**3GPP TSG-RAN WG2 #128 *R2-24XXXXX***

**Orlando, Florida, USA Nov. 18th – 22th, 2024**

Agenda Item: 8.3.2.1

Source: OPPO(rapporteur)

Title: Draft Summary of [POST127][029][AI mobility] RLF Simulation Assumption (OPPO)

Document for: Discussion, Decision

# Introduction

This draft intends to kick off following email discussion:

* [POST127][029][AI mobility] RLF Simulation Assumption (Oppo)

 Intended outcome: Agreeable simulation assumption

 Deadline: extra long

Although RAN2 concluded on definition and metric aspect for RLF, but rapporteur believes some further clarification on definition, methodology and metrics of RLF prediction are necessary in order to discuss simulation assumptions thoroughly. So the email will cover these aspects apart from simulation assumption for RLF prediction.

# Discussion

## Definition clarification

Since RAN2 agreed @127 meeting that SINR is the measurement quantity, so it is clear that L1 SINR results of RLF detection reference signal are the input of the model for both direct and indirect RLF prediction. As agreed, the output of the model is different between indirect and direct RLF prediction.

For indirect RLF prediction, the output of the model is L1 SINR of RLF detection RS in the future. And then RLF event is derived. In order to predict future L1 SINR, it makes sense to follow methodology as temporal domain case A.

The assumed T310 is 1000ms in 36.839, which is proposed by [1][2][3]. If prediction window is equal or longer than T310, then only predicted L1 SINR will be used for derivation. But if prediction window is shorter than T310, then actual measured L1 SINR will be also involved in the assessment procedure. Based on this analysis and RAN2’s agreement @ RAN2#127 meeting (please refer to Annex), here is recommended definition of indirect RLF:

**Indirect RLF prediction: The L1 SINR results are predicted based on historical actual L1 SINR results of the serving cell by following intra-frequency temporal domain case A and then RLF event at one time instance is determined based on predicted L1 SINR and optional historical actual L1 SINR results without further AI/ML models.**

**Q1: Can you confirm the further clarification of the indirect RLF prediction?**

|  |  |  |
| --- | --- | --- |
| Company | Opinion: Yes or No | Comments |
|  |  |  |
|  |  |  |

For direct RLF prediction, to make it clear here is an example to illustrate it:



Figure 2.1-1 interpretation 1 of direct RLF prediction

In Figure 2.1-1 t0 is current time instance. The direct prediction model predicts that an RLF event may occur during a window between t1 and t2 with x% possibility based on actual L1 SINR results in OW. Note t2 is always bigger than t1 because otherwise the window becomes a time instance.

This interpretation means the model also need conclude that during time window between t0 and t1, there is no any possibility that RLF event will occur. It sounds difficult for model to do so considering any output will be not 100% sure i.e. with a possibility. So, another interpretation is that there is no gap between t0 and t1 i.e. the window is prediction window itself as illustrated in Figure 2.1-2.



Figure 2.1-2 interpretation 2 of direct RLF prediction

**Direct RLF prediction: A RLF event within a time window is predicted with possibility x% directly, where 0<x<=100, based on historical actual L1 SINR results of serving cell in observation window. As for the time window:**

**Interpretation 1: the start of the window could be after current time instance i.e. t1>=t0**

**Interpretation 2: the start of the window is current time instance i.e. t0=t1**

**Q2: Any view on the recommended direct RLF prediction definition, especially on the two interpretations of time window?**

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

## Metrics clarification

Apart from SINR, 5 agreed metrics for indirect RLF prediction seems a bit messy for clean comparison. Actually, some of them are redundant with each other. The classic metric for model targeting classification e.g. RLF event, is F1 score. The key concept of F1 score is 3 counters in following table:

|  |  |  |
| --- | --- | --- |
| true event\predicted event | Negative | Positive |
| Negative | n0 | n1 |
| Positive | n2 | n3 |

Table 2.2-1 counter for F1 score

Counter n3 counts a case where a predicted RLF event really occurs i.e. it is “true RLF prediction”. Because a RLF event can be predicted indirectly at one time instance instead of within a window, it will be very difficult for the predicted RLF occurs exactly at the same time instance as the corresponding true RLF. So, some time gap between predicted RLF event and true RLF event should be tolerated. The tolerated time difference is actually agreed as “time difference of true RLF and predicted RLF”. Let’s call it as RVTD(RLF event timing difference). Since predicted RLF event could occur either before or after the corresponding true RLF event, RVTD should be an absolute value i.e. >0. Counter n3 will count only when RVTD is less than or equal to a predefined threshold.



Figure 2.2-1

For indirect RLF prediction the counter 3 i.e. “true RLF prediction” could be defined as:

**Counter n3: it increases by 1 when a true RLF event occurs around a predicted RLF event with RVTD, whose range is [0, maximum RVTD]**

Similarly counter n1 can refer to “false RLF detection” and counter n2 can refer to “missed RLF detection”. Based on the definition of counter n3, counter n1 and n2 could be defined as:

**Counter n1: it increases by 1 when no true RLF event occurs around a predicted RLF event with RVTD, whose range is [0, maximum RVTD]**

**Counter n2: it increases by 1 when no RLF event is predicted around a true RLF event with RVTD, whose range is [0, maximum RVTD]**

To reflect how many RLF events are falsely detected, the false RLF detection rate could be defined:

False RLF detection rate = 1-n3/(n1+n3), where n3/(n1+n3) is called “Precision” usually

To reflect how much RLF event is missed, the missed RLF detection rate could be defined:

Missed RLF detection rate = 1-n3(n2+n3), where n3/(n2+n3) is called “Recall” usually

While the classical F1 score could be then defined as:

F1 score = 2\*Precision\*Recall/(Precision + Recall)

**Observation 1: The agreed metric F1 score can already cover the other agreed metrics i.e., true RLF prediction, missed RLF detection, false RLF detection, time difference of true time RLF and predicted RLF**

For direct RLF prediction, the definition of counter n3’(corresponding to counter n3 of indirect RLF prediction) should be different because of the different output of the model from indirect RLF prediction. One proposal from [4] is to check whether a true RLF falls within the time window of the predicted RLF event. By combining RAN2’s agreement w.r.t. to possibility threshold, we can have following definition:

**Counter n3’: it increases by 1 when a true RLF event occurs within the time window of predicted RLF event whose possibility is higher than a predefined threshold**

Based on this definition, counter n1’ and n2’ can be defined as:

**Counter n1’: it increases by 1 when no true RLF event occurs within the time window of predicted RLF event whose possibility is higher than a predefined threshold**

**Counter n2’: it increases by 1 when a true RLF event occurs, but it doesn’t fall in the time window of any predicted RLF event whose possibility is higher than a predefined threshold**

It is clear that the counter definitions are different between direct and indirect RLF prediction due to the different nature of their definition. But the formula of the rest metrics on top of counters are the same. By taking the comprehensive metric i.e., F1 score, the performance of model with direct or indirect RLF prediction is comparable.

**Q3: Do you agree above definitions of true RLF prediction, false RLF detection and missed RLF detection for direct (i.e. counter n3’,n1’ and n2’) and indirect RLF prediction (i.e. counter n3,n1 and n2) as well as RVTD?**

|  |  |  |
| --- | --- | --- |
| Company | Opinion: Yes or No | Comments |
|  |  |  |
|  |  |  |

**Q4: Do you agree only F1 score based on the counters in Q3 is used as metric for evaluating both direct and indirect RLF prediction?**

*Note for indirect RLF prediction, SINR is another metric.*

|  |  |  |
| --- | --- | --- |
| Company | Opinion: Yes or No | Comments |
|  |  |  |
|  |  |  |

## Simulation assumption

In RAN2#126 meeting RAN2 agreed:

3 FR2 study will be prioritized for RLF prediction

Since RLF detection only occur within serving PCell. So, the scenario addressed for RLF use case is RLF within a FR2 PCell

**Q5: Can you confirm the scenario for RLF use case is RLF within a FR2 PCell?**

|  |  |  |
| --- | --- | --- |
| Company | Opinion: Yes or No | Comments |
|  |  |  |
|  |  |  |

In RAN2#127 it was confirmed that simulation assumption for FR2 RRM measurement prediction is reused as baseline. During the post email discussion [POST127][030] [AI mobility] RRM simulation assumptions, it is agreed that the TX and RX beam number is {8,16,32} and {4} respectively. It seems no reason not aligning these parameters between RRM measurement and RLF prediction.

**Q6: Do you agree updated TX and RX beam number for RRM measurement prediction (FR2) is also applicable for RLF prediction?**

|  |  |  |
| --- | --- | --- |
| Company | Opinion: Yes or No | Comments |
|  |  |  |
|  |  |  |

Apart from baseline simulation assumptions and additional ones mentioned in **Q6**, more parameters are proposed at RAN2@127meeting. Among the contributions, there are 5 kinds of simulation parameters:

Case 1: parameters related to channel model e.g. channel blockage [1] [8]

Case 2: parameters related to RLF procedure [1] [5] [8] [3][2]

Case 3: parameters related to HO procedure [3]

Case 4: parameters related to UE’s characteristics e.g. speed, trajectory etc. [6]

Case 5: Interference related parameters [1]

Parameters in case 3 can be saved because RAN2 agreed to simulate RLF without HO procedure. Parameters in case 4 can be also saved because company can report UE’s characteristics parameters, which is already in the template of the spreadsheet after [POST127][027][AI Mob] Simulation table (MediaTek). The intention for parameters in case 1 is to trigger RLF event easier. But now RAN2 agreed that UE will not change serving cell after it is dropped, triggering RLF event is not difficult any more. Considering company would like to reuse existing data set as much as possible, no change of such parameter for channel model is preferred.

**Observation 2: parameters in case 2, case 3 and case 4 are not necessary.**

For parameters in case 2, some companies [5][7][9] propose parameters to enable easier RLF triggering e.g. a short T310 timer. Again, since no HO procedure will be simulated, such change is also not necessary. And other company [1][2][3] believe typical RLF parameters are sufficient, which could be based on configuration in 36.839:

|  |  |
| --- | --- |
| Parameter | Value |
| Qin threshold | -6db |
| Qout threshold | -8db |
| Sample rate (TIndication\_interval) | 20ms  |
| Qin evaluation period | 100ms |
| Qout evaluation period | 200ms |
| T310 | 1000ms |
| N310 | 1 |
| N311 | 1 |

Table 2.3-1 RLF parameters

Although those parameters are from a simulation assumption for LTE system, it is rapporteur’s understanding that they are also applicable for 5G NR system apart from Sample rate. When DRX is not used, TIndication\_interval is max(10ms, TRLM-RS,M) according section 8.1.6 of 38.133:



If taking TSSB as example, sample rate could be 10,20,40,80 ms. In [POST127][030][AI mobility] RRM simulation assumptions (OPPO), we agreed to change sampling period of FR2 from 20ms to 80ms. If 10ms is used for sample rate of RLF prediction on FR2, likely we can’t reuse existing data set as much as possible. Because of this, rapporteur recommend to use 20ms.

**Q7: Do you agree the RLF parameters in table 2.3-1?**

|  |  |  |
| --- | --- | --- |
| Company | Opinion: Yes or No | Comments |
|  |  |  |
|  |  |  |

The difficult part is how to simulate interference. RAN2 agreed not to simulate traffic but also keep some FFS i.e. “full buffer and assumption that all the cells are fully loaded. We will not simulate traffic”. Before we dive deeper into the discussion, let’s first look at the definitions of those concepts.

Full buffer model is a kind of traffic model that is widely used by 3GPP. The model assumes that UEs always have data to transmit or receive, and their buffers are always full. This reflects a scenario where the system is operating at full capacity, and it is used to study the system's performance under saturated conditions. Compared to other traffic model (e.g., FTP and VoIP) which need to stimulate the stochasticity of data arrival and storage, full buffer model is much simpler. It can be seen as a simple assumption applied and does not need additional work for simulation.

“All the cells are full loaded” refers to a case that all cells are using the same wireless resources (i.e., resource block) for DL transmission. A UE will receive interference from all neighbour cells. In this way, interference is more common and RLF will be easier to trigger.

If we adopt that “full buffer and assumption that all the cells are fully loaded”, we do not need to model a scheduler to model the wireless resource allocation procedure, which is rather complicated. The simulation workload can be saved.

Based on the above analysis, it seems we can directly remove the FFS in the agreements.

**Q8: Do you agree to use the full buffer assumption and the scenario that all the cells are fully loaded for interference modelling and no resource scheduler is needed?**

|  |  |  |
| --- | --- | --- |
| Company | Opinion: Yes or No | Comments |
|  |  |  |
|  |  |  |



Figure 2.3-1: ways to generate interference

From figure 2.3-1(a), we can see that a UE in cell 1 get more interference than a UE within other cells if only those 7 cells are considered. It is a kind of artificial boundary effect, which is not representative of how radio signals behave in practice. Therefore, in simulation, we often use wrap-around to smooth transition of signal strength across the entire simulation area. As we consider 2-tier deployment, the majority of interference comes from the surrounding six cells. An example of wrap-around would be that UEs in cell 6 receive interference from cells 1, 5, 7, 4a, 2b, and 3b, as shown in figure 2.3-1 (b). There is also another way to generate interference as shown in figure 2.3-1(c), where a 3-tier scenario is modelled to get the interference surrounding cell 6. It has the same effect as wrap-around and also use the surrounding six sites for interference analysis. Besides the adjacent 6 sites, interference can also be seen from two co-site cells that are not the serving cell (as we consider 3 sectors/cells per site).

**Q9: Do you agree that interference in simulation comes from co-site cells and surrounding 6 sites of serving cell, i.e., interference comes from 20 cells?**

|  |  |  |
| --- | --- | --- |
| Company | Opinion: Yes or No | Comments |
|  |  |  |
|  |  |  |

We have agreed in RAN2#126 that FR2 study will be prioritized for RLF prediction. It means beams must be considered. From 38.213, RLM using the associated SS/PBCH block when the associated SS/PBCH block index is provided by RadioLinkMonitoringRS. That means RLM-RS can come from beam(s) that are not serving. Given that different companies may use different number of DL beams and different ways to index beams, even if we give a fixed beam pattern for RLM, the SINR may deviate much among companies. Therefore, a simple way is to use the serving beam for interference analysis.

**Q10: Which option would you prefer to determine the serving signal of RLM:**

* Option 1: a fixed beam pattern that may differ from the serving beam, e.g., beams indexed with {0,2,4,6}.
* Option 2: serving beam

If you prefer option 1, please also provide the preferred number of beams (RLM-RS) and the way to choose them. If you have any other options, please list them in the comments.

|  |  |  |
| --- | --- | --- |
| Company | Opinion: 1 or 2 | Comments |
|  |  |  |
|  |  |  |

If answer to Q8 is yes i.e. no scheduler is modelled, we also need to discuss how to determine the interference from each cell. From the previous discussion, there are three ways on the table.

* Option 1: the interference comes from the beam with maximum RSRP of neighbor cells
* Option 2: the interference comes from fixed beam(s) of neighbor cells.
* Option 3: the interference comes from randomly chosen beam(s) of neighbor cells.

Option 1 is easier to generate a small SINR but does not strictly follow the current specification. While option 2 aligns more with existing protocols, the modeling of it faces the same issue as we discussed in **Q9**. It would be hard to compare companies’ results as we may use different number of DL beams and indexing method. Option 3 can model the randomness of interference but can lead to possibly large fluctuation of SINRs, making AI hard to learn and predict.

**Q11: Which option would you prefer to determine interference signals of neighboring cells:**

* Option 1: the interference comes from the beam with maximum RSRP of neighbor cells
* Option 2: the interference comes from fixed beam pattern of neighbor cells.
* Option 3: the interference comes from randomly chosen beam(s) of neighbor cells.

If you prefer option 2 or 3, please provide the preferred number of beams of each cell used for interference. If you pick option 2, please also provide ways to choose the beams. If you have any other options, please list them in the comments.

|  |  |  |
| --- | --- | --- |
| Company | Opinion: 1, 2, or 3 | Comments |
|  |  |  |
|  |  |  |

**Q12: Do you believe there are any other important factors we should consider for simulating interference? If so, please provide them below.**

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

# Conclusion

# Reference

[1] R2-2406424 Discussion on simulation assumption for RLF prediction ZTE

[2] R2-2406976 Discussion on Simulation Assumption and Methodology for RLF prediction CMCC

[3] R2-2407492 Discussion on simulation assumptions and evaluation methodology for RLF prediction Samsung

[4] R2-2406705 Discussion on RLF prediction assumptions Xiaomi

[5] R2-2407093 AI-ML based RLF predictions Ericsson

[6] R2-2407211 R19 NR AIML A8341\_RLF\_prediction\_simulation\_assumptions Interdigital

[7] R2-2407481 RLF Prediction Aspects Nokia

[8] R2-2407389 Simulation assumption for RLF prediction KDDI

[9] R2-2406312 Discussion on RLF use case OPPO

# Annex: RAN2 agreements in #127 meeting

Table 5-1

**Agreements**

- Both direct and indirect are allowed. Companies should indicate what they used and what inputs they are using

- Output for indirect: predicted SINR. Based on predicted SINR the time instance the RLF occurs can be determined without further AI/ML models.

- Output for direct: probability of RLF within an window

- Companies should report the prediction time window they have used in the simulations

**Agreements on KPI**

For indirect

- SINR difference, missed RLF detection, false RLF detection, F1 score, time difference of true time RLF and predicted RLF, true RLF prediction.

Direct

- missed RLF detection, false RLF detection, F1 score.

- Whether the RLF will happen based on the following methodology. Above a probability threshold we assume that RLF will happen. This is then compared with true RLF.

- For the time being we don’t need HO procedure simulation in RLF simulation

- FFS full buffer and assumption that all the cells are fully loaded. We will not simulate traffic.

- Simulation results are not expected before February