3GPP TSG-RAN WG2 Meeting #128 R2-240xxxx

Orlando, USA, 18 – 22 November 2024

**Agenda item: 8.1.3**

**Source: Nokia (Rapporteur)**

**Title: [POST128][019][AI PHY] NW side data collection (Nokia)**

**WID/SID: WI\_CODE - Release XX**

**Document for: Discussion and Decision**

# 1 Introduction

This document is the report of the following email discussion:

* [POST128][019][AI PHY] NW side data collection (Nokia)

Intended outcome: Discuss the motivation and specification complexity for the three radio conditions.

Deadline: Long

# 2 Contact Points

Respondents to the email discussion are kindly asked to fill in the following table.

|  |  |  |
| --- | --- | --- |
| Company | Name | Email Address |
| Nokia (Rapporteur) | Jerediah Fevold | jerediah.fevold@nokia.com |
| ZTE | Fei Dong | dong.fei@zte.com.cn |
| Apple | Peng Cheng | pcheng24@apple.com |
| OPPO | Jiangsheng Fan | fanjiangsheng@oppo.com |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# 3 Discussion

An email discussion was triggered to discuss the following event-based logging mechanisms.

🡺 Focus on the following three radio condition event-based logging

1. L3 serving cell measurement based (e.g. X1/X2 similar to A1/A2)

2. Beam based events (e.g. beam becomes top-1 beam and number of measurements is less than configured value)

3. L1 beam level measurement

Additionally, companies discussed how long to log after an event has been triggered, e.g., log a configured number of samples after an event triggers, log a configured number of samples per beam, or log periodically after an event has been triggered.

The purpose of AI/ML data collection is to develop one or more datasets which capture a representation of scenarios, e.g., radio conditions, changes in the best beam, and different locations in the cell, which a UE might encounter as it traverses through the network. To capture a complete dataset, especially considering rarer scenarios, event-triggered logging could aid in identifying useful measurement logging occasions and reduce overhead of transmission of redundant samples. Therefore, for each event-triggered logging initiation event and event-triggered logging termination event, the mechanism should be evaluated against that goal.

The questions that follow are general in nature to promote discussion about the benefit of each event to the resulting dataset. The evaluations could be, for example, in terms of the following: redundant sample reduction and data collection coverage, e.g., covering unique measurement scenarios. For each event type discussed, it is asked whether the event helps accomplish the goal of data collection and what deficiencies it has. In the end, a combination of events would be considered holistically to perform together to build complete datasets.

## 3.1 Motivation for Event Triggers

We find it useful first to discuss the motivations and goals of event-based measurement logging. Coming to a common understanding will help us determine the set of events that will help collect a complete and representative dataset for training AI/ML beam management models. The events discussed in this email discussion could be evaluated against our common understanding of the criteria to help identify any gaps.

**Question 0**: Beyond collecting a complete and representative dataset for training AI/ML beam management models, is there any other motivation and goals for event-based measurement logging?

|  |  |
| --- | --- |
| Answers to Question 0 | |
| Company | Technical Arguments |
| ZTE | From NW perspective, the basic motivation of event triggered measurement is to save the air interface resources of sending collected data by avoiding collecting the useless data. |
| Apple | We identify benefit to both NW side and UE side:   * NW-side   1. Allow NW flexibility to only collect data in interested / useful area.      + For example, NW may have to fall back to non-AI beam management when the UE’s radio condition is poor (e.g. located in cell edge) because beam predication performance is downgraded in low SINR region. In this case, it is useless to collect data when UE’s radio condition is poor.   2. Allow NW flexibility to collect non-uniform data under different interested or issued area.      + The distribution of UE in one cell is generally non-uniformed. According to simulation result provided in R1-2405163, 50% of the cells host 80% of the UE population in 18% (or less portion) of their spatial coverage area. Thus, NW may need to collect more data in region with higher distributed UEs while less data in region with less distributed UEs.   3. Allow NW flexibility to offload data collection to multiple UEs for better model generalization (by introducing more UE diversity).      + For example, NW may configure different event threshold to different UEs, so that different UEs collects data in different SINR region. It is helpful for model generalization by introducing more UE diversity (i.e. avoid model being impacted by some specific UEs with too good/bad performance). * UE-side   1. Benefit for UE power saving by avoiding collecting/reporting useless data to NW (e.g. L1 measurements under poor radio condition).   2. Benefit to reduce probability of AS buffer becoming full by avoiding logging useless data.      + Note that if following current AS buffer mechanism of logged MDT/QoE, the NW only know UE’s supported minimum AS buffer size (64KB) rather than UE’s actual AS buffer size. Thus, NW can’t accurately estimate when UE’s AS buffer is full.     Meanwhile, we understand that event triggered logging should be one optional feature on top of baseline of periodical logging, i.e. NW should have flexibility to disable it (i.e. rely on periodical logging if NW is interested in data of all possible area). |
| OPPO | Based on RAN1 simulation observation, beam prediction performance will degrade when the beam RSRP is rather low, but this still can not motivate to only collect high RSRP samples. On the one hand, the network can do the training dataset filtering from the complete samples across one cell; on the other hand, beam management functionality is applicable to all conditions a UE encounters within the serving cell. Considering dataset under all conditions can help NW to train a well generalized model, which matches better for the real deployment.  More addition, it’s not an easy thing for NW to set a threshold for the data logging event once for all as the environment is diverse across different cells, the threshold requirement can be quite different for different cell which is complicated for NW implementation. Even if for a single cell, the physical environment can be different for different directions. One threshold is usually not enough, which means NW may trigger another data collection with different threshold. In this sense, it seems event-based logging saving some overhead for a specific UE during the first round, but second or third round data collection even cause more overhead from system perspective no matter which UE is requested.  Based on above, we prefer to check the motivations with RAN1 first as BM use cases is RAN1 centric, all training assumptions are under RAN1 scope. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Summary 0**: TBD.

**Proposal 0**: TBD.

## 3.2 Measurement Event Triggers

### L3 serving cell measurement-based events

R2-2409945 (Apple) proposed the following:

Proposal 8: Support the following two radio condition based event triggered logging:

• Event X1: when L3 serving cell measurement becomes better than absolute threshold (similar to A1).

• Event X2: when L3 serving cell measurement becomes worse than absolute threshold (similar to A2).

**Question 1**: Do L3 serving cell level measurement events help accomplish the goal of data collection for training network-side beam management AI/ML models? If not, state any deficiencies or ways to augment the event to mitigate them.

|  |  |  |
| --- | --- | --- |
| Answers to Question 1 | | |
| Company | Yes/No | Technical Arguments |
| ZTE | Yes with comments  (The event can be further considered) | According to our understanding about the motivation of event triggered measurement ,we think event by using the L3 serving cell measurement result is valid.  Normally, when the L3 measurement for a serving cell is worse than a threshold which means the UE have reached at the edge of the cell and handover is expected to be happened, assuming UE at the edge but the handover is not happened in time, the RLF will be triggered even worse, the sample collected under this scenario is no longer valid or helpful for the NW side model training.  In this sense, it is reasonable that UE performs collecting the data when L3 serving cell measurement is better than an absolute threshold.  Then the event X can be modified as below:  The data logging is performed only when L3 serving cell measurement ~~become~~ is better than an absolute threshold, |
| Apple | Yes (Proponent) | The main intention is to allow UE to collect data only in NW interested/intended area, which is typically identified with specific radio conditions (i.e. RSRP/SINR range). Meanwhile the L3 event is more stable, less UE complexity with cell specific time-to-trigger (TTT) and manageable spec work.    In addition, we want to clarify below 2 aspects:   * On the issue of possible mismatch between L3 and L1 measurement raised in online discussion of RAN2#128   + Following our motivation in Question 0, event triggered logging is mainly used to allow UE to only log data in a subset of NW interested area (e.g. in good radio condition). When the trigger condition is met, the UE needs to log L1 measurements of all NW configured beams, i.e. the NW can still get a complete L1 measurement dataset for training in its interested area. Thus, L3 measurement triggered event is sufficient.   + L3 measurement is derived from multiple L1 measurements with NW configured RRC parameters. Thus, if any concern on mismatch, NW can adjust these RRC parameters (e.g. *nrofCSI-RS-ResourcesToAverage, absThreshCSI-RS-Consolidation, L3 filter coefficient*) to mitigate mismatch. * On the wording suggested by ZTE, we are fine with the change. But we want to clarify the wording is actually copied from same wording used in event A1/A2 in section 5.5.4/6.3.2 of TS 38.331:   “Event A1: Serving becomes better than absolute threshold;  Event A2: Serving becomes worse than absolute threshold;” |
| OPPO | Yes with comments | We have the similar feeling that A1-like event is more suitable considering beam prediction performance is usually better when beam RSRP is good enough. |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Summary 1**: TBD.

**Proposal 1**: TBD.

**Question 2**: What is the specification impact of implementing logging based on L3 serving cell measurement events?

|  |  |
| --- | --- |
| Answers to Question 2 | |
| Company | Technical Arguments |
| ZTE | Design the L3 serving cell measurement events for logging the data. No other specification impact can be foreseen. |
| Apple | We foresee limited spec work:   * Most existing procedure text related to logged MDT event L1 in Section 5.5a.3.2 (Measurements logging) of TS 38.331 can be reused. * We assume that ASN.1 of existing L1 event can also be reused (i.e. need to introduce trigger threshold, hysteresis and TTT):   eventL1 SEQUENCE {  l1-Threshold MeasTriggerQuantity,  hysteresis Hysteresis,  timeToTrigger TimeToTrigger  }, |
| OPPO | Similar view with Apple |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Summary 2**: TBD.

**Proposal 2**: TBD.

### L1 Beam-based Events

R2-2409908 (Qualcomm) proposed the following:

Proposal 4: RAN2 is requested to consider at least the following events for training data collection for network-side model training,

• Event 1. Based on the number of samples to be collected and reported across different beams, UE triggers the measurement collection and logging if a beam becomes the top-1 beam and the logged number of measurements is less than the configured value.

• Event 2. Based on the change of the top-1 beam: UE is configured to log the measurement when the top-1 beam changes. UE can additionally be configured with the number of samples to be logged and its periodicity

These events are triggered when the top-1 beam changes, i.e., “if a beam becomes the top-1 beam” or “based on the change of the top-1” beam. The key difference between these events is the termination condition. These termination conditions will be discussed in another question.

**Question 3**: Does triggering logging based on the change of the top-1 beam help accomplish the goal of data collection for training network-side beam management AI/ML models? If not, state any deficiencies or ways to augment the event to mitigate them.

|  |  |  |
| --- | --- | --- |
| Answers to Question 3 | | |
| Company | Yes/No | Technical Arguments |
| ZTE | Maybe No | My understanding both events is to prevent UE from continuously collecting the data when the mobility status of the UE is static in order for power saving. However, in the real deployment, the beam measurement result is still fluctuated even though UE mobility status is static, from NW perspective, those data is still useful for NW to train the model regardless of the top-1 beam have been changed or not. |
| Apple | No | We think it has below issues:   1. Only considering Top-1 beam is too restricted in FR2. As we know, the L1 beam change in FR2 is quick and may be caused by channel time fading and/or blockage by other object(s). Thus, top-1 beam measurement is not stable. And that is why existing beam management procedure needs UE to report top-K beams and NW can use TCI to dynamically change beam. So, only considering top-1 beam may lead to misleading information to NW. 2. As mentioned by ZTE, the proposed event is intended to prevent UE from continuously collecting the data when the UE is static (i.e. when top-1 beam is not changed). However, the L1 beam quality is generally fast time-changing in FR2. It is possible that although top-1 beam is same during one period but its measurement has changed a lot, which means some useful data are missed to collect. Correspondingly, at least AI/ML based BM case 2 (i.e. temporal prediction) will not work. 3. The proposed event has no RSRP threshold which means that the UE may log useless data when its radio condition is poor (e.g. right before RLF). 4. It will bring high UE complexity. As discussed in L1 measurement event for Rel-19 LTM and MIMO, time to trigger (TTT) is necessary to make event trigger stable. Then, it means UE needs to simultaneously maintain multiple TTT for multiple beams (as top-1 beam may change, UE has to track multiple candidate beams to determine when any of them becomes top-1 beam). |
| OPPO | Maybe No | We understand triggering logging based on the change of the top-1 beam is too strict condition as many valuable samples will be dropped, we doubt the dataset quality collected under such condition. For BM, even if the top 1 beam has not changed, it’s still a good input for model training. |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Summary 3**: TBD.

**Proposal 3**: TBD.

**Question 4**: What is the specification impact of implementing events based on a change in the top-1 beam?

|  |  |
| --- | --- |
| Answers to Question 4 | |
| Company | Technical Arguments |
| Apple | We foresee a lot of specification work. Below are just some examples:   1. Whether to capture the proposed event in RAN1 spec (i.e. via timer + counter similar to BFD) or RAN2 spec (i.e. via TTT)? Please note that similar discussion happened in Rel-19 LTM and MIMO, and it was controversial. LTM took 3 meeting to make decision. 2. If we capture it in RAN2 spec (via TTT), RAN2 need to discuss the UE behaviour on when to reset / continue TTT of different beams because different beams’ TTT may not start / stop at the same time. For example, if 1st strongest beam has met its TTT but 2nd strongest beam is in middle of its TTT, whether the UE can regard top-1 beam has changed, or the UE needs to wait 2nd strongest beam meets its TTT? 3. In current spec, filtering of L1 measurement is up to UE implementation. However, if we specify L1 measurement triggered logging, it is not clear whether the filtering can still be up to UE implementation, which may need RAN1 input. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Summary 4**: TBD.

**Proposal 4**: TBD.

### L1 beam level measurement

From the discussion, ZTE proposed to consider “that both L3 and L1 beam measurements can be useful.”

Given the broad nature of possible L1 beam level measurement triggers and since L3-based triggers are being discussed in questions 1 and 2, this section can be used to discuss which types of triggers could be useful aside from those proposed to trigger on the top-1 beam changing. Therefore, one open-ended question will be asked to promote that discussion.

**Question 5**: Which types of triggers based on L1 beam level measurements could help accomplish the goal of data collection for training network-side beam management AI/ML models? Add specification impacts for identified options, if any.

|  |  |
| --- | --- |
| Answers to Question 5 | |
| Company | Technical Arguments |
| ZTE | For L1 beam level measurement, the event for triggering data collection can be as below:   * the RSRP value of top 1/K beams of set A are better than a threshold.   In the real deployment, if the top 1/K beam measurement result for the set A is worse than a threshold, then the abnormal case must be occurred, such as the beam failure, the deep fading, or obstacle..Those collected samples are not valid which is not helpful for the NW side model training. All those samples can be filtered out at UE side in order to save the air-interface resources for collected data reporting.  In this sense, we tend to suggest the following event triggered data logging:  *The data logging is performed only when the RSRP value of top 1/K beams of set A is better than a threshold.* |
| Apple | We think the event proposed by ZTE makes more sense by addressing the first 3 issues which we raised in Question 3.  However, we still have concern on UE complexity (issue 4 we mentioned in Question 3) and non-trivial specification work (similar spec work are required as we mentioned in Question 4 on introducing new L1 measurement trigger event). |
| OPPO | We have the similar feeling as Apple, the event is just a trigger for data logging, not the collected data content, L1 event may touch too much UE implementation aspect which may need to consult RAN4. If event-based logging is justified, L3 based event is more stable and easier for UE implementation. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Summary 5**: TBD.

**Proposal 5**: TBD.

## 3.3 Measurement Logging Behaviours

### Logging periodically after a data collection event trigger

R2-2409547 (OPPO) proposed the following:

Proposal 5: During the period that radio condition-based logging event fulfills, UE performs data logging periodically.

Proposal 6: For periodic logging or event-triggered periodic logging, data logging interval is configured by the network, the value range of data logging interval is pending on RAN1 inputs.

**Question 6**: Should periodical logging after a measurement event is triggered be supported?

|  |  |  |
| --- | --- | --- |
| Answers to Question 6 | | |
| Company | Yes/No | Technical Arguments |
| ZTE | Yes |  |
| Apple | Yes | In our understanding, L3 measurement event is intended to allow UE to collect data only in NW interested area. And when it is in NW interested area, the UE still needs to periodically log data according to NW configured interval to provide time-changing beam information, which is useful at least for AI/ML based BM case-2 (i.e. temporal predication). |
| OPPO | Yes (Proponent) |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Summary 6**: TBD.

**Proposal 6**: TBD.

Performing periodical measurement logging for a time duration was proposed by OPPO in R2-2409547.

Proposal 5: During the period that radio condition-based logging event fulfills, UE performs data logging periodically.

Proposal 6: For periodic logging or event-triggered periodic logging, data logging interval is configured by the network, the value range of data logging interval is pending on RAN1 inputs.

And performing periodical measurement logging of number of samples was proposed by Qualcomm in R2-2409908.

Proposal 4: RAN2 is requested to consider at least the following events for training data collection for network-side model training,

Event 2. Based on the change of the top-1 beam: UE is configured to log the measurement when the top-1 beam changes. UE can additionally be configured with the number of samples to be logged and its periodicity

**Question 7**: If periodical logging after a data collection event is triggered is supported, which parameters should be configurable to control the amount of data logged (e.g., time-based, sample-based)?

|  |  |  |
| --- | --- | --- |
| Answers to Question 7 | | |
| Company | Time-based,  Sample-based,  Other | Technical Arguments |
| ZTE | Other | The basic logic is that, when the event is met, then the data collection is performed, otherwise, the data collection is suspended. |
| Apple | See comments | Same understanding as ZTE: the baseline is the UE starts periodical logging when entering condition of event is met and the UE stops logging when leaving condition of the event is met.  Timer/sample are further optimization for UE power saving. We are open to discuss but we don’t think it is essential. So, we prefer to postpone the discussion after the baseline mechanism is clear. |
| OPPO | Other | Same understanding as ZTE and Apple |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Summary 7**: TBD.

**Proposal 7**: TBD.

### Event triggering based on number of samples previously collected

In R2-2409908 (Qualcomm) proposed the following:

Proposal 4: RAN2 is requested to consider at least the following events for training data collection for network-side model training,

• Event 1. Based on the number of samples to be collected and reported across different beams, UE triggers the measurement collection and logging if a beam becomes the top-1 beam and the logged number of measurements is less than the configured value.

That is, an event would only trigger a single measurement and could re-trigger up to a configured number of times or samples. For example, an event could trigger based on the top-1 beam changing to a hypothetical beam, A, which would capture a single measurement or a single set of beam measurements in the log. Later, the UE determines that beam C is the best beam, captures a measurement, and subsequently determines that beam A is once again the best beam. If the number of samples captured for a change to beam A is less than the configured value, single measurement or a single set of beam measurements could be captured, otherwise, the event would not be triggered.

**Question 8**: Should triggering an event, one or more times, based on having captured fewer than a configured number of samples based on the event criteria, e.g., the top-1 beam changed, be supported?

|  |  |  |
| --- | --- | --- |
| Answers to Question 8 | | |
| Company | Yes/No | Technical Arguments |
| ZTE | Other | The basic logic is that, when the event is met, then the data collection is performed, otherwise, the data collection is suspended. |
| Apple | No | Although we are open to discuss, we think absolute number of sample threshold is not reasonable. Different UEs may have different L1 beam measurement change / fading trends (e.g. due to UE in different moving speed, whether UE located in main-lobe of NW beam or in side-lobe of the NW beam). Thus, same absolute number of samples of different UEs may bring quite different level of useful information to the NW. |
| OPPO | No | For model training, what really matters is the dataset quality, not the number of samples. Sample-based event may cause UE missing some valuable training samples, in this sense, it’s not a good idea to consider more on overhead saving while leaving the dataset quality behind. |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Summary 8**: TBD.

**Proposal 8**: TBD.

# 4 Conclusion

TBD.