3GPP TSG-RAN WG2 Meeting #126 R2-24xxxxx

Fukuoka, Japan, May 20th -24th, 2024

Agenda Item: 8.1.2.2

Source: Intel, Samsung

Title: Report of [POST126][032][AI/ML PHY] LCM (Intel/Samsung)

Document for: Discussion, Decision

# Introduction

This report provides a summary for the following post-meeting email discussion:

* [POST126][032][AI/ML PHY] LCM (Intel/Samsung)

Intended outcome:

Phase 1: Agreable definitions (Samsung)

Phase 2: Reach common understanding of reactive/proactive framework for applicable functionality.

Deadline: long

The deadline for providing comments for phase 1 is June 11, 2024 Tuesday at 21:00 UTC.

Companies providing input to this email discussion are requested to leave contact information below.

|  |  |  |
| --- | --- | --- |
| **Company** | **Name** | **Email Address** |
| Apple | Peng Cheng | Pcheng24@apple.com |
| vivo | Boubacar Kimba D.A. | kimba@vivo.com |
| OPPO | Jiangsheng Fan | fanjiangsheng@oppo.com |
| Xiaomi | Xing Yang | Yangxing1@xiaomi.com |
| ZTE | Fei Dong | Dong.fei@zte.com.cn |

# Phase 1: definition of functionalities

In RAN2 #126 meeting, RAN2 discussed the following definition for functionality types and decided to have more discussion to identify the need of such definitions and whether further update is needed to clarify the definition [1].

|  |
| --- |
| Proposal 2: RAN2 agree the following definition for functionality types as a starting point.  - *Supported/identified functionalities:* this refers to functionalities that UE can indicate by using UE capabilities.  - *Configured functionalities:* this refers to functionalities that gNB can configure UE for model inference and performing measurements for training purposes?. Depending on proactive/reactive approach, configured functionalities may or may not be applicable upon configuration.  - *Applicable functionalities:* this refers to functionalities that the UE is ready to apply for model inference. It can be considered as candidates for functionality activation.  - *Activated functionalities:* this refers to functionalities that the UE starts predicting beam results via model inference. |

In this discussion, it would be good to discuss each functionality type.

## Supported functionalities

|  |
| --- |
| *Supported functionalities:* this refers to functionalities that UE can indicate by using UE capabilities. |

The moderator think that we can simplify the name from supported/identified functionalities to supported functionalities for convenience of discussion. Please comment if it is not ok.

**Q1: Do you agree that supported functionalities refer to functionalities that UE can indicate by using UE capability signaling and gNB/LMF can configure?**

If it is NO (or partially Yes), please add preferred definition (or additional definition) for it.

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comment |
| Apple | Partial Yes | We prefer not to couple the “supported functionalities” with “configured functionalities” (i.e. the 2nd half sentence “and gNB/LMF can configure” can be removed):   1. As discussed in Q2/Q3, it is not clear whether the “gNB/LMF can configure” means “configure for training” and/or “configure for inference”. This may bring additional ambiguity. 2. Since UE capability is RAN2 expertise, we think RAN2 have common understand what “indicate by using UE capability signalling” means (i.e. UE supports the feature and NW can configure the feature). So, the 2nd half sentence is redundant.   Thus, we suggest below change:  **supported functionalities refer to functionalities that UE can indicate by using UE capability signaling ~~and gNB/LMF can configure~~** |
| vivo | Yes | The original one is OK and agree with Apple that “gNB/LMF can configure” is not needed as the gNB/LMF should configure based on applicability/availability rather than supported. |
| OPPO | Partial Yes | We understand supported functionalities means the functionalities indicated via UE capability signalling, i.e. those functionalities are already implemented and tested from product engineering perspective, but it does not necessarily mean that NW can configure UE only based on supported functionalities, because RAN2 already agreed the understanding that UE may or may not have the model available for the supported functionalities. If NW configures a UE supported functionality but UE does not have the model associated with this functionality when receiving the configuration, what will happen from UE side? Trigger RRC re-establishment or delay the functionality until making the model available? Neither way is desirable from system point of view.  We notice that this situation for AI functionality is different than legacy functionality, i.e. non-AI functionality, for non-AI functionality, once UE reports the supported functionality via UE capability signalling, NW can be sure to configure any UE supported functionality if NW supports the functionality also, but when it comes to AI functionality, the working logic and assumption is totally different as the model availability is the additional aspect NW has to consider before making any configuration, otherwise, there is a risk to trigger RRC re-establishment or delay the functionality until making the model available as mentioned above.  So we suggest the following simple definition for supported functionalities:  **Supported functionalities refer to functionalities indicated via UE capability signaling.** |
| Xiaomi | Yes | We understand supported functionalities are static, which would not change dynamically. We would support to keep the ‘gNB/LMF can configure’ part. We don’t see the use case for NW to configure a functionality which is not supported by UE. |
| ZTE | Partial Yes | We have some sympathies with Apple’s suggestion.  In RAN2#125bis meeting, the below agreements have been achieved for the supported functionality reported by UE capability:  ***Which AI/ML-enabled Features/FGs and functionalities are supported should be standardized. The details wait for RAN1’s progress. “supported” means that the UE is capable of supporting the functionality and doesn’t mean neccesarily that the UE has the model available. FFS what functionality refers to.***  In our understanding, if there is no model is available at UE side for the supported functionality reported in UE capability , there is no need for gNB/LMF to configure it. In this sense, we would like to remove the ‘gNB/LMF can configure’ since it implies that NW may configure the functionality without any available models to the UE which is a bad implementation. |
|  |  |  |

## Configured functionalities

|  |
| --- |
| *Configured functionalities:* this refers to functionalities that gNB/LMF configured to UE. UE for model inference and performing measurements for training purposes? Depending on proactive/reactive approach, configured functionalities may or may not be applicable upon configuration. |

As commented during RAN2 discussion, the moderator changed “gNB can configure” to “gNB configured”. And, LMF is added to cover positioning use case. Please comment if the change is not acceptable.

Please note that the second part (“Depending on proactive/reactive approach, configured functionalities may or may not be applicable upon configuration.”) will be discussed in Q5 in Section 2.3.

**Q2: Do you agree that configured functionalities refer to functionalities that gNB/LMF configured to UE?**

If it is NO (or partially Yes), please add preferred definition (or additional definition) for it.

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comment |
| Apple | No | We doubt whether RAN2 really need this definition:   1. In 3GPP, we only specify UE behaviour under NW configuration. Thus, the definition suggested by moderator is obvious and doesn’t bring useful information. 2. As Q3 discussed, the configured functionalities may be understood as inference configuration or training configuration (and even performance monitoring configuration, although it was not mentioned in online discussion). So, the definition suggested by moderator may bring further ambiguity or further clarification. 3. We think at least stage 3 specification (either RRC or MAC) doesn’t need this definition because what NW configured to UE is always clearly specified in RRC and MAC spec. Thus, whether it is configuration for training or inference or monitoring should be crystal clear from text before and after in stage 3 specification. 4. In our understanding, the key controversial issue (or AI/ML specific issue) is the boundary among the following 3 terms: “supported functionality”, “applicable functionality” and “activated functionality”. Thus, it seems sufficient to define these 3 terms.   Thus, we think RAN2 don’t need definition of “configured functionalities”. |
| vivo | See comment | From our understanding, the term “configured functionality” is introduced for NW to configure the functionality in advance and activate it when needed.  For BM, the necessity of the definition can be FFS, depending on whether the configuration in advance is needed.  For positioning use case, the definition is not needed. We already agreed to reuse the existing LPP Location Information Transfer procedure (*RequestLocationInformation*/ *ProvideLocationInformation* messages) for providing the results of the UE-sided model inference operation, thus no configuration in advance is needed and LMF can directly request to activate the functionality. |
| OPPO | partially Yes | The suggested definition on configured functionalities does not make much sense as it’s too obvious if considering this terminology alone.  What really matters is the boundary among supported functionalities, available functionalities, configured functionalities, applicable functionalities and activated functionalities，because the boundary will decide how AI functionalities works in the whole picture, we understand the following relationship should be the baseline for our discussion:    For instance, UE supports 100 functionalities from UE capability signaling point of view, but currently 40 out of 100 functionalities, UE has the corresponding models available due to memory limitation; Then based on NW additional condition and UE inside additional condition, gNB knows 20 out of 40 functionalities is applicable based on UE reporting; then gNB configures 10 out of 20 functionalities to UE side and activate 5 out of 10 functionalities for this UE.  Based on above, we suggest the following definition for configured functionalities:  **configured functionalities refer to functionalities that gNB/LMF configured to UE, all configured functionalities shall be appliable and ready for activation from NW and UE perspective.** |
| Xiaomi | Yes with Comment | We understand configured functionalities can be used to trigger reactive applicable functionality report. However, if proactive applicable functionality report is used, NW can directly activate applicable functionality reported by UE. Configured functionality seems to be unnecessary in this case. We suggest to further clarify configured functionality is only used in reactive applicable functionality report. |
| ZTE | No | In our understanding, the configured functionalities is simply referring to the functionalities those have been configured to the UE. There is no need to over interpret this term since we have a lot of similar thing ( for example, configured TCI state, configured SCG, configured…., we should not always make a clear definition for those terms since we already have common understanding on what is configured). |

**Q3: do you agree that this can be used for both UE-side model inference and training purpose?**

If it is NO, please indicate your preference e.g. only model inference for now or any suggestion for RAN2 progress.

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comment |
| Apple |  | See our comment to Q2. |
| vivo | No | The discussion of data collection for UE-sided model training is still ongoing and controversial. The motivation to involve model training in configured functionality is not clear.  Suggest focusing on the model inference for now and model training can be revisited when needed. |
| OPPO | No | Only for UE-side model inference for now as inference is usually our focus for configuration. We can further clarify training part when it’s clear enough for inference part. |
| Xiaomi | Only inference for now | For training, we are not sure whether functionality is needed. It’s possible the data collection is achieved by legacy measurement frame work, which is not related to functionality explicitly. |
| ZTE | No | we do not see the configured functionality can be used for model training purpose. For example, the data collection for model training need both set A and set B information for the spatial beam management, but the configured functionality only can provide the set B information, how UE can use set B only information to perform the model training.  So the configured functionality is just for inference, if UE really think only set B information can be used for model training, it is up to UE implementation and no need to capture in the definition of the term. |
|  |  |  |

## Applicable functionalities

|  |
| --- |
| *Applicable functionalities:* this refers to functionalities that the UE is ready to apply for model inference. It can be considered as candidates for functionality activation. |

It is also noted that RAN2 made a following agreement regarding applicable functionalities.

|  |
| --- |
| The UE will indicate the gNB/LMF whether the AI/ML functionality is available/applicable. For a functionality to be applicable at least there should at least one model available within it. FFS other details on what is applicability/non-applicability. |

**Q4: Do you agree that applicable functionalities refer to functionalities that the UE is ready to apply for model inference and they can be considered as candidates for functionality activation/deactivation?**

If it is NO (or partially Yes), please add preferred definition (or additional definition) for it.

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comment |
| Apple | Yes with comments | We understand that “the UE is ready to apply for model inference” means the UE:   1. It has successfully completed model training compliant with UE’s internal status. And the trained model is already in UE device. 2. The dataset which was used for training the model is consistent with current inference configuration.    * Note that if it is not consistent, the UE can’t be regarded as “ready to apply” because current gNB is not aligned with UE and so can’t activate it.     Although we understand these are details which will be discussed in phase 2, it is better to confirm current definition can cover them. |
| vivo | Yes, with comments | The applicable functionalities are candidates for functionality configuration as well. In addition, deactivation is not needed as the deactivation may happen when the functionalities get inapplicable. Thus, suggest refining as:  **can be considered as candidates for functionality configuration (if needed) /activation.** |
| OPPO | partially Yes | Based on our picture raised in Q2, we understand applicable functionalities means these functionalities are ready to be configured and activated from both UE and NW point of view, but current definition only focuses on UE side, we think it’s not complete, so we have the following definition suggestion for applicable functionalities :  **applicable functionalities refer to functionalities that NW is ready to configure and the UE is ready to apply for model inference, and they can be considered as candidates for functionality activation/deactivation from both UE and NW point of view.** |
| Xiaomi | Yes | We agree with the principle proposed by rapp. How to determine the applicability can be up to UE. Because UE vendor may consider different UE implementations. |
| ZTE | Yes to have a definition, No for the current definition | We agree to have a definition of the applicable functionality. But the current definition is too complex to understand, we would like to make it more straightforward:  **Applicable functionalities refer to the functionalities those have available models and can be considered by UE and NW to be applicable for activation at the time beings.** |

In relation to configured functionalities, configured functionalities may or may not be applicable immediately upon configuration depending on proactive/reactive approach. For example, in one way, UE already provides applicable functionalities/applicability related information and gNB can configure applicable functionalities. The other way would be that UE provide applicable functionalities/applicability related information after receiving configured functionalities from gNB and hence, configured functionalities may not be applicable immediately upon configured functionalities. We can remove “depending on proactive/reactive approach” in the definition as it will be discussed further in Phase 2 and there is no need to add as a part of definition.

**Q5: do you agree that configured functionalities may or may not be applicable immediately upon configuration?**

If it is NO (or partially Yes), please comment what relationship you envision between configured functionalities and applicable functionalities.

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comment |
| Apple |  | See our comment to Q2. We think RAN2 only need to define “**supported functionalities**”, “**applicable functionalities**” and “**activated functionalities**”, i.e. no need to define “**configured functionalities**”. |
| vivo | No | NW should configure AI functionalities based on the applicability/availability indication from UE. Otherwise, the configuration may be useless if the functionality is always unavailable. |
| OPPO | No | We think the following scenario is invalid and should be deprioritised:   * The other way would be that UE provide applicable functionalities/applicability related information after receiving configured functionalities from gNB and hence, configured functionalities may not be applicable immediately upon configured functionalities.   Why a NW would like to blindly configured potential non-applicable functionalities to UE before knowing which functionality is applicable from UE point of view. It’s more like a try-and-error logic, which should be avoided from 3GPP system perspective. More addition, what will happen from UE side when a functionality is not applicable when configured by NW? Trigger RRC re-establishment or delay the functionality until making the model available or something else? Neither way is desirable from system point of view.  Based on our picture raised in Q2, we understand configured functionalities shall be applicable from both UE and NW perspective.  The following scenario should be the baseline for our discussion:   * in one way, UE already provides applicable functionalities/applicability related information and gNB can configure applicable functionalities. |
| Xiaomi | Yes | We agree with rapp configured functionalities can be un-applicable when it’s used to trigger reactive applicable functionality report. |
| ZTE | No | We do not want to couple the configured functionality and applicable functionality. To me, the configured functionality is dependent on the RRC configuration, the applicable functionality depends on whether the functionality is applicable to the current UE and NW additional conditions. The former one is related to a time period of RRC configurations, the last one is related to a time period of scenario change. These are two different things, we should not couple them. |

## Activated functionalities

|  |
| --- |
| *Activated functionalities:* this refers to functionalities that the UE starts predicting beam results via model inference. |

**Q6: Do you agree that activated functionalities refer to the functionalities that the UE uses beam prediction/positioning via model inference?**

If it is NO (or partially Yes), please add preferred definition (or additional definition) for it.

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comment |
| Apple | partially Yes | We think “**UE uses beam prediction/positioning”** is not necessary, especially considering CSI compression/prediction is still on the table of Rel-19. Since it is a high level definition, we prefer to make it more general, e.g.  **activated functionalities refer to the functionalities that the UE uses ~~beam prediction/positioning via~~ output of model inference** |
| vivo | Yes | Can be refined as  *Activated functionalities:* **this refers to functionalities that have been enabled for ~~the UE starts predicting beam results via~~ model inference.** |
| OPPO | partially Yes | The similar view as Apple, let’s make it more generic for now.  **activated functionalities refer to the functionalities that the UE is using ~~beam prediction/positioning via~~ the output from model inference** |
| Xiaomi | Yes |  |
| ZTE | Yes | We can make it more general like apple, vivo, oppo’s suggestion. |

## Available functionalities

During RAN2 discussion, there is a proposal on availability/available functionalities [2].

|  |
| --- |
| *Proposal 2 Introduce signalling for the UE to inform the gNB whether the AI/ML functionality is available for operation (e.g., whether there are trained models available within it). FFS whether the “availability indication” can be reported as part of the applicability-reporting information, or as a separate signalling.* |

**Q7: Do you agree that available functionalities should be separately considered from applicable functionalities?**

If yes, please comment what separate characteristics should be added/expected compared to applicable functionalities.

If no, please suggest how to consider available functionalities in RAN2 discussion (e.g. consider same as applicable functionalities or postpone to the next meeting)

|  |  |  |
| --- | --- | --- |
| Company | Yes/No | Comment |
| Apple | No (same as applicable functionalities) | Please note that above proposal was agreed as different wording after online discussion:  3 The UE will indicate the gNB/LMF whether the AI/ML functionality is available/applicable. For a functionality to be applicable at least there should at least one model available within it. FFS other details on what is applicability/non-applicability.  So, our understanding is that above agreement (esp. the highlighted part) has implied “applicable functionality” is same as “available functionality”.  Meanwhile, we also understand that the highlighted part implies the definition of “applicable functionalities suggested by moderator:  **applicable functionalities refer to functionalities that the UE is ready to apply for model inference and they can be considered as candidates for functionality activation/deactivation** |
| vivo | No | The available functionality is the same as the applicable functionality.  The intention to have separate definitions is not clear. |
| OPPO | Yes | Based on our picture raised in Q2, we understand available functionalities are the super set of applicable functionalities, not all available functionalities are applicable based on NW additional condition and UE inside additional condition, so we can have the following simple definition for available functionalities:   * **available functionalities refer to functionalities that the UE has the corresponding model(s).** |
| Xiaomi | No | The definition and usage of available functionality are not clear. Maybe we can further discuss in next meeting based on contributions if needed. |
| ZTE | Yes | We agree to have this definition for the future discussion, in our understanding, for one functionality reported as supported at UE side via UE capability, the main concern from NW is whether there is any available models for such functionality, whether the functionality is applicable to be activated at the time being. In this sense, only available functionality and applicable functionality does the matter for NW. So we support to have a clarification of the definition in terms of available functionality, the following is a suggestion on top of OPPO’s version:   * **available functionalities refer to functionalities that the UE has the model(s) to perform the inference.** |

# Phase 2

# Conclusion

# Reference

[1] R2-2405180 Functionality-based LCM for UE sided model Samsung

[2] R2-2405266 LCM for UE-side models for beam management Ericsson