3GPP TSG-RAN WG3 #119bis-e R3-231897

E-meeting, 17th – 26th April 2023

Agenda Item: 12.2.2.1

Source: ZTE (moderator)

Title: Summary of email Discussion on CB: # AIRAN2\_LB

Document for: Discussion and Approval

# Introduction

**CB: # AIRAN2\_LB**

**- How to support partial reporting mechanisms?**

**- Whether the indication that an indication in the new agreed request message that UE performance feedback is provided after handover event is in implicit or explicit way?**

**-** **The details of the trigger indication in the HO request message to indicate that UE performance feedback is requested after HO completion?**

**- The structure of UE performance feedback IE, and whether the UE performance feedback is reported by one-time or periodically?**

**- Discuss how to support validity**

**- Capture agreements and open issues**

**- Provide TP if agreeable**

(Moderator - ZTE)

Summary of offline disc [R3-231879](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/RAN3-120/EmailDiscussions/CB%20%23%20AIRAN2_LB/Inbox/R3-231879.zip)

Two phases of this email discussion:

- Phase 1 Deadline: Thursday April 20th, 13:00 UTC

- Phase 2 Deadline: Monday April 24th, 10:00 UTC, we will try to come up with agreeable TP in the 2nd phase discussion before online session.

Note: Following the chair’s guidelines and limitation of the number of questions, not all proposals in the submitted papers will be treated in this SoD. If something is strongly needed to be discussed, please use Section 3.7.

# For the Chairman’s Notes

**The impact on procedures for HO-ed performance:**

**Proposal 1: No additional explicit indication is required in the AI/ML INFORMATION REQUEST message (FFS on the name)** **that UE performance feedback is provided after handover if UE performance feedback is only considered as feedback.**

**Proposal 2: Introduce the pair Measurement ID (e.g., NG-RAN node1 Measurement ID and NG-RAN node2 Measurement ID) in the HO request message, to establish relationship with the AI/ML INFORMATION REQUEST message (FFS on the name).**

**To be continued:**

**Whether the following information in necessary:**

**- Handover timestamp and AI-based load balancing handover cause indication in the HO request message**

**- UE performance feedback IE to indicate the UE performance characteristics requested for the handover UE.**

**Proposal 3:**

**Discuss whether UE performance can be used as both input or feedback first.**

**The structure of UE performance feedback IE:**

**Proposal 4: A list of UE performance feedbacks is introduced into the AI/ML INFORMATION UPDATE message (FFS on the name).**

**Proposal 5: UE performance feedback can be reported through one-time reporting or periodic reporting.**

**Partial Reporting mechanisms:**

**Proposal 6: Introduce an indicator in the AI/ML INFORMATION REQUEST message (FFS on the name), that indicate partial reporting is allowed or not allowed.**

**Proposal 7: Introduce the failed measurement in the response message to indicate partial reporting result.**

**FFS on whether indicate the reason for failure per measurement.**

**Timing information:**

**Proposal 8: Inclusion of timing information in the request message is to specify the time window of the requested information.**

**To be continued:**

**FFS on whether the timing information is configured for one time reporting or periodic reporting.**

**The details of the timing information, e.g., Validity time, Requested time, Reporting periodicity, etc.**

**Proposal 9: The timing information specified in the request message pertains to predicted information.**

**FFS on whether the timing information can be used for other measurements.**

**Accuracy:**

**Proposal 10: No consensus on whether the accuracy information is necessary to transfer between requested node and requesting node.**

**Others:**

**Proposal 11: Postpone the discussion on Predicted TNL capacity indicator, predicted slice available capacity, predicted composite available capacity.**

**Second round:**

**Try to capture agreements into the TP.**

# Discussion

## The impact on procedures for HO-ed UE performance

In the previous meeting, it was agreed that The agreed class1 procedure (AI/ML INFORMATION REQUEST/RESPONSE, the name needs further discussion) is used to configure UE performance feedback reporting. And the discussion on indication in the agreed new request message and HO request message is to be continued.

[1] proposes the indication in AI/ML Information Request that UE performance feedback is provided after handover event should be in implicit way.

[3] compares the implicit indication with explicit indication, and propose that no event indication is needed in the agreed Class 1 AI/ML message requesting feedback. Measurement ID to be used as a common identifier between Class1 AI/ML message requesting feedback and in the Handover Request message.

[5] No additional explicit indication in the agreed new request message (e.g., AI/ML INFORMATION REQUEST) is needed. In the HANDOVER REQUEST message, a pair of Measurement IDs can be included to explicitly associate the handover procedure with a previous UE performance feedback request.

[10] proposes to add an *Event Index* IE in the AI/ML INFORMATION REQUEST message (FFS on the name) to let the receiver identify the event triggering the report (for instance a handover event) and the event reporting configuration for it. And the same Event Index IE is introduced in the AI/ML INFORMATION UPDATE and HO message.

[14] considers gNB can implicitly know the requirement of the source gNB and wait the specific UE handover based on the Request Characteristic IE in the AI/ML INFORMATION REQUEST message. And introduce the pair measurement ID in the Handover Request message, which is the same ID in the AI/ML INFORMATION REQUEST/RESPONSE message.

[17] propose no explicit indication is needed to indicate UE performance feedback is requested after handover completion. Measurement ID pair IE is introduced to associate the HO request to UE performance feedback configuration. And UE performance feedback IE is introduced in HO request to indicate UE performance feedback is required for the UE after HO completion.

[21] considers providing UE performance measurement items and reporting configuration in the agreed new request message for AI/ML information can be used as an implicit indication that UE performance feedback is provided after handover event. And the handover timestamp and/or additional AI-based load balancing handover cause indication in HO request message can be used as an implicit trigger indication that UE performance feedback is requested after HO completion.

Above all, with respect to the issue whether the indication that an indication in the new agreed request message that UE performance feedback is provided after handover event is in implicit or explicit way, it appears that two options have been proposed:

* **Option 1:** No additional explicit indication (except agreed IEs such as measurement ID and report characteristics) is required in the AI/ML INFORMATION REQUEST message (FFS on the name).
* **Option 2**: Introduce an explicit indication of the Event ID in the AI/ML INFORMATION REQUEST message (FFS on the name).

**Q1: Companies are invited to provide their views on which option above is preferred, regarding an indication in the new agreed request message that UE performance feedback is provided after handover event?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Option 1 or Option 2** | **Comments** |
| Ericsson | Option 2 | Adding an explicit Event Index is needed to enable the possibility of defining different events, each requiring reporting of specific metrics and with specific reporting timing characteristics.  It is true that for the LB use case only one event and one event-triggered metric has been defined so far, namely the event is a handover and the event metric is the UE Performance Feedback. Already for this use case, there would be the need to configure different handover events with different reporting timing characteristics. For example, for some handovers, e.g. of UEs on eMBB services, only few samples of the UE Performance Feedback could be collected because the UE performance is sufficiently detailed with few samples, while for other handovers, e.g. UEs on URLLC services, many more samples would be needed because the UE performance needs to be determined in details.  However, besides differentiation of the reporting timing characteristics, we notice that at this RAN3 meeting at least 2 more events were proposed in other papers.  R3-231619 proposes to use the AI/ML Information Request to configure the reporting of the Energy Cost after specific handovers. Hence the new event is a handover that triggers EC reporting  R3-231681 proposes that, together with the agreed metrics for UE Performance Feedback, the UE Trajectory is signalled in the AI/ML Information Update message. Hence a new handover event that triggers the reporting of the measured UE Trajectory needs to be defined.  The above are only examples that show that in the future we will need to define more events that trigger information reporting via the AI/ML Information Update procedure.  If we include only the Measurement IDs in the Handover Request message, it would not be possible to understand which measurements the handover needs to trigger because the Measurement IDs point at all the measurements that could be defined in a AI/ML Information Reporting process.  There is the need for a more granular event identifier, e,g, an index, that can be included in the AI/ML Information Request with the following characteristics:   * It can be associated to specific measurements and specific reporting characteristics * It can be included in the Handover Request message, so to inform the Target NG-RAN that the measurement associated to the index need to be reported after the handover with the associated reporting characteristics (e.g. duration, number of samples,…), via the AI/ML Information Update procedure.   Option 2 allows us to define different events via the same AI/ML Information Reporting Initiation procedure. |
| CATT | Option 1 | We do not see any necessity to introduce an event ID (or event IDs) or other indication that UE performance feedback is provided after handover event. Relying on measurement IDs, the measurement characteristics maps or both is sufficient enough. Generally we are open on introducing measurement IDs and/or bit maps into the Handover Request message. |
| Lenovo | Option 1 with comment | We have similar view as CATT. Especially if the report characteristics indicates the request of UE performance feedback, it can be only collected after the HO. We don’t think additional indication is needed. |
| Samsung | Option 1 | There is the common understanding that the UE performance feedback is for the HO-ed UE. And the real triggering is in the HO request. So there is no need to stress the performance feedback is requested after HO completion. The HO request can include feedback request indication in the HO request, and then the target node can report the information via agreed AI/ML related procedure. |
| NEC | Option 1 | No need to stress the performance feedback |
| Huawei | Option 1, see comment | We think that providing UE performance measurement items and reporting configuration in the agreed new request message can be used as an implicit indication that UE performance feedback is provided by the requested node after handover event driven by AI/ML LB decisions. No additional indications needed. |
| InterDigital | Option 2 | Agree with Ericsson |
| Nokia | Option 1 | We think that using measurement ID is sufficient to indicate that UE performance feedback is needed after a handover. For all the different use cases (AI/ML Energy Saving, Mobility Enhancement, Load Balancing) there is a single event to be monitored, a HO event. The source collecting the AI/ML information knows for which reason (e.g., energy saving or mobility) it collects the information so there is no ambiguity about it. Trying to define in the Class 1 procedure a number of possible events that may (or may not) happen in the future seems inefficient. |
| Qualcomm | Option 1 with comments | We agree that Measurement ID is sufficient to link Class 1 AI/ML message and the HO Request. However UE performance indication in Report characteristics is not sufficient to indicate that UE performance is needed after the HO event and not immediately.  UE performance is also used as AI/ML input for LB use case as pointed out in the TR37.817 Section 5.2.2.4 and not just for feedback.  Hence just by indicating UE performance in the Report Characteristics will not help the receiving node to understand that measurements are needed as feedback and not immediately.  The Class 1 message needs to inform the receiver that UE performance measurements should be sent only after the UE is handed over. |
| Deutsche Telekom | Option 2 | We share Ericsson’s view. |
| CMCC | Option 1 | The implicit indication is enough. When the UE performance is configured in Report Characteristics IE, it can implicit indicate the UE performance feedback is required after HO completion. No explicit indication is needed in AI/ML information request message. |
| ZTE | Option 1 | Basically, for the UE performance feedback, which is the HO-ed information we are discussing, the UE performance feedback is indicated in the Report Characteristics in the request message. Because the indication is indicated in the request message, so the target gNB can implicitly know the requirement of the source gNB and wait the specific UE handover.  In the case that different events need different configuration, the source NG-RAN node can trigger multiple request message to trigger corresponding configuration.  If UE performance feedback can be considered as input or feedback, another new indication to distinguish should be introduced. |

**Moderator’s summary:**

**Regarding an indication in the new agreed request message that UE performance feedback is provided after handover event, 9 companies support option 1 that no additional explicit indication is required in the AI/ML INFORMATION REQUEST message (FFS on the name), because the existing measurement ID and report characteristics IE can implicitly indicate the UE performance feedback should be provided after handover. And 3 companies support option 2 that it is essential to introduce an explicit indication of the Event ID in the AI/ML INFORMATION REQUEST message so that the possibility of defining different events, each requiring reporting of specific metrics and with specific reporting timing characteristics is enabled. And two company indicates by indicating UE performance in the Report Characteristics will not help the receiving node to understand that measurements are needed as feedback and not immediately.**

**Proposal 1: No additional explicit indication is required in the AI/ML INFORMATION REQUEST message (FFS on the name)** **that UE performance feedback is provided after handover if UE performance feedback is only considered as feedback.**

In addition, regarding the details of the trigger indication in the HO request message to indicate that UE performance feedback is requested after HO completion, it appears that two options have been proposed:

* **Option 1:** Introduce the indication (e.g., NG-RAN node Measurement ID) in the HO request message, to establish relationship with the AI/ML INFORMATION REQUEST message (FFS on the name).
* **Option 1a:** Introduce the pair Measurement ID (e.g., NG-RAN node1 Measurement ID and NG-RAN node2 Measurement ID) in the HO request message, to establish relationship with the AI/ML INFORMATION REQUEST message (FFS on the name).
* **Option 2:** Introduce the Event ID in the HO request message, which is the same Event ID in the AI/ML INFORMATION REQUEST message, to establish relationship with the certain event in the AI/ML INFORMATION REQUEST message (FFS on the name).
* **Option 3:** Introduce the handover timestamp and/or additional AI-based load balancing handover cause indication in HO request message.

**Q2: Companies are invited to provide their views on which option above is preferred, regarding the details of the trigger indication in the HO request message to indicate that UE performance feedback is requested after HO completion?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Which option?** | **Comments** |
| Ericsson | Option 2 | As commented in Q1, simply adding Measurement IDs in the Handover Request is not sufficient. Measurement IDs point at all the measurements that have been configured in the AI//ML Information Request.  Let´s assume, for example, that we want to configure in the future two different handover events in the AI/ML Information Request message. Event 1 triggers reporting of UE Throughput, while Event 2 triggers reporting of UE Throughput, Packet Error Rate and Energy Cost.  If the Handover request includes the Measurement IEs, how can the target NG-RAN know which measurements the event should trigger?  For this we need a more granular event identifier, which can be associated to a specific event configuration. This can be achieved with Option 2, where the event and its associated measurements/configurations is identified with a specific index. |
| CATT | Option 1, 1a and/or character bit map | First, we think that the actual metrics for requested UE performance e.g. Throughtput, Packet Error Rate, Data lose, should be included in the characteristics in AI/ML INFORMATION REQUEST message. Otherwise, vendors may have different understanding on what the UE performance means.  Then, if the request metric is supported, in the HANDOVER REQUEST message, the source NG-RAN node could also include the requested metrics via characteristics bitmap. In this case, the target NG-RAN node could know clearly what the source requested.We think this is a simple way to support UE performance feedback.  We are also open to consider use of measurement ID.In this case, different measurement ID may refer to different combination of metrics. |
| Lenovo | Option 1,  Option 1a (preferred) | Prefer using a pair of measurement ID though since in the legacy a pair of IDs are usually used to avoid any ambiguity |
| Samsung | Option 1 or 1a | The source node triggers the performance collection via measurement ID to help the target node knows the reporting configuration. Similar view as CATT, the source node needs to indicate the metrics that needs to collect. |
| NEC | Option 1 or 1a | Prefer using a pair of measurement ID though |
| Huawei | Option 3 | In our paper R3-231824 we proposed to introduce both the **handover timestamp** and **AI-based load balancing handover cause indication** in the HO request message, which are needed for the requested node to identify whether a handover is caused by the AI-based load balancing inference.  We think that the above two information could be used for an **implicit trigger indication** in the HO request message that UE performance feedback is requested after HO completion, while at the same time improving the effectiveness of the LB action, because the handover timestamp allows the requested node to prepare the resources timely, while the new HO cause allows the requested node to potentially reject the incoming AI/ML-driven HO request in case the conditions that previously led the requesting node to offload some UEs to the requested node have changed. |
| InterDigital | Option 2 | Agree with Ericsson |
| Nokia | Option 1 or 1a | We think that UE performance can be requested by using a measurement ID. If different types of UE performance are needed this can be done by using different instances of the procedure.  The detailed metrics to be collected in our view should not be included in the HO Request message but be part of the configuration in the class 1 procedure. In this way it is clear to the target node what kind of UE performance information to collect. |
| Qualcomm | Option 1 or 1a with comments | We also need an additional indication to inform the receiver that UE performance is needed as feedback. |
| Deutsche Telekom | Option 2  Alternatively, Option 1/1a | Option 2, if the intro of Event IDs would be agreed (see Q1).  Otherwise, preference on Option 1/1a. |
| CMCC | Option 1 or 1a with comment | The Measurement ID in HO request is used to associate the HO request with the UE performance feedback configuration to make the target node know which configuration is for the handover UE.  Besides, we also propose to introduce an UE performance feedback IE to indicate the UE performance characteristics requested for the handover UE. |
| ZTE | Option 1/Option 1a  Prefer Option 1a more | The purpose of the indication is to establish the relationship between the AI/ML information request message and the handover request message.  Option 1 is a much clearer and easier way to establish the relationship between the messages. |

**Moderator’s summary:**

**8 companies support option 1/1a**

**2 companies support option2**

**1 company can accept option 1/1a if Event IDs are not introduced in Q1**

**1 company support option3**

**1 company also propose to introduce an UE performance feedback IE to indicate the UE performance characteristics requested for the handover UE.**

**Proposal: Introduce the pair Measurement ID (e.g., NG-RAN node1 Measurement ID and NG-RAN node2 Measurement ID) in the HO request message, to establish relationship with the AI/ML INFORMATION REQUEST message (FFS on the name).**

**To be continued:**

**Whether the following information in necessary:**

**- Handover timestamp and AI-based load balancing handover cause indication in the HO request message**

**- UE performance feedback IE to indicate the UE performance characteristics requested for the handover UE.**

**[3] suggests that the agreed Class 1 AI/ML should indicate that measurements are requested as feedback for a specific AI/ML action. This is to prevent the receiver from sending measurements immediately. On the other hand, [14] suggests introducing the Request Type IE in the request message to trigger measurements after the UE(s) handover. The primary purpose of this indication is to inform the requested NG-RAN node whether the requested information is reported immediately or after a specific action, such as a handover.**

**Q3: Companies are invited to provide their views on whether the explicit indication is beneficial in the AI/ML INFORMATION REQUEST message (FFS on the name) in order to inform the requested NG-RAN node whether the requested information is reported immediately or after a specific action?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Ericsson | Not necessary, See comments | Both the proposals in [3] and [14] seems to point at the need to define an explicit event indication. We should then simply add in the AI/ML Information Request an event identifier that is associated to a set of measurements and reporting configurations, which then is also included in the HO Request. This allows the Target to understand that measurements should be signalled only after receiving the Handover Requests that are marked with the event identifier.  Note that the agreements taken about having data type agnostic procedures go against an explicit indication that the information to be reported is “feedback”. |
| CATT | No. | Specifying that these measurement objects should be collected after handover is sufficient enough. No need for additional IEs. |
| Lenovo | No | Same comment as for Q1 |
| Samsung | No | Same comment as for Q1. |
| NEC | No | Same comment as for Q1. |
| Huawei | No | We think that if the *Report Characteristics* IE in the AI/ML INFORMATION REQUEST message has the “UE Performance” bit set to 1, then it is obvious that such UE performance will be collected after the HO event only |
| InterDigital | No | No need |
| Nokia | No | Same comment as for Q1 |
| Qualcomm | Yes | Without the indication how can receiver understand UE performance is needed as input or feedback.  UE performance can be used as both input (TR37.817 Section 5.2.2.4) or feedback. |
| Deutsche Telekom | No | Not needed |
| CMCC | No | Not necessary |
| ZTE | Yes | We share the same view as QC.  If only indicate the requested measurement in the request message, the target RAN node will be confused, and does not understand whether the source NG-RAN node needs the requested measurement at this moment or after the decision is executed. |

**Moderator’s summary:**

**2 companies agree that the explicit indication is beneficial in the AI/ML INFORMATION REQUEST message (FFS on the name) in order to inform the requested NG-RAN node whether the requested information is reported immediately or after a specific action, while 10 companies disagree that.**

**Proposal 3: Discuss whether UE performance can be used as both input or feedback first.**

## The structure of UE performance feedback IE

[6][10][14] proposes the agreed new class2 non-UE associated procedure (AI/ML INFORMATION UPDATE, which name is FFS) can contain a list of UE performance feedbacks each is related to a particular UE. [13] consider that if a source node requests a target node for UE performance feedback information it should be capable to correlate a UE or a number of UEs to the UE performance feedback information provided by the target node.

It appears that two kinds of UE performance feedback IE in the AI/ML INFORMATION UPDATE message (FFS on the name) have been proposed, see yellow-highlighted part:

**Alternative 1:** A list of UE performance feedbacks in the AI/ML INFORMATION UPDATE message (FFS on the name).

9.1.3.FF AI/ML INFORMATION UPDATE (FFS on the name)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| IE/Group Name | Presence | Range | IE type and reference | Semantics description | Criticality | Assigned Criticality |
| Message Type | M |  | 9.2.3.1 |  | YES | ignore |
| NG-RAN node1 Measurement ID (FFS on the name) | M |  | INTEGER (1..4095,...) | Allocated by NG-RAN node1 | YES | reject |
| NG-RAN node2 Measurement ID (FFS on the name) | M |  | INTEGER (1..4095,...) | Allocated by NG-RAN node2 | YES | reject |
| *Skipped* | | | | | | |
| UE Performance List |  | *0..1* |  |  | YES | reject |
| >UE Performance Result Item |  | *1 .. < maxnoofUEPerformance >* |  |  |  |  |
| >>UE Performance | M |  | 9.2.3.X |  |  |  |
| >>UE Assistant Identifier | O |  | NG-RAN node UE XnAP ID  9.2.3.16 |  | YES | ignore |

**Alternative 2:** A single UE performance feedback IE in the AI/ML INFORMATION UPDATE message (FFS on the name).

9.1.3.FF AI/ML INFORMATION UPDATE (FFS on the name)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| IE/Group Name | Presence | Range | IE type and reference | Semantics description | Criticality | Assigned Criticality |
| Message Type | M |  | 9.2.3.1 |  | YES | ignore |
| NG-RAN node1 Measurement ID (FFS on the name) | M |  | INTEGER (1..4095,...) | Allocated by NG-RAN node1 | YES | reject |
| NG-RAN node2 Measurement ID (FFS on the name) | M |  | INTEGER (1..4095,...) | Allocated by NG-RAN node2 | YES | reject |
| *Skipped* | | | | | | |
| UE Performance | O |  | 9.2.3.Y |  |  |  |

9.2.3.Y UE Performance

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source UE XnAP ID | O |  | NG-RAN node UE XnAP ID  9.2.3.16 |  |
| Average UE Throughput DL | O |  | 9.2.3.4 |  |
| Average UE Throughput UL | O |  | 9.2.3.4 |  |
| Average Packet Delay | O |  | 9.2.3.w31 |  |
| Average Packet Loss | O |  | 9.2.3.w32 |  |

**Alternative 3:** An average UE performance feedback IE in the AI/ML INFORMATION UPDATE message (FFS on the name), the specific UE performance feedback configuration is introduced in the AI/ML INFORMATION Request message (FFS on the name).

9.1.3.CC AI/ML INFORMATION REQUEST (FFS on the name)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **UE Performance Reporting Configuration** | ifReportCharacteristicsFourthBitIsSet |  |  |  | YES | ignore |
| >Starting time | O |  | INTEGER (0..4095) | Corresponds to the starting time of the UE performance measurement after successful handover completion. Value 0 indicates that the measurement is started immediately upon successful handover completion.  Unit: [second]. |  |  |
| >Averaging window | O |  | INTEGER (0..60) | Corresponds to the measurement averaging window. Value 0 indicates single measurement.  Unit: [second]. |  |  |
| >>Direction | M |  | ENUMERATED (ul, dl, both, ...) |  |  |  |
| >>UE Performance Metric | M |  | BITSTRING  (SIZE(32)) | Each position in the bitmap indicates the object the NG-RAN node2 is requested to report.  First Bit = Averaged Throughput,  Second Bit = Averaged Packet Delay,  Third Bit = Averaged Packet Loss |  |  |

9.1.3.FF AI/ML INFORMATION UPDATE (FFS on the name)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| IE/Group Name | Presence | Range | IE type and reference | Semantics description | Criticality | Assigned Criticality |
| Message Type | M |  | 9.2.3.1 |  | YES | ignore |
| NG-RAN node1 Measurement ID (FFS on the name) | M |  | INTEGER (1..4095,...) | Allocated by NG-RAN node1 | YES | reject |
| NG-RAN node2 Measurement ID (FFS on the name) | M |  | INTEGER (1..4095,...) | Allocated by NG-RAN node2 | YES | reject |
| *Skipped* | | | | | | |
| UE Performance | O |  | 9.2.3.Y |  |  |  |

**Q4: Companies are invited to share their opinions on which structure of UE performance IE is preferred.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Alt1, Alt2, or Alt3** | **Comments** |
| Ericsson | Alt1 | Alt1 is similar to what used today in other procedures, e.g. Access and Mobility Indication and for this it is preferred by us. Alt 2 is also feasible.  Alt 3 could be considered in the future, but for the time being this alternative adds complexity without a clear identified benefit. With Alt 3 in fact, it is difficult to identify the impact of an action on UE performance, given that the performance is averaged. |
| CATT | Alt2 | This is tightly coupled with the design for mobility-triggered post-handover data collection. The case there should performed per-UE (since different UEs are handed over at different time and UE contexts at source are deleted at different time), so the simplest way is to include only one UE’s data per message.  Anyway, Alt1 is also acceptable for us. |
| Lenovo | Alt1 with comment | We also prefer a list of UE performance for each concerned UE. Regarding the implementation, we think the UE Assistant Identifier should be mandatory also, for the source gNB to understand the related handed-over UE. Note that even for a same LB action, gNB still needs to decide to handover which UEs, thus, it is beneficial for the gNB to understand if the handover decision for a particular UE is good or not. |
| Samsung | Alt 1 | Reporting with a list of UE performance metrics is more efficient. |
| Samsung | Alt 1 | It looks this one is more efficient. |
| Huawei | Alt2 | We think Alt2 is a good trade-off in terms of signalling complexity. To us, Alt3 adds too much complexity which seems not motivated at this point in time. We are open to discuss Alt1 also. |
| InterDigital | Alt 1 | Alt 2 is also acceptable |
| Nokia | Alt3 | To clarify our proposal, we support to report average UE performance per UE. We think that if a target calculates an average performance over multiple UEs, different target nodes may calculate this average differently.  Alt1 is also acceptable by replacing UE Assistant Identifier with Source UE XnAP ID. |
| Qualcomm | Alt1 or Alt2 |  |
| Deutsche Telekom | Alt 1 | Preferred against Alt 2 |
| CMCC | Alt 1 | When HO caused by load balancing, a group of UEs will handover in a certain time, then it is efficient to transfer the group of UEs’ performance information instead of single UE’s performance information. |
| ZTE | Alt 1 | As discussed before, the AI/ML INFORMATION UPDATE message is non-UE associated message. Hence, the UE Identifier IE should be included in the UPDATE message if the UE related information is reported via the message.  Moreover, since the AI/ML decision may be related to multiple UEs, the UE performance feedback is related to each UE which handover to the target NG-RAN node for AI/ML decision. |

**Moderator’s summary:**

8 companies support Alt1, 4 companies accept Alt 1 or Alt2, 1 company support Alt 3

**Proposal 4: A list of UE performance feedbacks is introduced into the AI/ML INFORMATION UPDATE message (FFS on the name).**

[1] advocates for on-demand reporting to be supported for UE performance feedback reporting, whereas proposal [6] supports periodic UE performance feedback reporting, and request for feedback on one-time UE performance reporting is unclear.

**Q5: Companies are invited to express their views on whether UE performance feedback can be reported through one-time reporting, periodic reporting, or both.**

|  |  |  |
| --- | --- | --- |
| **Company** | **One-time reporting, periodic reporting, or both** | **Comments** |
| Ericsson | Both | In R3-231602 and R3-231603 we explain that UE Performance Feedback can be configured to be reported after the triggering handover, either once or periodically, for a configured time duration. This allows feedback to be collected in a flexible way, i.e. it allows for flexible selection of the number of feedback samples |
| CATT | Prefer one-time at least for this release | We do not observe any necessity to split one post-handover measurement report into many messages. So one-time report is better.  If the source node wants a time-sequence of measurement result, it can be indicated in the AI/ML request message and the target node then provide the time sequence altogether in the report message. Seems no need for periodic reports. |
| Lenovo | At least periodic.  We can also accept one time if majority wants. |  |
| Samsung | Both | Both periodical and one-time reporting should be supported. The reporting type in the request is up to node’s implementation. |
| NEC | Both | Both periodical and one-time reporting are needed |
| Huawei | Both | We think that both types of reporting need to be supported since the beginning, to ensure enough flexibility in the signalling design |
| InterDigital | Both | Both periodic and one time are needed |
| Nokia | Prefer one time | UE performance is expected to provide feedback after an AI/ML Action. So, conceptually, it seems to us that if it refers to a single action one-shot reporting is more appropriate. |
| Qualcomm | Both | We should provide flexibility in specification to request for one time or periodic reporting. |
| Deutsche Telekom | Both | To have sufficient flexibility and to be future proof. |
| CMCC | Both | Both periodical and one-time reporting should be supported. The node can choose whether the UE performance feedback it requested is reported by one-time or periodically. |
| Orange | Both | Both mechanisms should be possible |
| ZTE | Both | Both periodical and one-time reporting are needed |

**Moderator’s summary:**

10 companies support that both periodical and one-time reporting should be supported, 2 prefer one-time reporting, 1 support periodical reporting.

**Proposal 5: UE performance feedback can be reported through one-time reporting or periodic reporting.**

## Partial reporting mechanisms

**In last meeting, it was confirmed that partial reporting is supported in the agreed AI/ML information procedure, and the solution to support the partial reporting is FFS.**

**Several companies support partial reporting and have proposed solutions to address it:**

**[1] suggests that the requesting node should indicate in the request message whether partial reporting is supported or not. If partial reporting is supported, the requesting node should indicate which objects must be reported. The requested node should then indicate which objects it can provide in the response message.**

**[7] does not suggest introducing any "partial success allowed" indicator in the request message, as it may not cover typical cases. Instead, it proposes relying on the method where the requesting node can cancel its request if it finds that the accepted items are too few to use. It also proposes introducing a list into the response message (and the failure message) with each item including a cause value and a bit map of failed characteristics.**

**[10] believes Partial reporting is not supported in the Resource Status Reporting Initiation procedure in TS38.423. The reporting node can use the AI/ML INFORMATION RESPONSE messages to explicitly indicate the measurements that can be provided and the list of failed measurements together with the corresponding cause indicating the reason for failure per measurement. Besides, the reporting node can use the AI/ML INFORMATION UPDATE messages to indicate potential issues preventing the reporting.**

[14] proposes to support partial reporting mechanisms by introducing the Partial Reporting Indication IE in the request message to indicate whether partial reporting is allowed by the NG-RAN node. It also suggests introducing a new cause value "Partial Reporting Failure" and the Successful Report Characteristic IE in the response message to indicate which requested measurement is successfully indicated.

**Additionally, some companies support partial reporting without any specification impacts:**

[4] suggests that partial reporting can be understood by the network implementation between the source NG-RAN node and the target NG-RAN node.

[13] proposes reusing existing Xn principles to support partial success over XnAP for NG-RAN AI/ML use cases.

[21] suggests that partial reporting can be supported without any new indication in the agreed new class 1/2 procedures for AI/ML purposes.

Moderators believe that the current discussion pertains to how the newly agreed procedures for AI/ML RAN can support the partial reporting mechanism, as opposed to exploring whether the existing principles of Xn are capable of supporting partial reporting. So, it is suggested that the participants shift their attention towards finding a solution that will aid in the support of the partial reporting mechanism.

**Q6: Companies are invited to express their views on how to support partial reporting mechanism, regarding whether the explicit indication in the request message is needed:**

* **Alt 1: Introduce an indicator in the AI/ML INFORMATION REQUEST message (FFS on the name), that indicate partial reporting is allowed or not allowed.**
* **Alt 2: Introduce a new IE to indicate which objects must be reported**
* **Alt 3: No explicit indication.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Alt 1, Alt 2, Alt3, or combination?** | **Comments** |
| Ericsson | Alt1 preferred | We believe that a simple flag stating if partial reporting is Allowed /Not Allowed enables the target node to either accept what it can or to reject if some measurements cannot be reported.  We see the benefits of Alt 2, but at this point in time we would like not to over-complexify the procedure. |
| CATT | Alt3 | Alt1 and Alt2 are optimisations.  In this release we can use a simple solution: if the requesting node finds that the requested node does not provide enough measurement, it releases the measurement. |
| Samsung | Alt 1 and 2 | For alt1, if the requesting node can not accept the partial reporting, the requested node may provide the unnecessary information. To deal with such case, it is better for requesting node to indicate whether the partial reporting is supported or not. If not support, when one of the request prediction can not be provided, the requested node sends failure message to the requesting node.  For alt 2, it is to tell the requested node which objects are necessary and which are beneficial. If partial reporting is supported and the requested node cannot provide some of the beneficial objects, the requested node can just provide the ones it is able to provide. But if the requested node cannot provide one of the “object must to report”, the requested node sends failure message to the requesting node. |
| NEC | Alt 3 | It can be implicitly indicated. If some measurement is missing, then the target node assumes these are not supported. |
| Huawei | Alt3 | We think that partial supporting is already supported by implementation, so there is no need to add further indications.  Partial reporting support can be **implicitly** indicated to the requesting node by reporting the set of successfully performed measurements within the new UPDATE message, meaning that if only part of the requested objects can be supported the requested node responds with an UPDATE message anyway. This also allows the requesting node to implicitly know which objects failed to be initiated and performed based on the actual content of the UPDATE message, with basically no standard impact to the current XnAP specification. |
| InterDigital | Alt 1 | Alt 2 may also be feasible |
| Nokia | Alt3 | We think that simple methods, as already supported over Xn are sufficient for AI/ML purposes. When AI/ML is deployed in the network, nodes are expected to have an understanding on each other’s support of AI/ML functionality to a certain extent. OAM may configure the nodes so that they have a general understanding of each other’s capability, so that requests for measurements are not completely bound to errors.  Regarding Alt1, in practice it is possible that a node places a measurement request, out of which some measurements are mandatory while the rest may be only optional. The binary indicator does not have sufficient granularity to indicate to a neighbour “which” are the mandatory measurements from the requesting node’s point of view and which measurements could be optionally received. Also, since it is binary it does not enable a node that may be receiving multiple measurement requests to prioritize provision of some of those, in the absence of enough resources to provide them all. So the benefits of this indicator seem unclear. |
| Qualcomm | Alt1 | We prefer to have an partial success indication, so that the receiver knows that if few measurements are not supported, then it has to fail the whole procedure. |
| Deutsche Telekom | Alt1 / Alt2 | Preference on Alt1 to have a simple solution, but we are open to discuss potential benefits of Alt2. |
| CMCC | Alt 1 and 2 | Alt 1 and 2 are feasible .  For Alt 1, the indicator can indicate partial reporting is allowed or not. If the partial reporting is allowed, Alt 2 can be used to inform target node the measurements it must report.  As the time is limited for this release, we are also OK with Alt 3 and leave the optimization for the next release. |
| Orange | Alt 1 | Alt 1 is a first simple optimization step |
| ZTE | Alt 1 | The indication here is to inform the requested NG-RAN node whether the requesting NG-RAN node can accept the partial measurement or not. The requirement can be informed during the configuration, rather than implicitly inferred during the reporting in the Update message. |

**Moderator’s summary:**

**8 support Alt1, 5 support Alt 2, and 4 support Alt 3.**

**Moderator think this one kind of optimizations to the AI/ML Information procedure, and this optimization is beneficial, so here I try to conclude to introduce an indicator in the AI/ML INFORMATION REQUEST message (FFS on the name), that indicate partial reporting is allowed or not allowed.**

**Proposal 6: Introduce an indicator in the AI/ML INFORMATION REQUEST message (FFS on the name), that indicate partial reporting is allowed or not allowed.**

**Q7: Companies are invited to express their views on how to support partial reporting mechanism, regarding how to indicate partial reporting result in the response message:**

* **Alt 1: Introduce *Successful Report Characteristic* IE (name is FFS) in the response message**
* **Alt 2: The list of failed measurements together with the corresponding cause indicating the reason for failure per measurement.**
* **Alt 3: No explicit indication.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Alt 1, Alt 2, Alt3, or combination?** | **Comments** |
| Ericsson | Alt 2 | The whole purpose of Partial Reporting is for the source node to learn which measurements are not supported/not available and why they were not successfully configured. Namely, the source node should be able to understand if the measurement was not configurable because it is not supported (hence it should not be requested again) or if it is not available temporarily (hence it can be requested later). For this reason, it is essential that a list of measurements that failed to be configured is provided, together with cause values per measurement.  Note that this is how Partial Reporting has been supported in TS36.423, Resource Status Reporting Initiation procedure. |
| CATT | (Alt1 +) Alt2 | In RAN3 specs we often deliver both a successful list and a failure list for robustness. Here we can deliver both a successful bitmap and failure bitmaps alike. |
| Samsung | Alt 1 | Due to the special feature for AI/ML that AI/ML capability is time-varying for a node, such as retraining, heavy computation burden, model availability, the requested node may provide different objects at different time point. With the consideration of the instability of AI/ML capability, there is no need to inform the requesting node about the cause of failure items. For example, the requested node informs the requesting node the one object is temporally unavailable, for the next request, the requesting node still needs to request the same object as it does not know whether the situation is fine now or not. In another case, the node has no model for one object prediction now and it tells the requesting node that it is unavailable. After that, it may be shared with one related model from OAM and it can do the corresponding prediction. But the requesting node thinks the requested node has no such capability, so that it will not request to get such object any more. Thus alt 1 is more suitable for AI/ML related information reporting. |
| NEC | Alt 3 | We must discuss what are the possible failure reasons and how the source node make use of these failure causes, otherwise I don’t see why we needs the failure causes. |
| Huawei | Alt3, open to discuss Alt2 | As previously commented, by simply reporting the set of successfully performed measurements within the new UPDATE message, the requesting node implicitly knows which objects failed to be initiated and performed based on the actual content of the UPDATE message, with basically no standard impact to the current XnAP specification.  But we see some value in Alt2, at least for the RESPONSE message enhancement which allows to indicate to the requesting node those measurements that are “not available” |
| InterDigital | Alt 2 | Will allow to distinguish between not supported and temporarily not available measurements. |
| Nokia | Alt3 | We can reuse principles in existing Xn: Resource Status Procedure. |
| Qualcomm | Alt 2 | Agree with E/// |
| Deutsche Telekom | Alt2 | Cause values may help the requesting node to better understand the (potentially time-varying) capabilities of the requested node. |
| CMCC | Alt 2 | Alt 1 may also be feasible, but it seems that Alt 2 is more clear to let source node understand which measurement is not available and why it’s not available. |
| Orange | Alt2 | Interesting as the information uselessly avoid requesting measurements that are not available, discriminating with those that are only temporally unavailable. |
| ZTE | Prefer Alt 1 | Introduce the Successful Report Characteristic IE in the response message. The NG-RAN node2 can respond with the Report Characteristic IE that indicates which the requested measurement is successfully indicated so that NG-RAN node1 understand whether the NG-RAN node2 can provide some of the requested information, which the requested measurements can be provided by the NG-RAN node2.  For Alt2, that is partly acceptable. But it is unnecessary to indicate the reason for failure per measurement, just one cause value “measurement is not available” |

**Moderator’s summary:**

**7 support Alt 2, 3 support Alt 1, and 4 support Alt 3.**

**Proposal 7: Introduce the failed measurement in the response message to indicate partial reporting result.**

**FFS on whether indicate the reason for failure per measurement.**

## Timing information

**[5] suggests to agree when exchanging prediction information over Xn interface, the two NG-RAN nodes should understand the prediction is about what time window/point in the future (i.e., validity time).**

**[7] suggest the existing “Reporting Periodicity” IE should be renamed into e.g. “Time Window Length” in order to cover one-shot report. Nevertheless, it is also used as the periodicity for periodical reporting.**

[9] [13] **[14] [18] propose that** the time of a prediction shall be indicated in the message requesting the prediction.

**Q8: Companies are welcome to provide their opinions on whether the inclusion of timing information in the request message is necessary to specify the time window of the requested information.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Ericsson | See comments | In our paper n R3-231619 we have described the following concepts:   * The point in time in the future to which a prediction refers should be named “prediction time” or “reference time of prediction” and not “validity time.” * The term “validity time” refers to the time interval, starting from when a prediction is obtained to the prediction time. During such time said prediction has a reasonable expectation to reflect the future state of the process or quantity being predicted.   We propose that for periodic reporting there is no need to specify either the prediction time or the validity time of the requested AI/ML assistance information.  This is because the prediction time is the time when the period expires, and the validity time is the length of the period itself. These parameters can be configured by the requesting node, hence it is totally controllable by the requesting node when predictions will be requested. We believe that it is of increased complexity to enable requests of predictions far into the future. Besides, the further in the future a prediction is requested, the more inaccurate it is, hence the increase of complexity would not deliver a good performance either.  We believe that for one time reporting, there is the need to configure a prediction time, in the form of a time lapse after which the prediction time occurs. This could, for simplicity, be configured as e.g. 1 second. |
| CATT | Yes | The requesting node should indicate at least:   * For what duration a prediction should be provided; * The periodicity of reporting, if periodic. |
| Lenovo | Yes | Same view as CATT |
| Samsung | Yes | Both reporting periodicity and prediction time to describe prediction is about what time window/point in the future are required.  Reporting periodicity is the time interval between two consecutive reporting.  Prediction time to describe prediction is about what time window/point in the future is to tell the requested node the prediction should be done for which period. |
| NEC | Yes | Same view as CATT |
| Huawei | Yes | In our paper R3-231823 we provided definitions for the timing information as follows   * **Validity time**: time period within which the requested prediction information in the AI/ML INFORMATION UPDATE (FFS on the name) is considered valid, hence the AI/ML prediction can be used by the requesting NG-RAN node. * **Requested time**: time duration of the requesting prediction information in the AI/ML INFORMATION REQUEST (FFS on the name), e.g. start time plus end time. * **Reporting periodicity**: reporting interval of the requesting prediction information in the AI/ML INFORMATION REQUEST (FFS on the name).   The requesting node expects the requested node to provide valid prediction information in the requested time, and possible reporting options are one-time and periodic reporting.  In the case of **one-time reporting**, because the requested node reports only once, the requesting node needs to explicitly indicate the requested time of the valid prediction information, and it may be considered that the requested time has the same meaning with the validity time. Therefore, we think the requested and validity time should be configured in the AI/ML INFORMATION REQUEST (FFS on the name).  For **periodic reporting**, we now include also the reporting periodicity and it needs to be discussed how reporting periodicity relates with the validity time of the predictions. E.g., if the reporting periodicity is shorter than the validity time of the prediction, then the prediction can be considered as valid by default, i.e. the validity time does not need to be indicated. But how to deal with the opposite case, i.e., the reporting periodicity is greater than the validity time of the prediction? |
| InterDigital | Yes | The requesting node should indicate at least:   * For what duration a prediction should be provided, and The periodicity of reporting, if periodic. * For a one-time measurement a time needs to be indicated |
| Nokia | Yes | This question is related to one of the questions in AIRAN1\_Stage2. Perhaps some coordination is necessary to avoid conflicts.  We agree with Ericsson’s clarification that validity time should not be confused with a prediction time. We provide a figure with our understanding on validity time and prediction time:  Shape  Description automatically generated with medium confidence  A node requests a prediction corresponding to a future time t. So, prediction time is the time when a prediction is made. If no validity time is indicated, it would mean that a prediction is only valid for the prediction time. Validity time gives a different angle to a prediction in that the requested information is meaningful not only for a specific time but over an interval of validity. We see value in that in the sense that a node receiving the prediction may utilize it over longer periods of time.  But we think that the prediction time and validity time need to be provided both for one-time reporting and for periodic reporting. Validity time is not necessarily the same as reporting periodicity (T) and we should not try to correlate validity time (ML Model parameter) with periodicity (reporting procedure parameter) . |
| Qualcomm | Yes | For Periodic Reporting, we need to set a defined time for the node to send reports in a periodically. The sender cannot send the reports infinitely. |
| Deutsche Telekom | Yes | We think based on the explanations given especially by Ericsson, Huawei, and Nokia, a set with 3 parameters can be applied to describe the required timing information. |
| CMCC | Yes | We think the timing information is helpful when the requesting node ask for the prediction, e.g., the request node only needs the resource status related information for the midnight for AI-based load balancing. Another potential scenario is that the node requests the prediction information periodically, e.g., the predicted recommended cell activation/deactivation information at the midnight of everyday for AI-based energy saving.  For details, we propose to use start time and end time/duration to specify the time window. |
| Orange | Yes | We think that the 3 parameters 1/validity time, 2/requested time and 3/reporting periodicity (as indicated by Huawei above) are relevant. |
| ZTE |  | The node requesting a prediction, itself, knows the necessary duration of prediction, and then it indicates the other node the start time and end time of prediction. |

**Moderator’s summary:**

**All company agree the Inclusion of timing information in the request message is necessary to specify the time window of the requested information. One company think that there is the need to configure a prediction time for one time reporting rather periodic reporting.**

**Proposal 8: Inclusion of timing information in the request message is to specify the time window of the requested information.**

**To be continued:**

**FFS on whether the timing information is configured for one time reporting or periodic reporting.**

**The details of the timing information, e.g., Validity time, Requested time, Reporting periodicity, etc.**

**Q9: If the answer to Q7 is yes, companies are invited to express their views on whether the timing information specified in the request message pertains solely to predicted information or if it includes all types of information that is being requested.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Predicted information, or all kinds of information** | **Comments** |
| Ericsson | Prediction information | The prediction time is obviously applicable to a prediction. We do not understand what is meant by a prediction time pertaining to all the information that is requested. |
| CATT | Prediction at least | For measurement objects other than prediction, it should be discussed case by case whether this IE can be reused. |
| Lenovo | Prediction at least, and see comment | Not sure if the timing information in the question also includes periodicity information.  For periodicity information, it also applies to other information request such as UE performance feedback, or EC (if agreed). |
| Samsung |  | Reporting periodicity for all the information.  Prediction time to describe prediction is about what time window/point in the future for prediction information only. |
| NEC | Prediction information | The prediction time is obviously applicable to a prediction. |
| Huawei | Prediction, but open to discuss other information | If “other information” includes the UE performance feedback or actual/current EC we think we should discuss this at a later stage, on a case-by-case basis. At least predictions should be considered since the beginning |
| InterDigital | Prediction information |  |
| Nokia |  | We are not sure of the intention of this question. We think that the prediction time and validity time should be provided in the request message both for one time and periodic reporting of predictions. |
| Qualcomm |  | Question is unclear.  If the question is whether to provide end time for periodic reporting, then the answer is Yes. We need to provide end time for periodic reporting of any measurements requested in the Class 1 procedure.  If the question is regarding validity time, then it should be provided only for predicted information. |
| Deutsche Telekom | Prediction information, others tbd. | We are open to discuss handling of timing information for other measurements on case by case, but the primary focus is certainly the prediction information. |
| CMCC | Prediction information at least | Whether the measurement information needs the time information should be discussed per case, but at least the time window is needed for prediction information. |
| Orange | Prediction information | Case by case for others. |
| ZTE | Prediction at least | Other information also needs the timing information. Especially to indicate the target NG-RAN node the timing information of the feedback measurement. |

**Moderator’s summary:**

**All companies believe that** **the timing information specified in the request message pertains to predicted information.**

**Proposal 9: The timing information specified in the request message pertains to predicted information.**

**FFS on whether the timing information can be used for other measurements.**

## Accuracy

[5] suggest to discusses the pros/cons and down select from the following options for a requesting NG-RAN node to understand the accuracy of received prediction information

a. Option 1: required prediction accuracy in prediction request

b. Option 2: prediction accuracy in prediction report

c. Option 3: via requesting actual measurement

[7] [18] propose **prediction accuracy could be provided to the requesting node when requested.**

[9] thinks there is no need to transfer the prediction accuracy information over Xn interface for the receiving node.

[13] believes that trigger unsuccessful outcome if the NG-RAN node providing predictions can’t satisfy the minimum requirements placed by the requesting node.

**Q10: Companies are welcome to provide their opinions whether the accuracy information is necessary to transfer between requested node and requesting node.**

**If yes, which option is preferred:**

* **Option 1: Inclusion of accuracy information in the request message (FFS on the name) to specify required accuracy of prediction**
* **Option 2: Accuracy information together with each predicted information in the update message (FFS on the name).**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Ericsson | No | We have explained in our paper in R3-231619 that an AI/ML model cannot provide an accuracy for an inferred prediction. At best, the AI/ML model can provide the average accuracy measured during model testing, where model testing is a process of model training carried out with a specific set of testing data inputs.  Hence, the accuracy the model is able to provide is strictly related to the data used for model testing. Such accuracy could be very high, but if the model is given as inputs data that were not included in the model testing data set, the accuracy of the prediction may be considerably lower.  Hence relying on the accuracy the model can provide may be harmful and misleading.  The only way to calculate the accuracy pertaining to a prediction is to compare the prediction with the ground truth values, i.e. measured values. Such measured values can be requested and obtained by the requesting node. It is true that the requesting node will have to go through a learning period to evaluate the goodness of the predictions of a neighbour node, but it is also true that after this period, the goodness of such predictions will be known with certainty. On the contrary, relying on the accuracy a model can provide is in itself unreliable and does not produce any trustable result. |
| CATT | Option 2 | We stated repeatedly that AI/ML mechanisms can provide some accuracy, e.g. by the output from a few models.  And such accuracy keeps changing over time so it cannot be used as a threshold in the initiating procedure. It doesn’t work. |
| Lenovo | Option 1 (preferred), or Option 2 | For accuracy/confidence information about the prediction, we give our analysis in paper R3-231433. In short, we believe it is useful for the prediction receiver to understand the likelihood a prediction will be true. In a reasonable RAN implementation, it will impact how a RAN node weights the prediction result when use it for a RAN decision.  Comparing Option 1 and 2, we prefer Option 1 since it can avoid any unnecessary AI inference and information exchange. |
| Samsung | Option 2 | The accuracy parameter may provide reference to the receiving node, so that the receiving node can adjust the decision about how to refer it accordingly, such as setting policy based on the high-accuracy inference results and taking low-accuracy results as additional reference. In addition, the accuracy can be obtained in testing stage, which can be regarded as the feature of this model. As the model may change, e.g. fine-tuning to adjust model, the accuracy is not a static value, so the current accuracy of the model can be sent to the requesting node along with the prediction information. |
| NEC | Option 2 | The target node needs the accuracy to determine the behaviour for load balancing. |
| Huawei | No | We need to make clear whether transmitting prediction accuracy (along with the prediction itself) is to measure the performance of the AI/ML model or the accuracy probability of prediction information.  If the purpose is to measure the performance of the AI/ML model, **it doesn't seem necessary**: as long as the AI/ML model is not modified, the performance of the AI/ML model that inferences the prediction information should be constant. Not to mention that there is no unified standard for measuring the performance of the AI/ML model when different nodes have different AI/ML models or the same AI/ML model is (re-)trained based on different data sets.  If the purpose is to measure the accuracy probability of the prediction information, also in this case **we think this is not needed**. We believe that it is not useful to provide the prediction’s accuracy/confidence (along with the prediction) over Xn since accuracy is not a deterministic information by default but a prediction itself. Moreover, it is not clear to us which could be the behavior of the receiving NG-RAN node when provided with such kind of information, will it trust the prediction or not? Eventually, the prediction accuracy can only be calculated by comparing the prediction with the ground truth in the prediction time. But there could also be other means to evaluate an AI/ML prediction’s accuracy, e.g. via the UE performance feedback or via other use case specific means |
| InterDigital | No | Basically agree with Huawei and Ericsson |
| Nokia | Option 1 | First of all, we would like to clarify that it is hard to monitor model accuracy but we think that a confidence level can be provided instead to provide some information on the prediction quality. This can be provided as an error margin e.g., x% of the average (predicted value). A node requesting a prediction needs to know some information about the predictions that can be provided by a neighbour since low quality predictions may negatively impact node’s performance.  We think that it is redundant to send accuracy/confidence information with every prediction since this would be a model related parameter that is not expected to change per inference output. Therefore we think that a required confidence can be provided by the requesting node to indicate to its neighbours what is the required accuracy/confidence level that it expects in the received predicted information.  This question is related to one of the questions in AIRAN1\_Stage2. Perhaps some coordination is necessary to avoid conflicts. |
| Qualcomm | NO | We think it is very hard for the predicting node to achieve high accuracy due to feedback data unavailability, or model change or input parameter change etc. If the predicting node knows that accuracy of the predicted data is not upto the mark, then predicting node shall decide not to send the predicted data and send the predicted data only when the prediction accuracy is high.  BTW this question is also discussed in [CB # AIRAN1\_Stage2](https://www.3gpp.org/ftp/tsg_ran/WG3_Iu/TSGR3_119bis-e/Inbox/Drafts/CB%20%23%20AIRAN1_Stage2)  Lets have a converged view on this between both the CBs. Better to discuss it in Stage 2 CB. This CB already contains many other topics. |
| Deutsche Telekom | Option1  (preferred) or Option2 | As also stated by other companies, there is a similar question in CB # AIRAN1\_Stage2, so both CBs need to be aligned to get consistent feedback.  Having accuracy/confidence information about prediction data would definitely help the receiving RAN node in its decision process, but as stated by some companies, it is hard to get actual accuracy/confidence values if input data sets for predicting node are different to training data. But do we assume homogeneous model deployment resulting in same accuracy/confidence levels across neighbouring nodes? If not, we should think about how such info can be provided: via Xn or e.g. via OAM taking care of training results? Please consider that models may be exchanged and/or updated during online training. |
| CMCC | Option 2 | By providing the prediction accuracy to the requesting NG-RAN node, the requesting node could have an understanding of “how good” these predictions are and make use of the prediction result in different ways. Provide the accuracy information when requested by the requesting node seems reasonable. |
| Orange | Option 1 | We believe that accuracy/confidence information about prediction data would be beneficial for the receiving RAN node to take a decision |
| ZTE | No | Accuracy information is an important piece of information that is highly relevant to AI/ML models. However, different models may have different evaluation methods to calculate accuracy, and it is not always represented as a value from 0 to 100. Additionally, the predicted information is generated by the NG-RAN node that is being requested, and it is up to that node to decide whether the predicted information is good enough to be sent back to the requesting NG-RAN node. Given these factors, we don't believe it is necessary to transfer accuracy information along with the predicted information. |

**Moderator’s summary:**

5 companies disagree that the accuracy information is necessary to transfer between requested node and requesting node, while 8 companies think it is beneficial, and 4 companies prefer option2, 4 company accept option1.

**Proposal 10: No consensus on whether the accuracy information is necessary to transfer between requested node and requesting node.**

## Other predicted information

[3] [13] [18] propose to agree the predicted radio parameters including predicted TNL capacity indicator, predicted slice available capacity, predicted composite available capacity group, should also be exchanged in the predicted resource status information between the neighbour NG-RANs.

**Q11: Companies are welcome to provide their opinions whether following predicted information can be agreed to transfer between the neighbour NG-RANs now, or discuss these information later:**

* **Predicted TNL capacity indicator**
* **Predicted slice available capacity**
* **Predicted composite available capacity**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Comments** |
| Ericsson | Not for now | We should first make sure that the definitions of the current predictions is clear, before adding more information to the discussion. For example, we need to clarify if the Radio Resource Status includes per SSB area predictions, or per slice predictions. |
| Lenovo | Ok to postpone |  |
| Samsung |  | They are beneficial for LB decision. We are fine to postpone the discussion. |
| Huawei | No strong view | The proposed measurements are already available and specified, so it should not be an issue to also include them in the set of the already agreed predictions that could be exchanged over Xn. However, at this point in time, we should make it clear which is the purpose for this new information and which use case(s) will eventually benefit from them |
| InterDigital | Ok to postpone |  |
| Nokia | Not yet | We think that these predictions can be postponed to later since there doesn’t seem to be an immediate need to introduce them for the current Rel.18 use cases. |
| Qualcomm | No strong view | We are fine to go with the majority. |
| Deutsche Telekom | Not yet | Mentioned parameters could be useful for the decision process, but we are fine with postponing the discussion. |
| CMCC |  | We think it could be easy to support them since they are existing measurements and beneficial to transfer them. |
| ZTE | Not yet | Postpone the discussion |

**Moderator’s summary:**

**Proposal 11: Postpone the discussion on Predicted TNL capacity indicator, predicted slice available capacity, predicted composite available capacity.**

## Others

Please use this section to provide additional input on postponed issue, if strongly needed.

|  |  |
| --- | --- |
| **Company** | **Comments** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# References

|  |  |
| --- | --- |
| 1. [R3-231205](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231205.zip) | Discussion on Xn impact of LB (Samsung) |
| 1. [R3-231209](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231209.zip) | TP to 38.423 for partial reporting of AI/ML information (Samsung) |
| 1. [R3-231260](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231260.zip) | Feedback for NG-RAN AI-ML (Qualcomm Incorporated) |
| 1. [R3-231378](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231378.zip) | AIML Load balancing (NEC) |
| 1. [R3-231433](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231433.zip) | Discussion on prediction accuracy and time information (Lenovo) |
| 1. [R3-231434](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231434.zip) | Discussion on UE performance feedback collection (Lenovo) |
| 1. [R3-231465](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231465.zip) | Discussion on partial success and validity time (CATT) |
| 1. [R3-231469](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231469.zip) | (TP for 38.423) Updates on non-UE associated messages to support AI/ML (CATT) |
| 1. [R3-231515](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231515.zip) | Discussion on validity time and prediction accuracy (China Telecommunication) |
| 1. [R3-231602](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231602.zip) | AI-ML Event triggered UE performance reporting (Ericsson, InterDigital, Deutsche Telekom) |
| 1. [R3-231603](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231603.zip) | (TP for AI/ML BLCR to TS 38.423) AI-ML Event triggered UE performance reporting (Ericsson, InterDigital, Deutsche Telekom) |
| 1. [R3-231616](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231616.zip) | (TP for AI/ML BLCR to TS38.423) Partial success for the AI-ML Assistance Data Reporting procedure (Ericsson, InterDigital) |
| 1. [R3-231656](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231656.zip) | (TP for TS 38.423) LB and AI/ML Information Exchange over Xn (Nokia, Nokia Shanghai Bell) |
| 1. [R3-231680](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231680.zip) | Further discussion on remaining issues on procedures for AI (ZTE) |
| 1. [R3-231681](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231681.zip) | (TP to 38.423 and 38.420) AIRAN impact on Xn Interface (ZTE) |
| 1. [R3-231795](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231795.zip) | Discussion on AIML UE performance feedback (CMCC) |
| 1. [R3-231796](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231796.zip) | (TP for 38.423)Procedure for AIML related Information (CMCC) |
| 1. [R3-231798](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231798.zip) | Remaining issues on predicted information (CMCC) |
| 1. [R3-231259](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231259.zip) | Xn enhancements for NG-RAN AI-ML (Qualcomm Incorporated) |
| 1. [R3-231432](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231432.zip) | Miscellaneous Xn interface issues (Lenovo) |
| 1. [R3-231824](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231824.zip) | (TP for AIML BLCR for TS 38.423) Remaining open issues for load balancing (Huawei) |
| 1. [R3-231620](https://ericsson-my.sharepoint.com/personal/angelo_centonza_ericsson_com/Documents/Documents/3GPP%20Related/3GPP_ETSI/RAN3/Docs/R3-231620.zip) | (TP for AI/ML BLCR to TS38.423) AI-ML Threshold Based Events (Ericsson) |

# Conclusion, Recommendations [if needed]

If needed.