3GPP TSG-RAN WG2 Meeting #116-e DRAFT\_ R2-2111512

Online, 1 – 12 November 2021

**Agenda item: 8.14.2.2**

**Source: Huawei, HiSilicon**

**Title: [AT116-e][043][eQOE] QoE report handling at QoE pause (Huawei)**

**WID/SID: NR\_QoE-Core - Release 17**

**Document for: Discussion and Decision**

# 1 Introduction

This document captures the input and report for the following offline discussion:

* [AT116-e][043][eQOE] QoE report handling at QoE pause (Huawei)

      Scope: Reply to SA4s questions

      Intended outcome: Report, TP for LS out.

      Deadline: Tuesday W2 (CB online only if not possible to agree offline)

The following questions were asked by SA4 in their LS in [1]:

|  |
| --- |
| In light of the above issue, and before SA4 is able to decide on our preference among the three options described in your LS, SA4 kindly asks RAN2 to respond to the following questions:   1. What is the expected typical duration of a temporary stop – e.g., in the order of minutes or perhaps much longer, say hours? As per-session QoE reports are typically sent relatively seldom (at the end of each session or say every few minutes for longer sessions), we would expect that a temporary stop lasting about half an hour should not require additional AS layer storage beyond the supported buffer size limitation, e.g., 64 kB as indicated for Option 2. 2. In case a temporary stop can last for a very long time (e.g., hours), are there any mechanisms already defined or being considered at the RAN side to ensure that subsequent resumption of delivery of potentially a large volume of buffered QoE reports, upon recovery from RAN overload, will not trigger RAN overload recurrence? 3. Will pausing of QoE reporting during RAN overload effectively help the RAN, given that the average QoE load per application is <100 bits/sec? |

## 1.1 Companies contact details

|  |  |
| --- | --- |
| **Company** | **Contact details (name, e-mail)** |
| vivo | panxiang@vivo.com |
| Apple | pnuggehalli@apple.com |
| Qualcomm | jianhua@qti.qualcomm.com |
| LGE | SangWon Kim, sangwon7.kim@lge.com |
| ZTE | Liu.yansheng@zte.com.cn |
| Ericsson | cecilia.eklof@ericsson.com |
| Nokia | malgorzata.tomala@nokia.com |
| Lenovo | hchoi5@lenovo.com |
| CATT | nichunlin@catt.cn |
| OPPO | liuyangbj@oppo.com |
| Samsung | s90.jeong@samsung.com |
| Intel | Ziyi.li@intel.com |
| ITRI | tjtsai@itri.org.tw |
| CMCC | hanxingyu@chinamobile.com |
| China Unicom | gaos30@chinaunicom.cn |

# 2 Discussion

## 2.1 Question 1

Firstly, SA4 asks the following question [1]:

|  |
| --- |
| 1. What is the expected typical duration of a temporary stop – e.g., in the order of minutes or perhaps much longer, say hours? As per-session QoE reports are typically sent relatively seldom (at the end of each session or say every few minutes for longer sessions), we would expect that a temporary stop lasting about half an hour should not require additional AS layer storage beyond the supported buffer size limitation, e.g., 64 kB as indicated for Option 2. |

**Summary of companies views from Tdocs:**

[2], [4], [6], [7], [8] indicate it is hard to predict the duration of the overload situation and it may vary from several minutes to an hour or longer, e.g. depending on the cell size, UE density or scenario where the overload happens (e.g. mass event, hot spot areas etc.). [3], [5] indicate that typically overload should not last very long or in case they last long, the network may release the QoE configurations instead of pausing it. [10] indicates that RAN overload may last long, but in such case the network may release some QoE configurations to decrease impact on UE memory.

**Rapporteur’s understanding:**

Most of the companies agree that the duration of the overload situation is hard to predict and it depends on many factors. [3] indicates that RAN overload control timers and parameters are usually in the order of seconds or minutes, so such time should be assumed. However, it should be considered that the timers can be restarted or UAC can be reapplied for each consecutive access attempt in case the overload lasts longer than that. [5] and [10] indicate that in case of longer overload, the network may release QoE configurations, but that seems to go against the SA5 requirement who underlined that QoE reports gathered during an overload are useful.

**Proposed reply:**

|  |
| --- |
| RAN2 would like to indicate the duration of the overload situation may vary from minutes to hours depending on multiple factors such as the cause of overload, area and time where it occurs, cell size, UE density etc. |

**Comments from the companies on the proposed reply to Question 1:**

|  |  |
| --- | --- |
| **Company** | **Comments (agree/disagree, reason, what to add/modify/remove etc.)** |
| vivo | agree |
| Apple | agree |
| Qaulcomm | Agree. |
| LGE | agree |
| ZTE | agree |
| Ericsson | Not agree. The network takes lots of actions at overload, so that the situation improves. Therefore, overload situations can be considered as short. |
| Nokia | Agree, with conclusive addition: “It is not possible to give any guidance about the duration of the Pause, as it depends entirely on the network configuration and load. “ |
| Lenovo | Partly and we agree with Ericsson. The question from SA4 was about “expected typical duration” and we should respond SA4 accordingly. Therefore, we suggest to add a further sentence saying that the expected typical duration is in the range of minutes. |
| CATT | Agree. But we may have an estimation on typical duration for different heavy level overload |
| OPPO | Agree |
| Samsung | Agree |
| Intel | Agree with Lenovo. As SA4 also replied in the LS “the RAN overload event which triggered the temporary stop may be accompanied by poor service quality causing the user to terminate the service and its associated application.” For some cases, we think the typical RAN overload situation can be short, as the services themselves (not only QoE measurement, but also corresponding application service) may be terminated. |
| ITRI | Agree |
| CMCC | Agree with Ericsson. |
| China Unicom | Not agree. We share the same view with Lenovo and Intel, the APP or service will be terminated if RAN overload solutions last for a long time. Thus we think “expected typical duration” can be short, such as seconds or minutes. |

**Summary of Q1:**

10 companies agree with the reply proposed by the rapporteur that the duration of the overload situation may vary from minutes to hours depending on multiple factors such as the cause of overload, area and time where it occurs, cell size, UE density etc. 5 companies believe that the typical duration of an overload will be in the order of minutes, but that goes against the views from other companies in the table above and in the contributions. The rapporteur propose then to have the following answer:

|  |
| --- |
| RAN2 would like to indicate the duration of the overload situation may vary from minutes to hours depending on multiple factors such as the cause of overload, area and time where it occurs, cell size, UE density etc. RAN2 could not get consensus on the expected typical duration of a temporary stop. |

## 2.2 Question 2

Second question from SA4 in [1], is:

|  |
| --- |
| 1. In case a temporary stop can last for a very long time (e.g., hours), are there any mechanisms already defined or being considered at the RAN side to ensure that subsequent resumption of delivery of potentially a large volume of buffered QoE reports, upon recovery from RAN overload, will not trigger RAN overload recurrence? |

**Summary of companies views from Tdocs:**

[2] proposes to introduce a new mechanism where the AS layer informs the APP layer in the event of impending overrun of available PDCP/RLC layer memory by incoming data from the APP layer.

[3], [4], [5], [6], [7], [9] indicate the network already has means to avoid the issue of RAN overload recurrence, e.g. move some of the UEs to non-overloaded cells/frequencies, utilize UAC, release some QoE configurations, assign lower priority to SRB4 where QoE is reported, resume QoE configurations gradually etc..

[8] indicates “there is no such mechanism defined, but RAN2 is discussing how to handle the pausing/resumption gradually”.

[10] proposes to assume that only limited number of reports should be stored to avoid overload recurrence.

[11] proposes that the pause/resume mechanism can act selectively per QoE configuration

**Rapporteur’s understanding:**

[2] seems to address the issue from the UE perspective, which does not seem to be related to the question from SA4 and has not been discussed/agreed in RAN2 so far. [8] indicated the situation from before the start of the meeting and now the partial resume was agreed. Also, it should be noted that the resume could be done per UE as well. In [10], it is indicated that we could limit the number of reports stored, but it seems that thanks to the means mentioned by other companies already address the issue sufficiently. Other than that, the companies seem to agree the network has already sufficient means of ensuring that QoE resume does not cause a RAN overload recurrence.

**Proposed reply:**

|  |
| --- |
| There are already several mechanisms that can be used to prevent triggering RAN overload recurrence due to QoE resume, e.g. the network may:   * move some of the UEs to non-overloaded cells/frequencies or utilize Unified Access Control * release some QoE configurations * assign lower priority to SRB4 where QoE is reported * perform QoE resume gradually, i.e. indicate resume for different UEs or QoE configurations at different time |

**Comments from the companies on the proposed reply to Question 2:**

|  |  |
| --- | --- |
| **Company** | **Comments (agree/disagree, reason, what to add/modify/remove etc.)** |
| vivo | OK for bullet 1,3,4.  For bullet 2， We prefer to confirm that RAN can release QoE configuration autonomously when RAN overload. SA5 emphasized that QoE reports are useful for the operators and therefore where possible the QoE reports shouldn’t be discarded during a pause. The principle seems to apply to all the related behavior during RAN overload, i.e., the QoE configuration should not be discarded by RAN autonomously during RAN overload. |
| Apple | Agree with Vivo; we can keep bullets 1,3, and 4, and drop bullet 2. |
| Qualcomm | Generally ok. But we want to clarify about addressing the issue from UE perspective. In [2], we are not intended to do any optimization, but want to mention anyway there should be flow control between upper layer and AS layer by UE implementation, just like today’s DRB traffic handling. Otherwise the application data will be lost if Layer-2 buffering is full. |
| LGE | We can also mention the UE based solution, e.g. limited memory size allocated to the QoE logging and the timer based discard as in MDT logging. |
| ZTE | We are fine for all bullets.  For bullet 2, we think this is an optional behaviour which can be used when overload occurs.  In addition, i think i need to further explain our understanding about the number of QoE reports during paused period.  In section 16.4 Metrics Reporting, TS 26.114(MTSI), SA WGs describe that :  *In order to bound the resources used by metrics reporting, the* ***minimum values*** *for the* ***Measure-Resolution*** *and* ***Sending-Rate*** *are specified to be* ***5 seconds*** *and* ***30 seconds*** *respectively.*  The above content means that, based on SA WGs definition, the most frequently QMC scanning rate is 1time per every 5 seconds and the most frequently QMC reporting rate is 1 time per every 30 seconds. It is clear that QoE reporting is not a high frequency behaviour. Hence, a super large buffer is not necessary. |
| Ericsson | We are fine with all bullets. |
| Nokia | We support bullet 1 and 2, as the latter is valid option in RAN for handling overload.  We are not sure about statement “assign lower priority to SRB4 where QoE is reported”. SRB4 is used by default (so far the only) option as reporting, so in case overload or no, QoE reports are passed on SRB4 and by default in RAN it is SRB with assigned lower priority. |
| Lenovo | All bullet points are fine with us. |
| CATT | Agree bullet 1,3,4. For bullet 2, I have same view as vivo. We cannot release the configuration due to overload. Otherwise we don’t need the suspend/resume method. |
| OPPO | Agree bullet 1,3,4 also. In case of RAN overload, we should rather use pause means than release the configuration. |
| Samsung | We are okay with all the bullets, and want to note that they are just the options gNB can choose (i.e., gNB is not required to follow all of the options.). Regarding bullet 2, RAN2 already agreed:  *From RAN2 point of view, the UE shall follow gNB commands and, NG-RAN can in principle release by RRC the application layer measurement configuration towards the UE at any time, e.g. if required due to load or other reasons (Note that other WGs are responsible to define the normal system procedures for release and which nodes are responsible etc).*  Therefore, we are fine with bullet 2. |
| Intel | Agree with all. |
| ITRI | All bullets are fine with us. For bullet 3, we have sympathy with Nokia’s opinion that SRB4 is assigned with lower priority by default for QoE reporting, and we think it needs more clarification。 |
| CMCC | Agree 1,3,4. And for bullet 2, our understanding is QoE configuration release is a result, not a method. If the server and client want to keep the configuration, then it is not proper for RAN to release it liberally. |
| China Unicom | Agree with bullet 1,3,4. |

**Summary of Q2:**

All companies agree with bullets 1, 3 and 4. For bullet 3, it was noted that the wording can be improved as it is not yet clear whether the priority of SRB4 will be hard-coded in specifications or configurable.

For bullet 2, some companies raised a concern, but as indicated by other companies, it was already that the network may release a QoE configuration at any time, e.g. in case it needs to do so to avoid overload.

It is then proposed to keep all the bullets with the following modifications to cater for the comments:

|  |
| --- |
| There are already several mechanisms that can be used to prevent triggering RAN overload recurrence due to QoE resume, e.g.:   * the network may move some of the UEs to non-overloaded cells/frequencies or utilize Unified Access Control * in some cases, the network may decide to release some QoE configurations to avoid/address overload * QoE reports will be provided over the radio bearer with a lower priority (i.e. SRB4) QoE resume can be performed gradually, i.e. the network may indicate resume for different UEs or QoE configurations at different time |

## 2.3 Question 3

Third and last question from SA4 in [1], is:

|  |
| --- |
| 1. Will pausing of QoE reporting during RAN overload effectively help the RAN, given that the average QoE load per application is <100 bits/sec? |

**Summary of companies views from Tdocs:**

[2], [3], [4], [7] indicate that due to a possibility of a UE having multiple QoE sessions, the load may be bigger than what SA4 indicated.

[6] indicates that another intention of pause mechanism is to ensure that the report is not discarded during overload, but that it is stored and sent out after the overload situation passes. [4] also indicates the problem may be not with the average throughput, but with the peak throughput when sending large reports.

[5] indicates that pausing of QoE measurements reporting is one of the functionalities that a RAN overload protection application may choose to address overload.

[4], [8], [9], [10] are admitting that the load will be insignificant and suggesting that the usefulness of pause mechanism is limited based on this and it should be reconsidered whether to support it.

**Rapporteur’s understanding:**

The companies seem to have mixed feelings about whether pause/resume mechanism is needed, at least in Rel-17. It should be noted that even with many QoE configurations configured at the UE, most likely no more than a few QoE measurement sessions will be running at the same time at the UE, which will still generate very little load. Based on this, such mechanism may be seen as low priority for Rel-17. On the other hand, the support for this mechanism is currently included in the WI scope and it is not for RAN2 to decide whether to remove it, so this would require RAN plenary discussions.

**Proposed reply:**

|  |
| --- |
| RAN2 would like to indicate there can be multiple applications running at the UE and generating QoE reports simultaneously, so the load generated by QoE may be larger than what SA4 indicated. On the other hand, RAN2 admits the generated traffic would still be low, hence pausing of the reports may not help RAN after all. Since the pause/resume mechanism is currently in the NR QoE WI scope, RAN2 plans to continue to work on it, unless another decision is made by RAN plenary. |

**Comments from the companies on the proposed reply to Question 3:**

|  |  |
| --- | --- |
| **Company** | **Comments (agree/disagree, reason, what to add/modify/remove etc.)** |
| vivo | agree |
| Apple | Agree with the answer but we can just answer SA4’s question without providing additional colour. So we suggest dropping the last sentence. And instead of saying “admit” (which implies it is our fault!), we can say “believes” (in the second sentence). |
| Qualcomm | From RAN2 point of view, we cannot predict or assume whether the generated traffic could be low or large, which depends on the overload duration and the applied service type. And in case of low QoE traffic, the existing L2 (PDCP and RLC) buffering mechanism can be used for lower priority traffic scheduling. So we prefer to change the part of reply to:  “RAN2 cannot assume the generated traffic could be low or large, which depends on the overload duration, number of QoE sessions, service type. In case of low QoE traffic, pausing the reports may not help after all and not necessary. The existing Layer-2 buffering mechanism can be used for lower priority bearer data scheduling.” |
| ZTE | Though we do not think this is an appropriate time for RAN2 to further discuss the pause/resume QoE reporting function since SA4 and SA5 have different views on this function. We are fine to leave the final decision by RAN plenary. |
| Ericsson | Agree with the comments by Apple. Could also add “..may not help RAN so much after all”. |
| Nokia | agree |
| Lenovo | In the online discussion we already agreed to support selective pause/resume so the last sentence can be removed. In any case the use of pause/resume functionality is left to NW and its usefulness was not questioned by SA4/SA5. Therefore, we suggest the following reply:  “RAN2 would like to indicate there can be multiple applications running at the UE and generating QoE reports simultaneously, so the load generated by QoE may be larger than what SA4 indicated. Therefore, RAN2 thinks pausing of QoE reporting can effectively help the RAN to mitigate RAN overload.” |
| CATT | Agree. We should emphasis the peak load of the QoE report at some time point since some application session will stop in many UEs at same time. For example, in gym, for real time video application, the application of all the UE will stop at almost same time and the RAN also in overload status due to many UE is online |
| OPPO | We insist on the opinion that QoE pausing could help the network for alleviating the network overload situation. Lenovo’s reply is preferred. |
| Samsung | Agree |
| Intel | Agree with Apple’s revision. We also think last sentence can be dropped and RAN2/RAN can further send LS to SA4 and SA5 when decision is made. |
| ITRI | Agree |
| CMCC | Agree with Lenovo’s version. |
| China Unicom | Agree with Ericsson. |

**Summary of Q3:**

10 companies agree with the neutral reply suggested by the rapporteur while 4 companies would prefer to clearly indicate the usefulness of this mechanism to SA4. Since the majority prefers a more balanced reply, it is proposed to adopt the version suggested by the rapporteur with modifications to cater for some comments which seemed agreeable, i.e.:

|  |
| --- |
| RAN2 would like to indicate there can be multiple applications running at the UE and generating QoE reports simultaneously, so the load generated by QoE may be larger than what SA4 indicated. On the other hand, RAN2 believes the generated traffic would still be low, hence pausing of the reports may not help RAN so much after all. |

# 3 Conclusions

Based on the discussion, RAN2 is requested to agree on the reply LS proposed in [12].

# References

1. S4-211290, LS Reply on QoE report handling at QoE pause, Source: SA4
2. R2-2109567 QoE pause and resume handling Qualcomm Incorporated discussion
3. R2-2109833 Further discussion on QoE report handling at QoE pause Lenovo, Motorola Mobility
4. R2-2110608 Discussion on SA4/SA5 reply for QoE pause Huawei, HiSilicon
5. R2-2109868 Pause and resume of QoE measurements Ericsson
6. R2-2109985 Discussion on start and stop of QoE measurement vivo
7. R2-2110101 Discussion on QoE measurement pausing and resuming OPPO
8. R2-2110721 QoE stop and pause Nokia, Nokia Shanghai Bell
9. R2-2110990 Discussion on buffer for NR QoE start and stop ZTE Corporation, Sanechips
10. R2-2109662 QoE measurement configuration and general aspects Intel Corporation
11. R2-2110996 Discussion on QoE collection start and stop CATT
12. R2-2111513 Further reply on QoE report handling at QoE pause

# Annex – Draft reply LS to SA4

**3GPP TSG RAN2 Meeting #116-eR2-210xxxx**

**Online, 1 – 12 November 2021**

**Title:** Further reply on QoE report handling at QoE pause

**Response to:** S4-211290

**Release:** Rel-17

**Work Item:** NR\_QoE-Core

**Source:** RAN2

**To:** SA4

**Cc:** SA3, SA5

**Contact Person:**

**Name:** Dawid Koziol

**E-mail:** [dawid.koziol@huawei.com](mailto:dawid.koziol@huawei.com)

**Send any reply LS to: 3GPP Liaisons Coordinator,** [**mailto:3GPPLiaison@etsi.org**](mailto:3GPPLiaison@etsi.org)

**1. Overall Description:**

RAN2 thanks SA4 for their LS in S4-211290 and for the replies to RAN2 questions contained therein.

For the following issue raised by SA4 in the reply LS,

“*However, SA4 wishes to point out that fully-reliable resumption of QoE reporting by the application layer, upon receiving a restart directive, may not always be possible in the current Rel-17 QoE architecture. For example, the application layer entity responsible for the reporting may no longer be running at the time of the indicated restart, since the RAN overload event which triggered the temporary stop may be accompanied by poor service quality causing the user to terminate the service and its associated application. Possibly, SA4 is able to address this in the future release.”*

RAN2 think it can be addressed by RAN implementation. In the event that RAN overload condition is associated with minor deterioration in RAN capacity i.e. in a state whereby the RAN can guarantee the QoS of the service traffic with best effort but will pause QoE reporting. In the event that RAN overload is associated with severe deterioration in RAN capacity, for which the RAN cannot guarantee the QoS of the service traffic and predicts that the service may be stopped by the user or by the RAN node (RAN can initiate service release in case of RAN overload), the RAN node should resume QoE report to avoid QoE data loss. This can be implemented by RAN implementation.

RAN2 discussed the additional questions posed by SA4 and would like to provide the following replies:

* **SA4 question 1:** “What is the expected typical duration of a temporary stop – e.g., in the order of minutes or perhaps much longer, say hours? As per-session QoE reports are typically sent relatively seldom (at the end of each session or say every few minutes for longer sessions), we would expect that a temporary stop lasting about half an hour should not require additional AS layer storage beyond the supported buffer size limitation, e.g., 64 kB as indicated for Option 2.”

**RAN2 reply to Q1:** RAN2 would like to indicate the duration of the overload situation is hard to predict as it may vary depending on its cause. However, overload situations lasting more than 30 minutes are definitely feasible. Moreover, they may often happen in situations where QoE collection is especially useful, e.g. during mass events such as concerts, football matches, conferences etc.

* **SA4 question 2:** “In case a temporary stop can last for a very long time (e.g., hours), are there any mechanisms already defined or being considered at the RAN side to ensure that subsequent resumption of delivery of potentially a large volume of buffered QoE reports, upon recovery from RAN overload, will not trigger RAN overload recurrence?”

**RAN2 reply to Q2:** There are already several mechanisms that can be used to prevent an overload to be caused by the surge of QoE reports upon QoE resume, e.g.:

* + QoE reports are sent over SRB4 which can be set to priority lower than any user data and will not impact user experience and data QoS in that situations.
  + QoE pause can be sent by the network gradually to different UEs, i.e. it does not have to be sent to all UEs at the same time.
  + If needed, QoE pause can be also sent per QoE configuration of the UE, so that QoE reports can be resumed gradually per service and not for all services of the UE at the same time.
* **SA4 question 3:** “Will pausing of QoE reporting during RAN overload effectively help the RAN, given that the average QoE load per application is <100 bits/sec?”

**RAN2 reply to Q3:** On one hand, considering the average throughput generated by QoE, the already available mechanisms (i.e. putting QoE reports on lower priority SRB4, having a possibility to release a QoE configuration if needed) can be sufficient and QoE pause mechanism is rather an optimization. On the other hand, the problem may be not with the average throughput, but rather a peak throughput required at the time of sending large QoE reports. Furthermore, there can be multiple applications running at the UE and generating QoE reports simultaneously, leading to further increase of traffic caused by QoE.

**2. Actions:**

**To SA4 group.**

**ACTION:** RAN2 respectfully asks SA4 to take the above information into account for their further discussions and to provide further feedback on QoE pause/resume mechanism.

**3. Dates of next TSG-RAN WG2 meetings:**

RAN2#116-bis-e 17 – 25 January 2022 Online

RAN2#117-e 21 February – 3 March 2022 Online