TRAN2#129bis

|  |  |  |
| --- | --- | --- |
|  | UE sided model | Network sided model |
| Measurement event prediction | Yes, | No (up to NW implementation) |
| Temporal domain case A | Yes, all sub cases | Yes |
| Temporal domain case B | Yes , all sub cases | Yes |
| Frequency domain | Yes , all sub cases | Yes |
| Spatial domain | Pending | Pending |
| L3 beam level measurement prediction | Yes, | Yes |

* L3 Beam-level prediction specification impact will be studied

**Agreements**

1. The general LCM framework for beam management can be the baseline (where applicable) for AI mobility, such as the following aspects:

* Data collection for model training
* UE capability
* Applicability reporting
* Inference configuration and reporting
* Performance monitoring and management

2 Only functionality-based LCM is considered for AI/ML mobility (i.e. we don’t support model based LCM)

* Start by discussing spec/ran2 impact of the agreed cases/subcases and then discuss functionality granularity
* The UE applies inference configuration for the applicable RRM measurement prediction and measurement event prediction upon receiving configuration via RRC. As baseline, no dynamic signaling (e.g. L1/2) is required. FFS when the UE can perform inference

Agreements

* For RRM prediction, UE sided model, the UE can be configured with periodic or event triggered reporting of predicted and/or actual RRM measurements. FFS details
* For event prediction, UE-sided model, the UE can be configured with event-triggered reporting based on prediction and/or actual measurements. FFS details
* Baseline is that this applies to all A1-6 events, unless technical problems are identified. Baseline quantity is RSRP.

**Agreements**

1. The inference configuration and reporting for AI/ML mobility can be based on RRM measurement framework
2. For AI/ML for mobility, UE can report the applicable functionalities to NW via UAI or RRCReconfigurationComplete message.
3. FFS whether association ID is required and should provide motivation on per use case bases
4. As a baseline, use the AI ML PHY NW-side data collection procedures, configuration and reporting framework.
5. Study UE side data collection configuration, taking AI ML PHY relevant procedure as baseline. Postpone UE side data collection transfer discussion until further progress is made in AI ML PHY in R20.
6. Model transfer will not be discussed in this study item

7 For UE sided model, the performance monitoring considered in AIML air interface can be studied as a baseline for AIML aided mobility, i.e. NW-side and UE-sided performance monitoring.

FFS metrics per use case

RAN2#130:

**Agreements**

1 The scenario of intra-frequency spatial domain prediction (in cell dimension) is not considered by the UE-side model in 5G. RAN2 assumes that a network-sided model requires no specification impacts. Intra-frequency spatial domain prediction (in cell dimension) involves measuring one or more cells as input to the model to derive L3 filtered cell-level measurements for the same time instance of another cell(s).

2 The specification impact of intra-frequency spatial domain prediction in beam dimension will be studied by RAN2 for both UE-side and network-side model.

**Agreements**

1. Use the beam management agreements as the baseline, including the following aspects (details can be updated after further progress is made in AI/ML PHY BM use case)

• Consider Option A and Option B like scheme (if/when AI/ML PHY makes further progress)

• Upon receiving a full inference configuration, the UE sends the initial applicability report in RRCReconfigurationComplete. UAI can be sent to update applicability.

• Upon receiving one or more full inference configuration(s) via RRCReconfiguration message, UE shall maintain all the full inference configuration(s) no matter the full inference configuration is applicable or inapplicable until the network releases it explicitly.

• Support the explicit reporting of applicability/inapplicability in the initial report and subsequent reporting when the applicability changes.

• Together with inapplicability reporting, UE further indicates a simple cause value of inapplicability (FFS pending AI/ML PHY progress).

• No prohibit timer is introduced.

1. Associated ID should be optionally configurable for training and inference (i.e. it may not mandatorily required for all training/inference configurations). FFS for WI phase further details (what absence means, which use case, whether it is per cell or multiple cells/frequency, terminology).

**Agreements**

1. The following potential “inference parameters” can be considered for inference/measurement configuration, which provides the necessary information for UE to examine the applicability, it may contain

• PW length (for temporal case A)

• Skipping pattern (e.g. SSB config that indicates SSBs transmitted) (optional). MRRT (for temporal case B ) not discussed in study item, but if RAN4 says otherwise can be considered in WI

• Measured and predicted beam pattern (optional). MRRS (for spatial domain prediction) not discussed in study item, but if RAN4 says otherwise can be considered in WI

• Measured and predicted frequency (For inter-freq domain prediction)

1. Model related choices (i.e., cluster-based vs. cell-based, L1 filtering, RRM sub-use cases, OW length, direct vs. indirect) can be up to UE implementation unless a requirement is identified to specify them.

**Agreements**

1. For network-side RRM measurement prediction, the legacy RRM measurement configuration and RRM measurement reporting framework can be used. For L1-filtered beam-level RSRP reporting can be configured by setting co-efficient to zero. FFS if there are specification impacts to support subcase 1 and 3 (if supported) and interference for cell level prediction.
2. To support cell/beam level prediction, one enhancements is to report RRM measurement results per cell at multiple time instances in one measurement report for NW-sided model.

**Agreements**

For the collection of data for the training of a network-sided model for RRM measurement prediction:

1. UE can be configured to log, at a certain logging periodicity, one or more of the following:

* L3 cell level measurements,
* L3 beam level measurements,
* L1-filtered beam level measurements (if sub-case 1 and 3 is supported)
* Cell ID (FFS CGI of serving cell. If CGI is unavailable, or for neighboring cells the UE logs PCI-ARFCN as a fallback)
* Time info (if as agreed by AI/ML and/or if needed)

1. The UE configuration is provided via the RRM measurement framework. Whether we reuse existing measconfig structure or we need a separate data logging configuration is for WI phase. Required enhancements are FFS.
2. The UE can be configured with a L3 event for determining when logging is to be performed. When the event conditions are fulfilled, it performs the logging with the logging periodicity.
3. The UE sends availability indication of collected logged data via UAI or RRCReconfigurationComplete message for HO case.
4. The availability indication can be triggered due to:

* full buffer being reached (if configured), (FFS if buffer is per use case)
* buffer threshold being reached (if configured), or (FFS if buffer is per use case)
* low power (if configured)

1. Upon sending the availability indication, UE indicates:

* Data is available
* Reason for the triggering of the availability indication (full buffer. Threshold)
* Low power indication

1. The UE sends the collected data upon explicit/on-demand request from the network (using UEInformationRequest/Response signaling)
2. The UE keeps the collected data upon HO, unless explicitly indicated to release it by the network (e.g., during HO).
3. The UE releases the collected data upon transitioning to IDLE/INACTIVE
4. For RLF case, capture in the TR that for mobility keeping the data during RLF can be beneficial. It can be up to WI phase if this is done depending on whether a simple solution can be found

For UE sided data collection – use AI/ML PHY request/configuration framework as baseline. FFS enhancements/or differences

* Capture in the TR that for mobility keeping the data during RLF can be beneficial. It can be up to WI phase if this is done depending on whether a simple solution can be found

**Agreements**

1. RAN2 confirms that UE will not be informed about any network-sided functionality management decision (e.g., selection, (de)activation, switching, fallback, etc.) for AI mobility use case
2. For performance monitoring of NW-side model, UE can provide the label data (i.e., actual measurement results) for gNB. The existing measurement/report configuration can be reused to configure UE to report the actual measurement
3. FFS on UE awareness and preference for NW sided model

**Agreements**

1 For UE-sided model, NW-sided monitoring, the performance monitor procedure contains

* NW sends monitoring configuration and inference configuration to UE.
* UE reports measurements and inference output based on the corresponding configurations to NW.
* NW performs monitoring and makes decisions, which can be fully NW-implemented.

2 For UE-sided model, UE-sided monitoring, the performance monitor procedure as baselined contains

* NW sends monitoring configuration to UE
* UE measures monitoring data and generates monitoring results by comparing inference output and monitoring data.
* (NW-decision): UE reports the monitoring result to NW, and NW makes the management decision.

1. Can further consider UE sided monitoring and (UE-decision) for some use cases: UE makes the management decision based on network configuration and sends the decision to NW. FFS which use cases.
2. For RRM measurement prediction
   1. RSRP differences can be used as the performance metric for monitoring.