, it may not reflect all the UL data which may need to be transmitted until the SDT procedure is completed. The UL data can arrive even after first message. |
| Qualcomm | Yes | It is important to let the Last serving gNB to know the UE small data status before deciding whether to relocate the UE context or whether UE needs more UL resources for subsequent data transmission. Therefore, the UE buffer status is one of the necessary information to assist the Last serving gNB to make decision.  In our view, only including the UE buffer status is not enough. Because UE buffer status only reflects the current status of the traffic UE may have which is one-shot traffic only. If UE’s traffic pattern is a multi-shot traffic or a small data arrival with gaps or more traffic arrival after the first uplink transmission, it is helpful for the Last serving gNB to make decision to configure multiple UL resources one time for the subsequent data transmission and to decide whether to relocate the UE context if UE can indicate traffic pattern in the UE assistance information. |

If assistance data is seen beneficial, there are a number of options foreseen. For example, the Retrieve UE Context Request message may contain a small data request indication or message for informing the Last Serving gNB of subsequent data transmission for uplink small data transmission. Such assistance information would seemingly fit naturally into the context relocation procedure itself as part of what impacts the context relocation decision.

Alternatively, the addition of a BSR or other indication may be present in the data forwarded by the Serving gNB in the SDT payload. However, limiting to BSR or similar UE indication may omit other node specific or system specific information. However, this information may be useful in general for the decision on how to handle particular UEs and their scheduling.

Note that it can be expected that the actual procedures and message exchange on Xn and other interfaces are discussed and defined in RAN3, although also based on informed assumptions of alternatives discussed in RAN2 for useful information exchange w.r.t SDT.

**Q3: Is there benefit of extending the Retrieve UE Context Request message to also contain assistance info e.g. requesting subsequent data transmission for uplink small data transmission?**

**Clarify and expand if other mechanism would be of interest if applicable.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Response** | **Comment** |
| ZTE | Yes | Per our comment above, one option is to include something similar to the BSR in the retrieve UE context Request message as the assistance information. |
| Huawei, HiSilicon | Yes | The complete minimum is to inform the old anchor gNB that the retrieve UE context procedure is initiated due to SDT. On top of that, at least the information about the estimated volume of data to be transmitted over SDT would be useful for the old anchor to make an informed decision about (not) relocating the anchor (e.g. size of data, single or more packets etc.). It would be useful for the gNB to know not only about the UL data in the UE buffer, but also whether the data is expected in DL (this can be known from upper layers in the UE, e.g. in case the UL data is sent using TCP). |
| Panasonic | Yes | As we mention in the Q2, other than BSR, UE may indicate its uplink traffic pattern in the UE assistance information instead, which helps anchor gNB to make the relocation decision. |
| ITRI | Yes |  |
| NEC | Yes | As the comment in previous question, anchor relocation is only applicable to one-shot SDT. Thus we need assistance information in Retrieve UE Context Request message. The information could be value of the remaining payload to be transmitted, e.g. BSR, or just indication of whether this is one-shot SDT, which can be FFS in RAN3. |
| Nokia | Up to RAN3. | See above in Q2. |
| Samsung | - | See comment to previous question. Detailed parameters/info exchanged between current gNB and anchor gNB is up to RAN3. |
| Qualcomm | Yes | See our comment in Q2. It is beneficial the Retrieve UE Context Request message also containing assistance information. |

### 2.2.3 UE DRB Configuration

Which RLC and PDCP DRB configuration is to be used by UE in RRC\_INACTIVE for SDT may have impact on the context relocation procedure as it may impact data forwarding and tunneling specified in RAN3. For example, if a UE specific RLC configuration is used, this needs to be provided to Receiving gNB. Alternatively, a default/common configuration may be used. Note here that similarly to RLC, the UE PDCP configuration may need consideration. The DRB configuration needs to consider both with and without context relocation.

**Q4a: What RLC configuration should be used and why for user data in SDT considering that context relocation may be performed?**

1. **The RLC configuration stored in UE Context, or**
2. **a default/common RLC configuration**
3. **other variant or combination**

|  |  |  |
| --- | --- | --- |
| **Company** | **Response** | **Comment** |
| ZTE | Option 1 was already agreed by RAN2#111e | No need to discuss this further since we already agreed to go with option 1. |
| Huawei, HiSilicon | Option 1 | This has already been agreed during the last meeting:  ”From RAN2 perspective, stored “configuration” in the UE Context is used for the RLC bearer configuration for any SDT mechanism (RACH and CG).”  This rule should be applied regardless of whether the UE context is relocated or not. |
| Panasonic | Option 1 |  |
| NEC | Option 1. | This has been agreed last meeting. Seems no need to discuss.  And if the intention is to discuss whether to have different options in case of with or without anchor relocation. We think that since the UE is not able to decide whether anchor relocation is to be performed or not, unified UE behavior is expected for both cases.  The question is where the RLC entities are located at the network side, the last serving gNB or the new gNB? In case of RLC entities located at the new gNB, the RLC configuration context needs to be forwarded to the new gNB even when anchor is not relocated. In case of RLC entities located at the last serving gNB, then how to forward RLC PDUs between the last serving gNB and new gNB should be specified by RAN3. It seems that the first option is simpler. |
| Nokia | Option 3. | We agree with ZTE that Option 1 was agreed in the previous meeting. However, we think that the implications of the agreement were not well discussed.  For instance, Option 1 means that the gNB where the UE performs SDT cannot decode the received RLC PDU before the context has been retrieved from the anchor gNB, which incurs quite a bit of delay for the data. Furthermore, it means that the context needs to be retrieved to the gNB in all the SDT cases (even without the anchor re-location in the NW) which seems like an unnecessary overhead.  Hence, we prefer to revise the agreement from the previous meeting a bit. It seems possible for the UE to use the stored RLC configuration when it performs SDT under the same gNB that sent the UE into INACTIVE (e.g., only in the cell the UE was put INACTIVE). In any case, the gNB should be able to decode the RLC packet at once to avoid the above issue. |
| Samsung | Option 1 | This has already been discussed and agreed in RAN2 #111e. |
| Qualcomm | Option 1 | Option 1 was agreed in the last meeting. |

**Q4b: Assuming that a UE PDCP configuration is dependent on the UE AS security context, along with DRB specific QoS, is a UE specific PDCP configuration to be assumed?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Response** | **Comment** |
| ZTE | Yes | * For the case of anchor relocation, given that anyway, some UE context has to be retrieved (i.e. RLC configuration per above), we think the PDCP context will also be transferred and the configuration will be established based on stored configuration * For the case of no anchor relocation, the PDCP layer will be terminated in the anchor and hence this is no need to transfer the PDCP configuration |
| Huawei, HiSilicon | Yes | RAN2 has already agreed that small data transmission is configured by the network on a per DRB basis. Small data should be then sent over the respective DRB using the stored DRB configuration. |
| Panasonic | Yes |  |
| NEC | Yes | As the UE is not able to decide whether anchor relocation is to be performed, unified UE behavior is expected for both cases, thus PDCP configuration in the UE context should be used. And in case of without anchor relocation, the PDCP entity should be located at the last serving gNB. |
| Nokia | Yes | Since the security context is known only to anchor gNB, it seems stored PDCP configuration can be used in all the cases (assuming the PDCP resides in anchor gNB in case of “without anchor relocation”).  However, it seems desirable not to require UE context retrieval in the “without anchor relocation” case as discussed by ZTE – see our response in Q4a. |
| Samsung | Yes |  |
| Qualcomm | Yes | The stored UE-specific PDCP configuration should be used. |

### 2.2.4 DL Data

SDT may in many cases result in DL data being sent as a result of the UE transmission. Since the data from UPF could be traversed through the Last Serving gNB (anchor node) and tunneled to the Receiving gNB before transmitted to the UE, additional delay may be introduced. In such cases it could be valuable to extend the time to release the UE, for example by extending T319 (timer started upon sending RRC Resume Request) in order to support transmission of DL data related to SDT. The maximum value currently is 2000ms.

**Q5: Is there benefit of that T319 should be extended for uplink small data transmission in RRC\_INACTIVE.?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Response** | **Comment** |
| ZTE | Yes (by the way, this is also being discussed in email #925). | We think this should be extended considering not just the RAN interface delays but more importantly the overall round-trip delay of the application layer (since the response packet in the opposite direction should also be covered in the SDT phase).  A maximum value in the order of 10000 ms should be considered in our view as mentioned in the email discussion #925 |
| Huawei, HiSilicon | Please see the comments | Extending and reusing T319 may not be appropriate, since it would impact also UEs sending Resume request in order to transition to RRC Connected state. Instead of this, we could either restart T319 every time the UE sends or receives small data or have a dedicated timer specific to SDT resume procedure.  It should be noted that this issue is related not only to context relocation, but even more to the subsequent data transmission topic, so perhaps it is more relevant to discuss this more in the other e-mail discussion. |
| Panasonic | No (the same question has been asked in email discussion #925) | Extending T319 will have negative impact to the legacy UE, as well as to the new UEs intending to perform RRC resume procedure for non-SDT purpose. |
| ITRI | No. | We share the same view as Panasonic that extending T319 will have negative impact for the resume procedure of non-SDT purpose. |
| NEC |  | This is also being discussed in email discussion #925. In general, we support to extent the value of the timer, and a new timer than T319 is better to avoid impact to legacy UE. |
| Nokia |  | Since due to the subsequent SDT data the SDT procedure length is arbitrary, it needs to be discussed if the current behavior is suitable. |
| Samsung | Yes ( also being discussed in email discussion #925) | Timer value of T 319 is separately configured for SDT and non SDT. |
| Qualcomm | Yes |  |

### 2.2.5 Other

For comprehensive understanding, there may be outstanding issues that may need input from other groups, or additional issues in this discussion that may need attention. Companies are invited to add those here. Please limit the input to issues that have been discussed or submitted to RAN2 -111e.

**Q5: Other issues relevant to this discussion?**

|  |  |
| --- | --- |
| **Company** | **Comment** |
| ZTE | In general, the details of this discussion can be fleshed-out in RAN3. In our view, the following aspects will need to be informed to RAN3:   * RLC bearer context needs to be transferred for SDT (both in case of anchor relocation and no-anchor relocation) * PDCP context needs to be transferred for anchor relocation case – same as legacy, but may not be needed for no-anchor relocation case (assuming PDCP layer is terminated in the anchor gNB in this case) * Target gNB to include some assistance information to source gNB to enable decision on whether to initiate SDT and whether to perform anchor relocation or not (RAN2 to decide the actual assistance information needed – for now we think BSR is enough) |
| Huawei, HiSilicon | We agree with ZTE that the details should be worked out in RAN3 (e.g. exact XnAP messages to be used etc.). On top of what ZTE indicated, we should also inform RAN3 about RAN2 preference about when the data should be sent from the receiving gNB (i.e. in parallel to context retrieval procedure/request or only after it is finalized). |
|  |  |
|  |  |

# Summary and Conclusion

TBD