**3GPP TSG-RAN WG3 Meeting #112electronic R3-212689**

**Online, 17th – 27th May 2021**

**Agenda Item: 18.3**

**Source: CMCC (moderator)**

**Title: SoD Data Collection Use Cases**

**Document for: Discussion and Decision**

# 1 Introduction

**CB: # 46\_DataColl\_UseCases**

**- Chair: suggest to structure discussion around 5 areas, splitting work among companies for the resulting TPs:**

**1) Common parts / overview / general descriptions (merging any agreeable parts from e.g. 2301, 2313, 1683, 2190) (NEC?/E///?)**

**2) ES/EE (merging any agreeable parts from e.g. 2030, 2031, 2523, 2507, 1669, 2315) (ZTE,CU,Len,Moto?/CMCC?)**

**3) Load Balancing / Load Prediction (merging any agreeable parts from e.g. 2032, 2033, 2504, 2524) (CMCC?/HW?)**

**4) Mobility / Traffic Steering (merging any agreeable parts from e.g. 2028, 2029, 2191, 2271, 2269, 2316) (CATT?/ID?)**

**5) Other use cases (if any / if agreeable) (from e.g. 1969, 2179, 2389)**

(CMCC - moderator)

There are some overlaps between CB#46 and CB#47，we found it is sensible to discuss the standard impacts with the specific solutions together, so after coordination with the moderator of CB#47\_DataColl\_StdImpact, this CB#46 will only focus on use case description and potential benefits. The solutions for the identified use cases and potential standards impact will be discussed in CB#47.

Under the chair’s guidance, the offline discussion consists of 5 areas, the discussion will be structured as follows:

- In the first round of offline discussion, it is proposed to collect company views on the general question for each area;

- In the second round of offline discussion, it is proposed to focus on TPs for use cases, including ES/EE, Load balancing/ Load Prediction, Mobility / Traffic Steering, as well as common parts and other use cases if agreeable.

# 2 For the Chairman’s Notes

**Agree to capture the following proposal:**

**Proposal 1: It is proposed to sufficiently progress the prioritized use cases on energy saving, load balancing, traffic steering/mobility optimization, i.e. at least by identifying their impact on the specifications, before considering any new use case.**

**Agree the following TPs for TR 37.817:**

**R3-21xxxx (in revision of R3-212030), the use case description of EE/ES**

**R3-212807 (in revision of R3-212524), the use case description of Load Balancing/ Load Prediction**

**R3-212868 (in revision of R3-212271), the use case description of Mobility/ Traffic Steering**

# 3 Discussion

**Round 1 proposal:**

**Proposal 1: It is proposed to sufficiently progress the prioritized use cases on energy saving, load balancing, traffic steering/mobility optimization, i.e. at least by identifying their impact on the specifications, before considering any new use case.**

**Proposal 2: Use 2030[6] as the baseline for use case description of EE/ES to work on the potential TP, companies’ comments should also be taken into account.**

**Proposal 3: Use 2524[23] as the baseline for use case description of Load Balancing/ Load Prediction to work on the potential TP, companies’ comments should also be taken into account.**

**Proposal 4: Use 2271[14] as the baseline for Mobility/ Traffic Steering to work on the potential TP, companies’ comments should also be taken into account.**

**Preparation for the TP:**

**ZTE, work on the potential TP for use case description of EE/ES based on 2030 [6] and merge agreeable parts in [1][2][17][21][22].**

**Huawei, work on the potential TP for use case description of load balancing based on 2524[23] and merge agreeable parts in [2][4][8][20].**

**InterDigital, work on the potential TP for use case description of mobility based on 2271[14] and merge agreeable parts in [2][12][18].**

**It is suggested that draft TPs are available by Monday 12:00 UTC 2021-5-24.**

## 3.1 Common

The proposals in this part (from [15] [16] [11]) are diverse and seems cannot to merge, TPs for prioritized three use cases are provided in [2] and could be discussed in following corresponding use case parts.

Companies are invited to provide views on whether following proposals are agreeable.

**Q1: Which following proposal(s) from [15] is agreeable for you?**

**Proposal 1: “Confidence Level” of a model inference output should be supported and provided to “Actor” node.**

**Proposal 2: “Validity Time” of model inference output should be supported and provided to “Actor” node.**

**Proposal 3: For ML enabled load balancing, “Model Training” can be located at CN/OAM or NG-RAN (CU). “Model Inference” can be located at NG-RAN (CU).**

**Proposal 4: For ML-enabled load balancing, input data of Model Training includes QoS requirement, traffic volume, cell capacity, cell resource status, and measurement data.**

**Proposal 5: For ML-enabled load balancing, output data of Model Inference includes prediction of traffic volume, cell capacity and cell resource status, action of UE (such as handover), etc.**

**Proposal 6: For ML-enabled load balancing, the prediction results of traffic, resource status, cell capacity, etc., should be exchanged between the source gNB CU and neighbouring gNB-CUs via Xn interface.**

**Proposal 7: For ML-enabled load balancing, handover decision at source cell can be decided by the prediction results of its own and neighbouring cells.**

**Proposal 8: NG-RAN can decide to switch off certain cells to lower energy consumption based on the prediction of cell load information.**

**Proposal 9: For ML-enabled energy saving, a gNB can send a message and avoid deactivating a needed cell of a neighbouring gNB, by utilizing the prediction result of cell load information exchanged between them.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| Futurewei | Proposal 1: No. Proposal 2: Yes with clarification.Proposal 3: No.Proposal 4-5: Yes.Proposal 6: Yes.Proposal 7: No.Proposal 8: No.Proposal 9: No. | Proposal 1: Not all ML models provide confidence interval (CI). If we want CI to be provided, then first we need to define how CI should be calculated for those ML methods that don't directly generate such information. Before the definition is done, we don’t see CI is appropriate.Proposal 2: need a clear definition of “Validity Time” if we want to use it. For example, in traffic prediction use case, the prediction result may be the traffic load for the next interval (which depends on the measurement interval in the input/output), thus, the prediction result is only good for the next interval naturally.Proposal 3: this proposal is not necessary. Regardless of use cases, where the model training and model inference should be performed belong to vendor decision.For proposal 4 and 5: while we agree to these two in principle, we have some comments. 1. Input data should be rich enough in order to generate output data. Proposal 4 does not cover all the needed measurements. For example, if the model suggests a UE handover, then the cell capacity of neighbouring cells need to be included as input data. It seems the relationship between Proposal 4 and 5 have not been carefully considered.
2. HO decision most likely does not based only on inference result as it will take into account other factors as well. It is cleaner to separate model inference and decision in any use case, while the 2 can reside on the same functional block.

Proposal 7: The decision of handover or not should be vendor-dependent, based on each vendor’s algorithm. We suggest rephrasing this proposal to something like : *“...the source cell can utilize the model inference results for itself and neighbouring cells together with other information to make HO decision...”*Proposal 8: Again, the decision-making is vendor dependent. Can be rephased like in Proposal 7.For proposal 9: two comments1) not sure if this is necessary; if the target gNB needs to remain active, it does not need this request message; its own AI model should be able to make the decision (note the traffic information is exchanged between the source gNB and the target gNB so they know each other’s situation well).2) while a gNB can send a message to neighbouring gNBs, whether to deactivate a particular cell on the gNB is up to the gNB's decision.**In general, we suggest separating action/decision-making from inference results/prediction; we should not define how the decisions will be made.**  |
| CATT | P1, P6, P7, P8, P9: yesP2: noP3–P5: cannot agree now (please see in the comment) | For P2, we don’t observe **exchanging** the “validity time” quite useful. It can be entirely up to implementation. Need further clarification on why exchanging this is necessary.For P3–P5, in our understanding load balancing encompasses mainly three sub-functions: load prediction, UE traffic and mobility prediction, and load balancing decision.The two “prediction” sub-functions can surely benefit from AI/ML (e.g. supervised learning), but once the “accurate” prediction result is provided, we doubt if it is really beneficial to make the load balancing decision function itself an AI-assisted one, e.g. by using reinforcement learning.P3–P5 are mainly about taking decision, so we cannot agree with them now. |
| Samsung | P1, P2: need to study case by caseP3, P6, P7, P8: YesP4, P5: Need more clarification/discussionP9: Need to discuss the exchange of predicted deactivation decision firstly | P1 and P2: Interface impact needs to be studied case by case. “Confidence level” and “validity time” can be considered for further detailed use case study.P4 and P5: input/output highly depends on the AI functionality. Suggest to discuss the AI functionality for load balancing first, and then to discuss the related input/output for each functionality.P9: this message to avoid deactivation is based on the exchange of predicted deactivation decision. So prefer to discuss whether to support exchange predicted deactivation decision firstly. |
| CMCC |  | Not sure P1, P2 are necessary for all AI models, and whether confidence level and validity time are useful and/or trustful.It is better to discuss P3-P9 in the standard impact study of corresponding use case. |
| Nokia | P1: Not sureP2, P4, P5: NoP3, P6,P7,P8, P9: Yes | P1: How can confidence level be useful to the Actor? If we consider confidence level over the predictions, this only tells us that prediction results are stable and within some range but it doesn’t tell anything about how good the prediction is (with respect to the real measurements. P2: How can you know in advance how long an ML model will be valid? P4, P5: Decisions on those details should be postponed to later stage.P3: We agree though we believe it would be good not to mix CN and OAM when it comes to training. Also, those options should be understood just as possibilities and not as restrictive options. |
| NEC |  | Overall proposals 1-9 are too detailed for use case description and look like solution. While we sympathize some of the ideas in proposals 1-9, they fill too limiting for use cases. Proposal 1 and 2: This could be useful for some outputs, but not for all. Looks like more intended for interim predictions, not for final decisions.Proposal 3: Such conclusion should be made after analysing all provided solutions.Proposal 4: Could be some of the inputs depending on ML model/algorithm, but not limited to. We could summarize later based on provided solutions.Proposal 5: Could be examples of outputs depending on ML model/algorithm, but not limited to. We could summarize later based on provided solutions.Proposal 6: Depends on solution. Not always necessary.Proposal 7: Depends on solution. Not always necessary.Proposal 8: It is possible. But this is not the only possible solution.Proposal 9: It is possible. But this is not the only possible solution. |
| ZTE | P1, P2: NoP3: Partly yes.P4,P5: Need more discussionP6,P7,P8, P9: Yes | For P1, P2, we think confidence level and validity time should be discussed case by case. Not all use case needs these two feature.For P3, from our understanding, for ML-enabled load balancing, model inference can also be located at OAM or NG-RAN(CU). Although if model training and model inference are both located at OAM, there is no impact in RAN, we still cannot preclude this deployment.For P4,P5, the input/output for each use case should be further discussed. It can not precluded that load prediction and UE trajectory prediction are useful for making load balancing decision. And the solution, input/output and standards impact need more discussion. |
| Intel | Yes for all | We are ok to discuss proposal 3-9 in standards impact or future solution in later phases.Regarding to “confidence level” or “accuracy”, compared with legacy mechanism, ML model outcomes may not be fully trusted, sometimes bad actions/prediction value may lead to poor system performance or even network shut-down. The detailed value range can be left for FFS.Regarding to “validity time”, the value of validity time can be considered case by case. However, in wireless communication network, the channel condition and other factors (such as number of connected UE, etc) may be changed unexpectedly. It is possible that the previous output from inference model may not be suitable for the time-being. More importantly, when prediction results is exchanged between network nodes, for example, load prediction value exchanged between source and target NG-RAN nodes, since the exchanged information is a fixed value referring to a status in the future, the target node have no idea about “when does this prediction value reflect” or “till when it can use this prediction value as assistance information”. It is the same for the “Actor node”.Hence, we think it is important to add “validity time” to every outcome from ML inference node. |
| InterDigital | P1, P2: need to study case by case, to see what is possible.P3: YesP4, P5 need more discussionP6 YesP7 Yes but could be rewordedP8,P9 yes | P1: Confidence level would be desirable as it is the training entity that can calculate it. Even though this would need to be evaluated case by case because there are multiple ways to calculate a confidence level, and all of them would take specific case characteristics/parameters as inputs.P2: Validity time only makes sense if there is a feedback loop from the actor to the model training entity. The model training entity cannot calculate this accurately otherwise.P3: At this time there is no reason to force model training and inference to a specific network node. For example if there are time sensitive cases that require UE involvement these would not benefit from this proposal. P4: these parameters/aspects are important but no reason to be limited to these. Also, in cases where particular traffic needs to urgently be prioritized, cell capacity should not play a role. So committing to this makes less sense.P5: Makes less sense and rationale is the same as for P4.P6/P8/P9: Yes but aren’t these obvious?P7: Proposed rewording: For ML-enabled load balancing, handover decision at source cell can take in account of the prediction results of its own and neighbouring cells. |
| Ericsson | Proposal 1: Needs clarification Proposal 2: Needs clarificationProposal 3: YesProposal 4: Not entirelyProposal 5: Not entirelyProposal 6: Yes.Proposal 7: YesProposal 8: YesProposal 9: No | For Proposal 1, we found useful to provide the accuracy of an output prediction, when available. Hence the presence of this data should be conditional to its availability, while the nature of the data should be an accuracy level, in case the output is a prediction.P2: perhaps we could further discuss the need of a prediction window. Normally we enable data retrieval (and the same could be applied to ML output retrieval) on a periodic basis, where ehte data will be valid until the next period. This would naturally provide info about the prediction windowP3: this could be an initial assumption.P4: QoS Requirements does not seem to be relevant. Needs more discussionsP5: We do not see how action of UE is an output of the ML model. The ML model in our view would return a load prediction on the basis of which actions on the UE would be taken. The ML model does not return the actual action on the UE. P8: this belongs to a different use case, which is the energy saving use case. P9: it should be rather that RAN nodes exchange their load predictions and as a consequence a cell activation/deactivation is taken by the node serving the cell.  |
|  | P1&2, not sure the intentionP3, partially agreeP4&5, in general yes, may see more info needed along with further studyP6&7&8, yesP9, not sure the intention | For P1&2, output of inference, in our understanding, could be a HO command or cell switch-off command, as existed in current spec, or maybe some policy for action to consider, then we are not sure if we need to introduce additional mechanisms just because the order comes from AI/ML inference…For P3, as proposed in our paper, we think “training” should be located in OAM and “inference” in RAN (CU)For P4, as the input for load balancing model, actually they might be also related with specific training method, but anyway, we think load status, load predication of neighbour node, radio quality etc. should be needed.For P5, similar as above, the output of load balancing inference could be a mobility command or just some load prediction for further decision.For P6, yes.For P7, yesFor P8, yes, actually this is somehow overlapped with energy savingFor P9, not sure the intention, we think prediction result could be exchanged between neighbour nodes, but activation/deactivation could be consequence of the output of inference. |
| Lenovo, Motorola Mobility | P1: OK for prediction taskP2: FFSP3: fine as WAP4: FFSP5: UnclearP6: see commentP7: OKP8: OKP9: no | P1: ok if the inference output is a prediction. For decision oriented tasks, concept of “Confidence Level” is not applicable.P2: FFS. If the inference output is a prediction, sometime related information could be applicable, such as a time point for which the prediction is applicable. Maybe we can leave it FFS for now. P3: fine as a WAP4: FFSP5: Unclear. In our understanding, one ML model is dedicated for one task. Does this proposal mean to support ML-enabled load balancing, multiple ML models for different tasks e.g. prediction of traffic volume, etc. are needed?P6: The relevant information exchanged between Xn interfaces shall be based on the demand of the peer node. It is upon the peer node’s implementation if e.g. traffic prediction is useful or not.P7: OK. This is one of the possibilities, basically a deterministic algorithm is used for HO decision and uses prediction result as input. P8: OK. This is one of the possibilities.P9: No. the decision of deactivating a cell should be in the hand of its own gNB. |
| Qualcomm | Agree all,as working assumption | It may be too early to decide such level of details. But, these proposals can be agreed as working assumption, and be used as starting point of future discussion. |

**Summary：For P1 and P2, some companies think that not all ML models need confidence level and/or validity time, and should be discussed case by case. For P3-P9, more discussion and clarification are needed since they are more related to solutions rather than use case description.**

**Q2: Which following proposal(s) from [16] is agreeable for you?**

1. It is proposed to sufficiently progress the prioritized use cases on energy saving, load balancing, traffic steering/mobility optimization, i.e. at least by identifying their impact on the specifications, before considering any new use case.
2. It is proposed to investigate standard support to enable machine learning at the network side to reduce the energy consumption at the UE, such as the introduction of new UE feedback information reflecting the UE’s energy consumption.
3. It is proposed to investigate a RAN-based Energy Efficiency solutions to achieve better energy consumption at the RAN, for example by means of exchange of this information.
	* + **Exchanging information between RAN nodes, describing if a certain RAN action is taken due to energy efficiency**
		+ **Exchanging of RAN node energy-related information between RAN nodes**
		+ **Exchanging performance feedback related to a certain energy efficiency action taken in another RAN node**
		+ **UE traffic prediction**
		+ **UE mobility prediction**
		+ **UE performance feedback**
4. RAN3 to discuss the following improvements related to Traffic Steering:
	* + Improved Mobility Load Balancing decisions using UE traffic related predictions and RAN load predictions.

**Proposal 5: It is proposed to consider the Link Adaptation use case as low priority.**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| Futurewei | Proposal 1: YesProposals 3 and 4: Yes, with clarificationProposal 2 and 5: No | Proposal 2: The discussions so far were for RAN power saving, not for UE. So UE power saving should be considered a new use case. As a new use case, it should have lower priority and should be postponed. Proposal 3: 1) Item 3: To clarify: what does "performance feedback” mean? 2) Item 4: To clarify: what does UE traffic prediction mean, per UE? If yes, this would require a new use case to predict UE traffic?3) Item 5: similar as the above, UE mobility prediction is not included in the use cases being proposed so far.In general, we believe ML-based RAN energy saving benefit can be evaluated at OAM or other-higher level entity instead of exchanging all the information between RAN nodes.Proposal 5: This proposal is not necessary. If other use cases score higher based on agreement, then link adaptation use case becomes lower priority naturally. |
| CATT | P1, P4, P5: yes.P2: no.P3: cannot agree now. | For P2, we don’t think UE power saving belongs to the agreed “prioritised use cases”.For P3, similar with the comment in the tabular above, need to clarify the benefit of AI-assisted MLB decision, assuming that AI-assisted predictions are already supported. |
| Samsung | P1: NoP2: NoP3: Prefer to identify AI functionality first, and then to study input/output and interface impact.P4: YesP5: Partly yes | P1: prefer to identify the new potential use case in parallel with studying the standard impact of agreed use cases. P2: UE power saving seems not belong to the agreed use case. And it is more relevant to other WGs. Slightly prefer to down-prioritize it.P3: the input/output and interface impact highly depends on ML model functionality. Suggest to identify ML model functionality firstly.P5: it seems that link adaptation is highly relevant to other WGs. Slightly prefer to down-prioritize it.  |
| CMCC | P1: YesP2-P4: cannot agree now | For P2, agree with CATT and Samsung that UE power saving does not belong to the agreed “prioritised use cases”. It is better to discuss P3-P4 in the standard impact study of corresponding use case.  |
| Nokia | P1:YesP2, P3, P4, P5: No | P2: This SI is about introducing ML in the RAN. To our understanding it is not about studying UE improvements but RAN improvements.P3, P4: We should not focus on UE predictions and extensive new UE measurements that may have high UE impacts. This is a RAN3 led SI where minimal input can be expected from other groups, e.g., RAN2, based on LSs.P5: We should focus on the approved use cases and reach some sufficient baseline description before studying new ones. We have already agreed on the prioritized use cases. |
| NEC |  | Proposal 1. It is fine for us to concentrate on merging prioritized use cases:- Network energy saving- Network load balancing- Mobility optimization- Traffic steering.For us it is also fine to define supporting use cases:- UE position prediction- Traffic prediction- Load prediction.Also, it is important to distinguish first group above and second group above. For example, Network load balancing is different from Load prediction. Mobility optimization is different from UE position prediction.Proposal 2. No problem for us if time allows given the priorities mentioned in our comment to Proposal 1.Proposal 3.This depends on solutions. Better first to see solution description, then decide which information should be exchanged.Proposal 4.Traffic steering is prioritized, so we support this use case. Second half of the proposal is solution. Better to see full solution description before decision.Proposal 5.This looks like RAN1/RAN2 topic. |
| ZTE | P1,P5: YesP2, P3, P4: No | For P2, we agree with SS and CATT. UE power saving seems not belong to the agreed use case, and it seems out of RAN3-led SI scope.For P3,P4: The input/output, impacts and solution for each use case should be further discussed. |
|  |  |  |
| Intel | Proposal 1: Not fully agreed.Proposal 2: No.Proposal 3: Agree with commentProposal 4: AgreeProposal 5: No. | Proposal 1: In RAN3 #110e meeting, we agreed that energy saving, load balancing and traffic steering/mobility optimization as the starting point, and other use cases are not precluded. New use cases can be considered and discussed by companies’ contributions.Proposal 2: the proposal is not quite clear how to measure UE energy consumption and how to report it to the network.Proposal 3: Regarding to “UE traffic prediction”, “UE mobility prediction”, and “UE performance feedback”, from our understanding, besides UE reporting such prediction results to the network, the network node can also use historical uplink and downlink information to predict for each accessed UE, where such training and inference are in the network side. Hence, UE prediction information does not necessarily to be reported by UE.Proposal 5: As commented for proposal 1, we are not going to exclude any other use cases. |
| InterDigital | P1, P2: YesP3: NoP4: Maybe | P1: It is best to identify additional use cases in parallel.P2: In order to achieve energy savings at UE level from a RAN perspective, it is necessary for the RAN to receive this kind of feedback from the UE.P3: Proposal is too general and the information that needs to be exchanged should be evaluated in a case-by-case basis.P4: UE traffic related predictions and RAN load predictions are good to improve traffic steering. However, proposal is again too general.P5: This should be deprioritised  |
| Ericsson | Yes to All | Comments on P2: Energy efficiency for the UE is NOT relying on an ML algorithm at the UE, but rathe ron an ML algorithm at the RAN, which helps with energy efficiency improvements at the UE.Th ecomments that this use case is out of scope are not correct, the SID states:*Study high level principles for RAN intelligence enabled by AI, the functional framework (e.g. the AI functionality and the input/output of the component for AI enabled optimization) and identify the benefits of AI enabled NG-RAN through possible use cases e.g. energy saving, load balancing, mobility management, coverage optimization, etc*Namely, we need to identify the benefits of AI enabled NG-RAN (which is what the use case studies) through the energy saving use case. The use case agreed for prioritisation is not specifically narrowed to RAN node energy savingP3: we need to start from a general list of inputs. We can capture this as FFS, but at least we would need to capture something. |
| Huawei | P1, yesP2, maybe notP3&4, in general yesP5, yes and left to next release | For P1, yes.For P2, we are a bit hesitated towards that direction, sometimes to maximize the energy saving at network side may not lead to UE’s energy saving, and how to evaluate UE’s energy efficiency is another topic actually.For P3, in general yes; at least some prediction info (traffic, mobility, performance) are needed, and some action order (e.g. switch on/off command) is also needed.For P4, anyway, for traffic steering, some prediction info should be needed, including traffic prediction and load predictionsFor P5, maybe Link adaptation could be left to next release. |
| Lenovo, Motorola Mobility | P1: fine in principleP2: not nowP3: based on demandP4: okP5: ok | P1: fine in principle. But just to be clear that we consider PSCell change is also in the scope of traffic steering/mobility.P2: Not now. The proposal implies RAN2 impact, we shall focus on the RAN side for the time being.P3: The relevant information exchanged between Xn interfaces shall be based on the demand of the peer node. It is upon the peer node’s implementation if e.g. traffic prediction is useful or not.P4: OK, and in case of any UE impact, RAN2 should be involved. P5: OK |
| Qualcomm | P2, P5: NoP1, P3, P4: OK | P2: This study is for RAN improvement, instead of UE.P5: We should not exclude this use case for now. |

**Summary：It seems that only P1 obtained sufficient support. For P2, as pointed by companies that UE power saving does not belong to the agreed “prioritised use cases”. P3 and P4 are more related to solutions and cannot agree now. For P5, some companies deem that the proposal is not necessary. Since we have already agreed on the prioritized use cases, then link adaptation use case becomes lower priority naturally.**

**Moderator’s proposal:**

**Proposal 1: It is proposed to sufficiently progress the prioritized use cases on energy saving, load balancing, traffic steering/mobility optimization, i.e. at least by identifying their impact on the specifications, before considering any new use case.**

**Q3: Which following proposal(s) from [11] is agreeable for you?**

Proposal 1: RAN3 is proposed to consider splitting the use cases into two parts: “AI-assisted decision” where the output is the decision itself (i.e. to turn on/off a cell, to hand over a UE from this cell toward that cell), and the “AI-assisted prediction” where the output is merely the prediction result used as an intermediate input for decision making.

Proposal 2: If proposal 1 is agreed, RAN3 is proposed to focus firstly on “Load prediction” “UE location prediction”, “Service data prediction” and “AI-assisted handover / data offloading decision”.

Proposal 3: If proposal 1 is agreed, RAN3 is proposed to confirm that “cell/carrier/symbol/etc turning on/off decision” and “radio resource configuration / scheduling” can be affected by the aforementioned predictions.

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| Futurewei | Proposal 1-3: No | There is no need for these proposals. Whether to use ML output or inference to make decision directly or just as part of the decision-making is implementation dependent. |
| CATT | Slightly prefer splitting the use cases into “AI-assisted decisions” and “AI-assisted predictions” (or identically, “tool-box use cases”). | An alternative option: includes only the decisions (both AI-assisted and non-AI-assisted conventional ones) into the “use case” section, but add a new section parallel to the “use case”, maybe namely “tool box”. |
| Samsung | Slightly prefer to classify the solution as “ML-generated solution” and “ML-assisted solution” | For one specified use case, the solution can be classified based on ML model functionality, as “ML-generated solution” and “ML-assisted solution”. For example, for use case of load balancing, one functionality of ML model is to predict load, and ML-assisted solution is load balancing strategy obtained based on the predicted load; or ML model functionality is to generate load balancing strategy, and ML-generated solution is the load balancing strategy obtained from ML model directly. |
| CMCC | Proposal 1-3: cannot agree now | Not sure P1 is necessary. |
| Nokia | P1, P2, P3: No | About P1, P2 and P3: We don’t see it necessary to split use cases in different categories. We have identified the use cases we will study. In our view “load prediction”, “UE location prediction”, “service data prediction” and “AI-assisted handover/data offloading decision” are enablers to study different use cases and can be used transparently for the identified use cases. |
| NEC |  | Proposal 1.No problem to have two categories of use cases. However, naming could be improved to avoid confusion with prediction/classification.First category: Output is directly used by RAN.Second category: Output is used by other ML models as input.Proposal 2.Prioritized use cases are:- Network energy saving- Network load balancing- Mobility optimization- Traffic steering.First, these use cases should be merged.For us it is also fine to define supporting use cases:- UE position prediction- Traffic prediction- Load prediction.Proposal 3:Such decision is better to make after analysing provided solutions to use cases. |
| ZTE | Prefer to regard “load prediction” and trajectory prediction” as the tool-box use case to study. | Predicted load and predicted UE trajectory can be the information for some use case, such as load balancing, mobility optimization, etc, to make specific decision.It is essential to add a new section as tool-box use including load prediction and trajectory prediction. |
| Intel | Proposal 1: Agree with the intention, while we can capture in another way.Proposal 2: No. Proposal 3: Agree. | Proposal 1: The reason to consider splitting the use cases/solutions into “AI-assisted decision” or “AI-assisted prediction” is caused by using different ML model types in certain use cases. If the detailed ML algorithm for one use case belongs to reinforcement learning, then the outcome from the inference could be “AI-assisted decision”. Otherwise, one may get “AI-assisted prediction” from ML algorithms belonging to supervised/unsupervised learning. Hence, instead splitting use cases into two directions, we suggest considering three types of ML models, which is more common and acknowledged by the industry. Proposal 2: Same as comment to proposal 1/5 in Q2. |
| InterDigital | P1,P2,P3: No | Although we feel the approach is nice, we do not see the need for Proposal 1. Therefore, the other proposals fall by themselves. |
| Ericsson | P1, P2, P3: No | We do not see the need, nor the gain from a standard point of view, to split use cases in AI-Assisted decisions or AI-assisted predictions. We will study the use cases and determine if the AI output is a prediction or not and we will specify the information flow on the interfaces of relevance accordingly |
| Huawei | No | For P1, we are not sure. We understand the intention, but the load prediction, UE location prediction, service data prediction, etc are tools, they are not use cases. With this understanding, P2 is not needed. Regardless of P1, P3 seems to us is a general description whose intention could be reflected in the TP. |
| Lenovo, Motorola Mobility | Ok for the sake of make the relevant discussion clear. Maybe no need to capture in the spec.  | P1: ok, we share the same view, which could facilitate the relevant discussion.P2: shall also consider AI based PSCell change, which we believe is in the scope of the mobility and traffic steering.P3: sure, that is one of the possibilities.  |
| Qualcomm | P1, P2, P3: No | These are the model implementation details and the impact to common architecture is not clear. So, we prefer not to agree for now. |

**Summary: It seems that none of the proposals obtains enough support.**

## 3.2 EE/ES

Several papers provide views and potential TPs on the use case of energy saving. Moderator would like to suggest companies to discuss the potential TP for EE/ES based on 2030[6], where detailed use case description, simulation results and potential benefits are provided, and merge agreeable parts in [1][2][17][21][22].

**Q4: Do you agree to work on the potential TP for use case description of EE/ES based on [6] and merge other TPs?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| Futurewei | Yes. |  |
| CATT | Yes |  |
| Samsung | Yes |  |
| CMCC | Yes |  |
| Nokia | Yes | Small clarification perhaps in the text that we don’t consider AI but rather ML (ML techniques, ML Energy Saving, ML algorithms, etc.). |
| NEC |  | Yes. We prefer “Network Energy Saving” as a name for the use case. |
| ZTE | Yes |  |
| Intel | Yes | We suggest having a more general use case description and solution description.1. From solution direction point of view, we suggest not to limit the scope of prediction only to traffic load, as described in [6]: “AI algorithms predict the state of the next period , specially the traffic load,”. Instead, we think the second paragraph in section 5.2.1 of [17] is more general, including not only load prediction, but also other factors, such as cell utilization, etc, which may also impact energy saving at NG-RAN nodes.2. As agreed that we will not discuss the detailed ML algorithms for use case, we suggest not to capture the detailed ML algorithms in [22], such as “RF” or “LSTM”. |
| Ericsson | Partially yes | The TP is fine but the last part of the text mentioning the simulation results is not ok. |
| Huawei | In general yes | In general, we are ok with moderator’s suggestion, anyway, more or less those TPs have some common points. We just need to make updates according to common understandings. |
| Lenovo, Motorola Mobility | Yes |  |
| Qualcomm | Yes |  |

**Summary: All companies support to use 2030[6] as the baseline for EE/ES, companies’ comments should also be taken into account during the second-round discussion.**

**Moderator’s proposal:**

**Proposal 2: Use 2030[6] as the baseline for use case description of EE/ES to work on the potential TP, companies’ comments should also be taken into account.**

## 3.3 Load Balancing/ Load Prediction

Moderator would like to suggest companies to discuss the potential TP for load balancing based on 2524[23] and merge agreeable parts in [2][4][8][20].

**Q5: Do you agree to discuss the potential TP for use case description of load balancing based on [23] and merge other TPs?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| Futurewei | Yes. |  |
| CATT | Generally yes, but | Not quite sure of the following bullet in [23]:* The load balancing based handover highly depends on the measurement report from the UE. The UE measurement configuration and measurement reports may cause amount of signaling overhead over Uu interface, UE power consumption and data interruption of running service.

We are not sure the current MLB mechanism *does* cause “amount of signalling overhead”, so maybe we needn’t mention it here. The remaining two bullets are already sufficient enough. |
| Samsung | Partly yes | Same view with CATT. It is unclear why “measurement based HO” causes large overhead.*-The load balancing based handover highly depends on the measurement report from the UE. The UE measurement configuration and measurement reports may cause amount of signaling overhead over Uu interface, UE power consumption and data interruption of running service.* |
| CMCC | Yes |  |
| Nokia | Yes, but | The use case description from TP in [23] is OK for us. We also have a problem with the second bullet but for an additional reason. It creates an impression that the use case will be about optimizing UE measurement configuration which should not be the case.  |
| NEC |  | Yes for use case description part. Solution part should be discussed separately. |
| ZTE | Partly yes | Generally, use case description in TP [23] is fine for use. However, same concern as CATT and Nokia, the second bullet seems the ML-enabled load balancing has the optimization for UE? |
| Intel | Yes | The use case discussed in [4] is UE trajectory prediction, which is a separate use case than load balancing/load prediction. Hence, we suggest considering [4] as a separate use case, and not merge it in the use case description of load balancing. |
| InterDigital | Yes, but | We agree that this second bullet needs clarification |
| Ericsson | Yes | However, we also do not see that UE measurements have much to do with mobility load balancing and would like to remove bullet 2 |
| Huawei | ok | See comments to Q4 |
| Lenovo, Motorola Mobility | Yes |  |
| Qualcomm | Yes |  |

**Summary: All companies support to use 2524[23] as the baseline for Load Balancing/ Load Prediction, companies’ comments, e.g. removing bullet 2, needs to be taken into account during the second-round discussion.**

**Moderator’s proposal:**

**Proposal 3: Use 2524[23] as the baseline for use case description of Load Balancing/ Load Prediction to work on the potential TP, companies’ comments should also be taken into account.**

## 3.4 Mobility/ Traffic Steering

Moderator would like to suggest companies to discuss the potential TP for mobility based on 2271[14] and merge agreeable parts in [2][12][18].

**Q6: Do you agree to discuss the potential TP for use case description of Mobility/ Traffic Steering based on [14] and merge other TPs?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| Futurewei | Yes. |  |
| CATT | Yes, but | Two comments:1. It is OK to capture use case related to CHO/DAPS. However, maybe the CHO/DAPS part could be discussed after we have reached consensus on normal mobility.2. UE mobility prediction and/or UE traffic prediction is not mention, so merging with other contributions seems necessary. |
| Samsung | Partly yes | Prefer to merge CHO/DAPS into mobility management, as CHO and DAPS are two mechanisms of mobility management.Load balancing is agreed as a separate use case, so it is better to put load balancing related description in load balancing use case part.Coverage optimization is a standalone use case. Whether to involve it to be the agreed use case can be discussed later. So prefer to not involve the coverage optimization part in mobility description. |
| CMCC | Yes |  |
| Nokia | Yes, but | In [14] coverage optimization is described. However, since coverage optimization is not in the agreed use cases, a use case description is not needed for it. Also, DAPS could be considered at a later stage when the baseline mobility use case is sufficiently studied. Finally, given that this is a RAN3-led SI we do not think it is feasible to introduce extensive Uu impacts. This comment is related to the following part of the TP:“The relevant information needed from the UE may include, amongst others: * UE radio conditions
* UE location
* UE trajectory
* UE connectivity status
* UE capabilities

Other UE contextual information” |
| NEC |  | No, we believe it is too detailed and limiting, more like a solution. Instead, we would propose to consider Mobility optimization use case in [2] R3-212316 or [18] R3-212316 as a starting point for merging. |
| ZTE | Yes, but… | Load balancing is regarded as a standalone use case to discuss. No need to duplicate the description.Coverage optimization and CHO/DAPS Resource optimization can be further discussed later when the prioritized use cases are studied sufficiently. Besides, agree with CATT, trajectory prediction should be mentioned in the description, because trajectory prediction, we think, is the important part for ML-enabled mobility management. |
| Intel | No | We think mobility/traffic steering use case description should use [18] as the baseline.1. The description in [14] is overlapped with use case description for load balancing.2. CHO and DAPS optimization should be considered as a new use cases, rather than mobility and traffic steering optimization.In this case, we suggest to the potential TP should be based on [18]. |
| InterDigital  | Yes but | We of course agree to use [14] as baseline 😊Given the feedback on coverage optimization, we are ok to remove this to be handled at later time.Also agree that load balancing is covered in another use case.As for the Nokia comment on Uu impacts listed in [14], we understand the concern and we can simplify the text here, we think that the RAN3 study will not go deeper than recognizing potential Uu impacts, going into details like listed in [14] would come in later follow up SI/Wis with RAN2 participation.Looking at the other contributions, areas that I think should be included UE mobility prediction (which would be a major input in data forwarding optimization) and traffic steering.  |
| Ericsson | Yes but | We would like to remove the parts on CHO and DAPS. Being this a generic TP, we should only focus on a general description of mobility optimisation vi aAI, and remove any specific example (like CHO and DAPS). There might be many more mobility optimisation cases and we need to converge on them before agreeing them for a TP.  |
| Huawei | ok | See comments to Q4 |
| Lenovo, Motorola Mobility | Yes |  |
| Qualcomm | Yes |  |

**Summary: Majority companies support to use 2271[14] as the baseline for Mobility/ Traffic Steering, companies’ comments, e.g. removing coverage optimization, keeping or removing remove the parts on CHO and DAPS, needs to be taken into account during the second-round discussion.**

**Moderator’s proposal:**

**Proposal 4: Use 2271[14] as the baseline for Mobility/ Traffic Steering to work on the potential TP, companies’ comments should also be taken into account.**

## 3.5 Other use cases

Some new use cases are proposed to discuss, including

1. URLLC [3]
2. PSCell change [10]
3. Network slicing, QoE optimization [15]
4. L1/L2 Beam Management Configuration and Operation Optimization，Multi-user MIMO Configuration and Optimization， Scheduler Tuning and Parameter Optimization [19]

Companies are invited to provide views on whether to study above new use cases.

**Q7: Do you agree to study above new cases?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| Futurewei | No. | 1) Suggest we postpone these topics to later stages.2) Some use cases under 4 belong to other RAN groups; should they be handled by RAN3? |
| CATT |  | Bullet 2 seems already covered.Neutral for the rest. |
| Samsung | Yes for URLLC | URLLC is a potential use case to maintain the stringent QoS requirements, as network variations (e.g. channel state, mobility, etc.) make it challengeable to keep performance in a stable near-perfect level. AI can help to do prediction or generate optimized policy with high resiliency to variations.PDCP duplication optimization is one of the potential directions for AI-aided URLLC, as complexity dramatically increases due to up to 4 RLC entities now. AI can help to set accurate/adaptive configuration & selection among RLC entities to realize trade-off b/w QoS performance and resource efficiency. |
| CMCC |  | Prioritized use cases should be studied first, and then new use cases could be studied if time allowed.  |
| Nokia | 1, 2: No QoE Optimization: NoNetwork slicing: Yes4: Maybe | URLLC: URLLC use case seems interesting but it is difficult to evaluate it without UE impacts (and heavy RAN2 involvement). Perhaps some more information is needed to evaluate the RAN3 aspect.PSCell change could be considered as part of the mobility use case (not as a separate use case) but after we sufficiently study the basic mobility mechanism. QoE optimization: RAN visible QoE is covered by Rel-17 WI objective, but remains a new topic. So, it is preferable to avoid parallel discussions in QoE WI and AIML SI. Use cases in 4 could be interesting alternatives to the agreed L3 use cases. But it is a bit obscure to us at the moment how we can evaluate them and what will be the RAN3 impacts since they are pretty much focused in the DU. |
| NEC |  | We support the following prioritization based on the previous agreements and interest in supporting use cases from several companies.Prioritized use cases are:- Network energy saving- Network load balancing- Mobility optimization- Traffic steering.First, these use cases should be merged.For us it is also fine to define supporting use cases:- UE position prediction- Traffic prediction- Load prediction.If there is time after above use cases, we are open to consider more use cases. |
| ZTE | No | We should focus on the prioritized use case(e.g. ES, Load balancing and mobility management) in R17.QoE optimization and Network slicing can be considered as potential use case to discuss later. |
| Intel | Yes | Detailed solutions or proposals can be further discussed in the following meetings. |
| InterDigital |  | We agree that these are all important use cases to complete over time.1. URLLC is interesting, would be open to consider
2. PSCell is an advanced mobility topic and could be considered after baseline mobility (but of course some of this may come free with other mobility enhancements)
3. QoE is too new would prefer to look at this once the paint is dry on R17 QoE definitions, Network slicing possibly, but still considering RAN slicing in R17.
4. I think this is a great area of optimization, however it depends heavily on RAN1 input, so it not practical for this study. Probably should be looked at in R18 with RAN1 inputs.
 |
| Ericsson | NO | As described in [16]´, for the sake of reaching good progress and some usable study outcome, we should first prioritise and study to a good extend the prioritised use cases. We can then decide which extra use cases we will tackle. |
| Huawei | Maybe not | Maybe there is no need to rush into new use cases, let’s focus on the three use cases on the table and reach consensus which could be referred as base line for further discussions of other use cases. |
| AT&T | Yes | We are OK to consider these use cases which impact other WGs (e.g. RAN1/RAN2) with less priority, however it would be good to at least capture them at a high level in the SI, especially if they can utilize the common framework defined for other use cases. |
| Lenovo, Motorola Mobility | Yes | We actually believe PSCell change is in the scope of Mobility optimization/Traffic steering. We would expect the solution used for HO, e.g. prediction information exchanged over Xn, can be to a large part reused for AI assisted PSCell change. |
| Qualcomm | Yes for QoE, no for other in this release | A key advantage of AI/ML algorithm is to make decision based on input from multiple sources (e.g. multiple protocol layers, multiple nodes). QoE is a good use case for us to study the architecture for such kind of cross-layer optimization. |

**Summary: It seems that none of the new use case obtains enough support.**

# 4 Conclusion, Recommendations

**Proposal 1: It is proposed to sufficiently progress the prioritized use cases on energy saving, load balancing, traffic steering/mobility optimization, i.e. at least by identifying their impact on the specifications, before considering any new use case.**

**Proposal 2: Use 2030[6] as the baseline for use case description of EE/ES to work on the potential TP, companies’ comments should also be taken into account.**

**Proposal 3: Use 2524[23] as the baseline for use case description of Load Balancing/ Load Prediction to work on the potential TP, companies’ comments should also be taken into account.**

**Proposal 4: Use 2271[14] as the baseline for Mobility/ Traffic Steering to work on the potential TP, companies’ comments should also be taken into account.**

# 5 Reference

1. R3-211669 Machine Learning Use Case for BS Power Saving Futurewei
2. R3-211683 TP to TR 37.817 Use case description NEC
3. R3-211969 Discussion on Use Cases for RAN Intelligence Samsung, Verizon Wireless
4. R3-212028 AI based UE Trajectory Prediction ZTE Corporation, China Unicom, Lenovo, Motorola Mobility, CMCC
5. R3-212029 Solution to AI based UE Trajectory Prediction ZTE Corporation, China Unicom, CMCC
6. R3-212030 AI based Energy Saving ZTE Corporation, China Unicom, Lenovo, Motorola Mobility
7. R3-212031 Solution to AI based Energy Saving ZTE Corporation, China Unicom
8. R3-212032 AI based Load Prediction ZTE Corporation, China Unicom, Lenovo, Motorola Mobility
9. R3-212033 Solution to AI based load prediction ZTE Corporation, China Unicom
10. R3-212179 AI based PSCell change Lenovo, Motorola Mobility
11. R3-212190 Discussion on use cases for AI in RAN CATT
12. R3-212191 Discussion on UE Location prediction CATT
13. R3-212269 Data Forwarding Optimization Use Case for AI InterDigital
14. R3-212271 Mobility Optimization Use Case for AI InterDigital
15. R3-212301 Use cases for AI/ML enabled NG-RAN Intel Corporation
16. R3-212313 Overview of AI/ML use cases Ericsson
17. R3-212315 AI/ML for energy efficiency use case discussion Ericsson
18. R3-212316 AI/ML traffic steering use case discussion Ericsson
19. R3-212389 Use Cases and Requirements for Artificial Intelligence in RAN AT&T
20. R3-212504 AI-based load balancing CMCC
21. R3-212507 AI-based energy saving CMCC
22. R3-212523 Further discussions on detailed procedure and potential spec impacts of energy saving Huawei
23. R3-212524 Further discussion on AI/ML assisted load balancing Huawei