**3GPP TSG-RAN WG2 Meeting #128 R2-24xxxxx**

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

Agenda Item: 8.3.5

Source: OPPO

Title: Draft summary of [AT128][017][AI mob]Simulation assumptions (OPPO)

Document for: Discussion, Decision

# Introduction

The draft covers following offline discussion:

1, Generalization issue i.e. the simulation combination of UE speed

2, Open simulation assumptions for measurement event and RLF prediction and SLS.

# Discussion

## Generalization issue

During online discussion RAN2 agreed:

|  |
| --- |
| **Agreements**   1. Reuse the evaluation methodology in TR38.843 for generalization study, i.e., the generalization performance is evaluated with the following cases,  * *Baseline:* The AI/ML model is trained using the dataset with Configuration #B and tested using the dataset with Configuration #B. * *Generalization Case #1 (GC#1):* The AI/ML model is trained using the dataset with Configuration #A but tested using the dataset with Configuration #B. * *Generalization Case #2 (GC#2):* The AI/ML model is trained using mixed datasets with both configurations and tested using the dataset with Configuration #B.   2 Companies can choose which case they compare with and should report it with simulation results.  3 Generalization issues on RRM measurement prediction are prioritized.  4 Start the study with generalization issue with RRM measurement prediction in temporal domain. Companies can chose to study frequency domain prediction cases and report what they have simulated. |

* Study generalization over UE speeds
* The simulation assumption of FR1 temporal domain case B is reused for generalization study with 3 UE speeds i.e. 30Km/h, 60Km/h and 90Km/h. FFS on combinations
* The simulation assumption of FR2 temporal domain case A is reused for generalization study with 3 UE speeds i.e. 60Km/h, 90Km/h and 120Km/h. FFS on combinations

In order to understand better, the following table list all the simulation combinations for 1 baseline UE speed i.e. 30km/h for FR1 temporal domain case B:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Training @Dataset: 30km/h | Training @Dataset: 60km/h | Training @Dataset: 90km/h | Inference @30km/h | Inference @60km/h | Inference @90km/h |
| Baseline | Yes |  |  | Yes |  |  |
| GC#1 | Yes |  |  |  | Yes |  |
| GC#1 | Yes |  |  |  |  | Yes |
| GC#2 | Yes | Yes | Yes | Yes |  |  |
| GC#2 | Yes | Yes | Yes |  | Yes |  |
| GC#2 | Yes | Yes | Yes |  |  | Yes |

Table 2.1-1 simulation combinations

If the UE speed in baseline is changed to e.g. 60km/h, then more training and more inference are needed. Here is the statistics:

|  |  |  |
| --- | --- | --- |
| Number of Baseline UE speed | Number of models to be trained | Number of inferences |
| 1 | 2 | 6 |
| 2 | 3 | 9 |
| 3 | 4 | 12 |

**Question 1: Do you share rapporteur’s understanding of the simulation combination?**

|  |  |  |
| --- | --- | --- |
| Company | Yes/no | Comments |
| Ericsson | Yes |  |
| Huawei | Yes | But if we change the baseline to 60 km/h, then the number of combinations does not increase. It only increases if we add more speeds for baseline. |
| vivo | Yes |  |
| Samsung | Yes |  |
| Qualcomm | Mostly agree. Please see comments | We were wondering, for Baseline, whether we can consider the Training dataset to be from 60kmph. The errors upon generalization can be lower compared to the case when the Baseline is 30kmph, since 60kmph is intermediate between 30kmph and 90kmph. |

If company think there is too much simulation combinations, one potential way is to set less UE speed e.g. only one UE speed as baseline.

**Question 2: Do you agree one potential way to limit the simulation combination is to limit the number of UE speed as baseline? If yes, please indicate whether 1 UE speed is sufficient. If no, please provide your opinion in detail.**

|  |  |  |
| --- | --- | --- |
| Company | Yes/no | Comments |
| Ericsson | Yes | We could consider the lowest and the highest speed as input and for inference. |
| Huawei | Yes | Our preference is to focus on GC#2 as it corresponds better to real life scenario. |
| vivo | Yes |  |
| Samsung | Yes | 1 UE speed (30km/h) is enough.  Also agree with HW that we can focus on GC#2 considering the real situation. |
| Qualcomm | Yes, 1 UE speed is sufficient as Baseline |  |

## Measurement event prediction (part 1)

After post email discussion [1], here is part of proposal 9 by removing direct prediction specific parameters. In addition, in current TR the set of UE speed is 60/90/120 km/h for 2nd study goal, we need respect that and hence 30km/h is removed from 3rd column. The main discussion point is baseline value in 2nd column. Note the removed parameters related to direct prediction will be discussed as part 2 in section 2.5.

**Proposal 1: To agree the baseline value for the listed parameters for intra-frequency temporal domain case A and open for more values for some of the parameters as indicated in the table below:**

|  |  |  |
| --- | --- | --- |
| Parameters | baseline value | Note |
| A3 event offset (db) | 2 |  |
| TTT (ms) | 320 | Open for one shorter value |
| UE speed (km/h) | 90 | Open for ~~30 ,~~ 60 and 120km/h |
| OW length (ms) | N/A | Up to implementation |
| PW length (ms) | 400 | Open for more values |
| Max ETD (ms, note1) | 80 | Open for more values |
| ~~Event occurrence Window Length (ms, note 2)~~ | ~~N/A~~ | ~~Up to conclusion under question 2~~ |
| ~~Probability threshold (%, note 2)~~ | ~~80%~~ | ~~Open for more values~~ |

*Note1: parameters for indirect prediction*

Table 2.2-1 Parameters for measurement event prediction of temporal domain case A

**Question 2: Do you agree with the content listed in the table 2.2-1?**

|  |  |  |
| --- | --- | --- |
| Company | Yes/No? | Comments |
| Ericsson | Yes |  |
| Huawei | Yes |  |
| Samsung | Yes |  |
| Qualcomm | Mostly agree. Please see comments | **A3 event offset (dB):** 3dB (for higher RRM RSRP MAE values) |

For temporal domain case B, here is proposal 10 in [1]:

**Proposal 3: To agree the baseline value for the listed parameter for intra-frequency temporal domain case B and open for more values for some of the parameters as indicated in the table below:**

|  |  |  |
| --- | --- | --- |
| Parameters | baseline value | Note |
| A3 event offset (db) | 2 |  |
| TTT (ms) | 320 | Open for one shorter value |
| UE speed (km/h) | 30 | Open for 60 and 90km/h |
| OW length (ms) | N/A | Up to implementation |
| PW length (ms) | N/A | Up to implementation |
| Max ETD (ms, note1) | 40 | Open for more values |
| MRRT | 50% | Open for more values |

Table 2.2-2 Parameters for measurement event prediction of temporal domain case B

In table 2.2-2, one more parameter i.e. MRRT is added in the same table by combining proposal 11 as below:

Proposal 11: For intra-frequency temporal domain case B, RAN2 is invited to discuss whether MRRT=50% could be baseline value

**Question 4: Do you agree with the content listed in the table 2.1-2?**

|  |  |  |
| --- | --- | --- |
| Company | Yes/No? | Comments |
| Ericsson | Yes |  |
| Huawei | Yes, but… | For PW we also could agree on one baseline value (e.g. 400 ms) and allow for more. |
| vivo | Yes with comments | Add another factor on the skipping pattern, i.e., example 2 as baseline and open for example 1. |
| Samsung | Yes with comment | PW length is up to MRRT. We understand that 50% MRRT means PW length (= 1 SSB period = 40msec for FR1). |
| Qualcomm | Mostly agree. Please see comments | **A3 event offset (dB):** 3dB (for higher RRM RSRP MAE values) |

Another issue for temporal domain case B is the filtering options for input L3 RSRP of the model. There are 3 options listed in the proposal 12 as below:

**Proposal 12: For intra-frequency temporal domain case B company can report following filtering options for input L3 RSRP measurement in sub-use case 2:**

**Filtering option 1: L3 filtering is based on its L1 filtered result and the immediate last skipped measurement result;**

**Filtering option 2: L3 filtering is based on its L1 filtered result i.e. no L3 filtering if the immediate last result is skipped;**

**Filtering option 3: L3 filtering is based on the L1 filtered result and last actual measurement result i.e. the skipped result(s) in between is ignored.**

**For indirect prediction, the skipped result refers to predicted L3 RSRP measurement result previously by the RRM measurement prediction model**

**For direct prediction, the skipped result refers to skipped L1 measurement result**

It could be difficult to reach consensus about the options. But it will be helpful to understand preference from company so that RAN2 can know which option is preferred by the majority company.

**Question 5: Which option are your favourite option?**

|  |  |  |
| --- | --- | --- |
| Company | Option 1/2/3 | Comments |
| Ericsson | Option 1 |  |
| Huawei | Option 3 | We have agreed to use L3 measurements as the output, so option 2 is not preferred. Option 1 will result in error propagation, so this leaves us with option 3. |
| vivo | Option 1 | Option 1 may cause prediction error accumulation in sub-use case 2, but OK for sub-use case2 1 and 3 as the input is L1 RSRP. |
| Samsung | Option 3 | For option 1, the output from the previous prediction would be used as part of input for the following prediction. I.e., the predicted value at T2 is used to calculate the L3 filtered value at T3 that is input data for the prediction at T4. In this case, the prediction error would be propagated/accumulated over time.  The option 2 means no L3 filtering in practice. |
| Qualcomm | It should be left up to companies to choose |  |

## RLF prediction (part 1)

Here is the parameter for RLF prediction by combining proposal 18 and proposal 23 in [1] by removing direct prediction specific parameters. Note the removed parameters related to direct prediction will be discussed as part 2 in section 2.5. In addition, for indirect prediction OW length and PW length should be also assumed.

**Proposal 18: To agree on following parameter for RLF prediction:**

|  |  |
| --- | --- |
| Parameter | Value |
| Qin threshold | -6db |
| Qout threshold | -8db |
| Sample rate (TIndication\_interval) | 20ms(FR2)/40ms(FR1) |
| Qin evaluation period | 100ms |
| Qout evaluation period | 200ms |
| T310 | 1000ms |
| N310 | 1 |
| N311 | 1 |
| Max ETD (ms, note1) | 20ms(FR2)/40ms(FR1) |
| ~~Event occurrence Window Length (ms, note 2)~~ | ~~Under discussion in question 13~~ |
| ~~Probability threshold (%, note 2)~~ | ~~80%~~ |
| The number of beams for FR1 fixed beam pattern | 1 |
| The number of beams for FR2 fixed beam pattern | ~~4~~ |
| PW length (ms) | 400 |
| OW length (ms) | Up to implementation |

Table 2.3-1 Parameters for RLF prediction

**Question 6: Do you agree with the content listed in the table 2.3-1?**

|  |  |  |
| --- | --- | --- |
| Company | Yes/No? | Comments |
| KDDI | No. See comments | We suggest to consider blockage in a channel model to ensure realistic performance evaluation of RLF prediction. We see some companies want to consider blockage. Considering the workload, option1 can be chosen.  **Proposal: RAN2 to choose from the following options in order to collect more RLF event.**  **Option1. Blockage in a channel model is allowed to consider and how to model the blockage can be up to companies.**  **Option2. Parameters related to blockage in a channel model are added below.**  Considering the workload, Model A can be selected (clause 7.6.4 of TR 38.901).  For Blockage model A of RLF prediction, unify required parameters of Blockage model A, or determine the rules for describing parameters.   * + - The number of blocker *K* (the standard number is *K* = 4).     - Scenarios for deciding non-self-blocking regions parameters given in Table 7.6.4.1-2 of TS 38.901.     - Scenarios for deciding spatial correlation distance given in Table 7.6.4.1-4 of TS 38.901.     - Whether the blocker has movement speed *v.*     - Are there any other missing configurable parameters? |
| Ericsson | Partly | We think ETD can be 80 ms, the prediction needs to be very precise otherwise. |
| Huawei | Yes | We could agree on one ETD value, but it might be worth checking more of them (up to companies) as it will give us an idea of what performance we get for each “accuracy”. Blockage will make things complicated and we already agreed not to do HO, so we can force some RLFs anyway. |
| vivo | Question on the PW length | The current T310 is 1000ms and the PW length is 400ms. But we think the PW length should be equal/ larger than T310. |
| Samsung | Question on the number of fixed beam pattern in FR2. | We have a question about the number of fixed beam in FR2.  First, we would like to clarify that SINR measurement is based on SSB (not CSI-RS) in our simulation. Then, from our understanding, when the UE measures the SINR of a certain SSB from the serving cell, other neighbor cells also transmit SSB with one beam. Thus, we think the number of beams for fixed beam pattern for neighbor cell should be also 1 in FR2. Can’t understand how neighboring cell can transmit SSB with four different beams simultaneously. |
| Qualcomm | Ok, except for the number of beams for FR2 fixed beam pattern | It is not clear what it means to have a fixed number of beams for FR2 fixed beam pattern. |

## System level simulation (part 1)

The issue left for SLS is the handover model for both temporal domain case A and case B. To facilitate the discussion, this section focuses on case that measurement event is predicted indirectly so that the predicted time instance i.e., t1 in the context is clear.

For temporal domain case B, if the last measurement results to derive the measurement event is actual measurement, then there is the time to report the measurement result i.e. t0 and the time of event occurrence i.e. t1 is the same, then there is no ambiguity about handover model because legacy handover model can be adopted. But if the last measurement result is predicted one, then t1>t0. In this case there are two options:

 

Figure 2.4-1 Option 1 for case B Figure 2.4-2 Option 2 for case B

**Option 1:** UE report measurement report @t0 as illustrated in Figure 2.3-1 when it is predicted. Network start handover preparation once the measurement report is received. Since t1>t0, handover command could be received @ max(HO prep, t1-t0)

**Option 2:** UE delay to t1 to report measurement result as if the predicted measurement event occurs there. Network starts the preparation after receiving the measurement report. And handover command will be received after handover preparation.

|  |  |  |
| --- | --- | --- |
| Case B | Pro | Con |
| Option 1 | (partial) handover preparation time can be saved so that HO CMD can be received early | Mixed the study between case A and case B |
| Option 2 | The evaluation is purely for 1st study goal i.e. decouple from case A | the benefit of the model is wasted |

Table 2.4-1 analysis between 2 options

**Question 7: Which option of handover model for temporal domain case B do you prefer?**

|  |  |  |
| --- | --- | --- |
| Company | Option 1/2 | Comments |
| Ericsson | Out of these options, option 1 | No benefits of predictions with option 2. We could also leave it to company implementation. |
| Huawei | Option 1 | There is no need to wait for the actual measurement in case the prediction already indicates that the event will be met. |
| vivo | Option 1 with modification | The handover is sent at max(t0+HO Pre, t1-TTT) |
| Samsung | Option 2 | Now we think that there can be two features related to temporal domain prediction.  1. Report of predicted future event is to improve HO KPI by far future prediction  2. Measurement skipping at UE-side is to reduce UE’s measurement overhead by near future prediction  In Case B scenario, we understand the first and second feature can be used together. However, we don’t think both features should be always used/configured together. We don’t need to combine these two features for different goals as one feature.  Considering that Case B is for the 1st goal, only the second feature for measurement skipping can be assumed in Case B to evaluate the impact of measurement skipping on HO KPI correctly. |
| Qualcomm | Please see comments | Our understanding of the options is that in Option 1, the UE sends a report when it predicts that the measurement event occurs without waiting to send at the time when the event occurs, which is Option 2. In that case, our preference is Option 1, because otherwise the benefits of prediction are wasted. There should be some accuracy requirements that the predictions must meet so that the network can rely upon the prediction report and act, i.e., initiate HO preparation. |

As for handover model for temporal domain case A, there are two options in the summarized proposal 25:

Proposal 25: As for simulation base on temporal domain case A, RAN2 conclude one of the two options to decide exactly when to transmit handover command:

**Option 1:** if there is an actual measurement event occurring (@ t2) before the predicted measurement event (@t1), then network will transmit handover command based on actual measurement event ,or otherwise on predicted measurement event(@t1).

**Option 2:** network transmit handover command purely based on actual measurement event regardless whether an actual measurement result(@t2) is earlier or later than predicted measurement event((@t1))

Another option is proposed in [3] as following:

Proposal 1(**Option 3**): For AI mobility, HO preparation starts when an event is predicted to happen (i.e., t0), and HO command is sent when A3 entering conditions are met based on actual/real measurement and an event is predicted to be met for the duration of TTT. Using the timing from the figures: HO command is sent at t3, where t3=t0+max(HO prep time, t1-t0-TTT), provided that entering conditions of the event are met based on real/actual measurement at t3.

 

Figure 2.4-3 Figure 2.4-4



Figure 2.4-5 Figure 2.4-6

|  |  |  |  |
| --- | --- | --- | --- |
| Case | Option 1 | Option 2 | Option 3 |
| Case 1: Actual event is earlier than predicted event | Figure 2.3-4 @ t2 | Figure 2.3-4 @ t2 | Figure 2.3-6 @ t1-TTT |
| Case 2: actual event is later than predicted event | Figure 2.3-3 @ t1 | Figure 2.3-5 @ t2 | Figure 2.3-6 @ t1-TTT |

Table 2.4-2 when HO CMD is received

For all 3 options, if HO preparation can’t be completed before the time supposing to receive HO CMD, then HO CMD is delayed until HO preparation is finished.

**Question 8: Which option of handover model for temporal domain case A do you prefer?**

|  |  |  |
| --- | --- | --- |
| Company | Option 1/2/3 | Comments |
| Ericsson | Out of these options, option 1 | Regarding option 3, the formula should not contain TTT, since it is already taken into account when the event prediction is triggered (at t1). The event prediction is triggered when the cell quality is predicted to be higher than a certain threshold for the TTT time. So, the predicted event is triggered (at t1) TTT ms in advance. |
| Huawei | Option 3 | It seems there may be different interpretations of whether TTT is considered or not. Last meeting it was clarified online that event prediction time indicates the time after TTT has passed. Our point with option 3 is that we should analyze the maximum gains we can get from the predictions, i.e. saving HO preparation time and TTT (under the condition that the entering conditions of the event are met and the prediction indicates they will keep on being met until TTT expiry). |
| vivo | Option 3 | The table 2.4-2 is not aligned with the description for option3? The HO CMD is t3=t0+max(HO prep time, t1-t0-TTT) |
| Samsung | Option 2 | First, AI-based prediction should be additional feature that can be used on top of existing HO framework. Thus, HO based on actual measurement event should be baseline.  Also, we think HO command without actual MR (based only on prediction) is risky. If UE’s moving speed or direction change after the prediction, when NW make HO command at the predicted timing, UE would not be able to perform HO successfully.  Given that prediction accuracy can not be guaranteed in real field, we prefer to be conservative on HO decision and support option 2.  For option 3, we can avoid the risk of early HO a little bit more than option 1 by sending predicted MR after A3 entering condition is first met. However, in that case, we understand there should be two different prediction report from UE (One at t0 for early HO preparation and another one at t1-TTT for early HO command), which seems inefficient and has not been discussed until now. |
| Qualcomm | Please see comments | If network receives an event prediction report, the network need not wait for an actual measurement report before transmitting the HO command to the UE, assuming that the event prediction report is transmitted by the UE if accuracy requirements are met, as mentioned in our response to Q7. |

## Direct prediction

### Common issue

The common issue for both measurement event prediction and RLF prediction is how to interpret the time window where a measurement event/RLF can be predicted. Rapporteur put two interpretations on the table for temporal domain case A:



Figure 2.5.1-1 Interpretation 1 of time window



Figure 2.5.1-2 Interpretation 2 of time window

In the summary of post email [1] some company propose there could be multiple time windows in the prediction window. Rapporteur think it is not aligned with the agreement in RAN2#127 bis meeting as following:

For direct measurement event prediction, the model output is the probability of event occurrence within a time window



Figure 2.5.1-3 intermediate time windows

But it seems gives a way to achieve interpretation 1 by the model. It can be assumed that model can predict several time windows in future, among which there are different probability. Then there must be a time window with highest probability and hence this time window should be chosen if the probabilty is higher than preconfigured threshold. However, if a model can only predict one time window in future, then interpretation 2 is feasible.

**Question 9: Which interpretation of time window for direct prediction do you prefer?**

|  |  |  |
| --- | --- | --- |
| Company | Interpretation 1/2 | Comments |
| Ericsson | Interpretation 1 |  |
| Huawei | 1 | As commented previously in the e-mail discussion, with interpretation 2 the window where the event can actually happen is too wide and it is not clear how this can be useful. Direct prediction should be on par with indirect method where the event is supposed to be predicted correctly in case it falls within a short time window from the real event occurrence which corresponds to interpretation 1. |
| vivo |  | An simple approach is the RLF prediction is performed upon T310 and the prediction result is whether the T310 will expires. |
| Qualcomm | Interpretation 2 |  |

Regardless which interpretation is chosen, we still need figure the length of the time window for both measurement event prediction and RLF prediction. To align with indirect prediction, it could be equal to the parameter max ETD for interpretation 1 and equal to length of PW for interpretation 2.

Here is recommended value for time window length and probability threshold:

|  |  |  |
| --- | --- | --- |
|  | Measurement event prediction | RLF prediction |
| Time window length (Interpretation 1) | 80ms | 20ms |
| Time window length (Interpretation 2) | 400ms | 400ms |
| Probability threshold | 80% | 80% |

Table 2.5.1-1

**Question 10: Do you agree with the parameters in table 2.5.1-1?**

|  |  |  |
| --- | --- | --- |
| Company | Yes/no | Comments |
| Ericsson | Partly | We think ETD can be the same for RLF prediction as for event prediction, i.e. 80 ms. |
| Huawei | Yes, but… | For interpretation 1 we also need to agree the distance between t0 and t1 which should be aligned with the PW length we agree for indirect prediction. |
| vivo |  | For interpretation 2, why the window is shorter than the T310?  For interpretation 1, the gap between t0 and t1 is also needed? |
| Samsung | no | Probability threshold can be up to company. Each company can use their own probability threshold to optimize the F1 score.  For fair comparison with indirect method, the length of time window used to determine True/False positive in direct approach should be equal to 2\*maxETD. |
| Qualcomm | Yes for Interpretation 1.  No for Interpretation 2. | Regarding Interpretation 2, we are confused since prediction window has been used only in the context of RRM measurement prediction, and hence for Indirect Event prediction. It is not clear what prediction window means for Direct Event prediction. |

### System level simulation (part 2)

For the handover model option 1 and option 3, it matters when the predicted measurement event occurs i.e., t1. For both interpretation 1 and 2. For direct prediction, a straight way is that t1 is in the middle of the time window. If option 2 is approved, then this issue is not valid any more.

**Question 11: Do you agree predicted measurement event occurs in the middle of the time window for the sake of handover modelling?**

|  |  |  |
| --- | --- | --- |
| Company | Yes/no | Comments |
| Ericsson | Yes |  |
| Huawei | Not sure | Do we really need to consider direct prediction for SLS? The modelling seems quite unclear. |
| vivo |  | Agree with HW to focus on the indirect prediction for SLS. |
| Samsung | no | Since it is unclear when the event is predicted to occur in Direct approach, we are not sure whether the option 1 & 3 can work with Direct approach. |

# Conclusion

# Reference

[1] Email discussion ([POST127bis][022][AI mobility] Simulation Assumption of measurement event/RLF prediction and SLS (OPPO)

[2] R2-2410800 Discussion on generalization aspects and simulation assumption ZTE Corporation

[3] R2-2410540 Discussion on simulation assumptions and generalization Huawei, HiSilicon discussion

[4] R2-2409867 Simulation assumptions on event/RLF/SLS and model generalization Xiaomi discussion