3GPP TSG-RAN WG2 #123-bis R2-23xxxxx

Xiamen, P.R. China, October 9th – 13th 2023

Agenda Item: 7.16.2.2

Source: Ericsson

Title: [Post123][059][AIML] Data Collection (Ericsson)

Document for: Discussion, Decision

# Introduction

This document is to address the following email discussion:

* [Post123][059][AIML] Data Collection (Ericsson)

Scope: Attempt to converge to agreements on outcome of discussion of R2-2308898, to have consolidated agreements.

Intended outcome: Report with agreeable proposals (agreeable as far as possible).

Deadline: Long

Let us observe that [R2-2308898](http://www.3gpp.org/ftp//tsg_ran/WG2_RL2/TSGR2_123/Docs//R2-2308898.zip) [1] mainly focuses on CSI feedback enhancement/Beam management uses cases, and on one-sided models. Hence, since the scope of this email discussion is to consolidate proposals in that document, in this email discussion we only address issues related to CSI feedback enhancement/Beam management and one-sided models.

Related to [R2-2308898](http://www.3gpp.org/ftp//tsg_ran/WG2_RL2/TSGR2_123/Docs//R2-2308898.zip), the following was captured in RAN2#123’s Chair Notes:

|  |
| --- |
| Chair: The proposals below are almost agreeable. It is a narrowing proposal (more specific than the physical entity mapping agreed) and is a reasonable baseline for further work:  Proposal 1 For training of NW-side models, RAN2 prioritizes discussion on the suitability of data collection frameworks for gNB-centric data collection.  Proposal 2 For training of NW-side models, the gNB-centric data collection implies that the gNB configures the UE to transfer data and initiates/terminates a data transferring session.  Proposal 3 For training of NW-side models, RAN2 evaluates the suitability of data collection frameworks for OAM-centric data collection  Proposal 4 For training of NW-side models, the OAM-centric data collection implies that the OAM initiates and terminates the data collection from the UE.  Proposal 5 If feasibility of OAM-centric data collection for NW-side models is assessed by RAN1, RAN2 considers enhancements to logged MDT, such as logging measurements in RRC Connected mode.  Proposal 6 For gNB-centric data collection for NW-side model, RAN2 to study a L3 data collection framework that allows the UE to measure and store a set of measurements (details up to RAN1) to be reported to the gNB upon request.  Proposal 7 For NW-side performance monitoring, RAN2 waits for RAN1 input on the need to enhance the L1 reporting configuration or the L3 RRC measurement configuration and reporting.  FFS Proposal 8 For UE-side model training, RAN2 considers (subject to RAN1 progress), the UE Assistance Information framework as a tool for the UE to request aid from the network in training at the UE.  Proposal 9 For UE-side performance monitoring at NW side, RAN2 to focus on impacts in layer-2, or layer-3 (possibly including some layer-1 related measurements) for reporting of the outcome of performance monitoring (e.g. performance monitoring results, (non)applicability of AIML functionality). Layer-1 details are left to RAN1.  FFS Proposal 10 The need of any enhancements to non-RAN data collection frameworks for UE-side models should be studied in SA WGs.  Proposal 11 For CSI/beam management use cases, RAN2 to agree to Table 1 in Annex A which maps LCM functions to the various existing data collection frameworks considering; the sidedness of the model, and the entity terminating/initiating the data collection. |

Regarding data collection, the following agreements/observations have been captured so far in Chair Notes:

|  |
| --- |
| **From RAN2#121-bis:**   * Observation: RAN2 may need to consider enhancements for AIML to existing functionality for data collection, e.g. for timing control (e.g. for MDT/RRM).   **From RAN2#122:**   * RAN 2 assumes that for the data collection in some scenarios (e.g., internal data up to implementation or the existing data are enough), possibly no RAN2 specification effort is needed in some scenarios, e.g. (not exhaustive):   - For model inference of UE-sided model, input data for model inference is available inside the UE.  - For UE-side (real time) monitoring of UE-sided model, performance metrics are available inside the UE. UE can independently monitor a model's performance without any data input from NW.   * P4a: For the latency requirement of data collection, RAN2 assumes:   - for all types of offline model training (i.e., UE- /NW-/ two-sided model training), there is no latency requirement for data collection  - for model inference, when required data comes from other entities, there is a latency requirement for data collection  - for model monitoring, when required monitoring data (e.g., performance metric) comes from the other entities, there is a latency requirement for data collection.   * P6a: RAN2 assumes that the analysis/selection of the data collection frameworks should focus on the RRC\_CONNECTED state (for both data generation and reporting). Analysis and potential enhancement on the non-connected state can be revisited when needed. * P5a: For the data generation entity and termination entity deployed at different entities, RAN2 assumes:   For CSI enhancement and beam management use cases:  - For model training, training data can be generated by UE/gNB and terminated at gNB/OAM/OTT server.  - For NW-sided model inference, input data can be generated by UE and terminated at gNB.  - For UE-side model inference, input data/assistance information can be generated by gNB and terminated at UE.  - For model monitoring at NW side, performance metrics can be generated by UE and terminated at gNB.  For positioning enhancement use case:  - For model training, training data can be generated by UE/gNB and terminated at LMF/OTT server.  - For NW-sided model inference, input data can be generated by UE/gNB and terminated at LMF and/or gNB.  - For UE-side model inference, input data/assistance information can be generated by LMF/gNB and terminated at the UE.  - For model monitoring at NW side, performance metrics can be generated by UE/gNB and terminated at LMF. |

Further the above assumptions were clarified/amended by RAN1 in the LS reply available in R1-2308730:

|  |
| --- |
| **From RAN1 LS reply in R1-2308730**:  RAN1 discussed Assumption 4 of Part A for CSI compression, CSI prediction, Beam management and positioning use case separately and made following clarification for each use case based on the Assumption 4 of Part A of RAN2 LS.   * For CSI compression ~~enhancement~~ ~~and beam management~~ use case: * For model training, training data can be generated by UE/gNB ~~and terminated at gNB/OAM~~/~~OTT server~~ * For ~~NW-sided model inference~~ NW-part of two-sided model inference, input data can be generated by UE and terminated at gNB. * For ~~UE-side model inference~~ UE-part of two-sided model inference, input data is internally available at UE~~input data/assistance information can be generated by gNB and terminated at UE~~. * For ~~model~~ performance monitoring at the NW side, calculated performance metrics (if needed) or data needed for performance metric calculation (if needed) can be generated by UE and terminated at gNB. * For CSI prediction ~~enhancement and beam management~~ use case: * For model training, training data can be generated by UE~~/gNB and terminated at gNB/OAM/OTT server~~. * ~~For NW-sided model inference, input data can be generated by UE and terminated at gNB.~~ * For UE-side model inference, input data~~/assistance information~~ is internally available at UE ~~can be generated by gNB and terminated at UE~~. * For performance~~model~~ monitoring at the NW side, calculated performance metrics (if needed) or data needed for performance metric calculation (if needed) can be generated by UE and terminated at gNB. * For ~~CSI enhancement and~~ beam management use case: * For model training, training data can be generated by UE/gNB ~~and terminated at gNB/OAM/OTT server~~. * For NW-sided model inference, input data can be generated by UE and terminated at gNB. * For UE-side model inference, input data~~/assistance information~~ is internally available at UE. ~~can be generated by gNB and terminated at UE.~~ * For performance~~model~~ monitoring at the NW side, calculated performance metrics (if needed) or data needed for performance metric calculation (if needed) can be generated by UE and terminated at gNB. * For positioning enhancement use case: * For model training, training data can be generated by UE/PRU/gNB/LMF ~~and terminated at LMF/OTT server~~. * For LMF~~NW~~-sided model inference (Case 2b, Case 3b), input data can be generated by UE/gNB and terminated at LMF ~~gNB~~. * For gNB-sided model inference (Case 3a), input data is internally available at gNB. * For UE-side model inference (Case 1, Case 2a), input data~~/assistance information~~ is internally available at UE ~~can be generated by LMF/gNB and terminated at the UE~~. * For ~~model~~performance monitoring at the ~~NW~~LMF side, calculated performance metrics (if needed) or data needed for performance metric calculation (if needed) can be generated by UE/gNB and terminated at LMF. * For ~~model~~performance monitoring at the ~~NW~~gNB side, calculated performance metrics (if needed) or data needed for performance metric calculation (if needed) can be generated by at least gNB.   Note: In RAN1’s answer to Assumption 4, RAN1 did not reply on the different NW entities for training (gNB/CN/LMF/OAM) as it is out of RAN1’s expertise that RAN1 cannot confirm.  Note: For the above replies for Assumption 1~4 in Part A, RAN1’s understanding is that “input data” in the RAN2 LS does not include assistance information that a model may additionally use as model input. In RAN1’s answer, RAN1 did not reply on assistance information, and informs RAN2 of related conclusions/agreements/observations in the Appendix. |

Related to data collection, it is also important to refer to the tables on the mapping of functions to entities that were discussed in [R2-2308286](http://www.3gpp.org/ftp//tsg_ran/WG2_RL2/TSGR2_123/Docs//R2-2308286.zip) and agreed in the last RAN2#123:

|  |
| --- |
| [R2-2308286](http://www.3gpp.org/ftp//tsg_ran/WG2_RL2/TSGR2_123/Docs//R2-2308286.zip) Report of [Post122][060][AIML] Mapping of functions to physical entities (CMCC) CMCC report Rel-18 FS\_NR\_AIML\_air  - Quite long discussion  - CMCC report that FFS items has support from 3 companies.  - Chair Comment: These options represent several possibilities. RAN2 would typically have selected a specific architecture option, and for a WI, specific option(s) need to be selected. Hope it is possible to further narrow down during the SI.   * P1-P6 are agreed, it is expected that FFS items for which support is not increased will be removed. |

# Discussion

In the following sections, the discussion is organized taking into account NW-side models and UE-side models and considering data collection for training, and data collection for performances monitoring.

## 2.1 NW-side models

Regarding NW-side models for the CSI feedback enhancement /Beam management use case, the following table from [R2-2308286](http://www.3gpp.org/ftp//tsg_ran/WG2_RL2/TSGR2_123/Docs//R2-2308286.zip) [2] was agreed:

|  |  |  |
| --- | --- | --- |
|  | **AL/ML functions (if applicable)** | **Mapped entities** |
| a) | Model training (offline training) | gNB, OAM, [FFS: CN, OTT server] |
| b) | Model transfer/delivery | OAM->gNB, [FFS: CN->gNB, OTT server->gNB] |
| c) | Inference | gNB |
| d) | Model/functionality monitoring | gNB |
| e) | Model/functionality control (selection, (de)activation, switching, fallback) | gNB |

Note 1: For a), only data collection part may be further discussed, how to perform the model training is up to implementation.

Note 2: For b), no model transfer/delivery is expected if the entity for model training and model inference is the same one.

Note 3: Whether/how OAM is to be involved may need to consult RAN3, SA5.

Note 4: Whether/how CN is to be involved may need to consult RAN3, SA2.

### 2.1.1 Training of NW-side models

For NW-side model training data collection, both the gNB and OAM may be the entities involved. In order to have a more focused discussion, it was discussed during RAN2#123 whether any prioritization is needed in the work.

* **Q1: For training of NW-side models, do you believe that there is the need to prioritize gNB- or OAM-centric data collection?**
  1. **Yes, gNB-centric data collection should be prioritized**
  2. **Yes, OAM-centric data collection should be prioritized**
  3. **No, both gNB- and OAM-centric data collection should be equally prioritized**

|  |  |  |
| --- | --- | --- |
| Company | Preferred Option (a,b,c) | Comments |
| OPPO | c) | In this early stage, we don’t find the strong motivation to do the down-scoping even if only RAN centric use cases are considered in R18. In R18 NG-RAN AI led by RAN3, both OAM and NG-RAN can do the model training, in which the three use cases, i.e. power saving, load balance, mobility, are also RAN centric. More addition, in MDT data collection framework, OAM can also collect many metrics from lower layer over the air, e.g. RACH report, so OAM can be a option for data collection. |
| ZTE | C | We agree the prioritization shall be discussed.  As rapporteur point out, the OAM/gNB can be a logical entity where the NW-side model is trained. In this sense, the gNB and OAM is a terminated point for data collection, which means, both gNB and OAM -centric data collection shall be prioritized. |
| Xiaomi | A or c | According to my RAN1 colleagues, RAN1 don’t have preference on gNB-centric or OAM-centric data collection. It can be up to RAN2. From RAN2 point of view, I understand it’s reasonable to collect data and train AI model at gNB, considering the use case is mainly within RAN and NW-side AI model is used by gNB. In OAM-centric data collection, additional signaling would be introduced to deliver AI model or data from OAM to gNB. So, we would prefer gNB-centric data collection. OAM centric data collection can be studied as well in this phase. |
| vivo | C | For the NW-side model for CSI/Beam use cases, prefer to reuse the functionality mapping in AI for NG-RAN. That is, the model training can be located at gNB or OAM, and both gNB- and OAM-centric data collection should be prioritized. |
| NEC | C | We support to discuss both gNB- and OAM-centric data collection in general. However, for different use case, gNB- or OAM-centric data collection may be more suitable than the other. |
| Apple | C | Let's not waste time on topic prioritization discussion, which is impossible to conclude at this stage. |
| Intel | C | As it was agreed in RAN2 #123 meeting that both gNB and OAM can be the model training entity, hence we think at this stage, both gNB- and OAM-centric data collection should be equally considered. |
| LGE | c) |  |
| Spreadtrum | C |  |
| Ericsson | c) | We are fine to study impacts on both the gNB and the OAM-centric data collection. |
| Fujitsu | c) | In this stage, both gNB and OAM can perform model training. There is no strong evidence that either gNB or OAM-centric has obvious benefits for data collection. Correspondingly, both OAM-centric and gNB-centric methods should be equally considered for AI/ML data collection. |
| Interdigital | c) |  |
| Huawei, HiSilicon | c) | Firstly, we think both methods are useful and different use cases may require different methods. RAN2 can firstly check data collection requirements per LCM component per use case, and then check possible enhancements to the existing data collection mechanisms.  Secondly, for the terminology “XX-centric”, we are a bit unclear about the meaning. Usually we use “XX-based” for describing solutions/methods, e.g. contention-based RACH resources, SSB-based Beam Failure Detection. In our view, “XX-centric” may imply that every operation is under the control of this node XX, which It's a little arbitrary for us (e.g. if we consider to apply user consent, the node XX may not be the central node). **So we suggest to use gNB-based/OAM-based data collection instead.** |
| Mediatek | c) |  |
| TCL | c) |  |
| CATT | C | Both are possible. At this stage, we cannot exclude either of the two. |
| Sharp | c) | At this stage, it is too early to preclude or deprioritize any option. Depending on the use-case suitability both options may be discussed |
| Qualcomm | c) | From the UE point of view, it should be one set of procedures, regardless of whether the end point is the RAN or OAM.  Additionally, we consider this section 2 to cover the NW-part of a two sided model, not only the Network-side model as defined by RAN1:  UE-side (AI/ML) model: An AI/ML Model whose inference is performed entirely at the UE  Network-side (AI/ML) model: An AI/ML Model whose inference is performed entirely at the network  One-sided (AI/ML) model: A UE-side (AI/ML) model or a Network-side (AI/ML) model  Two-sided (AI/ML) model: A paired AI/ML Model(s) over which joint inference is performed, where joint inference comprises AI/ML Inference whose inference is performed jointly across the UE and the network, i.e, the first part of inference is firstly performed by UE and then the remaining part is performed by gNB, or vice versa.  The email discussion moderator is asked to update the scope accordingly.  And, companies are invited to re-visit their answers under this assumption. |
| Lenovo | c) |  |
| China Unicom | C) | We are fine to study impacts on both the two solutions. Since neither of the two options has an overwhelming advantage, there is no need to converge to one solution during Rel-18 SI phase. |
| CMCC | c) | For NW-side model, we prefer to reuse the mechanism in AI for NG-RAN. |
| Nokia, Nokia Shanghai Bell | C | In our view, both approaches have their merits, and we are not at the stage of prioritizing one option versus the other. The prioritization can be done in the normative phase based on the timeline and the use-cases. For the SI phase, we think that RAN2 should consider both options and study them further to make an informed decision. |
| Samsung | c) |  |

Rapporteur summary:

* Option a: 1/23
* Option b: 0/23
* Option c: 23/23

Given the above outcome, Rapporteur proposes the following:

1. For training of NW-side models, both gNB- and OAM-centric data collection are considered in the study.

------------------------------------------------------------------------------------------------------------------------------------------------

#### 2.1.1.1 gNB-centric data collection

For gNB-centric data collection, it seems natural to assume that the gNB configures the UE to initiate the data collection and the gNB terminates the data collection.

* **Q2: For training of NW-side models, do you agree that a gNB-centric data collection implies that the gNB configures the UE to initiate the data collection procedure and the gNB terminates the data collection procedure?  
  If no, please provide your explanation.**

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comments |
| OPPO | Yes with some clarification | Usually, ‘session’ is not a concept defined/used by RAN2 even if one can see the wording ‘PDU session/Trace session’ in RAN2 spec, it’s still defined in CN domain. To remove the confusion that RAN2 may define the data collection session concept alone, it may be better to use ‘procedure/task’ instead of ‘session’ when we discuss gNB-centric data collection. |
| ZTE | Yes | The legacy framework/procedure is preferred. |
| Xiaomi | Yes | Following current logic, the UE measurement is under gNB control. |
| vivo | Yes |  |
| NEC | Yes | Maybe we can say the same gNB configures the UE to initiate the data collection and terminates the data collection |
| Apple | Yes with modified wording of Rapporteur |  |
| Intel | see comment | we agree that gNB is responsible to configure how the UE perform data collection. However, we are wondering what “data collection session” is referring to in the question. Does it refer to TCE session or a new session for data collection or others? Therefore, we propose with following updates:  For training of NW-side models, a gNB-centric data collection implies that the gNB configures and stops the UE to initiate the data collection ~~session~~ ~~and the gNB terminates the data collection session~~? |
| LGE | Yes |  |
| Spreadtrum | Yes |  |
| Ericsson | Yes | It seems natural to assume that when the training is gNB-centric, the gNB is in control of the data collection. |
| Fujitsu | Yes | Since it is NW-side model, so it is natural that NW (gNB) should initiate the data collection. |
| Interdigital | Yes |  |
| Huawei, HiSilicon | Yes |  |
| Mediatek | See comments | Agree with the principle that the network should control the data collection from UE to the network for network-sided model. The detailed mechanisms, e.g., control by configuration or explicit request is FFS. |
| TCL | Yes |  |
| CATT | Yes |  |
| Sharp | Yes |  |
| Qualcomm | See comments | This is agreeable for gNB-centric data collection if the same data collection procedure is reused for OAM-centric data collection. |
| Lenovo | Yes |  |
| China Unicom | Yes |  |
| CMCC | Yes |  |
| Nokia, Nokia Shanghai Bell | See comments | The intention of the question is a bit unclear to us. In our understanding, the gNB can use measurement and reporting framework to collect the data and this does not need to be visible to the UE. It is not clear what would be benefit of UE “knowing” that there is data collection session ongoing for NW-sided model. We would be ok with wording along the following lines:  **“For training of NW-side models, a gNB-centric data collection implies that the gNB makes decisions on the configurations provided to the UE and also acts as a termination point for the data collection protocol.”** |
| Samsung | Yes |  |

Rapporteur summary:

* Yes: 19/23

Large majority of companies seem to be ok with the rephrasing proposed by Oppo. Two companies (Nokia, Mediatek) propose to further study the data collection configuration to be provided to the UE.   
One company (Qualcomm) believes that the procedures for gNB-centric data collection may be reused also for the case of OAM-centric data collection, which is possible if the immediate MDT is considered as starting point for the OAM-centric data collection (see next questions).

Given the above outcome, Rapporteur proposes the following:

1. For training of NW-side models, the gNB-centric data collection implies that the gNB configures the UE to initiate the data collection procedure and the collected data are terminated at the gNB.   
   To further study the details of the data collection configuration.

---------------------------------------------------------------------------------------------------------------------------------------------

When it comes to RAN2 impact evaluation, L3 signalling reporting framework can be considered a natural candidate for gNB-centric data collection. While L1 measurement reporting can also potentially be used for this purpose, that does not have any RAN2 specific impact, so L1 measurement reporting should not be studied by RAN2.

* **Q3: Related to gNB-centric data collection for NW-side model training, do you agree that RAN2 should study the potential impact on L3 signalling reports?   
  If no, please provide your explanation and which alternative framework(s) (if any) can be used.**

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comments |
| OPPO | Comments | We’re fine that RAN2 shall not consider L1 signaling based data collection framework, but it’s also misleading to use the terminology ‘L3 measurements reporting framework’ as people in RAN2 may specifically refer to L3 RRM measurements reporting framework which may be not the intention here, so we suggest to use L3 signaling based data collection framework instead of L3 measurements reporting framework. More addition, L3 measurements reporting framework is mainly used for RRM purpose, which cannot be used to report multiple samples per time. MDT-like framework is better than RRM framework because MDT-like framework allows to report multiple collected/stored entries per time which can be considered as a baseline. |
| ZTE | Yes with comments | For gNB-centric data collection, we generally agree that RAN2 should focus on the L3 signaling based report framework. Among the current candidates, only the followings are L3 signaling based reporting framework;  1: L3 measurement (RRM)  2: Early measurement  3: UAI  To our understanding, the early measurement is mainly for the measurement for idle and inactive state which has been deprioritized before.   * **P6a: RAN2 assumes that the analysis/selection of the data collection frameworks should focus on the RRC\_CONNECTED state (for both data generation and reporting). Analysis and potential enhancement on the non-connected state can be revisited when needed.**   Regarding the UAI, it is not born as measurement/report framework, it can be deprioritized if there is no any strong motivation can be found.  Maybe we can improve our ambition level:  Regarding the gNB-centric data collection, RAN2 should study the potential impact on the L3 measurement at a higher priority. |
| Xiaomi | Yes | L3 signaling reporting is a natural choice. |
| vivo | OK to make the WA from RAN2 perspective | For CSI/Beam use cases, one straightforward way is utilizing the L1 measurement for data collection. However, if the data size exceeds the maximum threshold of the L1 report, L3 report can be considered by RAN2.  If RAN2 assume that L3 signaling report is used for gNB-centric data collection for model training, RAN2 can further discuss the principle and potential specification impact related to L3 signaling. |
| NEC | Yes | We agree to focus on L3 based report, but we do not think we need to limit to L3 signaling based report |
| Apple | Yes | As usual, RAN2 starts working on L1 measurement only when RAN1 sent LS to RAN2 (mostly on RRC or ASN.1 implementation).  On signaling of data collection, we think RAN2 is not in position to decide whether L1 measurement / signaling can be used (for one LCM purpose). Instead, RAN2 can only make conclusion on whether L3 measurement / signaling can be used (for one LCM purpose), and corresponding signaling change. |
| Intel | Yes |  |
| LGE | Yes | We are fine with the revised wording, but we believe L3 measurement reporting (RRM MR) framework can be a baseline for this question. |
| Spreadtrum | Yes |  |
| Ericsson | Yes | The use cases of AIML for beam management, and AIML for CSI feedback compression/enhancement studied in this Rel.18 SI should be driven by physical layer measurements (studied in RAN1). Hence the measurements expected to be used for the training of the model and for the inference are L1 measurements. However, for the case of gNB-centric data collection, it seems reasonable to consider as an option the L3 reporting framework, rather than the L1 reporting (e.g. UCI) which instead seems more suitable for the inference or for the performance monitoring of the gNB-side model. Using L3 reporting for training data collection seems to be a good option as large amount of data can be collected and transferred at once in comparison with L1 based reporting mechanisms.  Thus, RAN2 should study how to adjust the L3 reporting framework such that L1 measurements can be conveyed within it. |
| Fujitsu | Yes | We are fine with the modified version. |
| Interdigital | Yes | We agree with the views expressed by some of the companies above that collection/reporting of L1 measurements may be needed for some use cases, but as L1 measurements are not in the RAN2 domain, we can start working with the L3 measurements and signaling aspects while waiting for RAN1 input regarding the L1 measurements.  *Not a strong opinion on this, but “signaling reports” is a bit confusing. Can we just use just “signaling” or “signaling/reporting” instead (also in the subsequent questions)?* |
| Huawei, HiSilicon | Yes | We understand that L3 signalling reports here are quite general, and MDT is one example. We are fine to study it.  In addition, once we get the data collection requirements from RAN1, we could check more, e.g. which of existing data collection mechanisms are suitable for enhancements. |
| Mediatek | Comments | RAN2 can explore the potential impact on L3 signaling, provided that RAN1 justifies the need for data collection through L3 signaling per use case.  As Ericsson pointed out, data collection for inference and monitoring relies on L1 measurement reports. The need of L3 signaling for data collection for training should first be justified by RAN1, for instance, due to the requirement of data size. Even without L3 signaling, the network can still collect the L1 measurement reports of the UEs over an extended period to build the dataset for offline training.  We need RAN1’s inputs on the data collection requirements, specifically in data content, reporting type, data size and latency. |
| TCL | Yes | We share similar views with some of above companies, it is natural for RAN2 to focus on the L3 signaling for data collection reporting; and L1 signaling may be also useful for some use cases, which is better to be studied in RAN1. |
| CATT | Yes | We agree RAN2 should focus on L3 signalling framework, which may take immediate/logged MDT (can be utilized by gNB node), RRM measurement and UAI into consideration. |
| Sharp | Yes | L3 signaling can be discussed as baseline to initiate the discussion |
| Qualcomm | No | Since for inference, we will most probably have to rely on L1 reports, which also terminate at the gNB, we do not need another set of L3 reports that also have the same endpoints (UE and gNB).  So, there is no need for any new framework as L1 framework already exists and should be reused. |
| Lenovo | Yes with comment | We suppose MDT is OAM-centric method, and thus not considered here? Maybe they are somewhat equivalent in RRC connected state.  We have some sympathy with QC’s comment, maybe we need to identify first what is missing from the L1 measurement report and thus needs to be reported by L3 signaling. |
| China Unicom | Yes | We agree that RAN2 should study the potential impact on L3 signailing. Besides, QoE also belongs to L3 signaling based reporting framework, so it is required to enhance data collection mechanisms for QoE. |
| CMCC | Yes | We are fine with the revised wording. |
| Nokia, Nokia Shanghai Bell | Clarification needed | For the different use-cases that are being considered in Rel 18 SI, in their evaluations RAN1 has been considering L1 measurements (see the RAN1 LS reply) as inputs to the models that are being simulated. Indeed, we agree that this conclusion might not have an impact in the RAN2 specs. However, it is not clear to us why this motivates that we should study L3 enhancements.  In this context, the usefulness of L3 measurements is unclear to us. It would be good to know which use-case this question talks about. And what exactly would be the data content that we would be trying to get via enhanced or existing L3 measurement reports. |
| Samsung | Yes |  |

Rapporteur summary:

* Yes: 20/23
* No: 1/23
* Clarification needed: 2/23

Large majority of companies seem to be ok with the rephrasing proposed by Oppo. Some companies (e.g. Nokia, Mediatek, Qualcomm) believe that an alternative to L3 reporting is to use the L1 reporting for the gNB-centric data collection. Rapporteur highlights that RAN2 is not in the position at the moment to decide whether to adopt L1 or L3 reporting. That very much depends on RAN1 further inputs/progress on requirements. Additionally, if the L1 reporting will be adopted, that will likely not have large impact on RAN2 protocols.   
Hence, taking into account the views above, and for the sake of SI progress, Rapporteur suggests that RAN2 in the meantime studies the principles of a potential L3 mechanism for the gNB-centric data collection.

1. Related to gNB-centric data collection for NW-side model training, RAN2 studies the potential impact on L3 signalling for the reporting of collected data, taking into account RAN1 further inputs/progress.

---------------------------------------------------------------------------------------------------------------------------------------------

If the impact on L3 signalling reports is agreed to be studied for gNB-centric data collection of NW-side models, the Rapporteur suggests discussing the main principles that the L3 signalling report framework should support in order to fulfil the requirements of NW-side model training. Once such principles are assessed, RAN2 can discuss whether any enhancement to the current L3 measurement reporting framework is needed.

Taking into account proposals from different papers submitted to RAN2#123, the following principles may be envisaged for the L3 signalling reporting for NW-side model training:

1. The L3 signalling reporting framework for NW-side model training should not interfere with the ordinary L3 measurement reporting signalling used for RRM purposes and mobility.
2. The L3 signalling reporting framework for NW-side model training should allow the UE to store sets of measurements and then report them to the gNB in multiple RRC segments (which might be needed if the UE has collected lots of data).
3. The L3 signalling reporting framework for NW-side model training should allow the UE to report in a single RRC report multiple measurements taken at different points in time.
4. The L3 signalling reporting framework for NW-side model training implies that the UE may be configured to report measurements periodically.
5. The L3 signalling reporting framework for NW-side model training implies that the UE may be configured to report measurements upon fulfilling certain events.

Companies are invited to evaluate the above principles and provide their views.

* **Q4: Related to gNB-centric data collection for NW-side model training, which of the above principles should be considered for the L3 signalling reporting framework?**

|  |  |  |
| --- | --- | --- |
| Company | Option (a,b,c,d,e) | Comments |
| OPPO | b) and c) with clarification | Overall comments:  We can use ‘L3 signaling based data collection framework’ instead of ‘L3 measurements reporting’.  For a), RAN2 has not yet agreed to use L3 RRM reporting as the data collection baseline, so it’s too far to have the principle at this early stage. More addition, even if L3 RRM reporting may be considered as the baseline, it’s always followed by default that one new feature will not interfere with the legacy feature, which will be addressed during normative work as usual, so not so urgent for now to have bullet a);  For b), collecting measurements will narrow down the scope for future extension, we believe this data collection framework is not only aimed for CSI and BM use cases, which may be extended to collect metrics for future use cases, in this sense, we suggest to use more generic wording; Another point is that we think it’s too early to assume RRC segments are needed. Let’s take MDT as an example, although MDT reports may be reported via multiple RRC signalling, no explicit RRC segments is used, instead only one indicator is embedded into one report to indicate whether MDT report is still available at UE side and NW can choose to re-acquire or not to re-acquire the stored data at UE side in which one single message is triggered without RRC segment marking. Based on above, we suggest the following:  b) The L3 signaling based data collection report for NW-side model training should allow the UE to store multiple collected metric samples and then report them to the gNB in multiple RRC procedures (which might be needed if the UE has collected lots of data);  For c), the similar suggestions as bullet b):  c) The L3 signaling based data collection report for NW-side model training should allow the UE to report in a single RRC report including multiple collected metric samples taken at different points in time.  For bullet d) and e), it’s more related to normative work details, not so urgent to address. |
| ZTE | a,b (FFS), c,d,e(FFS) | For a: As we comment above, L3 measurement can be studied at a higher priority, the first rule we need to follow is that the RRM and mobility shall not be impact by the data collection for model training.  For b: Whether the RRC segment is supported depends on the data size requirement for model training at each report instance. It is not sure before the reception of LS for PART B from RAN1.  For c, in the legacy L3 measurement/report, the UE filter the historical L1 measurement result into L3 measurement result and then report to NW, theoretically speaking, it already support UE to collect the L1 measurement result at different timing points. In our understanding, consider the data collection for model training is not delay sensitive, there is no need for UE to frequently report the measurement result per measurement occasion which is power-consuming, in this sense, c can be supported  For d: this is legacy behavior and for AI based temporal beam prediction, the continuous data is very much important for model training. So the periodic data collection is needed.  For e: Even though this is also legacy, so far the motivation of the event triggered data collection for model training is not clear. We suggest to mark this as FFS. |
| Xiaomi | B,c,d,e | For a), it’s a bit unclear on ‘interfere’. This depends on the design of data collection. We understand it may be possible that UE can’t perform data collection and ordinary L3 RRM due to capability restriction. It’s too early to say no interference.  B and c are beneficial to reduce additional signaling.  D and e are beneficial to meet the delay requirement |
| vivo | b,c,d,e with comments | For a), agree with the above that the “interfere” should be clarified, does it imply that the model training data cannot be reported in RRM report?  For b), agree with the first part. Whether segmentation is needed can be revisited.  b) The L3 signalling reporting framework for NW-side model training should allow the UE to store sets of measurements and then report them to the gNB ~~in multiple RRC segments (which might be needed if the UE has collected lots of data)~~.  For c,d,e), in addition to reporting, “recording” can be added.  c) The L3 signalling reporting framework for NW-side model training should allow the UE to record and report in a single RRC report multiple measurements taken at different points in time.  d) The L3 signalling reporting framework for NW-side model training implies that the UE may be configured to record and report measurements periodically.  e) The L3 signalling reporting framework for NW-side model training implies that the UE may be configured to record and report measurements upon fulfilling certain events. |
| NEC | a/d/e | For b/c, we propose to also consider the data based report considering the large amount of data report for AIML training.  Even though we can consider L3 signalling based data collection, that should be based on a special SRB, to avoiding overwhelming the normal signalling transmission. |
| Apple | b,  c (with wording modification)  , d, e | For a), we think it is a general requirement for all new signaling (i.e. we should ensure the new signaling can work with legacy signaling), which doesn't need to capture explicitly. And it is not a specific requirement for data collection of AI/ML,  For b), we agree. And we are fine to keep segmentation, which we think it is obvious.  For c), we agree, but want to clarify that c) also allow the UE to store the measurements taken in different time points, right? Otherwise (if not allow UE to store), how the UE can report measurements taken in different time? The issue is that: if we review b) and c) together, it may be misunderstood that c) don't allow UE to store measurements because b) mentions "store" but c) doesn't. So, we suggest below change:  *c) The L3 signalling reporting framework for NW-side model training should allow the UE to store and report in a single RRC report multiple measurements taken at different points in time.*  For d) and e), although RAN2 generally don't make decision of signaling at this stage, latest TR 38.843 has captured both options in bottom of section 5.1:   * Performance metric including at least intermediate KPI (e.g., NMSE or SGCS) * UE report, including periodic/semi-persistent/aperiodic reporting, and event driven report. * Note: down selection is not precluded. * Note: UE may make decision within the same functionality on model selection, activation, deactivation, switching operation transparent to the NW.   So, we are fine to capture d) and e) as principles. |
| Intel | b, c, d, e | For a), it is not clear how to understand “should not interfere with the ordinary L3 measurement reporting”, does it mean the NW should not change configuration for existing L3 measurement reporting for AI/ML data collection purpose or others? |
| LGE | b,c,d,e | For a), we think RRM is the simplest extensible procedure since RRM can easily accommodate future expansions or modifications related to beam measurements.  For b) and c), both can be considered as solutions to store(collect) and report data set.  For d) and e), both are general solutions to trigger storing/reporting data. We can take both general methods in this stage. |
| Spreadtrum | b,c,d,e | a) L3 measurement report signaling for RRM and mobility maybe reused and extended for model training. In our side, the interference to those procedures seems to be acceptable.  b) and c) focus on data size. Both ways can be considered at this stage. RAN1 input is still needed.  d) and e) focus on the trigger of data report. In our view, Both ways can be considered at this stage. |
| Ericsson | All | All the principles listed should be taken into account when discussing potential enhancements to the L3 reporting framework. Related to a), we believe that it is somewhat inevitable to distinguish the ordinary measurement reporting framework from the AIML L3 reporting framework. Likely the configuration will be different, (since the UE may need to provide the results of data collection on CSI-RS/SSBs resources that are not used for the ordinary RRM L3 reporting) and also when to report the collected data and what to report in the collected data may differ.  Related to b), given that the data training does not have specific latency requirements, and that many measurement samples may need to be collected for quite some time in order to efficiently train a model, then b) may be important to consider because the UE may accumulate quite many data over time.  Related to c), that might be needed given that the individual measurements for CSI/beam management use cases may be taken at a different time granularity than the reporting itself. Related to d)e), those can be considered as part of a reporting configuration. |
| Fujitsu | See comments | We are fine with these points, but we think it is too early to list them as “principles” for discussion at this stage, the wording “principles” should be something in general, but these points are quite detailed. Option a) may be considered as one potential principle if Q3 are widely agreed by companies. |
| Interdigital | All | Regarding a), we agree that data collection related signaling, which is for offline training, should not impact measurements that are immediately used for mobility/RRM. We don’t have a strong opinion whether this needs to be captured explicitly. |
| Huawei, HiSilicon | b, c | For a, we wonder about the necessity of discussing it. As mentioned by some companies, the meaning of bullet a is unclear, and it can be further clarified.  For b and c, we are fine with the principle. For the wording “in multiple RRC segments”, it implies that RRC segmentation is used here. Currently, if UE wants to send a large report to NW, there can be two ways: multiple UL RRC messages (e.g. for logged MDT measurement reporting), or UL RRC segmentation (e.g. for UE capability reporting). So we suggest to change it to: **and then report them to the gNB via one or more RRC messages**. In study item phase, we may not need to discuss RRC solution details.  b) The L3 signallingmeasurements reporting framework for NW-side model training should allow the UE to store sets of measurements and then report them to the gNB in multiple RRC segments (which might be needed if the UE has collected lots of data).  c) The L3 signallingmeasurements reporting framework for NW-side model training should allow the UE to report in a single RRC report multiple measurements taken at different points in time.  For d and e, they are about network configurations, which seem straightforward. However, reporting type should be discussed before discussing configuration part, and it may corresond to RAN1 replies on data collection requirements. |
| Mediatek | See comments | For a) same comments as other company, not clear what the implication of ‘not interfere’ is. It may imply a new procedure, new configuration with existing procedure, or new type of report, etc.  For b) c) d) e), they are possibly true, but they depend on RAN1 reply on the data collection requirements in terms of data content, reporting type, data size and latency. We need to wait for RAN1’s LS. |
| TCL | All (a with comments) | We are generally fine with all mentioned principles.   1. It is a general principle when introducing a new feature, but until now, it is unclear whether the legacy L3 measurement reporting can be reused for data collection, especially for the model training case. Moreover, we think it may be possible to treat the information in the legacy L3 measurement reporting as training data. 2. and c): Which are possible to collect large size data, or decease the signaling overhead;   d) and e): The types of data reporting are able to be considered. |
| CATT | b, c, d, e with comments | b) could be agreed if we allowed the RRC segments for data collection of model training;  c) seems a kind of logged MDT in CONNECTED state, but the CONN logged MDT is unimplemented now;  d) and e) are the two measurement types (periodical and event triggered) similar as current RRM measurement.  So all these four potential principles maybe taken into consideration, but some may have premises that have not yet been agreed upon. |
| Sharp | b,c,d,e with comments | For a) it needs to be clarified what is meant by ‘should not interfere’. Our understanding is that backward compatibility is a default requirement for the design of any new signaling procedure. The existing L3-signaling reporting framework may be used as baseline to develop gNB-centric data collection framework of NW-side models.  For c, agree with the comments of some companies above that it needs to be clarified if option c allows the UE to store the measurements taken in different time instances as it is explicitly mentioned in option b.  For c, there are two interpretations as follows:   1. The L3 signaling reporting framework for NW-side model training should allow the UE to report in a single RRC report multiple measurement using filtering (e.g., as in L3 filtering). 2. The L3 signaling reporting framework for NW-side model training should allow the UE to report in a single RRC report multiple measurement without using filtering.   Therefore, we would like to clarify if this is with or without L3 filtering? |
| Qualcomm | Too early to discuss | This is assuming L3 reports are agreed, which is not the case yet.  Such L3 procedures create a large burden on the UE memory and processing. This burden is better placed at the gNB-side for gNB-centric data collection. The RAN can have access to unlimited cloud storage in a reliable manner. |
| Lenovo | All | The listed principles look generally fine to us. |
| China Unicom | b,c, d,e | For a), we think the new signaling and legacy signaling are not opposite, and It is almost a consensus that new signaling and traditional signaling can work together. So the necessity of this principle needs further clarification. |
| CMCC | b,c,d,e | For a), we think this can be further clarified.  For c), we are fine with Apple’s rewording.  For d) and e), these can be considered as starting point for triggering report, and we think other directions should not be precluded, e.g. aperiodic trigger. |
| Nokia, Nokia Shanghai Bell | See comments | In our view, the question should focus more be on agreeing with: "The current reporting L1-L3 data is sufficient for NW-side model training for Release 18 use cases?" In this context, we’d agree. Regarding enhancements of the existing frameworks (which are of course needed), we think that they would become clearer in Rel 19. Nevertheless, our view on the identified principles is as follows:  a) OK.  b)-c) We should start by studying whether or how the UE should store, report sets of measurements and then determine whether additional RRC segments are required. These are solutions to a).  c) This point seems like an implementation option for point b). Alternatively, b and c could be combined to study the support of multiple measurements made on  d) and e) can be determined in the work item or once we have requirements for event-based and periodical reporting. |
| Samsung | All | We think all the principles (above) should be considered/discussed. |

Rapporteur summary:

* Option a: 8/23
* Option b: 18/23
* Option c: 19/23
* Option d: 17/23
* Option e: 16/23
* Too early to discuss: 1/23

In general there is quite support to study the principles b,c,d,e. Companies not explicitly supporting some of the principles are pointing out that more inputs from RAN1 may be needed, and that these principles may not be applicable if a L1 reporting is agreed in RAN1. Related to a), that may depend on data collection configuration aspects. If that is separate from the ordinary RRM configuration/reporting procedures, then it seems important that that the two configuration/reporting procedures do not interfere. This can be further discussed once configuration aspects for data collection are clarified (see P2). Related to b), some companies are mentioning that the details of the reporting in multiple segments can be further discussed, and that an approach similar to the logged MDT can be used. Related to c), some rewording is proposed. In general, for all the listed principles, companies are mentioning that the RAN1 progress/requirements on data collection should be considered when studying the above principles.

Taking into account the above inputs, the following is proposed:

1. Related to gNB-centric data collection for NW-side model training, the following principles can be considered for the L3 signalling reporting framework, if used:
   1. The L3 signalling reporting framework for NW-side model training should allow the UE to store sets of measurements and then report them to the gNB in multiple RRC messages (e.g. similar to the logged MDT).
   2. The L3 signalling reporting framework for NW-side model training should allow the UE to store multiple measurements taken at different points in time and report them in a single RRC report.
   3. The L3 signalling reporting framework for NW-side model training implies that the UE may be configured to report measurements periodically or upon fulfilling certain events.
   4. The UE memory/processing power/energy consumption/signalling overhead should be taken into account.

Note: The above principles, for the L3 signalling reporting framework, can be revised depending on RAN1 progress/requirements

---------------------------------------------------------------------------------------------------------------------------------------------

Besides the above listed principles, Rapporteur would like to ask companies if there is any other principle that RAN2 should study for the L3 signalling reporting framework.

* **Q5: Related to gNB-centric data collection for NW-side model training, is there any other principle that RAN2 should take into account for the L3 signalling reporting framework? Please describe.**

|  |  |
| --- | --- |
| Company | Comments |
| OPPO | 1. The L3 signaling based data collection report for NW-side model training should have the flexibility to allow the UE to collect metrics in all RRC states or part of RRC states, which is under NW control.  * Because the data collected in different RRC state may be used by different target model training, one can not assume UE will always collect data in one RRC state for all use cases.  1. The L3 signaling based data collection report for NW-side model training should have the flexibility to allow the UE to collect metrics for specific feature/procedure/function, which is under NW control.  * Because different use cases may collect similar metrics but shaing a unified data collection framework, for instance, UE may collect L3 measurement during RRC establishment(including RRC setup/resume/re-establish cases) to optimize the performance during RRC establishment procedure while the L3 measurements collected during RRC connected state may be used to optimize the performance for UE mobility, UE should know from NW which procedure NW aims for when setting data collection metrics. |
| ZTE | None according to the current situation. |
| Apple | We agree with 1st point of OPPO (i.e. allow the UE to store metrics in all RRC states). Note that RAN1 has explicitly informed RAN2 in reply LS (R1-2308730):  *Regarding Assumption 3 of Part A,*   |  | | --- | | ***Assumption 3:***  *RAN2 assumes that the analysis/selection of the data collection frameworks should focus on the RRC\_CONNECTED state (for both data generation and reporting). Analysis and potential enhancement of the non-connected state can be revisited when needed.* |   *RAN1 confirms RAN2’s Assumption 3 for CSI compression, CSI prediction, beam prediction and Positioning use cases.*  *For positioning, it is noted that existing specification supports DL PRS measurement and UE positioning in both RRC\_CONNECTED and RRC\_INACTIVE state.* |
| Intel | For gNB-centric data collection and model training at gNB, since the collected data is used by gNB for model training. The collected data from UE should be visible and able to be used by gNB for model training directly. |
| Ericsson | Regarding the observation of enable training across different RRC states, doesn´t that conflict with the assumption that for CSI/beam management use cases we focus on RRC connected functionalities? |
| Qualcomm | Proposed principles:   * “There shall be one data collection procedure for the network side models as seen by the UE. The UE communicates with the gNB only.”   “The RAN carries the burden to pre-process and store data for gNB-centric data collection. The UE memory and processing power should not be used for this purpose.” |
| Lenovo | Duplicated data collection via L3 signaling and L1 signaling should be avoided. If gNB has already configured to UE to report some measurements via UCI, there is no need for gNB to configure the UE to report the same via L3 signaling. |
| Nokia, Nokia Shanghai Bell | Independent of L1 or L3, we think the following are important:  1) Future proofness and extensibility (to similar use-cases),  2) User security and privacy should be preserved,  3) Ensure that additional CP and/or UP air-interface traffic is minimized,  4) Allow prioritization depending on data content/characteristics and link conditions |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Rapporteur summary:

Very diverse principles have been proposed above by companies, spanning data collection across RRC states, data collection visibility, UE memory/processing requirements, privacy concerns, etc. Rapporteur proposes further discussing them, also taking into account RAN1 progress.

1. Related to gNB-centric data collection for NW-side model training, further principles for the L3 signalling reporting framework may be considered in RAN2, also taking into account RAN1 progress/requirements.

----------------------------------------------------------------------------------------------------------------------------------------------

#### 2.1.1.2 OAM-centric data collection

For OAM-centric data collection, it seems natural to assume that the OAM configures the UE to initiate the data collection session, and it terminates such data collection session.

* **Q6: For training of NW-side models, do you agree that an OAM-centric data collection implies that the OAM configures the UE to initiate the data collection procedure and the OAM terminates the data collection procedure?  
  If no, please provide your explanation.**

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comments |
| OPPO | Yes with clarification | In legacy, TCE in SA5/OAM domain is the node hosting the data collection procedure, so we suggest to make the wording more generic:  An OAM-centric data collection implies that one node in OAM domain configures the UE to initiate the data collection task/procedure and the node in OAM domain terminates the data collection task/procedure. |
| ZTE | Yes | Among the current candidates, only the followings are the OAM-Centric data collection:  1: Logged MDT  2: Immediate MDT. |
| Xiaomi | Yes |  |
| vivo | Yes, with comments | The management-based MDT framework can be reused as a baseline. To be specific, the OAM sends the data collection configuration to NG-RAN node. NG-RAN sends the configuration to the qualified UEs (e.g., UE capability, user consent) and the UEs feedback the data report when available. For the interaction between UE and NG-RAN, the configuration and reporting mechanisms for gNB-centric data collection can be reused. |
| NEC | Yes |  |
| Apple | Yes |  |
| Intel | see comment | In our understanding, there’s no direct interface between OAM and UE. According to current MDT mechanism, the configuration towards UE is transmitted by gNB, e.g. *LoggedMeasurementConfiguration*. For OAM-centric data collection, we expect this principle (no direct interface between OAM and UE) will not be changed.  Furthermore, similar to our comment to gNB-centric data collection, following changes are suggested:  For training of NW-side models, a OAM-centric data collection implies that the OAM configures and stops the UE to initiate the data collection via gNB ~~session~~ ~~and the OAM terminates the data collection session~~? |
| LGE | Yes |  |
| Spreadtrum | Yes |  |
| Ericsson | Yes | It seems natural to assume that when the NW-side model training is OAM-centric, the OAM is in charge of controlling the data collection. |
| Fujitsu | Yes, with comments | Generally, we agree that OAM should initiate and terminate the data collection procedure. Obviously, the data generation should be from UE.  But, for OAM initiated data collection request, if the gNB has enough data required for OAM, whether to configure UE to initiate the data collection procedure is FFS. Therefore, we prefer to delete the related wording at this stage. |
| Interdigital | See comments | Agree with the views expressed by Intel/Vivo that there is no direct interface between OAM and UE and the communication has to be done via the gNB similar to the MDT framework. |
| Huawei, HiSilicon | Yes, with comments | In our view, OAM-based data collection (we suggest to change the wording as commented to Q1) is equal to MDT. So we suggest to change the wording as below:  **an OAM-based data collection (i.e. MDT)**  If companies have other ideas except for MDT, we are interested in understanding more details. Otherwise, we seem to be discussing a very general concept, and this concept is closely related to other WGs (at least SA5). |
| Mediatek |  | Share the same understanding with Intel. |
| TCL | Yes |  |
| CATT | Yes |  |
| Sharp | Yes |  |
| Qualcomm | No | This is agreeable for OAM-centric data collection if this procedure is a superset of the gNB-centric data collection procedure. There is no UE-OAM interface. |
| Lenovo | Yes |  |
| China Unicom | Yes, with comments | Among the current candidates, QoE belongs to OAM-Centric data collection mechanisms , which can align with MDT for NW optimization. |
| CMCC | Yes | MDT framework can be considered as baseline. |
| Nokia, Nokia Shanghai Bell | Ok, with comments. | In our understanding, OAM cannot directly configure the UE (without gNB involvement). The statement hints more to signalling-based MDT and we do not see a need to limit ourselves to it. We think that management-based MDT should also be in scope. Please revise the wording in the proposals to capture that OAM configures the gNB which configures the UE. We agree that OAM can serve as a termination point for the protocol. We would suggest the following wording:  **“For training of NW-side models, an OAM-centric data collection implies that the OAM makes decisions on the configurations provided to the UE (via the gNB)and the OAM acts as a termination point for the protocol data collection.”** |
| Samsung | Yes |  |

Rapporteur summary:

All companies seem to be ok with the principles suggested in the question. However, some companies (Intel, Mediatek, Nokia, Qualcomm, Interdigital) suggest a rewording, since the original question may be misinterpreted that an UE-OAM interface is needed. Accordingly, the following is proposed:

1. For training of NW-side models, an OAM-centric data collection implies that the OAM provides the configuration (via the gNB) needed for the UE to initiate the data collection procedure and the collected data are terminated at the OAM entity. MDT framework can be considered as baseline.

-----------------------------------------------------------------------------------------------------------------------------------------------

Related to which framework to adopt for the OAM-centric data collection, the MDT seems to be a natural candidate, since the MDT was designed to allow the OAM to configure the UE to perform a data collection session and to terminate such data collection session (in the TCE). In particular, as showed in the endorsed tables in [R2-2302286](http://www.3gpp.org/ftp//tsg_ran/WG2_RL2/TSGR2_121/Docs//R2-2302286.zip), MDT consists of immediate MDT and logged MDT which have the following characteristics:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Logged MDT | TCE/OAM  (It can be utilized by gNB) | RRC\_IDLE/RRRC\_INACTIVE | <9kbyte | L3 cell/beam measurements, location info, sensor info,  timing info | 1. Procedure latency\*\*\*:    * Latency to enter CONNECTED state    * Latency to receive gNB request signaling (~20ms) 2. Air interface signaling latency\*\*\*\*:    * ~20ms (RRC) 3. Other latency:    * Forwarding latency between gNB and TCE | Upon gNB request after entering RRC\_CONNECTED | AS security via RRC message,  Privacy via user consent |
| Immediate MDT | TCE/OAM  (It can be utilized by gNB) | RRC\_CONNECTED | <9kbyte | L3 cell/beam measurements, location info, sensor info | 1. Procedure latency:    * Report interval:      + l20ms~30min for periodic report      + TTT for event triggered report 2. Air interface signaling latency:    * ~20ms (RRC) 3. Other latency:    * Forwarding latency between gNB and TCE | Event triggered report,  Periodic reporting | AS security via RRC message,  Privacy via user consent |

In a nutshell:

* Logged MDT allows the UE to collect and store multiple measurements, and then report such logged measurements to the gNB upon request, potentially in multiple RRC segments. However, the measurements included in the logged MDT are only performed by the UE in IDLE/INACTIVE state.
* Immediate MDT is based on the existing RRC measurement procedures for configuration and reporting. It works in RRC\_CONNECTED state, however, only a single measurement result can be conveyed in one measurement report, i.e., unlike logged MDT it does not support the possibility for the UE to collect and store multiple measurements taken at different points in time, and then report those measurements to the network in multiple RRC segments (which might be necessary when the UE has collected lots of data).

Given the above, the Rapporteur would like to ask the following question:

* **Q7: Related to OAM-centric data collection for NW-side model training, do you agree that RAN2 should study the potential impact on MDT?**
  1. **Yes, the study of immediate MDT should be prioritized.**
  2. **Yes, the study of logged MDT should be prioritized.**
  3. **Yes, both the immediate MDT and logged MDT should be equally prioritized.**
  4. **No, alternative collection framework should be studied. Please specify what.**

|  |  |  |
| --- | --- | --- |
| Company | Option (a,b,c,d) | Comments |
| OPPO | d) | Although CSI and BM use cases are related to connected mode, only one single sample can be reported via immediate MDT framework while Logged MDT allows to collect/store multiple samples which is also beneficial for model training purpose, so a combined solution, i.e. new framework, may be more suitable to leverage the pros for both frameworks. It should be noted that immediate MDT and logged MDT can work as the reference when designing the new data collection framework for model training. |
| ZTE | A | As the we mentioned in above, RAN2 have assumed the data collection shall focus on RRC connected mode. Then a) is an option that complies the RAN2 understanding.  However, we also need to dig the necessities about the logged-MDT, that is, whether the data collection at UE idle/inactive state is necessary.  For now, we think a should be prioritized. |
| Xiaomi | C | Both can be studied as baseline. |
| vivo | a) with comments | AI for NG-RAN may introduce enhancement on MDT in Rel-19 WI. Coordination is needed to avoid duplicated discussions. |
| NEC | d) | We think the OAM based data collection is a different framework from the traditional data collection from MDT in terms of the triggering for data collection and the characteristics of the data collection. And meanwhile, we may adopt some mechanism from immediate MDT and logged MDT.  With this said, we support to study a new framework for AIML data collection on top of MDT. This work can also be coordinated with SA5 and RAN3. |
| Apple | c) or d) (new MDT framework across all RRC states) | RAN1 has explicitly informed RAN2 to consider data collection for both CONNECTED state and INACTIVE state in reply LS (R1-2308730):  *Regarding Assumption 3 of Part A,*   |  | | --- | | ***Assumption 3:***  *RAN2 assumes that the analysis/selection of the data collection frameworks should focus on the RRC\_CONNECTED state (for both data generation and reporting). Analysis and potential enhancement of the non-connected state can be revisited when needed.* |   *RAN1 confirms RAN2’s Assumption 3 for CSI compression, CSI prediction, beam prediction and Positioning use cases.*  *For positioning, it is noted that existing specification supports DL PRS measurement and UE positioning in both RRC\_CONNECTED and RRC\_INACTIVE state.*  So, we think it is important to collect data in all RRC states. Regarding to specific solution, we are fine with both below ways:   * c) consider both immediate MDT (for data in CONNECTED state) and logged MDT (for data in INACTIVE/IDLD state) * d) introduce a new MDT type which allows the UE to store and report data collected in all RRC states (CONNECTED state and INACTIVE/IDLD state) |
| Intel | a | As replied in RAN1 reply LS R1-2308730, data collection should focus on only RRC\_CONNECTED state (except for positioning that INACTIVE state is already supported by existing PRS measurement and UE positioning mechanism).  Furthermore, as rapporteur analyzed in Q8, there are some commonalities between immediate MDT enhancement and gNB-centric data collection enhancement. It can also save effort/time for RAN2 to study.  Therefore, we think only immediate MDT enhancement should be prioritized.   |  |  | | --- | --- | | **RAN1 reply LS:**  Regarding Assumption 3 of Part A,   |  | | --- | | **Assumption 3:**  RAN2 assumes that the analysis/selection of the data collection frameworks should focus on the RRC\_CONNECTED state (for both data generation and reporting). Analysis and potential enhancement of the non-connected state can be revisited when needed. |   RAN1 confirms RAN2’s Assumption 3 for CSI compression, CSI prediction, beam prediction and Positioning use cases.  For positioning, it is noted that existing specification supports DL PRS measurement and UE positioning in both RRC\_CONNECTED and RRC\_INACTIVE state. | |
| LGE | d) | For b) logged MDT, considering motivation/efforts for specifying beam measurement mechanisms in the existing logged MDT as follows, we think logged MDT is not suitable for model training:   * Specific beam measurement settings and reports: Since the purpose of logged MDT is to perform measurement in RRC idle/inactive state, there are no beam specific configurations/reports. For example, measurement/report related to SSB/CSI-RS is not considered. * When/How to report the logged measurement results (for training) in connected state: In the current procedure, the UE can notify the network that the logged measurement result is available when entering connected state. The network can bring the logged measurement result through the UE information request/response procedure. In the connected state, additional consideration is needed to inform when it is available and when to report.   For a) immediate MDT, Immediate MDT utilizes the RRM procedure to report data. From the RAN2 point of view, we think immediate MDT can be analyzed with RRM analysis in 2.1.1.1 gNB-centric data collection. |
| Spreadtrum | c) | Both immediate MDT and logged-MDT can be considered. |
| Ericsson | a) | The immediate MDT is based on the existing RRC measurement procedures for configuration and reporting. Hence, if RAN2 decides to adopt the L3 RRC reporting framework for the gNB-centric data collection (with possible enhancements as discussed in Q3/Q4), then it will be easy to extend such L3 RRC reporting framework used for gNB-centric data collection to the OAM-centric data collection, i.e. to the immediate MDT.  On the other hand, if logged MDT is used, the impact in the specification will be higher since a dedicated logged measurement configuration and reporting is needed for the logged MDT. So RAN2 would need to study how to modify the logged measurement configuration/reporting. Additionally, the logged MDT works only for RRC IDLE/INACTIVE, and RAN2 should study the implication of extending this framework to CONNECTED mode (possibly involving also SA5).  Hence, if immediate MDT is agreed to be prioritized for gNB-centric data collection, RAN2 discussion will be much simpler, e.g. we would assume that then the principles valid for the gNB-centric data collection are to a large extent valid also for OAM-centric data collection. |
| Fujitsu | c) | We support to study both logged and immediate MDT enhancement for data collection, not only for CSI/BM, but for positioning. |
| Interdigital | C (with comments) | It’s correct that logged MDT is currently supported only for IDLE/INACTIVE. However, immediate MDT is limiting for data collection purposes as it is a one-shot measurement like in RRM measurements. Thus, if we rely solely on MDT, it implies some periodic reporting must be configured to enable data collection, which can be very inefficient from signaling and radio resource utilization. If we enhance immediate MDT to support also measurement logging, then it can be asked what is then the main difference from enhancing MDT to be supported in CONNECTED state. |
| Huawei, HiSilicon | a) | As we commented to Q6, we are fine to study MDT. For CSI/BM related use cases, only RRC\_Connected UEs will need to collect data, so immediate MDT can be studied. |
| Mediatek | c) and d) | We are open to studying both logged and immediate MDT, and even new mechanisms if enhancements to logged MDT and immediate MDT are deemed unsuitable |
| TCL | c) or d) | We are open for c and d):  Legacy MDT methods including logged MDT and immediate MDT are possible. However, due to some data collection requirements, such as large size training data, in order to decrease signaling overhead, we think some enhancements for legacy MDT are reasonable. And we are also open to introduce a new method used to support RRC\_connected UE to collect multiple samples. |
| CATT | c) | Since it has been confirmed by RAN1 that the Positioning enhancement case should also be considered in the RRC\_INACTIVE state, the logged MDT should also be prioritized together with immediate MDT. |
| Qualcomm | d) | We believe the purpose of data collection between MDT and AI/ML training is different enough that it warrants a study at the system level to identify the right solution end to end. |
| Lenovo | a) if only CSI/BM are considered here | a) if it’s only CSI/BM use case considered in this discussion  c) if we also consider positioning |
| China unicom | a) | Immediate MDT can be studied first. Besides, QoE also belongs to OAM-Centric data collection mechanisms , which can align with MDT for NW optimization. |
| CMCC | c) | Although logged MDT is only supported by RRC\_IDLE/RRC\_INACTIVE state, it can record more data for long time compared with immediate MDT, and we think at least beam info in idle mode can also be used for model training. In our understanding, immediate MDT is more simple and can be prioritized, but logged MDT should not be precluded at this stage. |
| Nokia, Nokia Shanghai Bell | a | According to TR 38.843 v1.0.0   |  | | --- | | *Data collection latency*:  For all types of offline model training (i.e., UE- /NW-/ two-sided model training), there is no latency requirement for data collection. For model inference, when required data comes from other entities, there is a latency requirement for data collection. For performance monitoring, when required monitoring data (e.g., performance metric) comes from other entities, there is a latency requirement for data collection.  At least for the use cases studied in this study item, it is assumed that the analysis/selection of the data collection frameworks should focus on the RRC\_CONNECTED state (for both data generation and reporting). Analysis and potential enhancement of the non-connected state can be revisited when needed. Note that existing specification supports DL PRS measurement and UE positioning in both RRC\_CONNECTED and RRC\_INACTIVE state. |     Between the options of immediate MDT and logged MDT, immediate MDT should be prioritized, as it is the RRC CONNECTED mechanism and align with the TR text proposal above. |
| Samsung | c) | Both the immediate MDT and logged MDT should be studied. |

Rapporteur summary:

* Option a: 8/22
* Option b: 0/22
* Option c: 10/22
* Option d: 7/22

A slight majority of companies prefer to study both the logged MDT and the immediate MDT. However, from the comments above and also from the comments Q10, it seems that the immediate MDT looks more suitable than the logged MDT, and that there are some strong concerns of the suitability of the logged MDT for this use case (see outcome of Q10). Hence, Rapporteur suggests first focusing on the study of the principles of the immediate MDT, and at the same time further investigating the suitability of the logged MDT framework (see P12).   
Some other companies are proposing discussing an alternative collection framework, i.e. option d). However, it seems that the majority of companies supporting option d) intends to study the potential enhancement to the existing immediate/logged MDT (which is the subject of the next questions), rather than studying a completely new framework.   
Given the above outcome, Rapporteur suggests focusing on the existing immediate/logged MDT:

1. Related to OAM-centric data collection for NW-side model training, RAN2 studies the potential impact on the immediate MDT, taking into account RAN1 further inputs/progress.
2. Related to OAM-centric data collection for NW-side model training, RAN2 to further discuss motivations of the suitability of the logged MDT framework or other frameworks.

---------------------------------------------------------------------------------------------------------------------------------------------

Irrespective of whether immediate or logged MDT is taken into account for the OAM-centric data collection of NW-side models, the Rapporteur suggests discussing what are the main principles that the MDT framework should support in order to fulfil the requirements of OAM-centric data collection. Once such principles are assessed, RAN2 can discuss whether any enhancement to Immediate or Logged MDT is needed.

For example, for Immediate MDT, we note that the current framework is based on the existing RRC measurement procedures for configuration and reporting. Hence, the following principles may need to be studied:

1. The Immediate MDT framework for NW-side model training should allow the UE to store sets of measurements and then report them in multiple RRC segments.
2. The Immediate MDT framework for NW-side model training should allow the UE to report in a single RRC report multiple measurements taken at different points in time.
3. The Immediate MDT framework for NW-side model training should allow the network to configure the UE to report measurements periodically.
4. The Immediate MDT framework for NW-side model training should allow the network to configure the UE to report measurements upon fulfilling certain events.

* **Q8: Related to OAM-centric data collection for NW-side model training, which of the above principles should be considered if Immediate MDT is used?**

|  |  |  |
| --- | --- | --- |
| Company | Option (a,b,c,d) | Comments |
| OPPO | a) and b) with clarification | The similar suggestion as Q4. No need to differentiate Immediate MDT and logged MDT when setting principles/requirements  For a) and b) suggest the following wording:   1. The OAM-centric data collection reporting for NW-side model training should allow the UE to store multiple collected metric samples and then report them in multiple RRC procedures. 2. The OAM-centric data collection reporting for NW-side model training should allow the UE to report in a single RRC report including multiple collected metric samples taken at different points in time.   For c) and d), postpone to normative work stage. |
| ZTE | a (FFS), b,c,d (FFS) | The similar suggestion in Q4.  a: Whether the RRC segments are supported depends on the requirement of data size for each report instance.  d: it is not still clear what is the motivation of event triggered data collection for model training. |
| Xiaomi | All | I understand these principles are beneficial to avoid signaling overhead and meet latency requirement. However, with such principles, MDT and L3 measurement report becomes similar. Therefore, I understand these principles are agreeable regardless MDT or L3 measurement report is used. |
| vivo | All with comments | Similar comments in Q4. And for the interaction between UE and NG-RAN, the configuration and reporting mechanisms for gNB-centric data collection can be reused as baseline.  a) The Immediate MDT framework for NW-side model training should allow the UE to store sets of measurements and then report them ~~in multiple RRC segments~~.  b) The Immediate MDT framework for NW-side model training should allow the UE to record and report in a single RRC report multiple measurements taken at different points in time.  c) The Immediate MDT framework for NW-side model training should allow the network to configure the UE to record and report measurements periodically.  d) The Immediate MDT framework for NW-side model training should allow the network to configure the UE to record and report measurements upon fulfilling certain events. |
| NEC | All with comments | As said, we support study a different framework for data collection.  However, if immediate MDT is studied as the start point, we may need to consider the RRC Idle/Inactive mode based data collection and report |
| Apple | All with wording modification in b) | Similar comments in Q4. We also support NEC raised point (i.e. need to consider the RRC Idle/Inactive mode based data collection and report).  For b), we suggest below change:  *b) The Immediate MDT framework for NW-side model training should allow the UE to store and report in a single RRC report multiple measurements taken at different points in time.* |
| Intel | a, b, c, d | Regarding to d), since event trigger reporting is now supported by immediate MDT, we are wondering whether this is for new event or existing event to be reused? If yes for new event, it would be good to clarify in the principle. |
| LGE | a,b,c,d | Similar answer with Q4.  For a) and b), both can be considered for solutions to store(collect) and report data set.  For c) and d), both are general solutions to trigger storing/reporting data. We can take both general methods in this stage. |
| Spreadtrum | All |  |
| Ericsson | All | RAN2 can take all of them into account in the study. Whether any of them should be disregarded very much depends on requirements that RAN1/RAN2 can further investigate.  Related to a) for example, that seems needed if we expect that the UE can be left collecting data from quite some time, so that it may accumulate quite large amount of data. Given that the data training does not have specific latency requirements, and that the AIML engine may require many measurement samples to efficiently train a model, then a) may be important to consider. Similarly related to b), might be needed given that the individual measurements for CSI/beam management use cases may taken at a different time granularity than the reporting itself. c)d) can be considered as part of a possible configuration. |
| Fujitsu | See comments | Similar to Q4. |
| Interdigital | All | Similar view as Q4. Also see comments to Q10. |
| Huawei, HiSilicon | A, b | Similar to Q4  For a, we suggest to modify the wording “then report them in multiple RRC segments” to: **and then report them to the gNB via one or more RRC messages**  For c and d, they are about network configurations, which seem straightforward. However, reporting type should be discussed before discussing configuration part, and it may corresond to RAN1 replies on data collection requirements. |
| Mediatek | See comments | Similar to Q4, those principles are possibly true, but they depend on RAN1 reply on the data collection requirements in terms of data content, reporting type, data size and latency. We need to wait for RAN1’s LS. |
| TCL | All |  |
| CATT | All with comments | Similar as we said in Q4, all these four potential principles maybe taken into consideration, but some may have premises that have not yet been agreed upon (e.g. allowing of RRC segments for data collection of model training). |
| Qualcomm | See comments | This is assuming already an MDT solution, which is not agreed yet. We need a system level design that is beyond RAN2 to study the best data collection scheme where equivalent features for collection may be supported. |
| Lenovo | All | They look in general ok to us |
| China Unicom | comments | Since there is no consensus regarding question 7, it is not yet time to determine the principles for Immediate MDT. |
| CMCC | All | Similar to Q4. |
| Nokia, Nokia Shanghai Bell | a, b -> Not ok (see comments).  c, d -> Ok | a) Delaying critical CP traffic should be avoided. Generally, we should only consider the use of additional RRC segments when it is clear that there is a need to do so, which would come in the WI stage.  b) While this approach could be valid, it seems like an optimization that could be discussed either in the WI or in a discussion about data transmission options. As a principle, we do not feel that it fits.  c)-d) are already supported by immediate MDT, but they could be enhanced for more efficient data collection procedures. |
| Samsung | All | We think all principles should be studied. |

Rapporteur summary:

* Option a: 16/22
* Option b: 17/22
* Option c: 16/22
* Option d: 15/22

Most companies seem to be ok with all proposals. However, some companies, similar to Q4, are pointing out that the need and the details of these principles may require further discussion, depending on RAN1 progress/requirements. Similar to the inputs to Q4, some companies require some rewordings, which are taken into account in the proposal below:

1. Related to OAM-centric data collection for NW-side model training, the following principles can be considered for the immediate MDT framework, if used:
   1. The Immediate MDT framework for NW-side model training should allow the UE to store sets of measurements and then report them in multiple RRC messages (e.g. similar to the logged MDT).
   2. The Immediate MDT framework for NW-side model training should allow the UE to store multiple measurements taken at different points in time and report them in a single RRC report.
   3. The Immediate MDT framework for NW-side model training should allow the network to configure the UE to report measurements periodically or upon fulfilling certain events.
   4. The UE memory/processing power/energy consumption/signalling overhead should be taken into account.

Note: The above principles, for the immediate MDT framework, can be revised depending on RAN1 progress/requirements

------------------------------------------------------------------------------------------------------------------------------------------------

Besides the above listed principles, Rapporteur would like to ask companies if there is any other principle that RAN2 should study related to the immediate MDT.

* **Q9: Related to OAM-centric data collection for NW-side model training, is there any other principle that RAN2 should take into account if the Immediate MDT is used? Please describe.**

|  |  |
| --- | --- |
| Company | Comments |
| OPPO | The similar suggestions as Q5.   1. The OAM-centric data collection report for NW-side model training should have the flexibility to allow the UE to collect metrics in all RRC states or part of RRC states, which is under OAM control. 2. The OAM-centric data collection report for NW-side model training should have the flexibility to allow the UE to collect metrics for specific feature/procedure/function, which is under OAM control. |
| ZTE | None according to the current situation. |
| vivo | For the interaction between UE and NG-RAN, the configuration and reporting mechanisms for gNB-centric data collection can be reused. That is, OAM- or gNB-centric data collection can be transparent to UE. |
| Apple | We agree with OPPO's 1st point: The UE is allowed to store data across different RRC states without pause. |
| Ericsson | Regarding the observation of enable training across different RRC states, doesn´t that conflict with the assumption that for CSI/beam management use cases we focus on RRC connected functionalities? |
| Qualcomm | Same as Q5:   * **“There shall be one data collection procedure for the network side models as seen by the UE. The UE communicates with the gNB only.”**   **“The RAN carries the burden to pre-process and store data for gNB-centric data collection. The UE memory and processing power should not be used for this purpose.”** |
| Lenovo | Similar understanding as Vivo. If we only consider CSI/BM here, and only consider RRC connected state, the principles for gNB to configure UE seems same in both gNB-centric and OAM-centric ways. |
| Nokia, Nokia Shanghai Bell | We suggest the following  1) Future proofness and extensibility (to similar use-cases),  2) User security and privacy should be preserved,  3) Ensure that additional CP and/or UP air-interface traffic is minimized,  4) Allow prioritization depending on data content/characteristics and link conditions. |
|  |  |
|  |  |
|  |  |

Rapporteur summary:

Very diverse principles have been proposed above by companies, spanning data collection across RRC states, data collection visibility, UE requirements, privacy concerns, etc. Rapporteur proposes further discussing them, also taking into account RAN1 progress.

1. Related to OAM-centric data collection for NW-side model training, further principles for the immediate MDT may be considered in RAN2, also taking into account RAN1 progress/requirements.

----------------------------------------------------------------------------------------------------------------------------------------------

On the other hand, for Logged MDT, it is already possible for the UE to store sets of measurements and then report them to the network in multiple RRC segments. However, it only works for UEs in IDLE/INACTIVE state (which is not in line with the RAN2 understanding that the CONNECTED mode use cases should be prioritized). Additionally, the logged MDT reports can be transmitted to the network only upon gNB request. Hence, the following principles may need to be studied:

1. The Logged MDT framework for NW-side model training should also work for UEs in RRC\_CONNECTED state.
2. The Logged MDT framework for NW-side model training should allow the network to collect the UE measurement reports periodically.
3. The Logged MDT framework for NW-side model training should allow the network to collect the UE measurement reports upon fulfilling certain events.

* **Q10: Related to OAM-centric data collection for NW-side model training, which of the above principles should be considered if Logged MDT is used?**

|  |  |  |
| --- | --- | --- |
| Company | Option (a,b,c) | Comments |
| OPPO | None | For OAM-centric data collection, no need to differentiate Immediate MDT and logged MDT when setting principles/requirements, the following suggestion can directly reflect the data collection requirement for OAM-centric data collection:   1. The OAM-centric data collection reporting for NW-side model training should allow the UE to store multiple collected metric samples and then report them in multiple RRC procedures. 2. The OAM-centric data collection reporting for NW-side model training should allow the UE to report in a single RRC report including multiple collected metric samples taken at different points in time. 3. The OAM-centric data collection report for NW-side model training should have the flexibility to allow the UE to collect metrics in all RRC states or part of RRC states, which is under OAM control; 4. The OAM-centric data collection report for NW-side model training should have the flexibility to allow the UE to collect metrics for specific feature/procedure/function, which is under OAM control. |
| ZTE | None | The logged MDT is applied for UE to collect the data during the idle/inactive state and report to NW by gNB request in RRC connected mode. It cannot be foreseen to enhance the logged MDT to work on RRC connected mode. |
| Xiaomi | All | I understand these principles are beneficial to avoid signaling overhead and meet latency requirement. However, with such principles, MDT and L3 measurement report becomes similar. Therefore, I understand these principles are agreeable regardless MDT or L3 measurement report is used. |
| vivo | None | Prefer to keep logged MDT only involving measurement logging by UE in IDLE and INACTIVE state. The analysis/selection of the data collection frameworks should focus on the RRC\_CONNECTED state (for both data generation and reporting). |
| NEC | All with comments | As said, we support study a different framework for data collection.  However, if logged MDT is studied as the start point, we may need to consider the data collection for the UE during state transition between RRC Idle/Inactive and connected in terms of data collection. |
| Apple | All | Similar view as NEC. We need to take RAN1 Reply LS into account. |
| Intel | others | We don’t think enhancement for logged MDT is needed, as discussed in Q7. |
| LGE | None | See Q7. Considering motivation/efforts for specifying beam measurement mechanisms in the existing logged MDT as follows, we think logged MDT is not suitable for model training. |
| Spreadtrum | b) c) with comments | For a) There is no need to make Logged-MDT work in RRC connected. Because immediate MDT can be executed in RRC connected. Continuous MDT can be achieved by the switch of logged MDT and immediate MDT.  For b) and c) it may be reasonable but still heavily depends on RAN1 input for data collection requirements. |
| Ericsson | All if logged MDT is considered | We do not believe that the logged MDT should be considered for the use cases of this SI, as explained in our replies above.  Spec impact seems much larger than the immediate MDT. Additionally, enabling RRC connected mode for the logged MDT may require further investigation and potentially coordination with SA5. Nevertheless, if the logged MDT is agreed to be studied, those principles should be considered, given that those principles are not currently addressed by the current logged MDT procedures. |
| Fujitsu | At least a) | According to RAN2 assumption, at least bullet a) is needed.  P6a: RAN2 assumes that the analysis/selection of the data collection frameworks should focus on the RRC\_CONNECTED state.  For bullets b) and c), if we can reuse the reporting scheme of logged MDT report, they may not be needed. |
| Interdigital | See comments | We have to look at this in conjunction with Q8, because we should discuss if there is a need to enhance both the immediate MDT to support logging and also enhance the logged MDT to support CONNECTED state, because the end result will be the same (i.e., logged measurements in CONNECTED state that can a periodic or event triggered). |
| Huawei, HiSilicon | None | For CSI/BM related use cases, immediate MDT can be studied, and we do not see a need of studying logged MDT. For positioning related use cases, in the RAN1 LS R1-2308730, they mentioned the following:  *For positioning, it is noted that existing specification supports DL PRS measurement and UE positioning in both RRC\_CONNECTED and RRC\_INACTIVE state.*  It is pending for RAN2 discussions whether to consider RRC\_Inactive state for data collection.  For now, we think the need of studying logged MDT should be clarified. |
| Mediatek | See comments | For a) we are open to enhance log MDT to support RRC\_Connected model. But we are not sure whether it is still considered as log MDT or new mechanism.  For b) and c) we need to wait for RAN1 inputs on data collection requirement. |
| TCL | All | As mentioned in Q7, if using the legacy logged MDT method, it needs to be enhanced, bullets a), b) and c) are possible from our perspective. |
| CATT | All with comments | Similar as we said in Q4, all these four potential principles maybe taken into consideration, but some may have premises that have not yet been agreed upon (e.g. allowing of the CONN logged MDT). |
| Qualcomm | See comments | This is assuming already an MDT solution, which is not agreed yet. We need a system level design that is beyond RAN2 to study the best data collection scheme where equivalent features for collection may be supported. |
| Lenovo | See comment | Logged MDT can be considered in case of positioning use case.  Not sure what does a) mean, how to use Logged MDT in RRC connected state? |
| China Unicom | Comments | Since there is no consensus regarding question 7, it is not yet time to determine the principles for Logged MDT. |
| CMCC | All | We are fine with these bullets as starting point if logged MDT is used. |
| Nokia, Nokia Shanghai Bell | None | Considering the RAN1 LS reply, and earlier agreements wherein we focus on RRC CONNECTED frameworks. Logged MDT has been designed for collecting measurements when the UE is RRC Idle/Inactive mode. In our view, enhancing logged MDT for RRC CONNECTED UEs would have high UE impact, while we think that enhancing immediate MDT would be sufficient for the Rel 18 use-cases. |
| Samsung | All (with comments) | We are open to study possible enhancements of logged MDT to support data collection for AI/ML. |

Rapporteur summary:

* Option a: 9/22
* Option b: 9/22
* None/skeptical: ~11/22

From the above views, it seems that many companies are challenging the usage of logged MDT itself. So before discussing principles for the logged MDT, further motivations on the suitability of the logged MDT is needed. This is proposed in the P9 above.

------------------------------------------------------------------------------------------------------------------------------------------------

Besides the above listed principles, Rapporteur would like to ask companies if there is any other principle that RAN2 should study related to the logged MDT.

* **Q11: Related to OAM-centric data collection for NW-side model training, is there any other principle that RAN2 should take into account if the Logged MDT is used? Please describe.**

|  |  |
| --- | --- |
| Company | Comments |
| OPPO | See answer in Q10 |
| Apple | The UE is allowed to store data across different RRC states without pause. |
| Qualcomm | Same as Q5 and Q9. |
| Nokia, Nokia Shanghai Bell | We think that logged MDT should not be a priority for study as we have agreed to focus on RRC CONNECTED mode. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Rapporteur summary:

Given the outcome of the above question, Rapporteur suggests postponing the discussion on the principles of the logged MDT. First RAN2 should assess the suitability of this framework.

1. Related to OAM-centric data collection for NW-side model training, the principles of the logged MDT framework will be discussed if its suitability is assessed.

----------------------------------------------------------------------------------------------------------------------------------------------

### 2.1.2 Performance Monitoring of NW-side models

On the topic of performance monitoring of NW-side models, RAN2 assumed in RAN2#122 that:

“- for model monitoring, when required monitoring data (e.g., performance metric) comes from the other entities, there is a latency requirement for data collection.”

Additionally, Rapporteur believes that also the signalling overhead could be a concern, especially if the monitoring should be done frequently.   
If L3 measurements are used for this purpose, the above requirements should be taken into account. Rapporteur would like to ask companies if there is any impact that can be foreseen in RAN2 protocols related to performance monitoring of NW-side model, taking into account the above expected requirements.

* **Q12: Related to performance monitoring of NW-side models, do you foresee any impact in RAN2 protocols that RAN2 should study? Please describe.**

|  |  |
| --- | --- |
| Company | Comments |
| OPPO | No yet. |
| Xiaomi | To assist performance monitoring of NW-side models, UE may need to collect and report monitoring data/assistance information to gNB. There may be spec impact to RAN2 on the signaling and procedure. |
| vivo | On this issue we can wait for RAN1 feedback on the requirement of data collection for model monitoring. E.g., Using L3 or L1 signaling for model monitoring relies on the content and size of the monitoring data.  Potential RAN2 impacts may include monitoring configuration and reporting.   |  | | --- | | Agreement  In CSI compression using two-sided model use case, further study the necessity, complexity, overhead, latency and potential specification impact on ground truth CSI report for NW side data collection for model performance monitoring, including:   * Scalar quantization for ground-truth CSI   + FFS: any processing applied to the ground-truth CSI before scalar quantization * Codebook-based quantization for ground-truth CSI   + FFS: Parameter set enhancement of existing eType II codebook, based on evaluation results in 9.2.2.1 * RRC signaling and/or L1 signaling procedure to enable fast identification of AI/ML modelperformance * Aperiodic/semi-persistent or periodic ground-truth CSI report. | |
| Apple | Not for now. We tend to think RAN2 study on data collection for inference and performance monitor should not start before reception of RAN1 specific requirement (i.e. RAN2 can only study data collection for offline training at this stage). |
| Intel | prefer to wait for RAN1 requirement and input regarding to performance monitoring. |
| LGE | Cannot see yet. It depends on the RAN1 decision. According to RAN1 LS, there would reports for calculated performance metrics or data needed for performance metric. |
| Spreadtrum | Some measurement configuration to UE and assistance information/report to NW maybe needed. But it still needs to wait for RAN1 requirement. |
| Ericsson | We are ok to wait for RAN1 inputs here. |
| Fujitsu | No. For performance monitoring of NW-side model, the L1 measurement and reporting can be considered. if the signalling overhead is a concern, it’s the scope of enhancement on L1 measurement and reporting. |
| Interdigital | We agree with the view expressed by other companies that we can wait for RAN1 input before discussing this further |
| Huawei, HiSilicon | Wait for RAN1 inputs. |
| Mediatek | Wait for RAN1’s reply. |
| TCL | Fine with wait for RAN1 input |
| CATT | We should wait for RAN1 feedback. The RAN2 impacts may include configuration and reporting. And the impacts may be different with different use cases. |
| Sharp | For evaluation and verification, the performance of NW side models the network may need to configure the UE to report targeted metrics for both on-sided and two-sided models. |
| Qualcomm | As others pointed out, it is too early to discuss now. |
| Lenovo | Wait for RAN1 input. |
| China Unicom | Yes. We are fine to wait RAN1 requirement and input regarding to performance monitoring. |
| CMCC | We prefer to wait for RAN1 feedback. |
| Nokia, Nokia Shanghai Bell | The question is re-iteration of the questions that we already sent as LS to RAN1 and RAN1 provided us the replies (Part 1) for performance monitoring which are listed below.   |  | | --- | | * For CSI compression use case: * For ~~model~~ performance monitoring at the NW side, calculated performance metrics (if needed) or data needed for performance metric calculation (if needed) can be generated by UE and terminated at gNB. * For CSI prediction use case: * For performance~~model~~ monitoring at the NW side, calculated performance metrics (if needed) or data needed for performance metric calculation (if needed) can be generated by UE and terminated at gNB. * For beam management use case: * For performance~~model~~ monitoring at the NW side, calculated performance metrics (if needed) or data needed for performance metric calculation (if needed) can be generated by UE and terminated at gNB. * For positioning enhancement use case: * For ~~model~~performance monitoring at the ~~NW~~LMF side, calculated performance metrics (if needed) or data needed for performance metric calculation (if needed) can be generated by UE/gNB and terminated at LMF. * For ~~model~~performance monitoring at the ~~NW~~gNB side, calculated performance metrics (if needed) or data needed for performance metric calculation (if needed) can be generated by at least gNB. |     Based on the above response, we foresee that no protocol level impact is expected. If we want to study any RAN2 impact, we need to discuss and analyze these responses. |
| Samsung | Ok to wait for RAN1 input. |

Rapporteur summary:

From the above replies, it seems too early to discuss this, and more inputs are needed from RAN1.

1. Related to performance monitoring of NW-side models, impact (if any) in RAN2 protocols needs to be further evaluated, depending on RAN1 progress/inputs.

------------------------------------------------------------------------------------------------------------------------------------------------

## 2.2 UE-side models

Related to UE-side models for the CSI feedback enhancement/Beam management use case, the following table from [R2-2308286](http://www.3gpp.org/ftp//tsg_ran/WG2_RL2/TSGR2_123/Docs//R2-2308286.zip) [2] was agreed:

|  |  |  |
| --- | --- | --- |
|  | **AL/ML functions (if applicable)** | **Mapped entities** |
| a) | Model training(offline training) | UE-side OTT server, UE, [FFS: gNB, OAM, CN] |
| b) | Model transfer/delivery | UE-side OTT server->UE, [FFS: gNB->UE, or OAM->UE, or CN->UE] |
| c) | Inference | UE |
| d) | Model/functionality monitoring | UE (UE monitors the performance, and may report to gNB), gNB (gNB monitors the performance) |
| e) | Model/functionality control (selection, (de)activation, switching, fallback) | gNB if monitoring resides at UE or gNB,  UE if monitoring resides at UE |

Note 1: For a), only data collection part may be further discussed, how to perform the model training is up to implementation.

Note 2: For b), no model transfer/delivery is expected if the entity for model training and model inference is the same one.

Note 3: Whether/how OAM is to be involved may need to consult RAN3, SA5.

Note 4: Whether/how CN is to be involved may need to consult RAN3, SA2.

### 2.2.1 Training of UE-side models

As shown in the table above, RAN2 has so far assumed that the UE itself or the UE-side OTT server will oversee the model training for UE-side models.   
It was discussed at length in RAN2 whether data collection protocols available outside RAN could be used for the purpose of UE-side model training when the UE-side OTT server is involved. However, no conclusion was made. At the same time, it is not clear what could be the impact in RAN2 protocols if the UE-side OTT server is responsible for the UE-side model training.

* **Q13: If the UE-side OTT server is responsible for the UE-side model training, do you agree that the way the OTT server collects data should not be studied in RAN2?  
  If no, please explain which data collection framework should RAN2 study when the UE-side OTT server is responsible for the UE-side model training.**

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comments |
| OPPO |  | For this scenario, SA work may be involved but RAN2 is not involved based on current observation. |
| ZTE | Yes | The OTT server collects data should not be studied in RAN. |
| Xiaomi | Yes |  |
| Vivo | Yes, with comments | Potential discussion on data collection framework for the UE-side model training at UE-side OTT server should involve SA WG(s), such as SA2. RAN2 should send Ls to SA2 to trigger the discussion in SA2. |
| NEC | Yes |  |
| Apple | Yes | Not in RAN scope, and no need to send LS to SA2 because SA2 is having such discussion. |
| Intel | Yes |  |
| LGE | Yes, but | To avoid affecting user data transmission (data performance), we can discuss whether the AIML data specific management is necessary. |
| Spreadtrum | Yes |  |
| Ericsson | Yes |  |
| Fujitsu | Yes | If OTT server makes use of only application-level data for the model training, there may not be any RAN2 involvement for data collection. |
| Interdigital | Yes |  |
| Huawei, HiSilicon | Yes | We think data collection between UE and OTT server has been discussed for lots of times in the past. There were no conclusions. At RAN2#122 meeting, the Chair had the following statement:  - Chair: there are strong objections from network vendors to include EVEX as an option for data collection to evaluate, RAN2 cannot decide to do this currently, suggest that RAN2 do not revisit this unless the situation has changed, e.g. by a TSG RAN decision.  So far, RAN has not made any progress on this specific topic and no Lses have been approved. So we think RAN2 should not revisit this. |
| Mediatek | No | Supporting UE-side data collection is a comprehensive scheme that requires the involvement of both RAN2 and SA2, and even SA3 for privacy and security issues, in order to enable RAN-related data to reach the OTT server. As we explained in R2-2308151 for UE-side data collection, the following issues need to be addressed:   * How can the assistance information and the configuration/condition information reach the OTT server for model training? (The UE side needs to know which configuration/condition the AI/ML model is targeting, especially for functionality-based LCM) * How is the data collection procedure controlled, and by which entity (RAN node, OAM, or OTT server)? * Through which tunnel (UP/CP) is the collected data transferred to the OTT server? * Is vendor-dependent information allowed to be delivered to the OTT server? And if it is allowed, how?   We also mentioned two types of UE-sided data collection methods: with and without RAN awareness.   * If data transfer from the UE to the OTT server is with RAN awareness, it is primarily a RAN2 task. In this case, the RAN node/OAM/CN controls the data collection procedure, and the procedure of data collection from the UE to the network needs to be studied. When UE data arrives at the RAN node, the RAN node/OAM processes those data, adds RAN-related information, and transfers the UE data to the OTT server through 5GS. * If data transfer from the UE to the OTT server is without RAN awareness, it is primarily an SA2 task. Since SA2 doesn’t consider RAN use case and support RAN-related data transfer over the application layer, RAN2 needs to send the new requirement to SA2.   Concluding that UE-side data collection doesn't require discussion in RAN2 simply because the OTT server isn't deployed in RAN would be too hasty.  Furthermore, EVEX is not the only solution for UE-side data collection. Excluding EVEX in RAN2 doesn’t mean exclude discussion of other solutions for UE-sided data collection in RAN2. |
| TCL | Yes | It seems like without RAN scope if the model is trained in UE itself or the UE-side OTT server. |
| CATT | Yes | Not in RAN scope. |
| Sharp | Yes, but | It depends on the type of data collected by the OTT server (e.g., if any RAN related data, AS layer/L1/L2/L3 info. or parameters are collected or not), if it is controlled by RAN and/or whether RAN is aware of this data collection process or not. |
| Qualcomm | No with comments | The strict answer to this question is that there may not be any RAN2 impact (yes to the question as asked).  However, this logic is incorrect: “We may not need RAN work to support UE-side model training => Thus, we don’t need 3GPP work”.  Companies in the last SA meeting could not open an SA2 work item to support RAN work because RAN has not asked for such support.  Even if an OTT server is collecting the data there will need to be a coordination between the UE-side and the network (in general). The UE will need assistance information, a valid data path with appropriate priority and QoS.  So, RAN2 needs to send an LS to SA/SA2 to study the best system level solution to collect data for UE-side model training (and, also for OAM-centric NW-side model). |
| Lenovo | Yes with comment | Maybe not led by RAN2, there could be other impacts e.g., model/functionality identification etc. |
| China Unicom | Yes |  |
| CMCC | Yes | Same view with Apple. |
| Nokia, Nokia Shanghai Bell | See comments | If the question is referring to an interaction between the OTT and the edge of the NW, then we agree, this is out of RAN2 scope.  However, the discussion of the use of a non-standardized or transparent container transmitted between the UE and the gNodeB is in RAN2 scope, and this end-to-end interaction requires more study. We agree with Mediatek’s analysis of the situation. Of course, the interaction between the OTT, NW, and UE is a complex one that involves many working groups. We do not see the need to disallow further discussion on what RAN2 impacts might arise. |
| Samsung | Yes |  |

Rapporteur summary:

* Yes: 21/23
* No: 2/23

There is a large majority of companies that are ok with the assumption that the the way the OTT server collects data should not be studied in RAN2. However, some companies (Nokia, Mediatek, Qualcomm) are mentioning that even if the training is done in an OTT server, RAN-awareness of the UE-side training and UE-side data collection should be discussed in RAN2. To this end, Rapporteur highlights that the question was on whether RAN2 should discussion how the OTT server collects data." It was not about precluding the discussion in RAN2 on whether RAN-awareness/gNB-UE coordination is needed to support UE-side model training. That is a separate topic addressed in the next question and proposal.

Hence, the following is proposed:

1. If the UE-side OTT server is responsible for the UE-side model training, the way the OTT server collects data should not be studied in RAN2.

------------------------------------------------------------------------------------------------------------------------------------------------

Even if the UE-side model training is performed by the UE itself or the UE-side OTT server, RAN2 may need to discuss whether any impact is expected in RAN2 protocols, e.g. in order to allow the UE to properly perform the training. To this end, Rapporteur notes that in the TR 38.843, RAN1 mentions “UE reporting to NW supported/preferred configurations of DL RS transmission”, “data collection initiated/triggered by configuration from NW”, “request from UE for data collection”, etc.:

In light of what is already captured above by RAN1 in the TR, the Rapporteur would like to ask companies if there is any other impact foreseen in RAN2 protocols related to UE-side model training, e.g. if the RRC UE Assistance Information can be reused and enhanced for the above purposes mentioned in the TR.

* **Q14: Related to UE-side model training, do you foresee any potential impact in existing RAN2 protocols, e.g. impact in the legacy RRC UE Assistance Information, that RAN2 should study? Please describe.**

|  |  |
| --- | --- |
| Company | Comments |
| OPPO | At least data collection initiated/triggered by configuration from NW may have RAN2 impact, but the details may have RAN1 dependency for CSI and BM use cases. |
| ZTE | We are open for UE to provide the UE assistance information to acquire their wanted measurement configuration for the UE sided model training, but it shall be discussed in normative phase. |
| Xiaomi | We would not prefer UE to do the model training due to limited capability on computation and storage. |
| vivo | UE may send the request for assistance information to the network, e.g., to categorize the collected data. And network may provide the assistance information to UE via dedicated signaling or broadcast SIB. |
| NEC | We think UE-side model training, the UE may need to report the information of the UE-side model via legacy message e.g. legacy RRC UE Assistance Information. |
| Apple | We agree that UAI may be used to send at least assistance information. However, as usual, such RAN2 work should be started only after reception if RAN1 LS. For now, we don't think RAN2 can do anything. |
| Intel | Not for now.  It is not clear at this stage what is the association between model and its data collection. It is possible that the network can defer the requirement for data collection configuration based on the UE reported model meta info, applicable condition, etc. For this case, there’s no need to let UE report additional duplicated preferred information to NW.  Hence, we prefer to wait until it is clear what the meta info will be. |
| LGE | Agree with OPPO |
| Spreadtrum | Depends on RAN1 input to analyze whether and/or what kind of training data is needed for UE side model training. And then the related signaling configuration and data collection assistance information can be considered. |
| Ericsson | When the UE needs to start/stop the training, this may be communicated to the network (e.g. via UEAssistanceInfo), and the network may then decide to allow the UE to start/stop the training based on e.g. availability of CSI/SSBs resources on which the UE needs to perform the training, on whether impacts on UEs ordinary measurements for RRM/mobility are expected etc. The gNB may also provide the necessary configuration needed for the UE to start performing the UE-side model training. Similarly when the UE has completed the training and needs to upload the collected data to the OTT server this may be communicated to the gNB so that the gNB can decide whether to allow the transfer of collected data, since that may have an impact on the UE/cell traffic, and hence in the system performances. |
| Fujitsu | Agree with OPPO. For the data collection of the UE-side model training, according to agreements from RAN1, either the initiated/triggered by configuration from NW or the request from UE (including supported RS set) have the RAN2 impacts which need to study. But which option is for the initiated/triggered data collection depends on RAN1 conclusion. |
| Interdigital | It is very likely that changes to UAI will be required/desirable. However, we agree with the view expressed by other companies that we can wait for RAN1 input before discussing this further. |
| Mediatek | If only UE-side model training is concerned, there is impact to RAN1 (and even RAN1), because model training is offline and implementation-specific.  But what important is still data collection to facilitate the model training at the OTT server.  Agree with vivo that UE may request assistance information for dataset categorizing.  If RAN node/OAM controls the data collection, the corresponding procedure is required. |
| TCL | We also agree with OPPO, based on the description“data collection initiated/triggered by configuration from NW”in TR 38.843, RAN2 may be able to study the corresponding specific impact. |
| CATT | UAI may be used for requesting the collection configuration or ground truth label (e.g. LMF generates location based on any positioning methods in Pos use case). |
| Sharp | For NW configured data collection, UE side model training and performance monitoring, the UE may need to report information related to the UE side model to the network. |
| Qualcomm | We agree with Oppo, Vivo, NEC, Fujitsu, and others, in that there may be a wider impact than UAI, especially as it relates to assistance information and needed configuration sent by the gNB. |
| Lenovo | Agree with many companies above UE may request gNB via UAI to provide assisting information /configuration that is useful for training at UE side. |
| China Unicom | Agree with OPPO. |
| CMCC | We are open to discuss it after receiving RAN1’s requirements. |
| Nokia, Nokia Shanghai Bell | We do not see any impact to RAN2 protocols related to data collection for offline UE-side model training with respect to RRC UE Assistance Information. For data collection for offline UE-side model training, there would be an impact to RAN2 protocols to carry the non-standardized data from the UE to the NW, and these are not possible via UAI or needForGap frameworks. |
| Samsung | Similar view to OPPO. |

Rapporteur summary:

20/22 companies believe that potential impact in existing RAN2 protocols should be studied to support the UE-side training. Many companies mention that some awareness at RAN level might be needed, e.g. for the UE to signal the need to perform training, for the gNB to provide necessary configurations to the UE to start performing training, etc. Companies also mention that further inputs/progress in RAN1 should be considered.  
The following is proposed:

1. Related to UE-side model training, RAN2 to study impacts in existing RAN2 protocols to enable RAN-awareness of UE-side model training (e.g. to initiate the data collection for UE-side model training, to provide the necessary configuration for data collection for the UE-side training, etc.), taking into account RAN1 progress/requirements.

------------------------------------------------------------------------------------------------------------------------------------------------

### 2.2.2 Performance Monitoring of UE-side models

Related to performance monitoring for UE-side models, the following is captured in the TR 38.843:

|  |
| --- |
| From TR 38.843:  For *performance monitoring* at the NW side, calculated performance metrics (if needed) or data needed for performance metric calculation (if needed) can be generated by UE and terminated at gNB. |

The methods for conveying the performance metrics or the data needed for performance metric calculations requires further discussion in RAN1 and RAN2. The Rapporteur believes that RAN2 could study the impact on RRC protocols for the reporting of calculated performance metrics or for the reporting of the data necessary for performance metric calculations.

* **Q15: Related to UE-side performance monitoring, do you foresee** **any potential impact in existing RAN2 protocols to convey data from the UE to the gNB with the calculated performance metrics or with the data needed for performance metric calculation? Please describe.**

|  |  |
| --- | --- |
| Company | Comments |
| OPPO | First, we’d like to clarify that UE sided model monitored by NW is one candidate option raised by RAN1, so it’s quite clear that this scenario may have RAN2 impact.  We also believe that L3 signaling is better and more future proof than L1 signaling for model monitoring metrics reporting. As for the reporting details, more RAN1 guidance is needed before RAN2 can go through the detailed discussion.  Another point is that RAN2 should clarify that NW based model monitoring can be applicable to both activated model and inactive model as RAN1 already made the following agreements:  *Agreement*  *For the purpose of activation/selection/switching of UE-side models/UE-part of two-sided models /functionalities (if applicable), study necessity, feasibility and potential specification impact for methods to assess/monitor the applicability and expected performance of an inactive model/functionality, including the following examples:*   * *Assessment/Monitoring based on the additional conditions associated with the model/functionality* * *Assessment/Monitoring based on input/output data distribution* * *Assessment/Monitoring using the inactive model/functionality for monitoring purpose and measuring the inference accuracy* * *Assessment/Monitoring based on past knowledge of the performance of the same model/functionality (e.g., based on other UEs)*   *FFS: Requirements for the assessment/monitoring to be reliable (e.g., sufficient data coverage during evaluation)*  *FFS: Additional aspects specific to the case where the inactive model has never been activated before, if any.*  So we suggest RAN2 to consider the following proposal for UE side model monitoring:  **Proposal: UE sided model monitored by NW can be applicable to both activated model and inactive model.** |
| ZTE | In general, both L3 signaling and MAC CE is applicable for this intention anyway from RAN2 perspective. However, so far as now, due to the lack of detail information about the model monitoring, we cannot identify any further impacts on the RAN2 protocol for UE to report the performance metrics and/or data for NW. |
| Xiaomi | As captured in TR, UE may need to report calculated performance metrics or data needed for performance metric calculation to gNB. Existing signaling may not support the required metrics or data. And the report procedure may also need to be enhanced. |
| vivo | See comments in Q12. |
| Apple | Not for now. We tend to think RAN2 study on data collection for inference and performance monitor should not start before reception of RAN1 specific requirement (i.e. RAN2 can only study data collection for offline training at this stage). |
| Intel | Without RAN1 further input on data content, typical data size and latency requirement, which are relevant to part B in RAN2 LS, though RAN2 might have some impact, we prefer to wait for RAN1 input and discuss based on reply to part B. |
| LGE | In case of monitoring, network may configure a threshold criterion to facilitate UE to perform model monitoring. However, if UE sends monitoring-related reports whenever conducting model monitoring operations, there would be a lot of power consumption in UE and signalling overhead. Note that several models can be activated at once for various functionality.  As many models are activated, many monitoring-related reports can be triggered.   * **RAN2 can consider when storing/reporting monitoring-related results**   Regarding to contents of the reports,   * **RAN2 can consider how to combine the monitoring-related results of Activated and Inactivated model-based results (align with RAN1 agreement)**   If the priority of the RRC message including monitoring-related information is always lower, it can lead to delays in LCM operation, such as model switching, (de)activation, fallback, etc.   * **RAN2 can consider adopting SRB depending on model-related information**   If the model is changed due to switching/update/etc, stored monitoring-related results can be removed in UE.   * **RAN2 can consider how to manage outdated monitoring information** |
| Spreadtrum | Prefer to wait for more RAN1 input and then to evaluate whether existing reporting framework needs to be enhanced or not. |
| Ericsson | We agree that discussion on performance monitoring (and inference) is very dependent on RAN1 progress. It could be that the metrics are conveyed in combination with the L1 measurements in L1 signalling, in which case very little impact is expected in RAN2 (maybe only in the configuration). Otherwise, metrics can be reported in RRC if more relaxed latency requirements are acceptable for performance monitoring. |
| Fujitsu | Yes.  For UE to perform model monitoring, the monitoring performance metrics may be configured between NW and UE. Besides, it is necessary for UE to report the calculated performance metrics to NW, or the requested data to NW for performance monitoring.  Therefore, it needs to clarify this discussion is about NW-side performance monitoring with UE-side model not about the UE-side performance monitoring.  For NW-side performance monitoring with UE-side model, at least the configuration for performance metric reporting has RAN2 protocol impacts. For the container for performance metric or data needed for performance metric calculation needs to be discussed in RAN1. And if the conclusion is the L1 signaling is the container, it falls to scope of RAN1 discussion. Otherwise, the RAN2 needs to consider the signaling to convey the data. |
| Interdigital | We agree with the view expressed by Intel. |
| Huawei, HiSilicon | Wait for RAN1 inputs. |
| Mediatek | Wait for RAN1’s reply on data collection requirement for model monitoring.  Even if the UE need to report calculated performance metrics, RAN2 needs to understand what the metrics would be. |
| TCL | In our understanding, for UE-side models, when monitoring is deployed in UE side, it may report the monitoring metrics or model output, which may be based on the model control locations, hence, at least the configuration for reporting of monitoring needs to be study. Similarly, if the monitoring resides NW side, it is natural to discuss the configuration and reporting of model monitoring. But it is safe to wait for RAN1 progress, and then RAN2 discusses any potential impact about model monitoring. |
| CATT | We should wait for RAN1 feedback. The RAN2 impacts may include configuration and reporting. And the impacts may be different with different use cases. |
| Sharp | As pointed by OPPO, UE sided model monitored by NW is one candidate option raised by RAN1 therefore we suggest that RAN2 should study signaling procedures considering network and device side performance aspects to monitor and report selective model and/or device specific or intermediate KPIs.  Further, the scenario for UE-side performance monitoring in which the UE independently monitors the performance (based on network (pre) configuration), makes decisions of model activation/ deactivation/updating/switching and may report it to the network can also be studied.  Where the monitoring metric and/or inference is calculated may also be considered as it may influence model performance monitoring and reporting signaling procedure. |
| Qualcomm | Wait for RAN1 input. |
| Lenovo | Wait for RAN1 input. |
| China Unicom | At least, UE should report the calculated performance metrics to NW, which has impact in RAN2. |
| CMCC | We prefer to wait for RAN1 input. |
| Nokia, Nokia Shanghai Bell | The reference provided from TR is confusing as it is for NW side and the section is discussing UE side. The intention is not clear here. Note that, the option for hybrid monitoring approach is also listed. We agree with Apple, Intel, HW, Qualcomm, and others that before RAN2 diving in signaling solutions, we should wait for RAN1’s LS response on part B. |
| Samsung | Wait for RAN1 input. |

Rapporteur summary:

* 8/22 see clear RAN2 impact on this.
* 14/22 prefer waiting for more RAN1 inputs/progress.

Given the above, the following is proposed.

1. Related to performance monitoring of UE-side models, impact (if any) in RAN2 protocols needs to be further evaluated, depending on RAN1 progress/inputs.

# Conclusion

**For the use case of CSI prediction/beam management with one-side model, the following is proposed:**

[Proposal 1 For training of NW-side models, both gNB- and OAM-centric data collection are considered in the study.](#_Toc146714672)

[Proposal 2 For training of NW-side models, the gNB-centric data collection implies that the gNB configures the UE to initiate the data collection procedure and the collected data are terminated at the gNB. To further study the details of the data collection configuration.](#_Toc146714673)

[Proposal 3 Related to gNB-centric data collection for NW-side model training, RAN2 studies the potential impact on L3 signalling for the reporting of collected data, taking into account RAN1 further inputs/progress.](#_Toc146714674)

[Proposal 4 Related to gNB-centric data collection for NW-side model training, the following principles can be considered for the L3 signalling reporting framework, if used:](#_Toc146714675)

[a. The L3 signalling reporting framework for NW-side model training should allow the UE to store sets of measurements and then report them to the gNB in multiple RRC messages (e.g. similar to the logged MDT).](#_Toc146714676)

[b. The L3 signalling reporting framework for NW-side model training should allow the UE to store multiple measurements taken at different points in time and report them in a single RRC report.](#_Toc146714677)

[c. The L3 signalling reporting framework for NW-side model training implies that the UE may be configured to report measurements periodically or upon fulfilling certain events.](#_Toc146714678)

[d. The UE memory/processing power/energy consumption/signalling overhead should be taken into account.](#_Toc146714679)

[Note: The above principles, for the L3 signalling reporting framework, can be revised depending on RAN1 progress/requirements](#_Toc146714680)

[Proposal 5 Related to gNB-centric data collection for NW-side model training, further principles for the L3 signalling reporting framework may be considered in RAN2, also taking into account RAN1 progress/requirements.](#_Toc146714681)

[Proposal 6 For training of NW-side models, an OAM-centric data collection implies that the OAM provides the configuration (via the gNB) needed for the UE to initiate the data collection procedure and the collected data are terminated at the OAM entity. MDT framework can be considered as baseline.](#_Toc146714682)

[Proposal 7 Related to OAM-centric data collection for NW-side model training, RAN2 studies the potential impact on the immediate MDT, taking into account RAN1 further inputs/progress.](#_Toc146714683)

[Proposal 8 Related to OAM-centric data collection for NW-side model training, RAN2 to further discuss motivations of the suitability of the logged MDT framework or other frameworks.](#_Toc146714684)

[Proposal 9 Related to OAM-centric data collection for NW-side model training, the following principles can be considered for the immediate MDT framework, if used:](#_Toc146714685)

[a. The Immediate MDT framework for NW-side model training should allow the UE to store sets of measurements and then report them in multiple RRC messages (e.g. similar to the logged MDT).](#_Toc146714686)

[b. The Immediate MDT framework for NW-side model training should allow the UE to store multiple measurements taken at different points in time and report them in a single RRC report.](#_Toc146714687)

[c. The Immediate MDT framework for NW-side model training should allow the network to configure the UE to report measurements periodically or upon fulfilling certain events.](#_Toc146714688)

[d. The UE memory/processing power/energy consumption/signalling overhead should be taken into account.](#_Toc146714689)

[Note: The above principles, for the immediate MDT framework, can be revised depending on RAN1 progress/requirements](#_Toc146714690)

[Proposal 10 Related to OAM-centric data collection for NW-side model training, further principles for the immediate MDT may be considered in RAN2, also taking into account RAN1 progress/requirements.](#_Toc146714691)

[Proposal 11 Related to OAM-centric data collection for NW-side model training, the principles of the logged MDT framework will be discussed if its suitability is assessed.](#_Toc146714692)

[Proposal 12 Related to performance monitoring of NW-side models, impact (if any) in RAN2 protocols needs to be further evaluated, depending on RAN1 progress/inputs.](#_Toc146714693)

[Proposal 13 If the UE-side OTT server is responsible for the UE-side model training, the way the OTT server collects data should not be studied in RAN2.](#_Toc146714694)

[Proposal 14 Related to UE-side model training, RAN2 to study impacts in existing RAN2 protocols to enable RAN-awareness of UE-side model training (e.g. to initiate the data collection for UE-side model training, to provide the necessary configuration for data collection for the UE-side training, etc.), taking into account RAN1 progress/requirements.](#_Toc146714695)

[Proposal 15 Related to performance monitoring of UE-side models, impact (if any) in RAN2 protocols needs to be further evaluated, depending on RAN1 progress/inputs.](#_Toc146714696)

# References

1. [R2-2308898](file:///C:\\Users\\mtk65284\\Documents\\3GPP\\tsg_ran\\WG2_RL2\\RAN2\\Docs\\R2-2308898.zip" \o "C:Usersmtk65284Documents3GPPtsg_ranWG2_RL2RAN2DocsR2-2308898.zip), Data collection for AI/ML, Ericsson

1. [R2-2308286](file:///C:\\Users\\mtk65284\\Documents\\3GPP\\tsg_ran\\WG2_RL2\\RAN2\\Docs\\R2-2308286.zip" \o "C:Usersmtk65284Documents3GPPtsg_ranWG2_RL2RAN2DocsR2-2308286.zip), Report of [Post122][060][AIML] Mapping of functions to physical entities (CMCC), CMCC