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][50]
* Option 2: Define the new procedure for the information used for AI/ML.[8][25][26][27][32][35][36][37][38]

**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, 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. |
|  |  |  |

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 seperately [39][40]:
  + AI/ML Data Collection Procedure (to collect historical inforamtion)
  + 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. |
|  |  |  |

## Network Energy Saving

Since, in RAN1 there is one study item “Study on network energy savings for NR”, [12][25][40] 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 (cell activation/deactivation) 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. |
|  |  |  |

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. |
|  |  |  |

## 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] |
| **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. |
|  |  |  |

## 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] |
| 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. |
|  |  |  |

## 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.

**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. |
|  |  |  |

## 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] 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.

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.

**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. |
|  |  |  |

## 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 Discussion on Data Collection and Signaling Support of AI/ML for NG-RAN (Intel Corporation) Discussion
48. R3-224874 Discussion of potential impacts on signalling procedures (China Unicom) Discussion
49. R3-224978 Data collection via MDT in split architecture (Beijing Xiaomi Mobile Software) Discussion
50. R3-224957 Signalling Procedure to support AI/ML in RAN (China Telecom) Discussion
51. R3-224560 AI/ML Network Energy Saving (Nokia, Nokia Shanghai Bell) Discussion
52. R3-224561 AI/ML Load Balancing (Nokia, Nokia Shanghai Bell) Discussion
53. R3-224562 AI/ML Mobility Optimization (Nokia, Nokia Shanghai Bell) Discussion