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 |
| vivo | Boubacar Kimba | kimba@vivo.com |
| Samsung | Seung-Beom | s90.jeong@samsung.com |
| Qualcomm | Rajeev Kumar | rkum@qti.qualcomm.com |
| Lenovo | Congchi Zhang/ Tapisha Soni | zhangcc16@lenovo.com, tsoni@lenovo.com |
| Huawei, HiSilicon | Dawid Koziol | dawid.koziol@huawei.com |
| CATT | Tangxun | tangxun@catt.cn |
| Ericsson | Marco Belleschi | marco.belleschi@ericsson.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. |
| vivo | The motivation for event-based measurement logging is for model finetuning/retraining. Specifically, for the initial model training phase, the NW may configure the periodic logging to the UEs at different locations/statuses and perform model training based on the collected data. During the model inference/monitoring phase, our simulation result shows that the model performance is poor when the beam-level input is at a low value and the NW would have the intention to collect more data to perform model finetuning/retraining. In this case, event-based measurement logging is essential to reduce unnecessary data logging, transfer, and post-processing. |
| Samsung | There may be some benefit for reporting overhead reduction, and UE buffer/power consumption saving with event-based logging. However, it has not been fully studied on data processing/categorization at UE side (i.e., event-based logging) i.e., whether it could be beneficial or degrade for model training. To ensure maximum performance, it is definitely better to collect all data by NW and then do pre-processing before training with full flexibility, rather than defining events without sufficient study/verifying the performance. Therefore, any event that may impact on the performance shall be well justified. Besides, we would like to note RAN1 has ongoing proposals and discussion on similar issues (e.g., data omission for NW-side data collection). So, we prefer to leave RAN1 to discuss together. |
| Qualcomm | We agree with ZTE, Apple, and others that the objective of events is to reduce the Uu overhead associated with the training data collection and power saving at the UE. We also agree with the rapporteur that the collected measurements should form a representative sample of RSRP in the beams of a cells. We want to clarify that a representative samples means that the data is uniformly collected across different locations in beams of a cell, which is necessary for data collection for training. However, this may not be achieved if the events are not defined properly (for example, L1/L3 threshold-based data collection, simple randomization, etc.) [1, 2]. For example, with L3 cell-level RSRP threshold, if the UEs are concentrated towards a few beams (SSB/CSI-RS), then the collected data will not form a representative sample. To compensate, the network would have to collect way more data than it needs, because it will need to drop a lot of redundant data points . Therefore, we believe even in this case, the network should be able to configure the UE with indices of SSB/CSI-RS for which training data collection is required.    Note that events defined in the standard currently have clear objectives. For example, the events are defined to determine need for the UE to start reporting measurements such that the network can take handover decisions. In the beam management use case, there can be the following objectives:   1. L1-RSRP prediction for beam management, 2. Beam management (e.g., beam switching) decisions.   Events should be defined (considering objective) such that it encompasses a representative measurements from different locations in the interested area/conditions (e.g., samples from beam centre or beam edge, or both). The events and measurement configuration should be defined to collect a sufficient number of samples in configured conditions / (L1/L3) ranges [to avoid oversampling (towards one or more conditions/labels) or skewing the training data].  [1] Yang, Yuzhe, and Zhi Xu. "Rethinking the value of labels for improving class-imbalanced learning." *Advances in neural information processing systems* 33 (2020): 19290-19301.  [2] Jain, Saachi, Kimia Hamidieh, Kristian Georgiev, Andrew Ilyas, Marzyeh Ghassemi, and Aleksander Madry. "Data Debiasing with Datamodels (D3M): Improving Subgroup Robustness via Data Selection." *arXiv preprint arXiv:2406.16846* (2024). |
| Lenovo | We also believe it is beneficial in the following sense:  (a) NW flexibility to collect data in specific area/by specific UEs/for specific scenario/conditions, and  (b) UE flexibility to collect/log only required dataset. |
| Xiaomi | We understand multiple models for different radio conditions may outperform single model for all radio conditions. NW may need to train multiple models using data collected from different radio conditions. The collected data may not be evenly distributed from radio channel condition point of view. Therefore, NW may only require data under certain conditions for the corresponding model training. These conditions can be described as event. By the event, UE only collects the data which NW is interested in when event is fulfilled. UE can save the power and storage by not collecting when event is not fulfilled.  Furthermore, we understand such event is optional configured by NW. If NW want UE to always collect data, NW doesn’t configure the event to UE. |
| Huawei, HiSilicon | We understand the main motivation behind introducing event-based logging is to limit the signalling overhead and reduce UE’s power consumption. However, whether this can in fact be achieved is very doubtful for several reasons:   1. The gains can only be achieved under the assumption that the network is only interested in data from part of the cell which we find very unlikely. 2. For R19 AI for PHY, only offline training is assumed, so it should be also noted that model training is not a continuous process, is only executed from time to time and requires input only from a subset of UEs in the network, thus the impact on capacity and UE power consumption will be small. 3. As indicated by OPPO, trying to pre-filter the measurement at the UE may turn out to be counter-productive in the end, e.g. the generalization performance for AI-based BM use cases may be very bad with such data. 4. The network already has a possibility to use existing mobility event reporting to configure the UE with the data collection only in case some radio conditions are met, i.e. the network first configures a certain mobility event and once the UE reports the event to be met, the network can configure data collection. This allows to achieve the benefits on the network flexibility in collecting certain data based on radio conditions and UE power saving on not collecting useless data without such event triggered logging.   Hence in our view both the motivation and benefit over the existing mechanisms is unclear. Before moving forwards with any new solutions, we believe we should at least check the motivation with RAN1, since they have better expertise in beam management and they are already discussing related issues as noted by Samsung. By checking with RAN1 we can at least verify whether what we intend to specify makes sense and can be useful. |
| CATT | Agree with OPPO to check the motivation with RAN1 first since it is mostly a PHY use case.  The latency requirement for data collection of AIML training is relaxed, so the data could be collected from multiple UEs within a long duration, the UE power and UE buffer are not the significant factors to be considered as only small amount of data needs to be collected from one UE. And since the training will not be frequently performed after the first few rounds of model training and model updates, the data filtering could be performed by the NW side based on NW implementation. |
| Ericsson | We agree with all the motivations presented by different companies above (ZTE, Apple, Vivo, QC, Lenovo, Xiaomi). In order for the network to generate model(s) that work(s) efficiently in an area (e.g. cell), it is important the diversity of the data collected so that the different radio conditions that the UEs may experience in the area are represented. By just leveraging on the periodic non-event driven data logging, with high probability the network will get redundant data for certain radio locations, and scarcity of data in other radio locations that instead would be important to further enrich the model. This will result in overhead over the Uu interface, and in UE power consumption because the UE might just collect, for most of the logging duration, radio measurements that the network has already acquired previously.  We would also like to emphasize that the motivations of the event-based data collection have been at length discussed in the past 2/3 RAN2 meetings, and we have already agreed two meetings ago in RAN2#127bis that “Event-triggered data logging will be supported.” and in RAN2#128 we further agreed to “Focus on the following three radio condition event based logging….”. Hence, we suggest Rapporteur mainly focusing on “the motivation and specification complexity for the three radio conditions”, which is the ultimate objective of this email discussion. |

**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. |
| vivo | Yes | For beam management, since the input and output are at the beam level, it’s straightforward that the logging event should also be at the beam level.  However, agree with Apple that the current L3 measurement configuration can provide the flexibility to achieve beam-level triggering, e.g., set the *nrofCSI-RS-ResourcesToAverage* as absent and the cell-level measurement quantity is the highest beam measurement quantity value. |
| Samsung | No | There has been no study on whether L3 serving cell measurement can help or not in RAN 1/2 study item. RAN2 should avoid introducing event without enough study and clear benefit. Otherwise, it just leads to biased training due to insufficient data. Besides, RAN1 has ongoing proposals and discussion on data omission for NW-side data collection. It is better to let RAN1 to carry on the discussion. |
| Qualcomm | Yes (but with further considerations) | We agree with Apple that L3-RSRP cell level measurement events are simpler for the UE implementation, while the UE can log the training data in the interested area. However, as mentioned in response to Q0, we want to highlight that for different objectives, the training data should encompass representative measurements from different locations in the interested area/conditions (e.g., samples from beam centre or beam edge, or both). The events and measurement configuration should be defined to collect a sufficient number of samples in configured conditions / (L1/L3) ranges [to avoid oversampling (towards one or more conditions/labels) or skewing the training data].  Additionally, as indicated in response to Q0, considering L3 cell-level RSRP threshold, if the UEs are concentrated towards a few beams (SSB/CSI-RS), then the collected data will not form a representative sample. To compensate, the network would have to collect way more data than it needs, because it will need to drop a lot of redundant data points. Therefore, we believe even in this case, the network should be able to configure the UE with indices of SSB/CSI-RS for which training data collection is required. |
| Lenovo | Yes | L3 events are more suitable to identify the scenario of interest, e.g., cell centre (A1) or cell edge (A2). |
| Xiaomi | Yes |  |
| Huawei, HiSilicon | No | In our view these events do not help to accomplish the purpose of data collection. This goal can be achieved without introducing any events.  The only thing these events can potentially help with is to limit the signalling overhead and limit the UE power consumption, but this is only under the assumption that the NW is only interested in the results from part of the cell which is unlikely. Furthermore, even this can be also be achieved with the current mobility framework, as we mentioned in our reply to Q0. |
| CATT | Yes with comments | A1-like event (serving becomes better than absolute threshold) seems more useful to trigger the collection of valid data than A2-like event. |
| Ericsson | Yes | In our understanding, the motivation of using the L3 measurements as triggering conditions is to reduce the potential fluctuations of L1-event driven measurements. From this point of view both the A1-like and A2-like approach make sense to us, as baseline, since they capture different radio scenarios. We also agree with QC, that the radio configuration should be at beam level so that the UE is enable to collect the L1 measurements only for those beam that are of interest for the NW-side training. Such details can be discussed during the stage-3. |
|  |  |  |
|  |  |  |

**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 |
| vivo | Similar view with Apple |
| Qualcomm | Similar view with Apple. |
| Lenovo | Similar view with Apple |
| Xiaomi | Design new events, similar as A1 and A3. But TTT may not be needed, since there is no robustness issues for data collection. |
| Huawei, HiSilicon | The impact on specifications depends on the final design of the feature, but what RAN2 needs to specify is at least:   1. Definition of the new measurement events (or agreeing to reuse existing events and clarifying how they are used for the new scenario). 2. Define the UE behaviour, i.e. how long are measurements gathered after triggering, what is the periodicity of logging after triggering, stop condition for logging etc.   We are not sure how we can reuse text in section 5.5a.3.2 which is for the UE in RRC IDLE/INACTIVE and it seems completely not applicable to UEs in RRC\_CONNECTED state which are under consideration for data collection case. Hence, a new procedure/section would rather be required. |
| CATT | Similar view with Apple |
| Ericsson | Since there are already precedents in the spec. both in logged MDT and in the RRM, we can get inspiration from that. Details can be sorted out during the stage-3 discussion, but spec. complexity is expected to be limited. |
|  |  |

**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. |
| vivo | No | The RSRP threshold of Top-1/K beam is sufficient for filtering the data logging. Agree with Apple’s concern on the specific Beam to trigger the data logging. |
| Samsung | No | Q3 deals with the proposal in R2-2409909 (i.e., event for top-1 beam change). What we understand for the proposal is   * Step 1) UE starts logging upon top-1 beam change. * Step 2) UE stops logging if the number of logged data samples reaches the configured threshold.   RAN2 agreed at least L1-RSRP and/or beam ID needs to be collected for SET A/B. However, such event leads to no report for both SET A/B, if there is no change of Top-1 beam. However, it is also important to collect data when Top-1 beam is kept. If the model is trained only with data when Top-1 beam changes, the model cannot work when Top-1 beam is kept. |
| Qualcomm | Yes | The proposal 2 is about the trigger and is not about the number of samples the UE should log once the event is triggered. When a certain beam (for example, a set A beam) becomes the top-1 beam, the UE starts collecting the training data (until the configured number of samples are logged). Of course, L1-RSRP can be considered further (as we also mentioned in our paper).  In ZTE’s response, ZTE mentioned that “both events are to prevent UE from continuously collecting the data when the mobility status of the UE is static in order for power saving.” This is not just about losing the valuable samples but about collecting the data when a certain beam is the Top-1 beam for label diversity.  We want to bring to ZTE’s and other companies’ attention that we did not propose these events for the purpose of “preventing UE from continuously collecting the data when the mobility status of the UE is static in order for power saving.”. We proposed them to avoid oversampling (towards one or more conditions/labels) or skewing the training data [1, 2]. If the training data set is imbalanced, (i.e., there are more measurements when one or more beam is/are Top-1/Top-K beam), the training data may become biased towards that beam / set of beam. This will impact the AI/ML model performance. To compensate for that, the network would have to collect more data than it needs and then drop a large number of redundant data points. Therefore, training data with balance labels are desired.  Furthermore, as seen from the below figure, the network may want UE to perform training data collection on one or more SSB / CSI-RS(when UE is connected to that SSB / CSI-RS). The network may configure the UE with indices of SSB/CSI-RS and L1-RSRP threshold for which training data collection is required.    We proposed considering Top-1 beam for the initial discussion. Of course top-K beam and L1-RSRP for logging the measurement can be considered, but it can be complicated. For example, for top-K beam, we need to specify how the L1-RSRP of these top beams are considered for logging the data (whether UE should consider L1-RSRP of one beam, two beam, or more number of beams? Whether it is average L1-RSRP among Top-K beams?)  The complexity can be added once we develop some understanding. |
| Lenovo | No | Agree with some companies that it could be too restrictive without clear motivation.   * If it is for time domain beam prediction, the samples before a beam becomes the best beam are valuable, which will not be logged with proposed events. * If it is for spatial domain beam prediction, we fail to understand why “best beam change” matters in this case. |
| Xiaomi | No | We believe the model shall not be per beam. Otherwise, there would be too many models. Therefore, the motivation of such event is not valid. |
| Huawei, HiSilicon | No | The situation is similar as for L3 based events, but in this case the results are pre-filtered based on the best beam of the UE instead of the UE being in cell center or cell edge.  Apart from the unclear motivation of this kind of event, we also agree with the issues raised by other companies, i.e. instability of L1 measurements and top-1 beam and complexity for the UE implementation. |
| CATT | No | A beam specific trigger is not that important, as during the logging period all beam measurement results can be recorded and used by network later to train a model. |
| Ericsson | Yes, but prefer approaches in Q1 and Q5. | The approach proposed in this question may have its merits, and we are not against it. However, we believe that an approach based on radio measurement thresholds (as in the previous and next question) may be more useful in practice, because they allow the network to configure radio thresholds such that the UE can specifically collect data just in those radio conditions where the NW model is weaker.  Since one of the objective of data collection is to achieve the diversity of collected data in an area, an event that is based on radio measurement thresholds (that could be mapped to measurement ranges that the training entity is missing) is more beneficial than an event based on which beam is top-1. The objective of reducing the data overhead and UE power consumption can be achieved by the network by setting the radio thresholds to those model operating conditions for which data are missing. |
|  |  |  |
|  |  |  |

**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. |
| Qualcomm | Solutions adopted for event-trigger reporting in MIMO; LTM; AI/ML performance monitoring for BM (if adopted in RAN1) can be reused for this purpose.  Furthermore, we also do not agree with Apple’s 3), because any network algorithm is going to be based on UE reports, whose filtering is left to UE implementation. |
| Huawei, HiSilicon | The specifications impact seems to be similar to that of L3 measurement based events, i.e.   1. Definition of new measurement events (no existing events to be reused) 2. Define the UE behaviour, i.e. how long are measurements gathered after triggering, what is the periodicity of logging after triggering, stop condition for logging etc. |
| Ericsson | Agree with Apple, designing this solution might require more efforts than other solutions based on radio measurement thresholds. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**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. |
| vivo | Based on the intention of model retraining/finetuning for the scenario that the beam-level input is at a low value, we think the RSRP value of top 1/K beams is worse than a threshold should be considered. |
| Samsung | There has been no study on whether L1 beam level measurement can help or not in RAN1/2 study item. RAN2 should avoid introducing event without enough study and clear benefit. Otherwise, it just leads to biased training due to insufficient data. Besides, RAN1 has ongoing proposals and discussion on data omission for NW-side data collection. It is better to let RAN1 to carry on the discussion.  BTW, regarding ZTE’s proposal, we do not think NW-side data collection configuration distinguishes SET A and SET B. i.e., UE does not know whether configured beam(s) is Set A or Set B. UE just measures all configured beams/resources and reports the results. After receiving reports, NW can map each beam to Set A or Set B depending its needs for training. Therefore, we think “set A” cannot be specified in the event evaluated by UE. |
| Qualcomm | As indicated in the below figure, the network may want the UE to perform training data collection within a region on one or more SSB / CSI-RS (when UE is connected to that SSB / CSI-RS). The network may configure the UE with indices of SSB/CSI-RS and L1-RSRP threshold for which training data collection is required.    We propose the following to consider for L1 beam measurements   * Top-1 beam L1-RSRP smaller than the threshold * Top-1 beam L1-RSRP larger than the threshold * Top-K beams and one or more beam L1-RSRP smaller than the threshold * Top-K beams and one or more beam L1-RSRP larger than the threshold   Furthermore, to have the diversity in the training data we should also consider different L1-RSRP bins. For example,   * Top-1/Top-K beam with different RSRP ranges (with different number of samples to be collected) |
| Lenovo | In general, we feel it is difficult to train an AIML model that works if some specific beams have good or poor RSRP. Also, the specification complexity raised by some companies also make sense.  In the example mentioned above,   * the RSRP value of top 1/K beams of set A are better than a threshold.   Using the collected data as filtered by the Set A beam quality, a trained AIML model will be applicable if Set A beams (which will the prediction output of the model) are good enough above a threshold. However, when the trained AIML model is deployed, UE will only measure Set B beams, thus cannot tell if the AIML model is really applicable. In a word, if the data filtering is applied to the output in a training data pair, it will cause difficulty w.r.t applicability determination later on. |
| Xiaomi | We are not sure whether L1 beam measurement based event is useful, which is in RAN1 scope. |
| Huawei, HiSilicon | We see no motivation for event-based data logging in general and compared to L3 event based logging, L1 event based introduces additional complexity with even less benefits due to instability of L1 measurements and top-1 beam. |
| Ericsson | We have in general similar view as QC, Vivo, ZTE and other. The top-1/K beam L1-RSRP above or below a threshold can be used as event for the starting of the data logging.  We are not sure about the complexity mentioned above by some companies. In our view, the spec. complexity should be similar as for the L3 measurements, e.g. it can be created a logging event configuration that the UE should monitor and then the related SSBs/CSI-RSs whose measurement results the UE should start collecting when the event is fulfilled. So the spec complexity should be limited as for the L3 measurements. Also related to the UE complexity, we are not sure that this would bring extra complexity, since the UE just needs to monitor the L1-RSRP of the top 1/K beam with no additional mechanisms on top of it. |
|  |  |

**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) |  |
| vivo | Yes | The data logging will be stopped when UE the buffer is full. |
| Samsung |  | Better to discuss this issue after the motivation of event-based logging is justified enough. |
| Qualcomm | Yes (but with modification) | Periodic reporting/logging (when an event is triggered) is the legacy method for reporting/logging more than one measurement for network decisions. We can have more than one sample logged periodically at the UE when an event is triggered. However, as we mentioned previously, the periodic logging should also not result in oversampling (towards one or more conditions/labels) or skewing the training data (which may degrade AI/ML model performance). Therefore, the network should also configure number of samples to be logged when an event is triggered (similar to L3 existing event-trigger reporting for L3 measurement / immediate MDT). |
| Lenovo | Yes |  |
| Xiaomi | No | We understand the event fulfilment is enough to trigger periodic logging. All the data is useful if event is fulfilled. What’s the motivation of periodical data logging after event is triggered? Less data would be collected due to the periodic gap. |
| Huawei, HiSilicon | Yes | In our view, in the event-based logging (if introduced), the UE should collect data periodically as long as the event is met. |
| CATT | Yes |  |
| Ericsson | Yes, but we should also agree that the logging continues until the event is fulfilled (first part of Proposal 5 in R2-2409547 above) | The periodicity of the logging is dictated by the periodicity of the reference signal in which the UE does the measurement. So the answer to this question is somewhat obvious, i.e. the UE logs the data upon measuring the associated reference signal whose periodicity is given by the configuration.  In our view, what is more important in the Proposal 5 in R2-2409547 is the first part of the proposal, i.e. the logging continues until the event is fulfilled. |
|  |  |  |
|  |  |  |

**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 |
| vivo | See comments | Logging periodicity/interval is needed as it should match the input pattern at NW side. |
| Samsung |  | Better to discuss this issue after the motivation of event-based logging is justified enough. |
| Qualcomm | Sample based | As we discussed in our response to other questions, it is very important to reduce the redundancy in the training data collection. In the last RAN2 meeting, RAN2 agreed not to support duration for training data collection. Following the L3 event-trigger reporting, to reduce redundancy in the training data collection, the UE should be configured with the number of samples the UE should log when an event is triggered. |
| Lenovo | See comments | Agree with some companies that “event to stop” will probably configured together with “event to trigger”, at least for L3 event based approach.  In this case, we may only need to configure the logging interval. |
| Huawei, HiSilicon | Other | Agree with ZTE and Apple. |
| CATT | Other | Logging interval should be configured similar as the event triggered logged MDT. |
| Ericsson | Other | Agree with ZTE and Apple. As mentioned in our previous reply, the basic is that the logging continues until the event is fulfilled. We can discuss the need to introduce a logging duration to reduce data overhead and UE consumption. |
|  |  |  |
|  |  |  |
|  |  |  |

**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. |
| vivo | No | The RSRP threshold of Top-1/K beam is sufficient for filtering the data logging. |
| Samsung |  | Better to discuss this issue after the motivation of event-based logging is justified enough. |
| Qualcomm | Yes | See response to Q3. This is useful to prevent oversampling (towards one or more conditions/labels) or skewing the training data.  *Note to the rapporteur and other companies: The answer to this question should assume that L1 based triggering is agreed. Answering “No” can also mean that if we do not support "triggering an event, one or more times, based on having captured fewer than a configured number of samples based”, which doesn’t make sense.* |
| Lenovo | No | The motivation is not clear. Also as commented in Q3. |
| Xiaomi | No | We don’t see the motivation to trigger data logging based on number of samples. |
| Huawei, HiSilicon | No | There is no need for additional complications for a mechanism whose benefit is anyway unclear. |
| CATT | No | The RSRP threshold of Top-1/K beam is sufficient for filtering the data logging. |
| Ericsson | No | Similar to our reply to Q3, an event that is based on radio measurement thresholds (that could be mapped to measurement ranges that the training entity is missing) appears to more beneficial than an event based on the number of measured samples.  What matters in the data collection is the diversity of collected data, not the number of measured samples. The objective of reducing the data overhead and UE power consumption can be achieved by the network by setting the radio thresholds to those model operating conditions for which data are missing. |
|  |  |  |
|  |  |  |

**Summary 8**: TBD.

**Proposal 8**: TBD.

# 4 Conclusion

TBD.