3GPP TSG-RAN WG3 #117-e R3-225014

Electronic Meeting, Aug 15th – 25th, 2022

Agenda Item: 12.2.2

Source: ZTE (moderator)

Title: [draft] Summary of CB: # AIRAN2\_Stage3

Document for: Approval

# Introduction

**CB: # AIRAN2\_Stage3**

**- Analyze and coverage on supporting input, output and feedback through existing procedures or new defined procedures**

**- Standard impact analysis on network energy saving, mobility optimization and load balancing**

**- Discuss the remaining issues, e.g., validity time**

**- Potential impacts over interfaces**

**- Capture agreements and open issues**

(ZTE - moderator)

Summary of offline disc [R3-225014](file:///D:\3GPP\RAN3%23114bis\TSGR3_114bis-e\Inbox\Drafts\CB%20%23%20AIRAN3_ES\Inbox\R3-225014.zip)

Two phases of this email discussion:

* Phase 1 Deadline: 23:59UTC, Thursday, 19th Aug (before online session starts).
* Phase 2 Deadline: 23:59UTC, Monday, 23th Aug. (before online session starts), we will try to come up with agreements in the 2nd phase discussion before online session.

# For the Chairman’s Notes

Propose the following:

Propose to capture the following:

**Agreement text…**

**Agreement text…**

**WA: carefully crafted text…**

Issue 1: no consensus

**Issue 2: issue is acknowledged; need to further check the impact on xxx. May be possible to address with a pure st2 change. To be continued…**

# Discussion

## Standard impacts on procedures for AI/ML function

Based on the TR37.817, majority companies in their contributions mention that for some information for three agreed AI/ML based use cases, (Network Energy Saving, Load Balancing, and Mobility Optimization) needs to be enhanced over Xn/F1/E1 interface.The information for AI/ML function is divided to input information, output information, and feedback information.

Since there is some overlapped information across these use cases (e.g., predicted resource status, current resource status, and UE performance, etc), there are two options as follows:

* Option 1: Enhance/reuse the existing procedure if the revelant procedure can be found. [16][21][23][24][38][39][40][44][45][46] [25][26][27]
* Option 2: Define the new procedure for the information used for AI/ML.[8][32][35][36][37][38] [50]

**Q1: Companies are invited to provide their views on which option above is preferred to support AI/ML function for the overlapped information (e.g. predicted resource status, current resource status, predicted UE trajectory and UE performance, etc )?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Which option?** | **Comment** |
| Lenovo |  | We are fine to go for either way, both have pros and cons as we discussed in our paper R3-224421. |
| Huawei | Option 1 preferred | Option 1 is simpler, technically we see pros and cons for both options, we are also open to further discuss new procedures, but in our understanding, we don’t see the need to introduce new procedure in E1/F1, maybe we could start from this point. |
| Qualcomm | Option 1 | Option 1 is preferred as mentioned in our paper. Resource Status Reporting procedure and Handover related procedures can be reused to transfer the AI/ML related data. If there a need to introduce new message, we are open to discuss them further. |
| Samsung | Option1 | According to the R17 discussion, most input/output/feedback information identified for AI/ML training or inference has a relevant existing procedure to collect. Hence, if applicable, taking existing procedure as baseline and enhancing it to support AI/ML functionality are convenient. |
| vivo | Option 1, FFS for Option 2 | It is straightforward to reuse the existing relevant procedures, e.g., reusing Resource Status request/response for the predicted resource status transfer.  If there is no relevant procedure, e.g., UE associated feedback, new procedures shall be introduced. |
| Ericsson | Option 2 | RAN3 has so far followed a modular design. That is the reason why the Resource Status Reporting procedures over Xn have been defined for a specific function, namely Mobility Load Balancing.  Mixing data used for AI/ML in a procedure that is today associated to a specific function (MLB) does not allow for a modular design. The same procedure needs to be used for different functions and subject to failures due to different reasons (nodes behaviours and processes). Troubleshooting problems becomes therefore more difficult. Dependencies between functions become more difficult.  Also, Resource Status Reporting procedures follow a structure that may not be optimal for AI/ML. For example, the Resource Status Request may need to be enhanced with a new “Cell To Report” list, indicating the cells for which AI/ML information are required, which might not be the cells for which MLB data are required. This makes the procedure more complex.  Also consider message sizes. Putting in the same message massive amount of information for MLB and AI/ML and on top signal these message frequently creates a big challenge at transport and implementation level. Smaller messages are easier to handle.  Finally, even in terms of compliance, it is more appropriate to define dedicated procedures. Reusing existing procedures would bring ambiguity to AI/ML compliance as support of the legacy procedures reused for AI/ML may be interpreted as compliance to AI/ML support.  We do not see the need for new procedures over F1 and E1. |
| CATT | Slightly prefer O2 | We have slight preference on option 2 while option 1 is also acceptable for us. |
| Intel | Option 1 for current status information, UE performance, feedback;  Option 2 for predicted information | For current information, e.g. current resource status, UE performance, other feedback information, since existing procedure already support some information exchanging, option 1 is preferred for current information.  However, for predicted information, since it highly depends on whether the requested NG-RAN node has such AI/ML capability to generate such information or not, a separate procedure for predicted information exchange is preferred.  Additionally, the nature of predicted information is also different from the current information, as discussed in R3-224770. The predicted data is available only after the results being generated from Model Inference, which is different from current information which are always available. The legacy procedure (i.e. periodically reported) may not be suitable for predicted information. |
| Nokia | Option 2 | We support defining new procedures to exchange AI/ML information. This option will make AI/ML functionality across different use cases simpler and it would also avoid extending each legacy procedure with new IEs and extra complication. We think it is a more clean way of introducing AI/ML functionality in the RAN. We agree with Ericsson on their analysis. |
| LGE | Option 2 | We think the existing procedure, e.g., Resource Status Reporting procedure, may not be appropriate for AI/ML, as Ericsson’s example above. |
| ChinaTelecom | Slightly Option 2 | We are fine with both options, but we are inclined to Option 2.   1. Goals of design: The existing procedures specified in XnAP, e.g., Resource Status Reporting procedures are applicable for SON MLB operation. It aims to acquire the long-term measurement result from the neighbor nodes. However, AI/ML procedures aim to acquire long term or short-term data or feedback information. Clearly, the design objective of MLB and AI/ML operation are different. 2. Extensibility: To support AI/ML Operation in more use cases, the procedure for AI/ML should have good extensibility. The signaling procedure we talked here is mainly leveraged for transmitting predicted information or dedicated AI/ML data, which involves timing information as well. Therefore, the most of the existing IEs in Resource Status Reporting procedures could not be directly reused in AI/ML operation. Too much data with different usage, e.g., load balance, ES, and Mobility included in Resource Status Reporting procedures makes the excessive amount of signaling message. Compares says, a new procedure is easier to achieve good extensibility.   Implementation &IoT issue: Here, we share the same view as Ericsson that troubleshooting problems becomes more difficult if MLB and AI/ML operation merged in a single signaling procedure. |
| Xiaomi | Option 2 or depends | Both are OK to us, considering AI/ML is new function in NG-RAN, option 2 may be more clear and future proof.  But we also acknowledge that some can reuse the existing procedure and some may need new procedure e.g. prediction related. |
| InterDigital | Slightly Option 2 | We think that Option 2 is better but have assumed option 1 in our discussion papers so either will work |

Following are the new procedures proposed by companies:

* Option 1: Define the new unified procedures for AI/ML function for input information, output information, and feedback information nformatio [39][40]:
  + AI/ML Data Collection Procedure (to collect historical nformation)
  + AI/ML Predicted Information Procedure (to transfer predicted information)
  + AI/ML Feedback Information Procedure (to retrieve feedback information)
* Option 2: Define two non-UE associated procedure for predicted information and UE performance [44][45][46]:
  + AI/ML Assistance Data Reporting Initiation (for handling the subscription mechanism)
  + AI/ML Assistance Data Reporting (for handling the collection of subscribed information)
* Option 3: Define two non-UE associated procedure, and two UE-associated procedure [16]:
  + Node Data Collection Initiation procedure and Node Data Collection Reporting procedure
  + UE Data Collection Initiation procedure and the UE Data Collection Reporting procedure
* Option 4: Define the new procedure only for predicted information [23][24]:
  + Predicted Information Request/Response/Update

**Q2: If you agree to define new procedures, companies are invited to provide their views on which stage3 signalling design option over interfaces above is preferred?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Which option?** | **Comment** |
| Lenovo | Option 4 | If we go for new procedures, it seems a cleaner option to define new procedure only for predicted information to distinguish from legacy procedures for the same .  A general question needs to be clarified if we call the new procedure “AI/ML xxx procedure” as in Option 1 and 2 is that, would that imply the carried information cannot be used for non-AI/ML purpose? E.g., the UE performance feedback after the HO. Similarly, would that imply other legacy procedures will not be used for AI/ML purpose, which doesn’t seem to be the intention. |
| Huawei | See comment | In our understanding, if we would like to introduce new procedure, this procedure should be common for all use cases (including possible new use cases in the future), and should be common for request/response for the data of different purpose, e.g. history info/predicted info etc. |
| Qualcomm |  | See our comment for Q1. Prefer Reuse.  But if there is a strong preference to use new message, we prefer a single Request/Response procedure for all the AI/ML use cases. |
| Samsung |  | If no relevant existing procedure, the detailed new procedure can be discussed case by case. |
| vivo |  | Revisit this issue when the Xn impact is concluded, e.g., feedback. |
| Ericsson | Option 2 | We believe that the new procedure to be defined should be use case and information agnostic. Namely, there should be a single procedure to setup and report all types of AI/ML data (inputs, outputs, feedback) and for all use cases.  The reason for this is that there is no gain at a functional level in creating different procedures for each use case and for each type of data. The only node that needs to know for which use case information is needed and the type of information that is needed is the requesting node. The reporting node only needs to know which data it should report. Therefore, there is no need to make a dedicated procedure per data type and per use case. The requesting and reporting nodes can communicate which data is needed via a single procedure.  We see that the new procedure could resemble the Resource Status Reporting Initiation and Resource Status Reporting over Xn. However, as explained in the previous question, we would like to have a dedicated procedure to make sure that resource status reporting is not impacted, given that this procedure has been used for many releases for MLB.  We also think that defining a procedure like in Option 2, which is based on the Xn: Resource Status Reporting is advantageous because it allows us to adapt to AI/ML, e.g. reporting periods values may be added depending on the use case, partial stops for specific cells may be adopted, etc. |
| CATT | O3 > O4 > O2 > O1 | For O1&2, We share the view of Lenovo that “AI/ML xxx procedure” is not a good name as they can be used for other purpose as well. That is the very reason why we call it “xxx Data Collection xxx procedure”.  O4 is acceptable for us anyhow. |
| Intel | Option 4 | As explained in Q1, the current data can be well supported by existing messages and procedures. The new procedure can be introduced for predicted information which can be commonly used by all use cases.  Regarding to Lenovo’s question, the data exchanged over new procedure does not prevent the received NG-RAN node to use it for other purpose, e.g. implementation, etc. |
| Nokia |  | In general, it seems premature to decide already at the very first meeting of this WI the needed procedures (whether we need 1 or 2 or 3 of those, whether they need to be UE associated or non-UE associated).  In our view, collecting data for Model Training may use existing reporting procedures. A lot of the currently identified input is based on existing information that can be exchanged over Xn interface between neighbours. The node providing information for Training may in some cases be completely oblivious on the purpose it provides the data.  When it comes to reporting of predicted information, we don’t think that existing resource status procedure can be re-used. For instance, the procedure should be able support longer reporting duration, possibly smaller granularity/period to make sure enough predictions are received. Also, it needs to be able to capture changes in inference output due to changes in the environment, currently not supported by existing resource status procedure. We agree that the “principle” of reporting predictions from neighbours should be based on a “subscription-based” solution where a node requests a certain prediction after which reporting is started, same for feedback information.  There may be need to exchange UE-associated information but at this point we haven’t identified the need to define UE-associated procedures. For instance, regarding UE Trajectory prediction we don’t think that a gNB will possibly keep trajectories for each UE, but some average will be calculated based on a number of UEs traversing a sequence of locations/cells. Thus, it seems at this point that non-UE associated procedures are sufficient. |
| LGE |  | We think the following non-UE specific procedures for reporting of AI/ML data are needed:   * The procedure that requests the AI/ML data to its neighbor nodes; * The procedure that reports the requested AI/ML data;   We also are not clear why a UE-associated procedure is needed. It is because the feedback information will be provided after the action (e.g., handover) is completed like SON report is sent over Xn. |
| ChinaTelecom | Option 2 | We prefer to define a unified procedure for one NG-RAN node to retrieve AI/ML related input, output and feedback data from other nodes, what exact the information contains in the input or output or feedback message, depends on the use cases. For instance, in case of Network Energy Saving, predicted energy efficiency and predicted resource status from neighbouring nodes are required. Besides, we think the new signaling could resemble the existing Resource Status Reporting procedure. |
| Xiaomi | Option 2 or option 4 | If a whole new procedure is needed, we prefer option 2.  If only define new procedure for predication, option 4 seems ok. |
| InterDigital |  | I think we might need to wait, for example how to handle feedback which would be post-handover and handling potential signalling as part of handover. I do agree that it should be at least a single procedure outside of current resource status procedure for feedback. |

## Network Energy Saving

Since, in RAN1 there is one study item “Study on network energy savings for NR”, [12][25][40] [51] propose to focus on the cell-level energy saving strategy (e.g., cell activation/deactivation), and avoid the overlapped discussion compared to RAN1 led SI.

**Prosposal: Regarding AI/ML based Energy Saving, cell-level energy saving strategy as a start point in RAN3, avoiding overlapped discussion for network energy saving in RAN1 SI.**

**Q3: Companies are invited to provide their views on the proposal above?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| Lenovo | Yes | It would be safer to keep the discussion on cell level to avoid overlapped discussion. |
| Huawei | Yes | We could start from cell level activation/deactivation; and of course, overlap with SI on ES should be avoided. |
| Qualcomm | Yes | We could start with cell level and avoid overlap with RAN1 |
| Samsung | Yes | The cell-level energy saving has been supported by the spec, and AI/ML can help to do the optimization. There is another SI network energy saving, and other granularities maybe will be studied in this SI. In current stage, it is better to take cell switch on/off as the starting point. Other granularities depends on the progress of network energy saving SI, whose corresponding AI for RAN impact can be delayed. |
| vivo | Yes | Follow RAN1’s conclusion. |
| Ericsson | See comment | First of all, cell level energy saving strategies are not limited to cell activation/deactivation. TR37.817 mentions “Efficient energy consumption can also be achieved by other means such as reduction of load, coverage modification, or other RAN configuration adjustments.”. for this reason we propose to remove (cell activation/deactivation) from the proposal.  We are fine to focus our efforts on cell level energy efficiency solutions, but we should also consider that any cell level energy saving action cannot be detrimental to UE energy consumption. Or at least the choice of cell level energy saving action that is detrimental to the UE should be taken in an educated way. This is why we suggest to also look at UE feedback when deciding which cell level energy saving action to take. |
| CATT | Yes |  |
| Intel | Yes | Agree with comments from above companies, overlapping with RAN1 new SI NES should be avoided. |
| Nokia | Yes | It is true that we should avoid overlapping discussions between groups. We should focus on cell level actions during this work. |
| ChinaTelecom | Yes | overlapping with RAN1 SI should be avoided |
| Xiaomi | Yes |  |
| InterDigital | yes |  |

Following information for AI/ML based network energy saving use case which has standard impacts are summarized based on contributions:

|  |
| --- |
| **Input Information** |
| 1. Predicted resource status information from neighbour NG-RAN nodes over Xn [1][10][25][37][40][44][48] |
| 1. Historical resource status [40] |
| 1. Current/Predicted Energy Efficiency from neighbour NG-RAN nodes [1][8][25] |
| 1. Current/Predicted energy state from neighbour NG-RAN nodes [1][8] |
| 1. Current/Predicted Energy efficiency score [44] |
| 1. Predicted overload status information over Xn/E1/F1 [25] |
| 1. Energy efficiency [23]: NG-RAN data energy efficiency, Network slice energy efficiency, including energy efficiency of eMBB, uRLLC and mIoT, PNF power consumption, including average power, minimum power, and maximum power, PNF energy consumption ,Energy state (high/low/active/inactive) |
| 1. Measurement frequency for reporting EE [37][51] |
| **Output Information** |
| 1. Predicted energy saving strategy (e.g., predicted cell switch-on/off decision) [25] |
| 1. Time for switch-off indication to show when the cell switch-off [25] |
| 1. Predicted time length indication that the cell will stay in the current activation/deactivation. [12] |
| 1. The predicted load transferring plan [25] |
| **Feedback** |
| 1. UE QoS parameters (Handover interruption time, UL Data Rate, UL Data Rate) [1][8] |
| 1. Per-cell total DL/UL UE throughput (i.e. “DRB.UEThpDl” and “DRB.UEThpUl” in TS 28.552)[16] |
| 1. Per-cell “Average delay DL air-interface” and the “Average delay UL on over-the-air interface” [16] |
| 1. RVQoE measurements [44] |
| 1. UE energy consumption [44] |
| 1. UE level performance (e.g., UL/DL throughput, packet delay, packet loss) [44] |

**Q4: Companies are invited to provide their views on which information listed above has the standard impact with priority order. And Elaborate more on the detailed standard impact.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Which?** | **Comment** |
| Lenovo | Input: 1, 3  Output: 9 or 11,  Feedback: 13 or 18, | 1: the predicted resource status can be carried in enhanced RESOURCE STATUS REPORT or new procedure depending on Q1.  3: new procedure is needed to exchange energy efficiency.  9 or 11: depending on if new procedure is used for transmitting the predicted e.g. cell activation/deactivation. At least it can be done in a way that when indicating the cell activation/deactivation as legacy, the gNB also indicate for how long the cell will stay in the activation/deactivation state.  13 or 18: After handover a UE due to energy saving, the UE performance after HO can be feedback to the source gNB. 13 and 18 look similar. |
| Huawei | See comment | Input: predicted resource status info 5, current/predicted EE 1  Output: predicted resource status info 5  Feedback: EE (energy efficiency), resource status info |
| Qualcomm |  | Input – 1, 6  Output – predicted resource status info, predicted cells to be deactivated/switched off and the duration  Feedback – 14  Energy Efficiency, Energy State and Energy Score are useful only when there is a standardized procedure on how to calculate them. Else the values provided cannot be interpreted correctly by the consumer.  UE Energy consumption is not needed to calculate a network energy efficiency. |
| Samsung | Yes for 1, 3, 6, 9, 10, 11, 12.  Yes but more description is needed: 13, 16, 18 | For 1, yes. It needs to collect predicted resource status information from neighbor node to generate ES decision to avoid local overload, switch on/off ping-pong.  For 2, it can be collected by existing resource status reporting procedure.  For 3, yes. Energy efficiency exchange can help to realize the global optimization.  For 4, it is better to start with cell switch on/off as Q3.  For 5, energy efficiency is more accurate.  For 6, yes. When predicted load status is high, nodes can send the predicted overload indication to peer node. In such mechanism, no request is needed so that the overload information can be transferred timely. For split architecture, DU and CUUP can send the predicted overload indication to CU.  For 7, the energy efficiency definition can refer to 28.310 as the Data Volume divided by the Energy Consumption.  For 8, the energy efficiency is kinds of resource status, and the detailed procedure can take the resource status reporting as the baseline.  For 9 and 10, yes. The energy saving strategy can be the action for a time point/period for future. For example, a node predicts it will be switched off in one minute. The node can exchange such as predicted cell switch-on/off decision with its neighbours to inform the action plan in advance, so the neighbour cells can take it as reference information to make proper decision (such as UE handover, load transferring, switch on/off and so on) to avoid the unnecessary handover, handover ping-pong, switch-off/on ping-pong, local overload etc.  For 11, yes. The predicted time information for the current activation/deactivation help the neighbor node to set SON decision to select target node to offload load or to handover UE.  For 12, yes. If deciding to switch off a cell, the existing load needs to be offloaded to neighbor node. AI/ML model generates the predicted traffic/load transferring action for a period for future. For example, a node predicts it will switch off in the future and generates the predicted handover strategy or load transferring plan in advance to avoid local overload and consecutive handover. If the target neighbor node can not accept the load, the node can make other proper candidate plans to guarantee the successful handover/transferring. Otherwise, the energy saving action may be delayed due to remaining load that has not transferred out successfully. Thus, it is beneficial for energy saving plan.  For 13, partial yes. The UL/DL data rate is fine, but handover interruption time is a little bit unclear.  For 14 and 15, the cell throughput is related to multiple factors, such as resource allocation policy, service type, UE mobility, etc. Thus per cell performance can not reflect the impact of energy saving decision directly.  For 16, yes but RVQoE should be the parameters of handed-over UEs due to energy saving decision. It can help to see whether the energy saving decision and related offloading policy are good or not.  For 17, UE energy consumption seems not in the scope of current energy saving. Now, it mainly focus on the base station energy saving.  For 18, yes but the UE level performance should be the parameters of handed-over UEs due to energy saving decision. It can help to judge the impact of energy saving decision. |
| vivo |  | Input: 1, 3  Output: none as all are internal output without standard impact.  Feedback: 14, cell level feedback is needed |
| Ericsson | Input: 1, 2, 3, 5  Output: 12  Feedback: 16, 17, 18 | With regards to Inputs it is worth mentioning that Input 5 Is simply an index expression of Input 3. We believe that expressing energy efficiency is sensitive over inter vendor interfaces, hence an indexed version of this metric would be easier to support as it does not reveal the exact energy efficiency of a node, but only how it improves or degrades.  With respect to Outputs: we consider output 9 as an internal output to the node running inference.  With respect to feedback: We see that feedback 13 is largely contained in feedback 18. We also see no technical reason to provide per cell level feedback (feedback 14, 15) because such feedback does not allow to check if the result of a per UE AI/ML action had a positive or negative effect. |
| CATT | Input:1  Output:9  Feedback:14,15,16 | 1: We think the predicted resource status from neighbor node could provide assistant information for the energy saving decision.  9: The main intention of AI for Energy saving is to make decision on cell on/off.  14, 15,16: This information could help the NG-RAN node understand whether the energy saving decision is good or not. |
| Intel | Yes for  Input: 1, 3, 4, 7;  Output: 10  Feedback: 14, 15, 18 | 1: the new procedure for exchanging predicted information can be used to carry 1 over Xn interface  3: current energy efficiency is carried over existing procedure, e.g. resource status report, while the predicted energy efficiency is carried over the new procedure. 7 is the same as 3, which gives the definition of how to calculate energy efficiency, as defined in TS28.554. Besides, power, energy consumption defined in TS28.552 (e.g. PNF power consumption, etc) can also be included as part of energy efficiency.  4: the energy state of neighbouring NG-RAN node is essential to for the requesting NG-RAN node to decide whether it can be selected as the target cell for the impacted UEs. The state of NG-RAN node can be exchanged over Xn interface together with energy efficiency.  7: it’s detailed definition of 3, as defined in TS28.554 and TS28.552.  10: the time for switch-off indication can be treated as validity time of when the energy strategy predicted strategy becomes valid. Since the output of model inference is a predicted information in the future, the corresponding time of when the behavior should taken place should be known by the node itself, as well as by the actors.  14/15/18: we believe those information can reuse the definition in TS28.552 and TS38.314. The performance information can be exchanged over Xn interface. |
| Nokia | Input: 1  Output: 9, 10 (12)  Feedback: 13,18 | 1.Predicted resource status information from neighbours can be sent over Xn.  2. We don’t support forcing a node to store historical data. In the availability of resource status reporting, historical load can be created at a node interested to “consume” this information.  4. Energy State seems redundant if the only options are that a cell is “on” or “off”. The way it is currently captured in the TR is confusing as it involves other states as well (active, high, low, inactive).  3, 7: We don’t support to exchange energy efficiency information between neighbours since in our view this is internal information of each gNB. Besides, the parameters required to calculate the Energy Efficiency, e.g. Data Volume (DV) and Energy Consumption (EC), other parameters may be needed to interpret variations in the Energy Efficiency KPI values from different network nodes. These can be classified into different classes dependent on demography, topography and climate and describe the network characteristics with regard to population density, geographical conditions and climate zones.  5: We agree with the intention of the energy efficiency score not to expose energy efficiency information but it is unclear it can be useful if a node cannot “exactly” interpret this information. An energy related score/cost could be useful, e.g., to capture what is the cost associated to adding traffic to a node or subtracting traffic from a node but could be difficult to be understood by the neighbours unless it is expressed in absolute values, e.g., additional energy consumed by adding X GBs of traffic to a cell.  6. Predicted overload could be determined at a node by using resource status from a neighbour and calculate a predicted load. No need to introduce a new predicted overload status information over Xn/F1/E1.  9,10: If a node predicts to switch off one of its cells it can indicate this information to its neighbours (using existing e.g., NG-RAN Node Configuration Update procedure), in which it can include a “Deactivation” indication and a time in the future or a time window when this will happen.  12: Seems to have the same effect as sending a predicted energy saving strategy (e.g., switch off a cell) in advance (9) with some timing indication (10).  13,18: Both seem related to feedback information related to UE performance for those UEs being handed-over from source to target. |
| ChinaTelecom | Input: 1, 3  Output: 9, 10  Feedback: 13,14 | 1: Predicted resource status information should be transmitted over Xn/F1 interface for ES decision making.  3: Current/Predicted Energy Efficiency is essential for network node to make analyze and optimization, and the information exchange need signaling enhancements.  For 9 and 10, the purpose of the ES use case is using AI/ML tool for making optimized ES decisions, the corresponding ES strategy and execution time is essential.  For 13 and 14, we think UE performance and cell throughput are important factors to be considered as feedback. |
| Xiaomi | Input: 1  Output: 9, 12  Feedback: 16, 18 | For input, we don’t see much value on the others except 1.  For output, we think Predicted energy saving strategy is the baseline, and it would be good to have the predicted load transferring plan.  For feedback, we think RVQoE can reflect the real performance of UE, however, RVQoE may not be used in some cases, where UL level performance metrics can be used. |
| InterDigital | Input 1, 3 or 5, 4  Output 12  Feedback 13, 16 | OK on use of 3 or 5 agree with Ericsson it is basically the same.  Agree with Ericsson that 9 is internal output 12 would be the external one  For feedback 13, but 18 is a similar set, probably should come to a conclusion on which parts of 13/18 are a good set – the agreed set would be feedback for the load balancing use case and probably the handover optimization case.  Feedback 16 yes as it would help understanding if the ES decision was good. |

## Load Balancing

Following information as input which has standard impacts for AI/ML based load balancing use case are summarized based on contributions (except the overlapped information e.g., predicted own resource status and predicted resource status from neighbouring NG-RAN nodes):

|  |
| --- |
| **Input Information** |
| 1. Predicted UE performance received from a target NG-RAN node [52] |
| 1. Predicted or measured UE traffic over E1 [11] |
| 1. Current and Predicted resource status information of neighbouring NG-RAN node(s) |
| **Output Information** |
| 1. Predicted resource status information of neighbouring NG-RAN node(s) [6] |
| 1. Predicted load balancing strategy [26] |
| 1. Indentification of an incoming handover for the purpose of AI based load balancing [36] |
| **Feedback** |
| 1. UE level performance metrics (e.g., UL/DL throughput, packet delay, packet loss) [36][45][52] |
| 1. RVQoE measurements [45] |

**Q5:** **Companies are invited to provide their views on which information listed above has the standard impact with priority order. And Elaborate more on the detailed standard impact.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Which?** | **Comment** |
| Lenovo | Input: 2  Feedback: 6 | 2: In case of CU-CP CU-UP split architecture, some enhancements are needed for CU-CP to receive either actual UE traffic measurement, e.g., data volume from CU-UP to make a prediction, or CU-CP requests the CU-UP to provide a prediction result. |
| Huawei | See comments | Input: predicted resource status information (output from neighbor node), 3  Output: predicted resource status information, 3  Feedback: UE level performance metrics as feedback info, 6 |
| Qualcomm |  | Input – 2, 3  Output – 4, Mobility Actions  Feedback – 6 |
| Samsung | Yes for 1, 2, 3, 5  Yes but more description is needed: 7, 8 | For 1, yes. Predicted UE performance from target node can help source node to choose proper target node to offload load to guarantee the UE performance during handover.  For 2, yes. CUUP has the information of UE traffic information. E1 impact should be studied.  For 3, yes. The current resource status of neighbour node can be collected by existing resource status reporting. The predicted resource status reporting can take the existing scheme as the baseline.  For 4, the node can collect the resource status of neighbor nodes and then do the prediction to provide the reference information for SON decision. However, there is no need to transfer the predicted status back to the neighbor nodes.  For 5, yes. Same as energy saving. AI/ML model generates the predicted traffic/load transferring action for a period for future. For example, a node predicts it will switch off in the future and generates the predicted handover strategy or load transferring plan in advance to avoid local overload and consecutive handover. If the target neighbor node can not accept the load, the node can make other proper candidate plans to guarantee the successful handover/transferring. Otherwise, the energy saving action may be delayed due to remaining load that has not transferred out successfully.  For 6, it is a little bit unclear about the identification of an incoming handover.  For 7, yes but the UE level performance should be the parameters of handed-over UEs due to load balancing decision. It can help to judge the impact of load balancing decision.  For 8, yes but RVQoE should be the parameters of handed-over UEs due to load balancing decision. It can help to see whether the related offloading policy are good or not. |
| vivo |  | Input: 2, 3  Output: none as all are internal output without standard impact.  Feedback: 7 UE level performance feedback. |
| Ericsson | Inputs: 3  Output: 4  Feedback: 7, 8 | Concerning Inputs: We believe it is not possible to express input 1 because the neighbor RAN is not able to derive the radio conditions and therefore the performance of the UE once handed over to the target cell  Concerning Outputs: We consider the predicted load balancing strategy (output 5) as a node internal output. Output 6 is also in a way node internal as it may come as the inference output of the IA/ML function |
| CATT | Input:2  Output :4  Feedback:7 | 2: predicted and measured UE traffic is useful when deciding which UE should be offloaded  4:This is natural the output of load balance use case  7:RVQoE provide assistant information on whether load balancing decision is proper or not. |
| Intel | Yes for (after number revision)  Input: 4  Output: 4  Feedback: 7 | Following what agreed in TR37.817. |
| Nokia | Input: 1, 3  Output: 4  Feedback: 7 (with a clarification) | 1: Predicted UE performance from the neighbour could be based on QoS performance prediction of the UE’s PDU sessions. For instance the target could predict the probability that a certain PDU session is dropped after a Handover. Or a probability with which the QoS requirements of a PDU session will be violated after the handover. In general this can be a prediction in a performance degradation that a UE may possibly experience after a handover to a target.  3: Current and Predicted Resource Status information is already in the agreed input from the neighbour in TR 37.817.  4: Predicted Resource status information of “neighbouring nodes” will be consumed internally so no standards impacts are foreseen.  7: UE level performance metrics (e.g., UL/DL throughput, packet delay, packet loss) “of handed over UEs” can be sent in the feedback as already agreed in the TR. |
| China Telecom | Input: 2, 3  Output: 4, 5  Feedback: 7 | 2: predicted and measured UE traffic is important and the E1 Interface case should not be excluded.  3: essential information to be considered for making LB strategies, and we suggest a new procedure to transmit the related information.  For 4 and 5, they are basic output for Load Balancing use case. |
| Xiaomi | Input: 1, 2 or 3  Output: 5  Feedback: 7, 8 | For input, we think UE traffic can be reflected by resource status or vice visa.  For output, we think Predicted load balancing strategy covers the other two.  For feedback, same as the comments in 3.2 |
| InterDigital | Inputs 3  Output: 4  Feedback 7, 8 | As stated in previous point, feedback 7 should be aligned with the UE QoS feedback for energy savings. |

## Mobility Optimization

Following information as input which has standard impacts for AI/ML based Mobility Optimization use case are summarized based on contributions (except the overlapped information as predicted own resource status, and predicted resource status from neighbouring NG-RAN nodes):

|  |
| --- |
| **Input Information** |
| 1. Predcited UE traffic over Xn/E1 [5][8][13][16] [27] |
| 1. Predicted QoS performanceo of UE. [53] |
| 1. Prediction regading an expected (incident) load from target NG-RAN node.[53] |
| 1. RSRP prediction from UE [20] |
| **Output Information** |
| 1. UE trajectory prediction over Xn (e.g., predicted serving cell, …) [8][13][16][27][41] |
| 1. Handover execution timing [5][23][35] |
| 1. Estimated Arrival Probability in CHO [5][16][23] |
| 1. Predicted resource reseveration time window for CHO [5] |
| 1. Predicted priority of selecting predicted target cell [23] |
| 1. Confidence level [23] |
| **Feedback** |
| 1. Existing SON reports [35] |
| 1. QoS parameters of handed-over UE [27] |
| 1. UE level performance metrics (e.g., UL/DL throughput, packet delay, packet loss) [16] [46] |
| 1. RVQoE measurements [46] |
| 1. Actual UE trajectory [13] [20][41] |
| 1. Actual UE traffic [13] [20] |

**Q6: Companies are invited to provide their views on which information listed above has the standard impact with priority order. And Elaborate more on the detailed standard impact.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Which?** | **Comment** |
| Lenovo | Input: 1 with clarification,  Output: 5, 10  Feedback: 12 or 13, 15, 16 | 1: measured or predicted UE traffic sent from CU-UP to CU-CP can be input for mobility optimization. In another scenario, the target gNB can receive predicted UE traffic over Xn interface from source gNB and use it for further mobility optimization.  5: the target gNB can receive predicted UE trajectory over Xn interface from source gNB and use it for further mobility optimization.  10: confidence level could be provided together with the prediction result.  15, 16: In one scenario, assuming the source gNB has made prediction on UE traffic and UE trajectory and then handover the UE to another gNB, the source gNB needs to understand if the previously made prediction is correct or not. Thus, some method is needed for the source gNB to get the actual UE traffic/trajectory after the handover. This can be considered as part of the AI/ML model performance monitoring. |
| Huawei | See comments | Input: predicted resource status information, predicted trajectory  Output: the timestamp of the HO, 6  Feedback: UE level performance metrics as feedback info, 13 |
| Qualcomm |  | Input – 1, predicted resource status information, predicted trajectory  Output – 1, predicted resource status information, predicted trajectory, Mobility Action and 6  Feedback – 13 and 14  RSRP prediction from UE, Actual UE trajectory, Actual UE traffic should are not needed for input and feedback. |
| Samsung | Yes for 1, 2, 5, 11, 12  Yes but more description is needed: 13, 14 | For 1, yes. UP has the knowledge of traffic, so that UP is a proper location for traffic prediction. UP can do the prediction to provide reference info to set resource allocation policy, and UP can transfer the traffic prediction results to CP to assist CP to set HO decision.  For 2, yes. The predicted QoS information of UE can help source node to choose the proper target node to guarantee the UE performance during handover.  For 3, it needs more clarification of expected (incident).  For 4, at current stage, it is better to not involve UE inference.  For 5, yes. The predicted trajectory helps target node for further mobility optimization. For high-mobility UEs, the collected data amount of a small-coverage cell (i.e. mmW cell) is small, which is not sufficient for AI/ML model to generate accurate predicted UE trajectory information. And the node capability to support AI/ML model is diverse, so there may exist some nodes with no ability for AI/ML model inference or the function for UE trajectory prediction. Thus, exchanging predicted UE information is beneficial for node to get the accurate predicted UE position information and set the proper further mobility optimization strategy to improve the handover robustness and efficiency.  For 6 and 8, they can be used internally to choose the candidate cells for CHO. If transferring the handover execution timing and predicted resource reservation time window to the target node, the target node may release the reserved resource based on the handover timing or the received time window. As the model inference result can not achieve 100% accuracy, the UE finds the condition is met just after node releasing the resource, so the CHO is failure.  For 7, it is already supported by current spec.  For 9, it can be used internally to choose the candidate cells and set the execution policy. It seems no need to transfer such information.  For 10, it needs to be discussed case by case.  For 11, yes, the SON report can be enhanced to carry the feedback information.  For 12, yes. QoS parameters of handed-over UE can help to judge whether mobility decision is good or not.  For 13 and 14, yes but should be the handed-over UE. Same reason as 12.  For 15 and 16, UE trajectory prediction and UE traffic prediction are help to set the handover decision. When the QoS/QoE performance of handed-over UE is good, it means the handover decision is good, so that the UE trajectory prediction and UE traffic prediction results are proper. Hence, to save the signaling, there is no need to transfer actual UE trajectory and actual UE traffic. |
| vivo |  | Input: 1 Predicted UE traffic over E1. 4, RSRP prediction from UE  For the 1 over Xn, the predicted UE traffic shall be internal output at the source node. 4 is quite beneficial to address HO to wrong cell issues via initial simulation.  Output: 7 Predicted Arrival Probability for CHO.  Feedback: 15, 16. |
| Ericsson | Inputs: see comments  Feedback: 11, 13, 14 | Concerning Inputs: The following are all inputs already captured in the TR which we consider valid: predicted and current resource status are inputs, UE information such as UE history information, UE measurements (not only RSRP), UE position.  Concerning Outputs: Output 7 is already available in the standard.  Output 6 does not make technical sense because an HO is triggered when an HO triggering event occurs. It is not possible to state that an HO shall be triggered in the future without validating that the HO triggering conditions are satisfied.  Concerning feedback: it seems that feedback 12 and feedback 13 are the same. Concerning Feedback 15, this is a very sensitive piece of information which is even difficult to express. Should it be expressed in terms of cells the UE goes through? If yes, then we already have that information in the form of the UE history. The source knows the UE history already because it can add to the UE history it has the HO target cell. |
| CATT | Input:1  Output:5, 7  Feedback:12,13,14,15,16 | 1: During handover or SN addition, providing the predicted UE traffic may help the target or SN to configure the radio resources better.  5: handover target node can make future UE trajectory prediction based on the received previous UE trajectory prediction.  7: CHO target cell may decide not to accept the CHO handover request in case of low Arrival Probability if overloaded.  12,13: QoS parameter or UE performance, such as UL/DL throughput, packet delay, packet loss, shall be considered after handover.  14: RVQoE can be seen as another type of UE performance parameter, so it shall be sent back to source.  15,16: agree with Lenovo. |
| Intel | Yes for:  Output: 6, 7, 8, 9, 10  Feedback: 11, 13, 16 | 6: handover execution timing is treated as validity time of the handover strategy  8: the time window for CHO could help to reduce the measurement complexity at UE side, as well as help to optimize resource reservation at the CHO candidate cell. |
| Nokia | Input: 2,3  Output: 5,7,8  Feedback: 11,12, 13 | 2: see, above.  3: Resource status procedure provides information on load at a gNB level. We think that having information on predicted load that a given UE is expected to cause (incident load) at a given node is useful to give a smaller granularity of control  4: We haven’t agreed that a UE can have AI/ML functionality.  Outputs 5,7, and 8 have been agreed in the TR 37.817 and we agree to them. Regarding UE Trajectory prediction, it may be internally consumed by a node (though could be agreeable to send it to a neighbour node). In our view UE Trajectory prediction is an average over a large number of trajectories providing sequences of locations/cells and corresponding time spent in those cells and not a “per UE” information.  6: Handover execution timing for normal Handover is not very meaningful since the handover will be executed when the event is met and this cannot be predicted (typically the horizon is too short for an accurate prediction).  9: On predicted priority, we are not sure what a target cell can do by knowing this information.  11: We support sending existing SON Reports. For sending those, existing procedures can be used.  12, 13: It seems from the corresponding references that both these are related to the UEs that have been handed over to a target gNB |
| LGE | Input: 1  Output: 5, 10  Feedback: 11, 12, 13 | 1: The target node or the secondary node that receives the predicated UE traffic can allocate the proper resource.  5: The target node uses the received UE trajectory prediction for further mobility optimization.  11: The existing SON reports can be used to carry the feedback information.  12, 13: After HO, based on QoS parameters and UE performance for the handed-over UE, we can check whether the AI/ML model is working well, for example, the source node may know whether the prediction for UE traffic or UE trajectory is accurate. |
| China Telecom | Output: 5, 6,7  Feedback:12,13 | 5, 7, 12 and 13 have been agreed in TR 37.817 and this time we still support them.  6: handover execution timing seems like validity time and should be considered as output. |
| Xiaomi | Input: either 1+ 3 or 2  Output: 5 can be baseline, others seems enhancements  Feedback: 12/13 or 14 | For the input, we think the predicted QoS performance may already consider the UE traffic and load of the NG-RAN node.  For the output, we think UE trajectory prediction can be a baseline, for other enhancements, we think 6,8 would be useful.  For the feedback, we don’t see much difference between 12 and 13, either 12 or 13 is fine, RVQoE is needed anyway. |
| InterDigital | Inputs 1  Output 5, 6, 7, 8  Feedback 11,13, 14, 15 | For outputs 6 is the estimated handover execution time, that is very useful for avoiding ping-pong and to optimize resource reservation (e.g. data forwarding). To address Samsung having the estimated handover execution time wouldn’t in a good implementation cause the release of resources but optimize data forwarding.  For Feedback, not sure what the difference is between 12 and 13 since 13 is QoS feedback… In any case this UE QoS feedback should be aligned with the other use cases |

## Validity time

Validity time has been discussed several times during SI phase, and it is left to be discussed in the normative phase. [2] thinks that validity time is a local node model output with no standard impacts unless an output is sent to another node in assistance, or an action is triggered to be initiate in another gNB or the UE. [8] think that validity time is included for each predictions as optional IE. [10][41] propose to include the information related to validity time (e.g., requested time and limited time, start time and duration) in the request message. [20] thinks validity time should be needed for the external output. [47] thinks validity time of the predicted data (model inference output) should also be included in the messages of requesting/responding/updating predicted data between the current and neighbouring NG-RAN nodes. In [45] it is argued that when an NG-RAN node requests predictions to another NG-RAN node, it is understood that the request includes the time horizon or validity time of such prediction in the form of the Reporting Period, hence the validy time is expressed via the reporting period.

**Q7: Companies are invited to provide their views on whether the validity time is needed for the predicted information or decision? If agree, should it be explicitly indicated in request/response/update message?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| Lenovo | Yes | It would be very strange, if the prediction receiving node doesn’t know the prediction is about what time point in the future.  It seems most reasonable for the requesting node to indicate some time information when requesting the prediction, e.g., the peer node should provide prediction result about what time point or time window. |
| Huawei | See comment | We are open to discuss the necessity of validity time, technically validity time could take effect when inference output is sent to neighbor node, however one could argue that new inference output could be generated before validity (i.e. validity time is not needed); while on the other hand, if further inference output could not be provided in time, some implementations have to be considered at target side. |
| Qualcomm |  | As mentioned in our paper, Validity time is needed. Any predicted information should have a validity associated to it.  Response and Feedback messages where predicted data is sent, a validity time should also be sent along. |
| Samsung | Yes, maybe can be discussed case by case | For prediction information, it should show for which future time interval that the information is valid. If applying the predicted information in a wrong time period, the network performance may downgrades.  Maybe it can be discussed case by case. Suggest to identify which inputs/outputs/feedback have the spec impact and then to study whether to need validity time for each of them. |
| vivo |  | The model output validity time is needed for the external output that does not associate with a timestamp. That is, if one output is for internal use or is already associated with a timestamp, the validity time is not needed.  It should be explicitly indicated in the response/update message. |
| Ericsson | See comments | Validity time needs to be discussed use case by use case. We believe that the validity time can be deduced from the Reporting Period. Namely, if a node reports predictions with a given periodicity, the prediction should be valid for that period, after which the prediction is updated. |
| CATT | Somehow | There are two types of prediction.  One is e.g. the next cell which the UE will move into. This type of prediction is used either at the node which makes the prediction itself (typical scenario) or instantly at the target cell. There is no need for any time delivered over RAN3 interfaces. (A validity time may anyhow be generated, but without RAN3 impact who cares?)  The other is time series, e.g. a list that 10s later the UE will be at Place A and 20s later the UE will be at Place B. The information “10s later” “20s later” should be delivered of course, but we don’t know whether it should be called “validity time”. |
| Intel | Yes | As discussed in our paper R3-224770, the validity time can be defined as the time when model inference output becomes valid in the system. The validity time is also important for the requesting NG-RAN node to use such information to predict for the corresponding timing of its own prediction. |
| Nokia |  | We think that validity time is model-dependent information. For different input information provided to Model Inference, its validity time will be different. For instance, a predicted UE trajectory over a cell granularity (input is a cell id) will have a much larger validity time than a predicted UE Trajectory over detailed location information (input is GNSS coordinates). But once the input/output of the model is fixed, the validity time of a prediction remains fixed. If a gNB subscribes to receive information about a ML Model supported by a neighbour, then for a given input and requested output the validity time will be fixed. This can be done without revealing any proprietary information about the model. Whether to include or not the validity time depends on the reporting procedures. Therefore we think that this topic can be solved/ revisited later, when we discuss reporting in detail. |
| LGE |  | Depending on that which information is predicted by the AI/ML model inference, the validity time may or may not be necessary. So, according to the predicted information, it can be discussed case by case. |
| China Telecom | Yes | We think validity time is necessary and should be investigated case by case. As for the predicted information, the current NG-RAN node and neighbouring NG-RAN node should be aware of its corresponding validity time to make better use of related information, otherwise, due to delay caused by data request and response processes, nodes that receive prediction information won’t know how to make use of that information. The decision information might be considered as internal output sometimes and not needful to be outputted with validity time together. Therefore, further discussions for validity time is needed. |
| Xiaomi |  | We’re fine to further discuss validity time case by case. |
| InterDigital |  | Our paper 2 defines a proposal which I will modify based on the discussion so far:  Validity time is a local node model output with no standards impact unless an output is sent to another node in assistance, *or an action is triggered to be initiated in another gNB or the UE.* The validity time may be able to be inferred from reporting periods or timestamps.  The part in italic here isn’t absolutely needed since it is not applicable to any of the current use cases but might be applicable in future use cases. |

## MDT enhancement

There is one Note in the objectives that specify MDT enhancement if needed.

[41] proposes to enhance the MDT procedure to support the consecutive AI/ML data collection for the certain time-series AI/ML model. When UE location information needs to be leveraged as input, model training or model inference in NG-RAN node shall collect the historical UE information including UE location information in the past period of time. And [47][54] thinks the existing MDT and RRM measurement procedures are re-used for data collection for AI/ML in NG-RAN without further enhancement.

[13] asks RAN3 to discuss mechanisms for the old NG-RAN node, that has made UE trajectory prediction before transferring UE to RRC Inactive/Idle state, to obtain logged UE trajectory information when UE enters RRC Connected state and reports to the new NG-RAN node. Such that the old NG-RAN node can understand how accurate the prediction was, and retrain the AI/ML model if necessary.

[49] finds that EM can directly send the m-based MDT activation to gNB-DU or gNB-CU-UP without gNB-CU-CP involvement, and the gNB-DU or gNB-CU-UP can send the MDT report to TCE without gNB-CU-CP involvement. And ask RAN3 to discuss how the gNB-CU-CP obtains the MDT measurement that is directly activated to gNB-DU or gNB-CU-UP by EM.

[56] argues to enhance the current MDT based activation to enable a more granular selection of UEs based on enhanced MDT Configuration information. It also proposes to discuss how to map AI/ML Feedback information to AI/ML Actions and report them over MDT

Moderator concludes the potential MDT enhancements proposed by companies below:

1. Enhance the MDT procedure to solve the issue how to support the consecutive AI/ML data collection for the certain time-series AI/ML model.
2. How the gNB-CU-CP obtains the MDT measurement that is directly activated to gNB-DU or gNB-CU-UP by EM
3. How the source NG-RAN node obtains logged UE trajectory information when UE enters RRC Connected state and reports to the new NG-RAN node.
4. How to enable a more granular selection of UEs based on enhanced MDT Configuration information in management based MDT
5. How to map AI/ML Feedback information to AI/ML Actions and report them over MDT

**Q8: Companies are invited to provide their views on whether the MDT enhancement proposed above is needed?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comment** |
| Lenovo | Yes | The raised MDT related issues seem reasonable and would be good to resolve from data collection and model performance monitoring point of view. The exact solution can be further discussed. |
| Huawei | Maybe not | From the discussions happened so far, no new measurement quantity is needed, which means that there is no need to enhance MDT. In addition, using MDT is for the purpose of training, the more important point for training is the amount of data collected, not the time spent for data collection. |
| Qualcomm | NO | For 1st bullet – As mentioned in our paper R3-224308, no new UE measurements are identified for AI/ML. Also MDT measurements from the UE can be reused for AI/ML purposes. We do not understand what is meant by consecutive AI/ML data collection.  2nd bullet – CU-CP can request for MDT information from DU or CU-UP as and when needed for AI/ML training. MDT information need not be sent to CU, whenever it is enabled in DU or CU-UP.  3rd bullet – UE Trajectory prediction can be done at NG-RAN based on UE history information and the HO related information. UE need not send the logged or predicted trajectory to the network. This simply complicates the procedure and needs RAN2 and SA3 interactions. |
| Samsung | Yes for 3 | For 1, yes, but it seems belong to RAN2 scope.  For 2, the current MDT procedure to collect location info is CUCP to send configuration to UE via RRC signaling. So there is no need of additional activation.  For 3, yes, the way for source node to get logged location info needs to be discussed. |
| vivo | Yes for 1 | The MDT enhancement shall focus on the new measurements or new formats. The consecutive AI/ML data is a new format of existing measurements. |
| Ericsson | Yes for 4 and 5 | 1. seems to be already possible by setting MDT traces for a given AI/ML process.  2. is based on a wrong assumption, which is that MDT is used to gather data at the RAN. This is not the case as MDT is used to gather data at the OAM, hence the issue is not valild.  3. is not justified by any use case. |
| CATT | 1. Slightly yes 2. Prefer no 3. Neutral | 1. In our understanding this bullet aims to collect more information than legacy MDT, e.g. periodical UE geographical coordination. Although its use may be limited to only Rel-18+ UEs, we may have to make such enhancement someday so why not earlier? 2. Well, we are RAN3, don’t bother the EM unless necessary…And another drawback is that the granularities in TS 28.552 are often too fine for RAN AI/ML use cases, increasing the signalling load. |
| Intel | Not at this stage | So far, we don’t see a strong motivation to enhance MDT procedure. Existing MDT procedure can be reused for AI/ML data collection.  But we are open to study UE selection to reduce overhead of data collection. |
| Nokia | 1,2,3: Solutions are not necessary to enhance MDT | 1. UE can provide the RAN with UE location information according to agreements. UE location information can be provided by RRM measurements, that do not need to be retrieved by MDT. Obtaining detailed location information in OAM  2. The agreed UE measurements in TR 37.817 are location information and radio measurements related to serving cell and neighbouring cells associated with UE location information, e.g., RSRP, RSRQ, SINR. These are RRC measurements configured by CU-CP. Configuring MDT to DU or to CU-UP is not needed here.  3.At this stage at least we can acknowledge the underlying problem, i.e. that the source node would need the further trajectory information but can't currently get it.  5.The issue on how to map AI/ML feedback to actions is more general than just limit it to MDT, which assumes AI/ML Training in OAM. Mapping AI/ML feedback to actions is a valid problem also when AI/ML Training is in the RAN, so maybe a broader scope for this is needed.  4. We are not yet convinced about this problem. If we solve the problem of mapping AI/ML actions to feedback, doesn’t this also help to identify UEs in a more granular way? |
| China Telecom | Maybe not | Agree with Huawei. At this stage we should focus on the signaling design and the details of input/output/feedback information for each use cases. We don’t see any need to enhance the existing MDT. We perfer this WI has no RAN2 impact in Rel-18 |
| Xiaomi | Yes for 2 | Some clarification for 2.  In our understanding, MDT measurements in DU or CU-UP can provide the throughput, packet delay, packet loss and UL measurements of the UEs, which can be used as Input data and feedback data as mentioned in the use cases in 3.2, 3.3 and 3.5, for these data, the more the better, it would be beneficial for AI training and inference. |
| InterDigital | Yes for 1 and 4, possibly 5 | 1.the issue of consecutive data does not relate to MDT only. It is intrinsic to ML that input data should follow a certain granularity (time gap between samples) and the data points should have continuity (consistency of the ) for proper training. The problem has therefore a broader scope. For connected UEs, the current measurement reporting framework is insufficient because it will either not collect data properly or it will increase signaling significantly. This is detailed in our paper [3]  However, the discussion on UE measurements in general has not been broached in this discussion we are fine with discussing further at a later time. |

## Others

**Q9: Any other essential issues are needed to be discussed in this meeting?**

|  |  |
| --- | --- |
| **Company** | **Comment** |
|  |  |
|  |  |

# Conclusion, Recommendations [if needed]

# References

1. R3-224231 AI/ML parameter Open Issue List Discussion (InterDigital ) discussion
2. R3-224232 Validity Time Discussion (InterDigital ) discussion
3. R3-224254 AI/ML Radio Measurement Discussion (InterDigital ) discussion
4. R3-224255 AI/ML UE location Discussion (InterDigital) discussion
5. R3-224256 Discussion on Mobility Optimization Model Outputs (InterDigital ) discussion
6. R3-224257 Discussion on Load Balancing Model Outputs (InterDigital Finland Oy) discussion
7. R3-224258 QoS Feedback (InterDigital) discussion
8. R3-224306 XN enhancements for NG-RAN AI/ML (Qualcomm India Pvt Ltd) discussion
9. R3-224359 Discussion on data collection enhancements and signaling support (NTT DOCOMO, INC.) discussion
10. R3-224422 Discussion on resource status prediction (Lenovo) discussion
11. R3-224423 Discussion on UE traffic prediction (Lenovo) discussion
12. R3-224424 Discussion on AI assisted network energy saving (Lenovo) discussion
13. R3-224425 Discussion on prediction and feedback transfer in mobility scenario (Lenovo) discussion
14. R3-224426 (TP for TS37.483 TS37.480) Support UE traffic prediction over E1 interface (Lenovo) other
15. R3-224491 BL CR to TS 38.423: Support for AI/ML in NG-RAN (Ericsson) CR0869r, TS 38.423 v17.1.0, Rel-18, Cat. B
16. R3-224655 Discussion on AI/ML deployment and Stage-3 impacts (CATT) discussion
17. R3-224656 TP on TS 37.483 for AI/ML (CATT) other
18. R3-224657 TP on TS 38.423 for AI/ML (CATT) other
19. R3-224658 TP on TS 38.473 for AI/ML (CATT) other
20. R3-224675 Discussion on remaining issues of Mobility Optimization (VIVO TECH GmbH) discussion
21. R3-224716 Discussion on stage 3 related impacts of mobility optimization (LG Electronics) discussion
22. R3-224774 Discussion on Common Aspects in Stage 3 of AI/ML based Use cases (Intel Corporation) discussion
23. R3-224775 Discussion on Stage 3 of AI/ML based network energy saving and mobility optimization (Intel Corporation) discussion
24. R3-224776 (TP for NR\_AIML\_NGRAN BL CR for TS 38.423) (Intel Corporation) other
25. R3-224850 Discussion on AI/ML based Network Energy Saving (Stage 3) (Samsung) discussion
26. R3-224851 Discussion on AI/ML based Load Balancing (Stage 3) (Samsung) discussion
27. R3-224852 Discussion on AI/ML based Mobility Optimization (Stage 3) (Samsung) discussion
28. R3-224853 Correction of AI/ML for NG-RAN (Samsung) CR0889r, TS 38.423 v17.1.0, Rel-18, Cat. F
29. R3-224854 Correction of AI/ML for NG-RAN (Samsung) CR1023r, TS 38.473 v17.1.0, Rel-18, Cat. F
30. R3-224855 Correction of AI/ML for NG-RAN (Samsung) CR0036r, TS 37.483 v17.1.0, Rel-18, Cat. F
31. R3-224881 Discussion on AI/ML energy saving strategy (China Unicom) discussion
32. R3-224882 Discussion on AI/ML resource status (China Unicom) discussion
33. R3-224892 Introduction of RAN AI/ML (Huawei) CR0894r, TS 38.423 v17.1.0, Rel-18, Cat. B
34. R3-224893 Introduction of RAN AI/ML (Huawei) CR1026r, TS 38.473 v17.1.0, Rel-18, Cat. B
35. R3-224894 Discussion on the support of mobility enhancements using AI&ML (Huawei) discussion
36. R3-224895 Discussion on the support of load balancing using AI&ML (Huawei) discussion
37. R3-224896 Discussion on the support of energy saving using AI&ML (Huawei) discussion
38. R3-224909 Stage 3 issues on Rel-18 AI ML for NG-RAN (CMCC) discussion
39. R3-224960 Discussion on the standard impacts of AI-RAN (ZTE) discussion
40. R3-224961 Discussion on standards impacts of load prediction across multiple AI/ML based use cases (ZTE) discussion
41. R3-224962 Discussion on standards impacts of trajectory prediction across multiple AI/ML based use cases (ZTE) discussion
42. R3-224963 CR to TS38.423 for the unified AI/ML procedure (ZTE) CR0896r, TS 38.423 v17.1.0, Rel-18, Cat. B
43. R3-224964 CR to TS38.423 to enhance the existing procedure for AIML function (ZTE) CR0897r, TS 38.423 v17.1.0, Rel-18, Cat. B
44. R3-224488 AI/ML Network Energy Saving (Ericsson) Discussion
45. R3-224489 AI/ML Load Balancing (Ericsson) Discussion
46. R3-224490 AI/ML Mobility Optimization (Ericsson) Discussion
47. R3-224770 Discuss
48. ion on Data Collection and Signaling Support of AI/ML for NG-RAN (Intel Corporation) Discussion
49. R3-224874 Discussion of potential impacts on signalling procedures (China Unicom) Discussion
50. R3-224978 Data collection via MDT in split architecture (Beijing Xiaomi Mobile Software) Discussion
51. R3-224957 Signalling Procedure to support AI/ML in RAN (China Telecom) Discussion
52. R3-224560 AI/ML Network Energy Saving (Nokia, Nokia Shanghai Bell) Discussion
53. R3-224561 AI/ML Load Balancing (Nokia, Nokia Shanghai Bell) Discussion
54. R3-224562 AI/ML Mobility Optimization (Nokia, Nokia Shanghai Bell) Discussion
55. R3-224308 UE impacts in NG-RAN AI/ML (Qualcomm India Pvt Ltd) Discussion
56. R3-224493, MDT enhancements to support AI/ML based processes (Ericsson)