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

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

Agenda Item: 8.1.4

Source: Mediatek Inc.

Title: Report of [POST125bis][020][AI/ML PHY] UE side data collection

Document for: Discussion, Decision

# Introduction

This contribution is aimed at reporting the discussion and results of the following post email discussion:

* [POST125bis][020][AI/ML PHY] UE side data collection (Mediatek)

 Intended outcome: Discuss new table capturing solution details and discussion fon control and visibility, privacy.

 Deadline: two weeks (Deadline May 3rd, 10:00 UTC)

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

|  |  |  |
| --- | --- | --- |
| **Company** | **Name** | **Email Address** |
| NEC | Xuelong Wang | xuelong.wang@emea.nec.com |
| Apple | Peng Cheng | Pcheng24@apple.com |
| BT | Salva Diaz | salva.diazsendra@bt.com |
| Nokia | Jerediah Fevold | jerediah.fevold@nokia.com |

# 2 Discussion

The following proposals were identified during the Rel-18 SI on AI/ML for NR Air Interface:

|  |
| --- |
| 1. UE collects and directly transfers training data to the Over-The-Top (OTT) server;

1a) OTT (3GPP transparent)1b) OTT (non-3GPP transparent)1. UE collects training data and transfers it to Core Network. Core Network transfers the training data to the OTT server.
2. UE collects training data and transfers it to OAM. OAM transfers the needed data to the OTT server.

RAN2 did not study or analyse these proposals and did not agree to requirements or recommendations. |
| 1a)  | 1b) |
| 2. Data collection via CN | 3. Data collection via OAM |

According to the consensus made during the online discussion in RAN2#125bis meeting,

=> need to better define what is control of data collection in MNO and visibility of data content in MNO.

=> understanding is that OTT is outside of MNO

We will clarify these methods in the following aspects:

* Inside/outside MNO’s network
* Termination Entity
* Controllability of data collection in MNO
* Visibility of data content in MNO
* Protocol layer for data transfer
* Privacy concerns

## 2.1 Terminology

According to the common understanding made in RAN2 that OTT is outside of MNO, it is inappropriate to refer to the term OTT in the context of solutions 1b, 2, and 3. To avoid confusion and enhance clarity, we propose replacing the term ' OTT server' with 'server for UE-side data collection' in these solutions. The term "UE-side data collection" refers to the data collection intended for use in UE-side model training. Please note that the term 'training' is omitted from the terminology to prevent confusion or debate regarding whether the locations for data storage and model training are identical or separate.

**Q1:** **Do companies agree to replace the term ‘OTT server’ with ‘server for UE-side data collection’ in the definitions/descriptions of different solutions?**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | Yes |
| Apple | Yes for solution 1b/2/3 with comments:As this is new terminology created by RAN2, we are still confused with whether it is: 1. a server owned by UE vendor, or
2. a server owned by MNO, or
3. a neutral server

Since **the privacy requirements for the 3 possible options are different**, we think it needs separate discussion. One may argue it is out of RAN2 scope, but please note that RAN1 has agreed and discussed extensively on different model transfer solutions (y and z1-z5) depending on whether training is UE-sided, NW-sided or neutral site. Thus, similarly, we think it is one important aspect RAN2 should discuss. Meanwhile, it is also part of “solution details” which are explicitly included as scope of this email discussion.  |
| Nokia | Yes, but maybe to be even clearer, we could use the term “server for training data collection for UE-side models”.  |
|  |  |

## 2.2 Inside/outside MNO’s network

One point that requires clarification is the location of the server for UE-side data collection relative to the MNO’s network. In solution 1a, the UE directly communicates with the server for UE-side model collection, typically via the application layer, without any interaction with the MNO's network infrastructure. Consequently, in this scenario, the server for UE-side data collection operates outside of the MNO's network, functioning as an OTT server. However, for solutions 1b, 2, and 3, the server for UE-side data collection has the flexibility to be positioned either inside or outside the MNO's network. It is noted that in solution 2, the server for UE-side data collection can be a NF in CN.

|  |  |
| --- | --- |
| **Solutions** | **Inside/outside MNO’s network** |
| **1a** | OTT server, i.e., outside MNO’s network |
| **1b** | Inside/outside MNO’s network |
| **2** | Inside/outside MNO’s networkNote: The server for UE-side data collection can be a NF in CN. |
| **3** | Inside/outside MNO’s network |

**Q2.1:** **Do companies agree that for solution 1a the server for UE-side data collection is outside of MNO’s network and is therefore classified as an OTT server?**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | Yes |
| Apple | Yes, as captured in Chair Note. |
| BT | Yes but for us, solution 1a is totally outside 3GPP scope. Consequently, it should be taken away from the WI phase. No need to discuss option 1a. |
| Nokia | Yes. We would also like to know if the following figure, an excerpt from TS26.531, Figure 4.2-1 is the intended representation of solution 1a.Additionally, the following note is provided in the same specification: “NOTE 5: Interactions at reference point R8 are beyond the scope of 3GPP standardisation.” |

**Q2.2:** **Do companies agree that for solutions 1b, 2 and 3, the server for UE-side data collection can be either inside or outside MNO’s network?**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | Solution 2, Yes. But, My understanding is that solution 1b) should still be sub-option of the solution 1 group (i.e. UE directly transfers data to server), so the difference from the solution 1a) is the server is “inside” MNO’s network and solution 1b is that the server is “outside” MNO’s network. We propose to preclude “outside” from 1b). In addition, for solution-3, it is a bit unclear if the UE-side data collection should be inside or outside MNO’s network.  |
| Apple | No for solution 1b, Yes for solution 2 with comments, not sure for solution 3.On solution 1b, we are not sure how the server can be inside MNO’s network. As we commented in Q1, we think two solutions can be: * a) UE vendor can deploy its owned server in MNO network or
	+ In this case, which Network entity UE vendor can deploy its server (e.g. in CN or RAN)?
* b) MNO deploy some servers (i.e. owned by MNO) which are rent to UE vendor for data collection?

Note that in case of understanding a), it may have 3GPP impact and cross-WG impact. Thus, RAN2 should make it clear. Before this issue is concluded, we think “inside” should be precluded from solution 1b). For solution 2, we agree it can be inside or outside MNO network. Meanwhile, we appreciate that Rapporteur provides example of “a NF”, but we are not whether it means “a NF” performs model training on behalf of UE vendor or it is just an intermedium entity for further forwarding dataset? Please clarify. For solution 3, we agree with NEC it is not clear. We noticed Rapporteur doesn’t provide example. We are wondering whether anyone can give a concrete example that the final UE server is within MNO network (e.g., if it is an entity in RAN, is it within DU or CU)?  |
| BT | No.From our understanding, the server for UE-side collected data on options 1a, 2 and 3 are inside the MNO and the ownership is for the MNO.A different discussion is where the algorithms for training, management and inference are allocated but we consider this is not asked here. |
| Jerediah [Nokia] | Maybe, but we first need to clarify what “outside” the network means, and whether that server could have, e.g., a VPN or N3IWF interface to the NW to communicate with NFs that are involved in DC. And it should be clarified whether there is any other difference in the interface between the server for UE-side data collection whether it is located inside or outside of the network. |

## 2.3 Termination Entity

The "(First)termination entity" refers to the entity that receives and stores data transmitted from the UE, **which possesses the authority to oversee the subsequent handling of this data**, such as data cleaning, forwarding, sharing, and analysis, among others, in compliance with privacy policies, security protocols, and any regulatory compliance requirements. The purpose of introducing the term ‘(first) termination entity’ is to emphasize the entity’s role in management and possession of the collected data.

According to the responsibility of the termination entity as define above, the termination entity for solution 1a) is the OTT server. The termination entity for solution 1b is the server for UE-side data collection. The termination entity for solution 2 is the CN. The termination entity for solution 3 is the OAM.

**Q3.1: Do companies agree that for solutions 1a the termination entity of UE-side data collection is the OTT server?**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | Yes |
| Apple | Yes |
| BT | Yes but as it was mention before, no need to discuss 1a  |
| Nokia | Yes. |

**Q3.2: Do companies agree that for solutions 1b the termination entity is the server for UE-side data collection?**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | Yes  |
| Apple | Yes |
| BT | Q2.2 indicates “for solutions 1b, 2 and 3, the server for UE-side data collection can be either inside or outside MNO’s network”. We do not consider “outside” the MNO as an option for that reason, we propose to clarify the sentence:Do companies agree that for solution 1a, the termination entity of UE-side data collection is inside the MNO?Using our proposed sentence, the answer is yes. |
| Nokia | Yes. We would also like to know if the following figure, an excerpt from TS26.531, Figure 4.2-1 is the intended representation of solution 1b, wherein the first termination entity is actually in the CN. Otherwise, we would like to discuss what is the difference between what is depicted in the figure vs. solution 1b as depicted in the diagram (shown in the introduction of Section 2, top-right) for solution 1b provided in this Email Discussion.  |

**Q3.3: Do companies agree that for solutions 2 the termination entity of UE-side data collection is the CN?**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | Yes |
| Apple | No. According to the nice figure provided by Rapporteur, only the first termination point is CN in Solution 2 (i.e. final termination entity is still “server for UE-side data collection”). Thus, we suggest below rewording: **for solutions 2 the first termination entity of UE-side data collection is the CN** |
| BT | For clarification, we propose the following sentenceDo companies agree that for solutions 2 the termination entity of UE-side data collection is inside the CN?Yes. That was the proposal done on [R2-2403492](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2403492.zip) |
| Nokia | Yes, however it is unclear which CN entity is the termination entity. We would also clarify that the termination in the CN is not the UPF. |

**Q3.4: Do companies agree that for solutions 3 the termination entity of UE-side data collection is the OAM?**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | Yes |
| Apple | No. Similar to Q3.3, we suggest below rewording:**for solutions 3 the first termination entity of UE-side data collection is the OAM** |
| BT | Similar to Q3.3, we propose:Do companies agree that for solutions 3 the termination entity of UE-side data collection is inside the OAM?Yes. The UE-side data collected is still under MNO control. |
| Nokia | Yes. |

## 2.4 Controllability for transfer of the collected data in MNO

To begin, we focus on the aspect of data transfer to address controllability.  Once these aspects are clearly understood, we can delve into more intricate details, such as how the MNO might configure measurement and reporting for the UE, among other considerations.

We define the controllability of data transfer for UE-side data collection within an MNO in the following dimensions:

* The capability of the MNO to control the data transfer to the server for UE-side data collection.
* The specific entity within the MNO to control the data transfer.
* The protocols or methods utilized by the MNO to control the data transfer.
* The MNO's ability to manage (e.g., allow/disallow, initiate/terminate, prioritize/de-prioritize, etc.) the data transfer.

Companies are invited to discussion on whether it is appropriate to start discussions on the controllability of data transfer, as outlined in the preceding four bullet points. Additionally, companies are requested to provide any definitions of controllability that may encompass critical aspects not already covered.

**Q4.1** **Do companies agree to start controllability discussion on data transfer as defined with the above four bullets? Please provide/clarify the definition on controllability if there are any important aspects missing.**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | Yes, but it would be important to clarify which entity within the MNO that should be responsible for the control of the data transfer for each solution if any.  |
| Apple | Yes with comments:We understand the 1st dimension is covered by the 4th one. So, suggest to remove 1st dimension.  |
| BT | Current proposals are ambiguous as they may have different understandings. We propose:* The capability of the MNO to control the data transfer to the server for UE-side data collection.
* The capability of the MNO to control the data transfer from the server for UE-side data collection.
* The specific entity within the MNO to control the data transfer to and from the server for UE-side data collection.
* The protocols ~~or~~ and methods utilized by the MNO to control the data transfer to and from the server for UE-side data collection.
* The MNO's ability to manage (e.g., allow/disallow, initiate/terminate, prioritize/de-prioritize, etc.) the data transfer to and from the server for UE-side data collection.
 |
| Nokia | Yes, but it seems that the fourth bullet is a more descriptive version of the first bullet. The list could be consolidated by deleting the first bullet. It should also be noted that there may be multiple protocols, e.g., for solution 3, there is an interface between OAM and the gNB, and between the gNB and the UE, each with a very different protocol. Furthermore, there would be an interface between the OAM and the server for UE-side data collection, which we need to determine to be proprietary or 3GPP-based. |

For solution 1a), the MNO has no specific controllability for transfer of the collected data for UE-side data collection. For example, the collected data might be transferred together with other UE traffic without differentiation.

**Q4.2:** **Do companies agree that in solution 1a) MNO has no specific controllability for transfer of the collected data for UE-side data collection? Otherwise,** **please clarify any controllability that the MNOs have.**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | Yes |
| Apple | We agree its intention, but we think the term “specific controllability” is not clear:1. We think NW implementation can already achieve NW controllability, e.g., by assigning a dedicated PDU session only for UE sided data collection. In this case, since the UE can’t mix other type of traffics with training dataset, it is at cost of radio resource inefficiency.
2. To alleviate the issue of radio resource inefficiency in 1), the controllability can be enhanced by limited spec change, e.g. SA2 introduce a new 5QI for dataset transfer of training. With it, the UE can mix other type of traffics with training dataset in one PDU session, and NW can differentiate it.

To make it clear, we suggest below rewording:**in solution 1a) although it may not fully optimize system performance, MNO can already have ~~has no specific~~ controllability for transfer of the collected data for UE-side data collection via NW implementation.**  |
| BT | Yes, but as mentioned before, 1a is a solution totally outside 3GPP. No need to discuss 1a |
| Nokia | Maybe. If the MNO knows the IP addresses of the OTT servers, traffic to the servers could be throttled. But there is, otherwise, no specific control over aspects of the data collection procedures. |

In solution 1b, it is recognized that the MNO retains a certain level of control over the data transfer to the server for UE-side data collection. The entity within the MNO responsible for this control might be a NF in the CN, such as the DCAF. The MNO exercises this control at the granularity of the PDU session, as governed by the SLA. However, since the initiation of the data collection procedure is triggered by the server as specific types of services, the MNO may not have the authority to directly control and manage the data transfer process.

**Q4.3: Related to solution 1b, can the MNO have control/management over the data collection for UE-side data collection, given that it may be managed by a NF within the CN, with control granularity at the level of the PDU session per SLA? Please clarify any other controllability that the MNOs might have in solution 1b.**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | Yes, what RAN2 can clarify is the CN control, but the details of CN implementation can be discussed by SA2.  |
| Apple | Yes with comments:It is our understanding that from controllability perspective, the main difference of solution 1b from solution 1a is that it has one more control in CN via a NF (e.g. NWDAF, DCAF). However, the issue of radio resource inefficiency is same as solution 1, and the deployment of NF for data analysis (e.g. DCAF or NWDAF) in practical network is not successful so far. Thus, to reflect above issue, we suggest below rewording: **Related to solution 1b, although it may not fully optimize system performance, MNO can have control/management over the data collection for UE-side data collection, given that it may be managed by a NF within the CN, with control granularity at the level of the PDU session per SLA.**   |
| BT | No. Considering Q3.2, our understanding is that UE-side data collected is inside the MNO. That means, the UE-side data collection belongs to the MNO. Therefore, it has control and awareness over collected data. |
| Nokia | Maybe. Depending on whether our interpretation of solution 1b matches what we have shown in our answer to Q3.2, the level of control could be considered to be more than just at the PDU session level.However, if 1b is simply a server attached to a UPF and has an IP address internal to the network or specific to a DN for data collection, then we would agree with the PDU session level of granularity. |

For solution 2, it is recognized that the MNO has full level of control over the data collection process. The entity within the MNO responsible for this control is a NF in the CN, such as NWDAF, LMF (for positioning), or other relevant NFs. There are two potential control methods:

* Option 1: The MNO may manage data collection through NAS signaling (or LPP for positioning).
* Option 2: Alternatively, the MNO may manage the data collection process from the CN to the UE via RAN nodes, utilizing RRC signaling.

This level of control enables the MNO to directly manage the data collection process as required.

**Q4.4:** **Do the companies agree that in solution 2, the MNO’s control over the data collection for UE-side data collection is characterized by full controllability? Please clarify whether it is managed by a NF within the CN through NAS signaling directly (option 1) or through RRC signaling via RAN node (option 2), or others?**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | Yes, what RAN2 can clarify is the CN control for this solution, but the details of CN implementation can be discussed by SA2.In addition , we think the potential control methods for solution-2 can be also appliable to solution-1b.From signaling perspective, we believe the NAS layer signalling should be prioritized.  |
| Apple | Yes but we have below question for clarification for option 2:* On RRC signaling, to help understanding, is it signaling like QoE (i.e. a transparent container in RRC message)?
 |
| BT | Yes to "the MNOs have full control over data collection configuration and collected data for UE-side model training, management and inference on the CN”.It is too early to decide the concrete signaling. |
| Nokia | Option 1: Maybe. It isn’t clear how the CN could provide detailed configurations to the gNB, though.If the approach is CN-based, then the initial 3GPP-based control mechanism for DC for BM could come from the CN over NAS. NAS messages would be required to configure the UE for data collection for BM and NAS messages would be required to transport the measurements to the AMF and then further through the CN.For positioning data collection, we agree that LPP would be the mechanism used to manage data collection from the UE.Option 2: No. Currently, there is a mechanism to pass a signaling-based MDT Trace Activation from OAM to the gNB through the AMF, but we do not think that is the intention of this solution 2. We see this as a RAN-based or OAM approach. And for this option, the CN would still need to configure the RAN node for physical layer data collection. |

For solution 3, it is recognized that the MNO has full level of control over the data collection process. The entity within the MNO responsible for this control is OAM. The MNO controls the data collection process through RRC signaling via RAN node. This level of control enables the MNO to directly initiate and terminate the data collection process as required.

**Q4.5: Do the companies agree that in solution 3, the MNO’s control over the data collection for UE-side data collection is characterized by full controllability, managed by OAM through RRC signaling via RAN node, and with the ability to directly manage the data collection procedure?**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | Yes, what RAN2 can clarify is the OAM control for this solution, but the details of OAM implementation can be discussed by SA5. Meanwhile, it should be noted that the interface between OAM and RAN is non standardized interface. It would be helpful to clarify how the OAM manage the data collection procedure through RRC signaling via RAN node. We assume the OAM management can reuse the MDT framework. The OAM triggers the UE-side data collection via RAN (RRC signalling). The detail is different from MDT but the triggering and request the RAN node to do/signal something to the UE will be similar. We may consider this as baseline, then can discuss further, e.g. whether new framework is needed or not.  |
| Apple | Yes |
| BT | Yes to "the MNOs have full control over data collection configuration and collected data for UE-side model training, management and inference on the OAM”. |
| Nokia | Yes. |

## 2.5 Visibility of data content in MNO

As a preliminary measure, the nature of the data content can be described by its format, type, value and others. Visibility refers to the extent to which the MNO is able to aware, access or even comprehend this data content.

Companies are invited to provide the definition on visibility and to what extent (aware, access or comprehend) the visibility is preferred.

**Q5.1 Companies are invited to clarify their understanding of data content visibility and specify their preferred level of visibility.**

|  |  |
| --- | --- |
| Company | Comment |
| NEC | In our understanding, the data content should be only visible by the termination point.  |
| Apple | First, it is not clear whether the “data content visibility” could comply with regulations in different countries, e.g. GDPR. It is beyond 3GPP scope to define those requirements/mechanisms closely related based on legal grounds. Second, assuming 3GPP is going to work on this requirement, we believe it is in remit of SA1 instead of RAN2. So, RAN2 should not define any requirement on behalf of SA1. Third, regarding partial and fully data content visibility, SA3 should be consulted as there is potential impact on UE privacy and security and potential contradict with regional regulations.Just to facilitate RAN2 discussion on clarifying solution details, we can compromise to start discussion on solution details based on **assumption of Rapporteur’s suggested 3 levels “visibility”**. However, we have strong concern that this assumption may be misused as requirement of “visibility”. Thus, **when capturing solution details, we requested to clearly clarify below points:** 1. **This is just an assumption for RAN2 study purpose, and the definition of “visibility” and the category of visibility should be discussed and defined in SA1.**
2. **RAN2 should not agree any definition and requirements on “visibility”.**
3. **Without proper definition and requirements from SA1 and guidance from SA3, RAN2 shall not conclude any solutions as it may contradict with regulations.**
 |
| BT | visibility for us mean awareness, access, and comprehend |
| Nokia | Visibility has at least two characteristics: knowledge that a certain type of data is being transmitted from the UE (e.g., by knowing that traffic to a certain IP address is always for training data collection); or knowledge about the contents of the data, e.g., the format of the data, or what a string of bits represents in terms of type and value.We think that both characteristics are necessary to enable full visibility from the MNO. |

As an initial step, we define the levels of data content visibility within the MNO as follows:

* No visibility: MNO is not aware of the collected data and cannot access the data content.
* Partial visibility: The MNO is aware of the collected data, has limited access/comprehension to some elements of the data content, allowing for limited access.
* Full visibility: The MNO is aware of the collected data, has complete access to all aspects of the data content, enabling thorough comprehension if needed.

For solution 1a), the MNO has no visibility of data content. For solution 1b), the MNO has no or partial visibility of data content depending on the SLA. For solution 2 and 3, the MNO is able to have full visibility of the data content.

**Q5.2: Do companies agree that in solution 1a) MNO has no visibility of data content for UE-side data collection?**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | Yes  |
| Apple | No. In solution 1a), it is still possible that part of data content compliance with regulations can be visible to MNO via offline engineering between particular pair of MNO and UE vendor. Thus, we suggest below change: **in solution 1a) MNO has no or partial visibility of data content compliant with regulations** **for UE-side data collection depending on offline engineering.**  |
| BT | visibility for us mean awareness, access, and comprehend |
| Nokia | Yes. While the MNO might be able to infer something about traffic to a certain IP address, the contents of the data would be unknown to the MNO. |

**Q5.3: Do companies agree that in solution 1b) MNO has no or partial visibility of data content for UE-side data collection depending on SLA?**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | No, we think the data content should be visible to the CN entity as solution 2.  |
| Apple | Yes with comments:1. We think the difference between solution 1a) and solution 1b) is just whether the offline agreement has 3GPP impact or not (i.e., offline engineering in solution 1a vs SLA in solution 1b). Please note that both solutions need offline negotiation between particular pair of MNO and UE vendor. And the efforts of offline negotiation should be equal for solution 1a) and solution 1b).
2. In the case of “MNO has partial visibility of data content”, which part of the data content can be visible to MNOs should be left to SA3 to decide, as it is impacting UE security and privacy.

And similar to Q5.3, we think the partial visibility of data content should also fulfill compliance with regulations. Thus, we suggest below change:**in solution 1b) MNO has no or partial visibility of data content compliant with regulations for UE-side data collection depending on SLA** |
| BT | No. We consider UE-side data collector is inside the MNO network. Consequently, MNO has full visibility |
| Nokia | No. We think that no visibility, partial visibility, and full visibility are all options for solution 1b data collection. We think that partial visibility should also be defined.We think that partial visibility means that some fields are decodable by the MNO and others are not decodable, e.g., proprietary measurement types. Additionally, we do not think that discussion about SLAs is in RAN2 scope, and we should instead discuss what type of access could be enabled from a third party. |

**Q5.4: Do companies agree that in solution 2 and 3 MNO is able to have full visibility of data content for UE-side data collection?**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | No, we think the data content should be visible to the CN entity for solution 2.It is FFS for Solution-3, which may be up to the discussion at SA5.  |
| Apple | No, it depends on whether/how the data content is 3GPP specified: * If whole data content is UE vendor’s proprietary format, MNO has no visibility because MNO can’t comprehend it and thereby can’t use the data.
* If some part of data content is UE vendor’s proprietary format (e.g. a container in RRC message) and other part is 3GPP specified format, MNO may have partial visibility.
	+ In this case, which part of the data content can be visible to MNOs should be left to SA3 to decide, as it is impacting UE security and privacy.
* If whole data content is 3GPP format (e.g. specified IE in RRC message), MNO may have full visibility.

Since RAN1 is still discussing specification of data/metric for data collection, RAN2 can’t assume all data contents are 3GPP specified. If necessary, we can send LS to RAN1 for confirmation.  And similar to Q5.3 and 5.4, we think the partial/full visibility of data content should also fulfill compliance with regulations. Thus, we suggest below change:**in solution 2 and 3 MNO ~~is able to have~~ has no, or partial, or full visibility of data content compliant with regulations for UE-side data collection depending on whether/how the data content is 3GPP specified.** |
| BT | No. We consider UE-side data collector is inside the MNO network. Consequently, MNO has full visibility |
| Nokia | Yes. However, a transparent container has been discussed in the past to enable a “partial visibility” solution, at least for solution 3. |

## 2.6 Protocol layer for Data Transfer

For solution 1a and 1b, the server for UE-side data collection receives data from the UE through the application layer, utilizing a UP tunnel for transmission. It should be noted that from user’s perspective, the data transfer from the UE to the server via UP tunnel is consider as traffic and is subject to data usage charges.

**Q6.1: Do companies agree that in solution 1a) and 1b) the data transfer from the UE to the server for UE-side data collection is through the application layer, utilizing a UP tunnel for transmission?**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | Yes |
| Apple  | Yes |
|  |  |
| BT | Yes.This data will be considered as user plane. Therefore, it is the end customer who must pay for it (perhaps without being aware of it).  |
| Nokia | Yes. Is it further implied, then, that the protocol for data transmission is proprietary, or are there standardized solutions, e.g., in the case of solution 1b? |

For solution 2, the UE collects training data and transfers it to the CN, which then forwards the data to the UE-side server. As a basic approach, the data transfer from the UE to the CN can be conducted through the NAS layer using CP tunnel. The necessity for a UP tunnel to facilitate data transfer from the UE to the CN depends on the data volume, which is based on the requirements provided by RAN1. The feasibility of support a UP tunnel falls under the purview of SA2.

**Q6.2:** **Do companies agree that in solution 2, the baseline method for data transfer from the UE to the CN, is through the NAS layer, utilizing a CP tunnel for transmission?**

|  |  |
| --- | --- |
| Company  | Yes/No (Comment) |
| NEC | Yes, the CP can be the baseline. On the other hand, RAN2 should keep the UP possibility on the table which is pending SA2 discussion.  |
| Apple | Yes |
| BT | Not in a position to answer this question without knowing the data volume required for different AI/ML algorithms. |
| Nokia | Yes. To properly differentiate between solution 1b and 2, we think that it is necessary to restrict solution 2 to NAS signaling between the UE and AMF. It is FFS whether additional signaling would be required between the AMF and gNB. |

For solution 3, the UE collects training data and transfers it to the OAM, which then forwards the data to the UE-side server. As a basic approach, the data transfer from the UE to the OAM via RAN node can be conducted through the RRC layer using CP tunnel. The necessity for a UP tunnel to facilitate data transfer from the UE to the OAM depends on the data volume, which is based on the requirements provided by RAN1. The feasibility of support a UP tunnel falls under the purview of SA5.

**Q6.3:** **Do companies agree that in solution 3, the baseline method for data transfer from the UE to OAM via RAN node is through the RRC layer, utilizing a CP tunnel for transmission?**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| NEC | Yes  |
| Apple  | Yes |
| BT | Not in a position to answer this question without knowing the data volume required for different AI/ML algorithms. |
| Nokia | Yes. We should reuse the legacy approach. |

**Q6.4: For solution 2 and 3, interested company please share their view on solution details of UP tunnel solutions.**

|  |  |
| --- | --- |
| Company | Yes/No (Comment) |
| Apple | We are open to discuss UP option for solution 2 and 3. However, we are not sure how it works for now. According to current NR definition of UP tunnel, it is between UE and UPF, i.e. **the terminated entity is UPF in UL**. It seems to be conflicted with option 2 (3GPP terminated entity is one NF in CN) and option 3 (3GPP terminated entity is OAM). We believe a new Network Architecture and new protocol stack is not in Rel-19 scope. |
| Nokia | With regard to Apple’s comment that for the CN approach, the UPF is the termination point of UP traffic, we think that the use of the UPF should not be considered a CN approach in particular. The CN approach should terminate in a CN entity other than a UPF prior to any data transfer to a server which trains UE-side models. |
|  |  |
|  |  |

## 2.7 Privacy Concerns

Privacy concerns in the communications ecosystem can vary significantly among different stakeholders, such as MNOs, network vendors, OEMs, and chipset vendors. Each entity has its own set of sensitive information that needs to be protected from unauthorized disclosure. Here are more detailed examples of privacy concerns for each stakeholder:

MNO Privacy Concerns:

* Network Information Disclosure: MNOs may inadvertently disclose sensitive network information, such as deployment strategies, network configurations, and performance metrics, to servers outside their network.
* Data Transfer Risks: When transferring data to external servers, there is a risk that the data may contain undisclosed information about the network or its users/customers. This could include subscriber identities, locations, website visited, phone calls, etc and other proprietary information that MNOs are obligated to protect.

Network Vendor Privacy Concerns:

* Sensitive Information Leakage: Network vendors may possess proprietary algorithms, system designs, and other intellectual property that are integral to their competitive edge. Unintentional disclosure of such information to third parties could undermine their market position and lead to potential legal issues.
* Implementation Details Exposure: The specific details of how network equipment is implemented, including software and hardware design, are crucial for maintaining the security and integrity of the network. If such information is disclosed, it could be exploited for malicious purposes or used by competitors to gain insights into the vendor's technology.

Chipset Vendor Privacy Concerns:

* Proprietary Technology Exposure: Chipset vendors develop specialized hardware and software that may contain trade secrets or patented technologies. here is a risk that shared information could be unintentionally disclosed to unauthorized parties, leading to privacy breaches.

OEM Privacy Concerns:

* User Information Disclosure: OEMs handle a vast amount of user data, including personal information, usage habits, and location data. There is a significant privacy concern if this information is disclosed to external entities without user consent, potentially violating privacy laws and damaging the OEM's reputation.

For all stakeholders, it is essential to implement robust data protection measures, such as encryption, access controls, and privacy policies, to mitigate these privacy concerns. Additionally, compliance with regulations like the General Data Protection Regulation (GDPR) and other local privacy laws is crucial for maintaining trust and legal standing in the market.

Companies are encouraged to clearly define and express their privacy concerns, even if these concerns may not be the primary focus of RAN2. There is a prevailing sentiment that companies have underlying fears that they have not been able to precisely define. This ambiguity regarding the specific nature of their concerns is not conducive to advancing our collective work on data collection. A well-defined awareness of the privacy issues at stake is important for facilitating meaningful progress in our study.

**Q7: What’s your privacy concerns on the data collection?**

|  |  |
| --- | --- |
| Company | Comment |
| Apple | As both UE vendor and chipset vendor, we have strong concerns on privacy from below aspects:1. Concerns on UE Privacy:
	* As UE vendor, we have strong concern on exposure of UE privacy to unknown 3rd entity. We believe it is always most sensitive issue to all UE vendors, and it may violate privacy laws and damage the OEM's reputation.
	* The UE privacy at least includes UE Identities (SUPI, IP address, etc.), personal information, UE location, and UE trajectory information, etc.
	* Thus, explicit UE consent for data collection is a must. Furthermore, we don’t think existing 3GPP specified UE consent mechanism/requirements in TS 33.501 Annex V (e.g. consent of MDT) is applied in this case because we understand it only specifies whether and how NFs in MNO domain check the user consent before processing the data. User consent mechanism for the scenario of sharing data outside of MNO domain has not been discussed in 3GPP yet. Without a proper mechanism on collecting user consent for sharing data outside of MNO domain, any solution will be risky to violate the user privacy as well as the regional regulations.
2. Concern on exposure of proprietary implementation to other vendors:
	* It is common understanding that chipset vendor can have some proprietary implementation on top of 3GPP specification. And UE differentiation is one of most important reasons why 3GPP can build successful ecosystem so far. We believe any challenging of this principle will challenge success of 3GPP.
	* As chipset vendor, we have strong concern on exposure of our proprietary implementation to other vendors, including other chipset vendors, UE vendors, NW vendors, MNOs, and any 3rd entity.
	* Thus, we do not agree any solution which has potential risk of exposure of proprietary implementation to other vendors. And we believe not exposure of proprietary implementation to other vendors should be one of the most important requirements for UE data collection design.
 |
| BT | It is the operator who is under regulatory restriction. The lack of control and knowledge over collected data may result on an undesired exposure of our customers personal data. Most likely, it will conclude on penalties for operators as we will break regional regulations. In consequence, MNOs requires tools to stop it.Option 1a is the only one that exempts operators from any responsibility, but that solution is totally outside 3GPP. It cannot be even considered as baseline.Collected data will expose our network design and network strategy. Any disclaim of that information will have a negative impact. |
| Nokia | An additional privacy concern on Network Vendor that is not covered here is that radio topology and radio specific settings can be disclosed, and standard should not enable this. |
|  |  |

# 3 Appendix

**Table 1 Comparative analysis among different data collection approaches**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Aspects** | **1a) OTT (3GPP Transparent)** | **1b) OTT (Non-3GPP Transparent)** | **2. Transfer via Core Network** | **3. Transfer via OAM** |
| **Termination Entity** | UE-side OTT server | UE-side OTT server | CN | OAM |
| **Inside/outside MNO’s network** | Outside | Inside/Outside | Inside/outside | Inside/outside |
| **Transport Tunnel** | UP tunnel (Note: data collection may be charged as normal traffic.) | UP tunnel (Note: data collection may be charged as normal traffic.) | CP tunnel (FFS: UP tunnel) | CP tunnel (FFS: UP tunnel) |
| **Protocol layer for data transfer** | Application layer | Application layer | NAS or RRC layer (FFS: transport layer of UP tunnel) | RRC layer (FFS: transport layer of UP tunnel) |
| **Data Collection Controller** | UE-side OTT server | CN (certain NF, e.g., DCAF) | CN | OAM/RAN |
| **Control Granularity by NW** | NA, the OTT server can directly request data from the UE. | Coarse e.g., based on SLA | Finer (e.g., per NAS procedure) | Finer (e.g., per RRC procedure) |
| **RAN Intervene**  | No | No | Possible  | Yes  |
| **Radio Resource Efficiency** | Low, as the network cannot optimize radio resource usage for data transfer | Low, as the network cannot optimize radio resource usage for data transfer | Higher, due to the possibility of RAN intervene | Maximum |
| **Data format** | Non-standardized  | Non-standardized | Standardized/non-standardized | Standardized/non-standardized |
| **Network Awareness of the data Content** | No | No | Yes, if the data content is standardized or disclosed to MNO.  | Yes, if the data content is standardized or disclosed to MNO. |
| **RAN configuration/condition acquisition** | Unclear whether the RAN configuration/condition acquired by the UE can be transferred to the UE-side OTT server.  | Unclear whether the RAN configuration/condition acquired by the UE can be transferred to the UE-side OTT server and how the CN can control it with limited intervene.  | Controlled by MNO | Controlled by MNO |
| **Spec Impact** | No | Limited | High | High |
| **Security and Privacy Risk** | High, managed by the OTT application, with potential risks if not 3GPP compliant | Lower, managed by the OTT application and NF based on SLA. | Minimum, NW can enforce security and privacy protection. | Minimum, NW can enforce security and privacy protection. |
| **Involved WGs** | No | SA2 | RAN2, SA2 | RAN2, SA2, SA5 |

# 4 Conclusion

*

# 5 Reference

1. [R2-2403967](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2403967.zip) Discussion on Definition of Different UE Side Data Collection Solutions
2. [R2-2402364](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2402364.zip) Data Collection for UE Side Model Training MediaTek Inc.
3. [R2-2403235](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2403235.zip) Data Collection for UE-Side Model Training Interdigital Inc.
4. [R2-2403473](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2403473.zip) Discussion on Data Collection for UE-side Model Training Futurewei Technologies
5. [R2-2403378](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2403378.zip) Consideration on UE Side Data Colection ZTE Corporation,Sanechips
6. [R2-2403492](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2403492.zip) AI/ML Data Collection Requirements T-Mobile USA, Verizon, Charter, NTT DOCOMO, Deutsche Telekom, Turkcell, BT, AT&T, Nokia, Telecom Italia, CMCC
7. [R2-2403230](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2403230.zip) Discussion on UE Side Data Collection Qualcomm Incorporated
8. [R2-2402375](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2402375.zip) Data collection for UE-side model training Samsung
9. [R2-2402962](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2402962.zip) Data Collection for Training of UE-side AI/ML Models Nokia
10. [R2-2403567](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2403567.zip) Discussion on UE side data collection China Unicom
11. [R2-2402171](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2402171.zip) Data Collection for UE Side Model Training OPPO
12. [R2-2402302](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2402302.zip) Consideration on UE side data collection CATT
13. [R2-2402316](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2402316.zip) Data collection for UE side model training Xiaomi
14. [R2-2402342](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2402342.zip) Discussion on UE side data collection Spreadtrum Communications
15. [R2-2402375](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2402375.zip) Data collection for UE-side model training Samsung discussion
16. [R2-2402478](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2402478.zip) Label Dataset Transfer for Positioning UE-sided model training Intel Corporation
17. [R2-2402489](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2402489.zip) Discussion on UE side data collection vivo
18. [R2-2402669](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2402669.zip) Data Collection for UE side Model training NEC
19. [R2-2402732](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2402732.zip) Discussion on data collection for UE-sided model Lenovo
20. [R2-2402864](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2402864.zip) Discussion on UE-sided data collection Apple
21. [R2-2403022](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2403022.zip) Discussion on data collection for UE-sided model training CMCC
22. [R2-2403122](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2403122.zip) UE-side data collection Fraunhofer HHI, Fraunhofer IIS
23. [R2-2403163](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2403163.zip) Discussion on UE-sided data collection for training Huawei, HiSilicon
24. [R2-2403230](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2403230.zip) Discussion on UE Side Data Collection Qualcomm Incorporated
25. [R2-2403567](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2403567.zip) Discussion on UE side data collection China Unicom discussion
26. [R2-2403573](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2403573.zip) UE side data collection LG Electronics discussion
27. [R2-2403658](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2403658.zip) Discussion on UE side data collection Indian Institute of Tech (M), IIT Kanpur
28. [R2-2403661](file:///C%3A%5CUsers%5Cpanidx%5COneDrive%20-%20InterDigital%20Communications%2C%20Inc%5CDocuments%5C3GPP%20RAN%5CTSGR2_125bis%5CDocs%5CR2-2403661.zip) UE-side Data Collection Ericsson