3GPP TSG-RAN WG3 Meeting #108-e R3-203957

E-meeting, 1 – 11 June, 2020

**Agenda item: 10.2.2.1**

**Source: Nokia (moderator)**

**Title: Summary of discussions on** **Mobility Load Balancing**

**Document for: Approval**

# 1 Introduction

This paper provides summary of discussions at RAN#108-e on:

**CB: # 1006\_Email\_SONMDT\_MLB**

**- Topics for discussion**

**- Misc. corrections**

**- Active UEs (which interfaces, how to encode)**

**- TNL capacity (range, which interfaces, granularity)**

**- HW capacity (which interfaces, granularity)**

**- RRC connections (range, which interfaces)**

**- SUL (whether to support, parameters)**

**- Misc corrections and alignment**

**- Can also discuss other issues based on contributions submitted**

(Nok - moderator)

I have separated out reporting per slice in a separate question in section 3.6 (might have been considered as "granularity" above).

Companies responsible for MLB TPs are proposed as follows:

* X2AP: CATT
* F1AP: Huawei
* E1AP: Nokia
* XnAP: Ericsson
* (no stage 2 TPs submitted)

# 2 For the Chairman’s Notes

Draft agreements (companies may still have comments):

**Number of Active UEs to be reported on X2, Xn and F1 interfaces, based on definition in TS 38.314 section 4.1.1.3.5 (Mean number of Active UEs per cell), with granularity 0.1.**

**Add TNL Capacity Indicator per node over X2/Xn (reflecting backhaul load), while cell load reflects fronthaul load in case of split architecture.**

**RRC connections metric to be removed from X2AP BL CR.**

**Continue discussion on reporting of SUL related parameters in Rel-17.**

**No additional metric reported per slice in Rel-16 (remove metric marked as FFS from E1AP BL CR).**

Additional question from miscellaneous corrections:

From 3676 proposal 4: Whether Number of RRC connections (on Xn) reflects PCell connections only, or both PCell and PSCell connections.

# 3 Discussion

## 3.1 Active UEs (which interfaces, metric definition, how to encode)

**Interface for reporting:**

For use case for reporting, one operator observes (3471): ***There is a big demand of MLB between LTE ad NR cells for initial EN-DC deployment***

**Question: Can priority be given to the EN-DC use case in Rel-16, and hence reporting on the X2 interface?**

|  |  |
| --- | --- |
| Company | Comment |
| ZTE | Fine to support this metric in X2AP for EN-DC. |
| CMCC | Xn should be also prioritized besides X2. The reason is it is a Rel-16 feature, when it is commercialized, Standalone network will be largely deployed, at least for china mobile, large scale SA network is planned soon. |
| Deutsche Telekom | What is meant in the question by “priority”? The number of active UEs should be reported on X2, but also on Xn and F1 interfaces. |
| Ericsson | We are fine to support the Mean Number of Active UEs as agreed by RAN2 in R2-2005434 (section 4.1.1.3.5 Mean number of Active UEs per cell), for X2 and Xn |
| CATT | Yes. We think it reasonable. |
| Nokia | In response to DT, by "priority" the idea indeed was X2 only in Rel-16... But no problem also to have Xn. But I think justification for F1 is more related to support of the definition from 38.314 (RLC/MAC buffers), and not for e.g. access control use in the CU because enough other information is already available on this interface. See our view on 38.314 below. |
| Samsung | We support the same metric in X2 and Xn in Rel-16. |

**Metric definition:**

Several companies propose using mean (and potentially also max) values defined in TS 38.314. Other companies question e.g. whether the current version of this specification is suitable as reference for RAN3 stage 3 specification.

**Questions: Could a mean value be sufficient for reporting in Rel-16? Does reference to TS 38.314 provide significant benefit?**

|  |  |
| --- | --- |
| Company | Comment |
| ZTE | At least support mean (defined in TS 38.314). The definition help Nodes to evaluate User plan load. |
| HW | We prefer to keep a single value (mean). Remember that the reporting period is quite short, so the benefit of adding more statistical reporting (min, max) may not be as important compared to if we would have longer reporting interval.  We are OK to refer to 38.314 but if the measurement is simple (a mean value) we think a simple definition in our spec is enough.  Note that we propose to align to 38.314 by using one value per cell, i.e. also no UL/DL differentiation |
| CMCC | At least support mean value per cell defined in TS38.314.  The main concern not to refer to 38.314 is the definition in 314 doesn’t consider PDCP buffer. However, since this metric will be measured and reported at the level of at least one second, there’s little chance that PDCP buffer has data to transmit while RLC/MAC has no data. So referring to 38.314 for RAN3 stg3 specs is enough in our opinion.  In addition, we identify that per DRB (QoS) level values for load reporting is beneficial. As specified in 38.314, such metric counts all DRBs mapped with the same 5QI for NR SA or with the same QCI for EN-DC, which is a proper and straight-forward way to differentiate various services in terms of QoS requirements. So we also propose to consider per QoS level load reporting on Xn/F1/X2.  Last, RAN2 spec is quite stable to refer to, although it is not approved by RAN plenary. It is planned to submit to last RAN plenary for approval, but failed just because of Rapporteur’s mistake to increment the wrong digit version number. |
| Deutsche Telekom | Preference for use of definition of mean number of active UEs given in TS 38.314.  DT also supports DCM’s views in R3-203471 to provide additionally the max number and having both values on a per cell and a per DRB per cell basis. |
| Ericsson | RAN2 has agreed to mean number of Active UEs in R2-2005434 (section 4.1.1.3.5 Mean number of Active UEs per cell). We would like RAN3 to re-use this value for the Active UEs measure, so to ensure we do not generate yet another version of an active UE statistics. Adding a mean value (as opposed to e.g. a maximum value) is the most useful enhancement as it allows to have a statistical representation of the number of active UEs in the cell. This can be compared with e.g. PRB utilisation to derive the status of the cell. |
| CATT | We do not have any strong opinion over mean/max, but we do not find any benefit to deliver any per-DRB load.  Whether TS 38.314 can be reference can be discussed days later maybe. |
| Nokia | Our understanding is that TS 38.314 is still not stable, and current definitions doesn't fulfil the goal of inter-vendor operability. In order to represent well high-load scenarios, to me the PDCP buffer should also be considered, or actually, even better, using (in)activity info (from E1) available in the CU-CP. If very detailed measurement definition is deemed beneficial, it is better to introduce it in Rel-17 in my view instead of using something that is still shaky.  The discussed metric is not for the purpose of statistics, but for load balancing and anticipation of access control. |
| Samsung | Support a single mean value by referring TS 38.314 in Rel-16. Others can be discussed in Rel-17. |

**How to encode:**

Different encoding schemes have been proposed, including granularity of 0.1 and 1.

**Question: Can granularity of 1 be considered sufficient?**

|  |  |
| --- | --- |
| Company | Comment |
| ZTE | Prefer 1 granularity than 0.1, the latter may introduce misunderstanding. |
| HW | We are open for both. The decimal may not be important for selecting proper load balancing. The more coarse we make the measurement, the more important it will be to consider defining the rounding method (e.g. floor) |
| CMCC | According to agreement by RAN2#109bis-e, ‘*The unit of mean number of active UEs is changed from integer to 0.1, in order to keep align with the equation*’, we think align with RAN2 is enough and the granularity is 0.1. |
| Deutsche Telekom | As mean numbers are computed, a granularity of 0.1 as used in RAN2’s definitions seems appropriate to allow assessment of low load scenarios. |
| Ericsson | The RAN2 definition of the granularity is of 0.1 and we propose to follow the same granularity. This allows first of all to use the same algorithm that would compute the measurement according to 38.314. Also, this helps understanding the situation of number of active UEs in a cell in a much better way. As the averaging period is seconds, it is very likely that the average number of UEs is not an integer number. Rounding the number of active UEs to the closest integer brings a considerable error in understanding the status of the cell. |
| Nokia | Granularity of 1 is more than good enough for load balancing and access control, including in low-load scenarios. We should not mix what is needed for statistics and analysis, with what is needed for distributed SON purpose. |
| Samsung | Agree with DT. |

## 3.2 TNL capacity

Main question is whether to include load TNL load information for both backhaul (S1-U/NG-U) and fronthaul links (F1-U) reported separately over X2 and Xn. It has been argued that the deployment architecture (split vs. non-split) should not be reflected over X2 and Xn interfaces.

**Question: Can a reasonable way forward be to add TNL Capacity Indicator per node over X2/Xn (reflecting backhaul load), while cell load reflects fronthaul load in case of split architecture (3319)?**

|  |  |
| --- | --- |
| Company | Comment |
| ZTE | Separate report backhaul and fronthaul helps load evaluation for split architecture. |
| HW | We think a value per cell is sufficient. The neighbour just need to know the limiting factor for TNL. Adding a value per node brings no benefit. |
| Deutsche Telekom | Both ways of reporting, i.e., per node or per cell level, are feasible for X2/Xn. The more important issue is to describe the meaning of the corresponding IE(s) as well as of the maximum offered TNL capacity in a proper way to avoid misinterpretation in a multi-vendor environment (see the examples given in R3-203235 and R3-203319). |
| Ericsson | At RAN3-107bis-e the following was agreed:  **On X2 and Xn, TNL capacity indicator to be included per cell.**  **On F1, TNL capacity indicator to be included per node.**  With that it was already determined that per cell TNL capacity would be reported over X2/Xn, while over F1 a single TNL capacity value is reported for the whole gNB-DU. We believe that the relevant capacity to report over X2/Xn is a single per cell value, representing the maximum capacity accessible over the transport network for a given cell. |
| CATT | Yes. We think it reasonable. |
| Nokia | For X2/Xn, the TNL load metric includes UL/DL TNL Offered Capacity. If I understand well the proposal to signal the "the limiting factor for TNL", we may further need to clarify what is signalled in terms of offered capacity if the node backhaul has significant load and represents the limiting factor. We also need to clarify what "limiting factor" means, because the backhaul may be heavily loaded while still having much available capacity with respect to the bandwidth of a single cell. However even in that situation a source node should in our view be informed of the backhaul load situation of the potential target node. |
| Samsung | Ok for the way forward. |

## 3.3 HW capacity (which interfaces, granularity)

No particular issue or FFS was detected by the moderator. Please comment if any change is needed.

|  |  |
| --- | --- |
| Company | Comment |
| Ericsson | The only comment would be regarding the HW capacity for TS38.463, where there is an FFS on whether such capacity is reported per slice. We believe that HW capacity should not be per slice as it is not considered viable to slice such capacity. |
|  |  |
|  |  |

## 3.4 RRC connections (range, which interfaces)

It is proposed to remove this metric from X2 (3676, 3900).

**Question: Can the RRC connections metric be removed from X2AP?**

|  |  |
| --- | --- |
| Company | Comment |
| ZTE | Fine to remove RRC connections from X2AP. |
| HW | Fine to remove |
| CMCC | Yes, to remove. |
| Deutsche Telekom | Ok to remove it from X2AP. |
| CATT | Yes. It can be removed. |
| Nokia | Fine to remove. |
| Samsung | Fine to remove. |

## 3.5 SUL (whether to support, parameters)

Reporting of radio load status for SUL is proposed by one company (3676). Postpone to Rel-17 is also proposed (3900).

**Question: Can reporting of SUL related parameters be postponed to Rel-17?**

|  |  |
| --- | --- |
| Company | Comment |
| HW | We think this would be beneficial. Even if the decision to use SUL is up to the target, the source can perform a better selection of candidates if he is informed hat e.g. SUL is overloaded.  If agreed, this should be added to measurements reflecting radio resources, i.e. PRB and CAC |
| CMCC | We’re supportive to introduce SUL related parameters, and we can try to converge. Since this is the last meeting for R16, it is also ok for us to postpone such discussion to R17. |
| Deutsche Telekom | Similar view as HW. If no consensus can be achieved in this meeting, consideration in Rel-17 should be envisioned. |
| Ericsson | In the offline discussion at RAN3#106, captured in R3-197581, the following was concluded:  ***Question 6: Is there any missing part in CAC (except for beam and slice granularity discussed in separate sections)?***  *[Status based on papers]*  *SUL support is proposed in [8] and [9].*  *Editor’s note: it seems already supported because the receiver can identify whether the cell is SUL or UL/DL cell by served cell info;*  *[Offline discussion]*  Therefore we already discussed the possibility of explicitly expressing SUL parameters and we concluded that SUL capacity/load can be represented as part of the current available capacity or utilisation. There is no benefit in knowing what is the exact capacity/load per SUL carrier because it is not known at the source cell whether the Ue will be served on SUL at the target. In fact, basing a load balancing decision on SUL available capacity for example could be erroneous if in the end the UE is served on NUL at the target.  For this reason we do not consider this to be a topic to be agreed in Rel16 or to be carried to Rel17. |
| CATT | We think it is reasonable to introduce SUL related parameters. However, considering it is the last meeting for Rel-16,we are OK to postpone to Rel-17. |
| Nokia | The editor's note cited above may not provide a full view of the situation, but we agree with Ericsson's conclusion that a source node should not take load balancing decision based on an assumption of how the UE will be served in the target cell (on which carrier). |
| Samsung | We are fine to continue the discussion in Rel-17. |
| LGE | It’s all right to discuss this issue in Rel-17. |

## 3.6 Reporting per slice (whether to support, parameters)

Current status: Available slice capacity is agreed on Xn and F1. TNL and HW capacity are FFS on E1. Other parameters (e.g. PRB metric, DRB usage) per slice are proposed in 3874, 3319. Benefits are seen, but also proposal to postpone to Rel-17 (3900).

**Question: Can a reasonable compromise be to agree TNL load on E1 in Rel-16, and postpone further reporting per slice to Rel-17?**

|  |  |
| --- | --- |
| Company | Comment |
| ZTE | Postpone all Slice reporting to Rel-17 seems not appropriate. It is because Rel-16 SON need provide elementary metric(s) for implementation to work. For NW slicing, it is ambiguity for load evaluation in case of multiple vendor scenario.  Even worse, it is also impossible to evaluate NW slicing load in uplink direction from downlink direction in each NW slicing even for single vendor.  Therefore, it is practical to support at least PRB per DL/UL as metric for NW slicing scenario. |
| HW | We already have CAC reporting per slice. This part does not need to be enhanced. CAC indicate available capacity, which makes it suitable for per slice reporting – it is possible to signal available resources for this slice, with similar principles as we did before (e.g. where network sharing was considered).  For TNL, this is also indicating available resources so at least it could possibly be considered straightforward to add. PRB usage and HW usage is a bit more difficult, since this indicates the number of used resources and is difficult to relate to the policy assigned to the node (usage is not reported relative to an available resource). Considering that this is the last meeting, we propose that we wait with adding this. |
| CMCC | Generally, we are supportive of slice related metrics, but considering it is related to SA5 NRM specification and we didn’t dig into the details in the past meetings. So we could have a try to converge, if not achieved, we could postpone this to Rel-17. |
| Deutsche Telekom | At least slice-specific PRB usage should be generally considered in Rel-16 (related to SA5’s approach w.r.t. RRM Policy in TS 28.541), whereas slice-specific TNL load incl. E1 is seen as a metric that could be postponed to Rel-17 (a more general discussion on slice-specific TNL load is required).  No slice-specific HW capacity info needed.  Note that current slice available capacity definition needs some further clarifications (see R3-203235 or R3-203874). |
| Ericsson | When the discussion on per slice CAC was taken one of the proposals on the table was to add PRB utilisation per slice. The decision was to capture per slice CAC and not to capture PRB utilisation because it is difficult to determine a PRB utilisation with respect to a maximum number of PRBs available to the slice. We believe that the CAC per slice is sufficient to derive the capacity available in a slice and therefore trigger load balancing actions.  Proposals to report per DRB resource utilisation usage are unclear and we do not see the benefit of them.  For E1 TNL capacity indication, it was not described how can traffic of a given network slice be allocated to TNL resources dedicated to this slice. How is this traffic identified as belonging to a given slice over the transport network? Due to the lack of clarity on how to technically achieve that, we believe we cannot agree now to these proposals.  For E1 HW load per slice we do not think this is needed because we do not see as viable to slice HW capacity. |
| Nokia | We agree that E1 HW load is not needed reported per slice. We see that Deutsche Telekom and ZTE are in favour of PRB usage reporting per slice in Rel-16, and would like to join that view. |
| Samsung | It is related to SA5 NRM specification. We should be very clear about that before moving forward. CAC reporting per slice provides available capacity which is suitable for load balancing per slice. Therefore, we see no problem to postpone this to Rel-17. |

## 3.7 Misc. corrections and alignment

The following miscellaneous corrections were found in submitted papers. Please flag below any proposals with which you disagree or which would require further discussion. Please also notify any missing miscellaneous corrections.

**From 3318:**

This is E1AP TP with outcome from RAN3#107bis-e, intended integrated in TP from the present meeting.

**From 3235:**

Proposal 4: Change the integer range *TNL Capacity Indicator* IE and *Number of RRC Connections* IE to starting from 0.

Proposal 5: Add procedure text for RESOURCE STATUS RESPONSE message in XnAP and X2AP.

Proposal 6: Correct the text description and the presence for the *Cell To Report List* IE in F1AP, XnAP and X2AP. (the presence of the *Cell To Report List* IE should be conditional to the *Registration Request* IE is set to "start")

Proposal 7: Correct the editorial mistake in the presence for *gNB-CU-UP Measurement ID* IE in RESOURCE STATUS REQUEST message in E1AP.

Proposal 8: Align the tabular text for RESOURCE STATUS RESPONSE and RESOURCE STATUS FAILURE to the procedure text in XnAP, X2AP and E1AP.

Proposal 9: Delete “partial stop” as a condition in the presence of NG-RAN node2 Measurement ID IE in RESOURCE STATUS RESPONSE message in XnAP.

Proposal 10: Correct the editorial mistake about Reporting Periodicity IE in F1AP and XnAP, and add text description of the Reporting Periodicity IE in X2AP and E1.

Proposal 11: Correct the figure of Unsuccessful Operation of Resource Status Reporting Initiation Procedure in E1AP.

**From 3676:**

Proposal 3: The Number of RRC Connections IE is proposed to be changed from “INTEGER (1..65536,...)” to “INTEGER (0..65535,...)”.

Proposal 4: It is proposed to clarify in the semantics description that the “RRC Connections” counts only the UEs configured with the current cell as the PCell.

Nokia: This metric is agreed removed from X2 but will remain on the Xn interface. In order to represent any interest for load balancing when reported by a node that may be SN for some or all of its served UEs, we believe that the metric should report the number of UEs configured with the reported cell as PCell or PSCell.

Proposal 8: The presence of Cell Measurement Result within the RESOURCE STATUS UPDATE message should be fixed to optional.

Proposal 9: Align the ASN.1 in the BL CR for XnAP with the tabular that the presence of the SSB To Report List IE and the Slice To Report List IE within the Cell To Report Item are optional.

**From 3900:**

Proposal 6: Add Cause value for Parameter adapting in TS38.423, reuse LTE as baseline.

# 4 Conclusion, Recommendations [if needed]

If needed

# 5 References

[1] R3-20xxxx, Title, Company