3GPP TSG-RAN WG1 Meeting #104-e R1-21xxxxx

e-Meeting, April 12th – 20th, 2021

Source: Moderator (Qualcomm)

Title: Email discussion for XR traffic models and KPIs

Agenda Item: 8.14.1

Document for: Discussion and Decision

# Introduction

This contribution is a summary on the email discussion on remaining open issues for traffic models and KPI’s for XR and Cloud Gaming.

[104b-e-NR-XR-01] Email discussion/approval on traffic model – Eddy (Qualcomm)

* 1st check point: April 15
* Final check point: April 20

# Outcomes of RAN1 #104b-e

Agreement:

Jitter for DL video stream for the case of a single stream per UE

* J is drawn from a truncated Gaussian distribution:
  + Mean: 0 ms
  + STD: 2 ms
  + Range: [-4, 4] ms (baseline), [-5, 5] ms (optional)
    - Note: The values are set to ensure that packet arrivals are in order (i.e., arrival time of next packet is always larger than that of the previous packet) rather than the real measurement
  + Other values can be optionally evaluated
* Note: The above parameters for random variable J are effectively identical to the following parameter values because air interface PDB (e.g., 10ms or 15ms) applies from the point when each packet arrives at gNB as agreed in RAN1#104-e.
  + Mean: 4 ms (baseline), 5ms (optional)
  + STD: 2 ms
  + Range: [0, 8] ms (baseline), [0, 10] ms (optional)
  + Other values can be optionally evaluated

Agreement:

Parameters of Truncated Gaussian distribution for packet size of DL video stream in case of single stream evaluation (note: these parameter values are those before the truncation):

* [STD, Max, Min]: [10.5, 150, 50]% of Mean packet size
* Other values that can be used for evaluation: [STD, Max, Min] = [4, 112, 88] % of Mean for single eye buffer, [3, 109, 91] % of Mean for dual eye buffer
* FFS: Whether and how to evaluate single eye and dual eye buffer
* Note: Companies report the values used in their simulation results.
* Note: There is no consensus that the [10.5, 150, 50]% of mean packet size is the best set of parameters

Agreement:

In case of single stream per UE in DL, a UE is declared a satisfied UE if more than X (%) of packets are successfully delivered within a given air interface PDB.

* The baseline X value is 99.
* Other values of X can be optionally evaluated, e.g., X < = 95, X=99.9.
* Additional combinations of (X, PDB) values can be optionally evaluated, e.g.,
  + (99, 7), (95, 13) for VR/AR
  + (99, 12), (95, 18) for CG
* FFS: Different values for I-frame and P-frame if evaluation of them is agreed.

Agreement:

On UL Traffic model and QoS parameters

* CG/VR: single stream (pose/control)
* Traffic model for Pose/control
  + Periodic: 4ms (no jitter)
    - Other values can be optionally evaluated.
  + Fixed: 100 bytes
    - PDB: 10 ms.
  + A UE is declared a satisfied UE if more than X (%) of packets are successfully delivered within the given air interface PDB.
    - The baseline X value is 99.
    - Other X values can be optionally evaluated are 90 and 95.

# Discussion: Round 2 (Apr 16-19)

## DL Traffic Model: Multiple Streams per UE

**Moderator proposal for 1st round of email discussion**

* In addition to single stream per UE in DL which is baseline, two streams are optionally evaluated.
* FFS details of traffic model, KPI per stream, and per UE KPI

**Summary of comments on the moderator proposal for 1st round of email discussion**

* Support optional evaluation of two streams: FUTUREWEI, CATT, OPPO, vivo, MTK, HW, ZTE, LG, QC, InterDigital, Samsung, AT&T, Intel, Apple, DCM
* No evaluation of two streams: Xiaomi, Nokia

**New moderator proposal**

Given the clear majority view, RAN1 confirm that two streams in DL are optionally evaluated.

**Question 1.**

Assuming that two streams is optionally evaluated for DL, please provide your comments on the following options. All of the options may not have the same level of interest/support/priority.

* Option 1: I-frame + P-frame
* Option 2: video + audio/data
* Option 3: FOV + omnidirectional stream

|  |  |
| --- | --- |
| **Company** | **Comment** |
| Huawei/HiSilicon | We support to evaluate multi-stream models. All the listed options can be evaluated, and can help RAN1 better understand XR/CG performance from different angles.  SA4 already provided the traces for VR2-1~VR2-6 (see S4aV200634 clause 4.2.1), which includes I/P frame/slice. Several companies have already given traffic models for I/P frame based on SA4 outcome.  In addition, in our initial simulation results (see R1-2102322 clause 3.1.2), we consider a simple scheduler which prioritize transmission of I-frame stream over P-frame stream, and found it can effectively increase the network capacity.  So we suggest at least “Option 1: I-frame + P-frame” can be agreed this meeting, so that more companies can provide corresponding simulation results for next meeting.  “Option 3: FOV + omnidirectional stream” reflects the traffic model of VR1: “Viewport dependent streaming”. The FoV stream and the omnidirectional stream may have different data rate and QoS requirement. RAN1 can continue studying this model. |
| vivo | If two streams are optionally supported in DL, support Option 1 with first priority, and Option 3 with secondary priority (The detailed traffic model trace files are expected to provide by SA4). The motivation of modeling audio/data stream for capacity evaluation should be justified, otherwise, there is no need to model it separately. |
| MTK | We share similar views with Huawei and vivo. “**Option 1: I-frame + P-frame” can be selected as first priority** since SA4 has provided rich input for the structure of I/P frame/slice based on real data measurement. **“Option 3: FOV + omnidirectional stream” can be selected as second priority** to reflect the traffic model of VR1: “Viewport dependent streaming”. Also, VR1 can be an important use case for XR applications. |
| DOCOMO | We share similar views with other companies. Our first preference is Option 1 and the second is Option 3. |
| Apple | We prefer Option 2 if prioritization is to be taken, of course all options can be evaluated depending on companies’ interest. Note the modeling of I-frame and P-frame may be quite complicated, e.g. as an error resilience scheme, I-frame can be inserted depending on higher layer feedback. |
| FUTUREWEI | Suggest prioritizing one of the options similar to the Huawei, vivo and MTK views so far. At least for the upcoming meeting some results may be then compared with fewer/prioritized cases. |
| InterDigital | Assuming 2 streams are optionally evaluated in DL, we think companies can provide results based on their interests regarding the 3 options. Our preference for prioritization is along the order of Option 1 followed by Option 2 and Option 3. |
| Ericsson | It is not clear to us how to proceed here. We have spent three meetings on agreeing on single stream traffic model for video. Agreeing on an additional traffic model for the second stream would go beyond what RAN1 can accommodate.  We propose that the definition of the second stream is up to the individual company. |
| OPPO | If two streams are optionally supported in DL, we prefer Option 2. Regarding Option 1, We share similar view with Apple |
| Nokia, NSB | We share concern raised by Ericsson. Every additional option requires significant effort from companies to analyze the real traces and then agree on the appropriate distribution and parameter for the size of the packet, inter-arrival time, etc.  Over three meetings already we are constantly discussing the main parameters for the baseline (relatively simple) traffic model. Our concern is that if multi-stream traffic is optionally adopted, we will spend a lot of time discussing multiple options that are mainly related to one or few standalone applications (e.g., audio for AR2 and FOV for VR1) instead of moving forward in this SI.  According to our understanding, it is not realistic to assume that both the gNB and the UE can always reliably distinguish between I and P frames (need for a deep packet inspection at both gNB and UE side decoding IP, TCP, and application headers of every packet; encryption, including end-to-end encryption, packet size only as an indirect criterium here, as there might be both aggregation/fragmentation in the network and/or other packets of similar size belonging to other applications and services).  Therefore, we don’t see much point in complicating the traffic model further and spending extensive time on the discussion towards all the many needed parameters. There are actually two different sub-models discussed here (namely, slice-based and frame-based), each having its own peculiarities. Option 2 and 3 in our opinion are very much application specific.  In addition to such an extreme diversity of proposed models (so a lot of time is needed to converge), we have serios doubts that RAN1 has all necessary data for all the required tens of individual parameters to agree on. E.g., it looks like that there are no data from SA4 or other study group to decide what should be the appropriate PDB and X values for the I/P-frame model.  **Therefore, we are not persuaded that a more complex traffic model but with freely/randomly selected parameters is anyhow closer to reality than a simpler model with parameters derived carefully over the recent meetings (e.g., based on SA4 conclusions or at least with a clear majority from individual companies’ findings).**  In case the majority really wants to optionally simulate a particular variant of a two-stream model, we highly encourage to limit the options to a reasonable amount that can be agreed fast, and also indicate how the parameters are to be selected before concluding that such a model is used in the evaluations. |
| xiaomi | From our understanding, to support the traffic model option1, some basic assumptions should be made, which may not be in RAN scope. For example, we need to assume cross-layer design such that I-frame and P-frame can be differentiated in RAN level. Also, we need to assume that different QoS e.g. PDB and packet error rate, needs to be set for I-Frame stream and P-frame stream, which is not yet confirmed by SA2 and/or SA4 (from SA2 and SA4 LS they seem to assume the same QoS for a video stream). Although RAN1 can make its own assumptions to generate the evaluation results, we are not sure whether the observation and enhancements based on these results are realistic or useful if these assumptions cannot be supported by the current NW.  For the other two options, there seems to be no technical issue to support video + audio/data or FOV + omnidirectional. The only concern is on whether the additional evaluation effort is worthwhile or not. |
| QC | If traffic model with two streams is optionally evaluated, then, it should be option 1. Option 1 seem to have relatively clearer reason than other options, in the sense that the impact of separate handling of I and P frame on performance could be measured quantitatively (of course this may need further discussion). |
| CATT | Since two streams modeling is optional, we don’t need to further discuss how it is modelled. In particular, there is no verification if traffic model for I-frame or P-frame alone in Option 1 is truncated Gaussian. For option 2, it does not seems to have separated encoding for voice and video. Thus, we don’t see the need to discuss the two stream traffic model. |
| Facebook | We share the same understanding as Apple/Oppo, the insertion of I frame is pretty adapative….You can have a very long sequence of P-frames …Hence, we think optional 2 would is the more reasonable selection. |
| Intel | We are open to all options but would like to discuss options 1, 3. One issue is the availability of raw data (from SA4) to guide understanding and modeling, especially for option 3. For option 1, some data is available but we can wait for 1 meeting cycle to check if data for option 3 is also becoming available from SA4. We think its beneficial to have even a preliminary working model in RAN1 which can be tuned as more raw data becomes available. |

**Question 2**

If two streams of I-frame and P-frame are optionally evaluated for DL, please provide your view on how to evaluate them, i.e.,

traffic model, KPI per stream, and per UE KPI. Below is one proposal on the table. Please provide your comment on this proposal if any.

Table. Two-stream model for video

|  |  |  |
| --- | --- | --- |
| Application | AR/VR/CG | |
| Two data streams, i.e. M1 = 2 | * Stream #1: I-stream * Stream #2: P-stream | |
| Option 1A: slice-based | Option 1B: frame-based |
| Packet modelling | Slice-level | Frame-level |
| Traffic arrival pattern | Both streams are periodic with the same FPS. | Follow the GOP structure, e.g. GOP size K = 8, 60 |
| Number of packets per stream at a time | * Stream #1: 1 * Stream #2: N-1 (N is the number of slice per frame, e.g. N = 8) | * Stream #1: 1 * Stream #2: 1 |
| Average data rate | Stream #1: Stream #2 = : (N-1) | Stream #1: Stream #2 = : (K-1) |
| is average size ratio between one I-frame/slice and one P-frame/slice, e.g. = 2. Other values can be optionally evaluated. | |

Note: the QoS requirement for each stream is separately discussed in the KPI part

|  |  |
| --- | --- |
| **Company** | **Comment** |
| Huawei/HiSilicon | Support.  The I/P frame model is derived based with SA4 outcome. As analyzed in our Tdoc (R1-2102320 section 2.2.2), according to SA4 outcome in S4aV200634 and S4aV200669, there are mainly two options for I/P-streams:   * Option 1A (slice-based I/P-stream model): For configuration VR2-1/2/3/4/6, each frame has N slices (e.g., N=8), including one I-slice and (N-1) P-slices * Option 1B (frame-based I/P-stream model): For configuration VR2-5, there is no slice setting and every 8th frame is intra coded (i.e., I frame). For a Group-Of-Pictures (GOP) with a size of K frames, the frame arrival pattern is: K = 1 means all I-frames, K=2 means IPIP, K=3 means IPPIPP, and so on. The GOP size could be different in different applications and settings. In SA4 outcome S4aV200634, VR2-5 chooses GOP size K=8.     Option 1A: Slice-based I/P-stream model    Option 1B: Frame-based I/P-stream model |
| vivo | For the details of two streams traffic model for DL, both frame-based and slice-based traffic models are adopted. It should be also noted that GOP structure is common in practical.  For the average data rate, we suggest to adopt a value of = 3 as the baseline while other values can also be evaluated, since = 3 is the typical average size ratio between I-frame and P-frame.  Regarding the KPIs for two streams, a UE can be declared as a satisfied UE if more than X (%) of packets are successfully delivered within the given air interface PDB for each stream, e.g   * Stream #1: X=99, PDB=10ms * Stream #2: X=95, PDB=10ms |
| MTK | **We support to agree the FL suggested table** (Table. Two-stream model for video). For traffic model, vivo has done a good analysis in their contribution (Table 2 & 3 below) and we suggest to directly use it:  Table 2. GOP-based multiple streams traffic model (FPS=60)   |  |  |  |  | | --- | --- | --- | --- | | **Traffic model** | I-frame | P-frame | Note | | **Date rate (Mbps)** |  |  | GOP length = 1 second | | **Packet size distribution** | Truncated Gaussian distribution | |  | | **Mean packet size (Bytes)** | 272177 | 90725 | The average ratio of I-frame size and P-frame size is around 3:1 | | **STD of packet sizes (Bytes)** | 40826 | 13608 | 15% of Mean packet size | | **Maximum packet size (Bytes)** | 408265 | 136087 | 1.5 \* Mean packet size | | **Minimum packet size (Bytes)** | 900 | 300 |  | | **Packet arrival interval (ms)** | 1000 |  | 1 I-frame and 59 P-frames in one second | | **Packet delay budget (ms)** | 10 | |  |   Table 3. Slice-based multiple streams traffic model (FPS=60)   |  |  |  |  | | --- | --- | --- | --- | | **Traffic model** | I-frame | P-frame | Note | | **Date rate (Mbps)** |  |  |  | | **Packet size distribution** | Truncated Gaussian distribution | |  | | **Mean packet size (Bytes)** | 28125 | 65625 | The average ratio of I-slice size and P-slice size is around 3:1, and each encoded video frame contains 1 I-slice and 7 P-slices | | **STD of packet sizes (Bytes)** | 4218 | 9843 | 15% of Mean packet size | | **Maximum packet size (Bytes)** | 42187 | 98437 | 1.5 \* Mean packet size | | **Minimum packet size (Bytes)** | 93 | 217 |  | | **Packet arrival interval (ms)** |  |  | each encoded video frame contains 1 I-slice and 7 P-slices | | **Packet delay budget (ms)** | 10 | |  |   For KPI per stream, due to the larger size of an I-frame, **the PDB for I-frame should be set larger than a P-frame, e.g., 8ms for P-frame and 12ms for I-frame, or, 5ms for P-frame and 15ms for I-frame**. For the **PER of I-frame and P-frame**, they **can be set equal** (Ex. 99%) since losing either one of I-frame or P-frame would request a retransmission of I/P-frame so the succeeding video frames can be decoded. Having said that, **we can also accept** Huawei and vivo’s suggestion **to have larger X for I-frame than P-frame** due to more video information contained in I frame (Ex. X=99 for I-frame, X=95 for P-frame).  For the value, SA4 provides input in R1-2102308 that   * As an example, intra-coded (I-frame) information typically requires 3 to 6 times the amount of data compared to inter-coded (P-frame) information   Our preferred value is 4, but we can accept to have = 3. |
| Ericsson | The definition of I- and P-frame traffic models are up to the contributing company. |
| QC | As noted in the question, **if** two streams (I-frame and P-frame) is evaluated, then, as shown as an **example** table, we can discuss about slice-based and/or frame-based modeling.  Regarding the KPI for each stream (PER requirement X, PDB), it is not very clear what value to use. The currently suggested X and PDB for I and P frame seem to be a bit arbitrary. This needs further discussion. |
| CATT | We need to verify the traffic distribution for I-frame or P-frame is truncated Gaussian. From theoretic point of view, this is a different distribution based on DCT-based compression and H.264 encoding |
| Intel | Thanks for the proposal, generally we support it. yes, its not clear how KPI for each stream should be set, but we can start with a preliminary assumption (then check results and fine tune). We are not sure if PDB should be different. PER requirement can be different to reflect priority of I-frame (reflecting the fact that a lost I-frame can affect N frames) |

## UL Traffic Model: AR

**Moderator proposal for 1st round of email discussion**

* For UL AR,
  + Baseline: two streams are defined as follows.
    - Stream 1: pose/control
      * Traffic model and QoS parameters are same as those for pose/control for UL CG/VR.
    - Stream 2: aggregated stream for scene, video, data, and audio.
      * Packet size: Truncated Gaussian distribution with the same parameter values with DL (Mean, STD, Max, Min)
      * Periodicity: 60 fps
      * Data rate: 10 Mbps (baseline), 20 Mbps (optional)
      * PDB: 60 ms
  + Single stream (Stream 2 above) and/or more than two streams can be optionally evaluated.
  + A UE is declared as satisfied only when each stream meets the following requirement.
    - In case of the above baseline (2 streams)
      * X value for Stream 1 follows that for pose/control of UL CG/VR
      * X value for Stream 2 follows that for DL video streaming for single stream case.
      * Other values can be optionally evaluated
    - In case of single stream (Stream 2 above) and/or more than two streams
      * FFS

**Observation from 1st round of email discussion**

* # streams
  + Single stream:
    - (3) Ericsson, Nokia, Samsung
  + Two streams as proposed by moderator
    - (11) : CATT, OPPO, Xiaomi, vivo, MTK, ZTE, LG, QC, InterDigital, Intel, DCM
  + Three streams (1): Apple
* PDB for different streams
  + Control/pose: 10ms
    - This is the same as for pose/control for CG/VR (agreed by all companies)
  + A stream aggregating scene, video, data, and audio
    - 10ms or 15ms (4) : OPPO, Ericsson, InterDigital, AT&T
    - 60ms (10): FUTUREWEI, CATT, Ericsson, Xiaomi, vivo, MTK, ZTE, LG, QC, Samsung,
  + Separate audio/video
    - 30ms: Apple
      * Note: audio/data flow proposed by Apple: 10ms periodicity, 0.756 Mbps/s or 1.12 Mbps (packet size is determined by periodicity and data rate)
* Jitter
  + Same jitter model as for DL video stream: Ericsson, MTK (optional), AT&T, Intel (further discussion)

**Question 3. Please provide your view on which option(s) out of the three options below should be baseline and which option(s) out of the three options below should be optional. Also please provide additional comments if any.**

**New proposal:**

On UL AR,

* Option 1: Two streams as defined below
  + Stream 1: pose/control
    - Traffic model and QoS parameters are same as for pose/control for UL CG/VR.
  + Stream 2: A stream aggregating streams of scene, video, data, and audio.
    - Packet size: Truncated Gaussian distribution with the parameter values same as for DL
    - Periodicity: 60 fps
      * Jitter (optional): same model as for DL
    - Data rate: 10 Mbps (baseline), 20 Mbps (optional)
    - PDB: 60 ms (baseline), 10/15 ms (optional)
    - PER requirements:
    - FFS separate streams for I-frame and P-frame
* Option 2: Single stream as defined below
  + Packet size: Truncated Gaussian distribution with the parameter values same as for DL
  + Periodicity: 60 fps
    - Jitter (optional): same model as for DL
  + Data rate: 10 Mbps (baseline), 20 Mbps (optional)
  + PDB: 60 ms (baseline), 10/15 ms (optional)
  + FFS separate streams for I-frame and P-frame
* Option 3: Three streams as defined below
  + Stream 1: pose/control
    - Traffic model and QoS parameters are same as for pose/control for UL CG/VR.
  + Stream 2: A stream aggregating streams of scene and video
    - Packet size: Truncated Gaussian distribution with the parameter values same as for DL
    - Periodicity: 60 fps
      * Jitter (optional): same model as for DL
    - Data rate: 10 Mbps (baseline), 20 Mbps (optional)
    - PDB: 60 ms (baseline), 10/15 ms (optional)
    - FFS separate streams for I-frame and P-frame
  + Stream 3: A stream aggregating streams of audio and data
    - Periodicity: 10ms
    - Data rate: 0.756 Mbps/s or 1.12 Mbps
    - Packet size: determined by periodicity and data rate
    - PDB: 30 ms
* In case multiple steams are evaluated for UL AR, a UE is declared as satisfied only when each stream meets the requirement that X (%) of packets are successfully delivered within a given air interface PDB.
  + X value for pose/control: follow X values for pose/control for CG/VR
  + X value for Stream 2 in Option 1: follow X values for DL video stream
  + X value for Option 1: follow X values for DL video stream
  + X value for Stream 2 in Option 3: follow X values for DL video stream
  + X value for Stream 3 in Option 3: follow X values for DL video stream

|  |  |
| --- | --- |
| **Company** | **Comment** |
| Huawei/HiSilicon | We suggest to first discuss on DL, and once agreements are made, we can adapt them to UL easily.  For AR UL video, separate model for I-frame and P-frame should also be considered as DL.  We suggest to add the following option 4 and also reflect this option in the last bullet about “X value”.  ==   * Option 4: Three streams as defined below   + Stream 1: pose/control     - Traffic model and QoS parameters are same as for pose/control for UL CG/VR.   + Stream 2: I-stream for video   + Stream 3: P-stream for video   + Note: For stream 2 and stream 3, the I/P-stream model for DL video is reused for UL video.   …   * In case multiple steams are evaluated for UL AR, a UE is declared as satisfied only when each stream meets the requirement that X (%) of packets are successfully delivered within a given air interface PDB.   + X value for pose/control: follow X values for pose/control for CG/VR   + X value for Stream 2 in Option 1/3/4: follow X values for DL video stream   + X value for Option ~~1~~2: follow X values for DL video stream   + ~~X value for Stream 2 in Option 3: follow X values for DL video stream~~   + X value for Stream 3 in Option 3/4: follow X values for DL video stream   ==  It seems the last main bullet about “X value” looks complicated, maybe we can simplify it as follows:   * + X value for pose/control: follow X values for pose/control for CG/VR   + X value for other cases: follow X values for DL video stream |
| vivo | From the perspective of power consumption and capacity evaluation, option 1 is preferred, since pose/control stream could have significant impact on power consumption. However, from the perspective of capacity evaluation only, option 2 is slightly preferred, since pose/control stream might have negligible impact on capacity. Therefore, both option 1 and option 2 can be evaluated with high priority. Other options can be optionally evaluated. |
| MTK | We support Option 1 and Option 4 proposed by Huawei. We can accept Option 2 for capacity evaluation if that’s the majority view, but we can not accept Option 2 for power evaluation due to the reason FL explained in email discussion that frequent pose/control can impact UL power consumption. |
| DOCOMO | We support Option 1 and Option 4 for power consumption and capacity evaluation. If majority companies prefer Option 2, we are also fine with it only for capacity evaluation. |
| Apple | We support Option 3 as the first preference as baseline, and Option 1 as the second preference as baseline. |
| FUTUREWEI | In our views a single stream should be baseline for evaluations since it may complicate capacity evaluations with no benefit. We also agree that the option 2 is useful for power evaluations. In this case, considering either of the options as mandatory and the other option as optional may not be fair from technical point of view.  Therefore, we propose to have two options as we suggested in the first round of discussion. There is no need to further classify the two options as mandatory or baseline. Companies may use either for power or capacity evaluations. |
| InterDigital | We support Option 1 as baseline since this would enable to adequately evaluate capacity and power consumption performance. While the impact of evaluating two streams (Option 1) on power consumption is clear (as explained by FL), we think that even for capacity there would be non-negligible impact with 2 streams due to the consideration of the proposed KPI that requires satisfying the respective X and PDB requirements in each stream.  We share similar view with vivo and MTK that Option 2 (single stream) is unable to adequately show the impact on both capacity and power consumption. In this regard, Option 2 and Option 3 can be optionally evaluated by companies. |
| Ericsson | For capacity evaluation, option 2 is baseline. For power evaluation, option 1 is baseline. |
| OPPO | Option 2 is the first preference and Option 1 is the second preference |
| Nokia, NSB | Since the pose/control stream is not a bottleneck for capacity evaluation we support Baseline to be: Option 2: Single video stream.  We also see the point from some companies, who believe that Option 1 is more realistic for the UE power consumption KPIs. Therefore, we also suggest supporting this option as optional, so that it can be selected when presenting the UE power saving results (PS gain, etc.).  We also propose to put the values for PDB for video [60ms] in square brackets and add an FFS: Clarify the PDB for the UL video stream. It is better that we report a value that has some references behind it in the TR.  In addition, we propose to remove three identical FFSs: “Separate streams for I-frame and P-frame” from all the options since the options already states how many streams are assumed (2, 1, or 3, respectively). In contrast, as also stated by Huawei, if companies see the benefit providing additional options, they better to be listed as Option 4, Option 5, etc. just to avoid confusion. Otherwise, we might immediately end up having at least 6 options supported. |
| Xiaomi | For the baseline, we are fine with either option1 or option2. |
| QC | We support Option 1 as a baseline.  **Option 1 should be a baseline.** Option 1 is simple yet complex enough to capture two different characteristics of pose and other traffics (data, video, etc.). And more importantly, Option 2 cannot accurately evaluate UE power consumption. Thus, option 1 is preferred as a baseline.  We think having **the same baseline option 1 for both capacity and power evaluation** is desirable from evaluation complexity point of view. For power evaluation results report, corresponding capacity evaluation needs to be done anyway, and to provide a capacity reference (#satisfied Ues/cell). The same results can be used for capacity evaluation report directly without additional evaluation. Thus, we think it is important to use the same option for capacity and power evaluation. |
| CATT | Option 2: Single stream should be the baseline. |
| Facebook | We support option 1 as the as the baseline. For capacity evulation, option 2 should be good enough. |
| Intel | We are okay with Option 1 as baseline for both capacity and power evaluations. |

# Discussion: Round 1

## DL Traffic Model: Single Stream

1. **DL packet size distribution**

RAN#104-e Working assumption

|  |
| --- |
| (Working assumption) Parameters of Truncated Gaussian distribution for Packet size (note: these parameter values are those before the truncation)   * 1. Mean: Derived from average data rate and fps as follows.      1. (average data rate) / (fps for video stream, i.e., # packets per second in our statistical model) / 8 [bytes]   2. STD      1. TBD   3. Max packet size      1. TBD   4. Min packet size      1. TBD      2. FFS whether or not to use this parameter |

Companies’ views in RAN1#104bis-e tdocs are presented in the table below.

|  |  |
| --- | --- |
| Huawei | STD: 15% of Mean packet size  Max packet size: 2 \* Mean packet size  Min packet size : 25% of Mean packet size |
| OPPO | For the distribution of packet sizes, confirm the working assumption of truncated Gaussian distribution by removing the brackets, i.e.,   * STD: 15% or 20% of Mean packet size * Max packet size: 1.5 x Mean packet size * Min packet size: 0.5 x Mean packet size |
| vivo | * STD   + 15% of Mean packet size derived above * Max packet size   + 1.5 x Mean packet size derived above * Min packet size   + 0.1 \* Mean packet size derived above |
| CATT | The truncated Gaussian distribution for VBR  Uniform distribution for CBR.  For the parameters of the statistical distribution for Packet size   * STD: 5% and 10% of Mean packet size for CBR and VBR, respectively. * Max: 1.2 time of and equal to Mean packet size for VBR and CBR, respectively. * Min: limited by the minimum IP packet size, i.e. 46Bytes. |
| MTK | * STD: 15% of Mean packet size derived above * Range: 1.5 × Mean packet size derived above |
| Futurewei | * STD: 15% of Mean packet size * Max packet size: 1.5 x Mean packet size * Min packet size: 0.5 x Mean packet size |
| Nokia | * STD: 2% of mean packet (frame) size * Max packet (frame) size: 1.1 x mean packet (frame) size * Min packet (frame) size: 0.9 x mean packet (frame) size |
| Ericsson | * STD: 15% of mean packet (frame) size * Max packet (frame) size: 1.5 x mean packet (frame) size * Min packet (frame) size: 0.5 x mean packet (frame) size |
| Intel | * for CBR configurations the frame-size variations are quite small   + the max/mean frame-size ratio is ~ 1.06   + the min/mean frame-size ratio is ~0.93   + the std/mean frame-size ratio is ~0.02 * for cVBR configurations the frame-size variations are larger   + the max/mean frame-size ratio is ~ 1.18 – 1.94 with smaller ratios corresponding to 8 slice/eye buffer case while large ratio corresponding to 1 slice case   + the min/mean frame-size ratio is ~ 0.24 – 0.48 with larger ratios corresponding to 8 slice/eye buffer case while small ratio corresponding to 1 slice case   + the std/mean frame-size ratio is ~ 0.07 – 0.14 |
| Qualcomm | * STD: 7.5% of Mean * Max packet size: 135% of Mean * Min packet size: 54.5% of Mean |
| Samsung | * STD: 15% of Mean * Max packet size: 150% of Mean |
| ZTE | * [Single eye packet size](#_Toc68618183)   + [STD = 4% \* mean, MAX = 112% \* mean](#_Toc68618184) * [Dual eye packet size](#_Toc68618185)   + [STD = 3% \* mean, MAX = 109% \* mean.](#_Toc68618186) |

**Summary**:

* From SA4 trace files: STD, Max, and Min depend on CBR, VBR, single vs. multiple (8) slices per frame, etc.
* Two companies address different values for CBR and VBR. However, moderator recommends to not evaluate separately CBR and VBR based on majority view.
* Average values from companies’ inputs
  + STD: 10.5% of Mean packet size
  + Max: 135% of Mean packet size
  + Min: 50% of Mean packet size

1. **Based on the discussions and proposals in tdocs for RAN1#104bis-e, the moderator makes the following proposal. Please share your comments.**

**Moderator proposal**: Parameters of Truncated Gaussian distribution for packet size of DL video stream in case of single stream evaluation (note: these parameter values are those before the truncation)

* Mean packet size: Derived from average data rate and fps as follows.
  + (average data rate) / (fps for video stream, i.e., # packets per second in our statistical model) / 8 [bytes]
* STD: 10.5% of Mean packet size
* Max: 135% of Mean packet size
* Min: 50% of Mean packet size

|  |
| --- |
| **DL packet size discussion for a single stream per UE**  **Moderator proposal for 1st round of email discussion**: Parameters of Truncated Gaussian distribution for packet size of DL video stream in case of single stream evaluation (note: these parameter values are those before the truncation)   * Mean packet size: Derived from average data rate and fps as follows.   + (average data rate) / (fps for video stream, i.e., # packets per second in our statistical model) / 8 [bytes] * STD: 10.5% of Mean packet size * Max: 135% of Mean packet size * Min: 50% of Mean packet size   **Observation from 1st round of email discussion**   * Option 1: [STD, Max, Min] = [10.5, 135, 50] % of Mean   + (6 companies) Xiaomi, vivo, MTK, InterDigital, Samsung, DCM * Option 2: [STD, Max, Min] = [10.5, 150, 50] % of Mean   + (9 companies) Ericsson, Nokia, vivo, MTK, LG, QC, InterDigital, Samsung, AT&T * Option 3: [STD, Max, Min] = [15, 150, 50] % of Mean   + (4 companies) FUTUREWEI, OPPO, MTK, HW * Option 4: [STD, Max, Min] = [10.5, 135, N/A] % of Mean:   + (1 company): CATT * Option 6: [STD, Max, Min] = [4, 112, N/A] % of Mean for single eye buffer, [3, 109, N/A] % of Mean for dual eye buffer   + (1 company): ZTE * Option 7: More than one options: Intel   Note: Open to a majority view: Ericsson, QC, MTK, HW, InterDigital  **New moderator proposal**: Take Option 2 as baseline while other options can be optionally evaluated, i.e.,  Parameters of Truncated Gaussian distribution for packet size of DL video stream in case of single stream evaluation (note: these parameter values are those before the truncation)   * STD: 10.5% of Mean packet size * Max: 150% of Mean packet size * Min: 50% of Mean packet size * Other values can be optionally evaluated, e.g.,   + [STD, Max, Min] = [10.5, 135, 50] % of Mean packet size   + [STD, Max, Min] = [10.5, 150, 50] % of Mean packet size   + [STD, Max, Min] = [15, 150, 50] % of Mean   + [STD, Max, Min] = [10.5, 135, N/A of Mean   + [STD, Max, Min] = [4, 112, N/A] % of Mean for single eye buffer, [3, 109, N/A] % of Mean for dual eye buffer * Note: Companies report the values used in their simulation results.   **New moderator proposal**: RAN1 will send an LS to SA4 about the RAN1 agreements w.r.t. XR/CG traffic models and KPIs. |

|  |  |
| --- | --- |
| **Company** | **Comment** |
| FUTURWEI | In our views, we suggest to rather base the proposal on the majority views rather than the average values as proposed. The average values may result in a pdf function of the truncated Gaussian distribution not symmetrical. Following the majority views then the parameters would be   * STD: 15% of Mean packet size (7 companies) * Max: 1.5 of mean (6 companies)   In regards to the Min size, 3 companies propose 0.5 mean and one company propose 0.545 which is close to 0.5. One company proposed to 0.9 of mean and another company propose 0.25. As such propose to use   * Min 0.5 of mean |
| CATT | We are OK with moderator’s proposal except the Minimum packet size. We understand that the maximum, mean and minimum values are derived by SA4 XR traffic model from most contributions. However, the minimum packet size was an potential issue since any type of traffic could have very small packet. For the minimum value set at 0.5 of mean, it will distort the truncated Gaussian distributions and bias the behavior of actual XR traffic generation. |
| OPPO | We share the same view as FUTURWEI. The average values are obtained from the perspective of compromise, rather than from the technical point of view.  Having said that, we can accept FL proposal for progress if majority companies support it |
| Ericsson | OK to go with average values – the precise values are not that important. Propose to set (min,max) symmetric around the mean, so that we get the average data rate we aimed for |
| Xiaomi | We are fine to accept FL proposal. But we suggest to send LS to SA4 including RAN1 agreements on traffic model. And SA4 can give response if they have any concern. |
| vivo | We are fine with the proposal for progress. |
| MTK | We share similar view with Futurewei but we can accept FL proposal if majority companies support it. |
| Huawei, HiSilicon | It’s relevant to make the distribution representative of something that companies have identified, rather than to be an average which does not particularly represent anything exactly. We’d be OK with the values that have been most identified by companies. |
| Nokia, NSB | We support the proposed approach and values. We also suggest to make the min and max values symmetric around the mean. By analysing the SA4 traces companies clearly demonstrated that even for a single VR2 application but different bit rate configurations, slice division, etc., the final values will be different. Therefore, assuming one particular configuration will show very biased result. |
| ZTE | We would like to share our opinion as follow:   1. We don’t find any support about STD = 15% \* mean and MAX = 150% \* mean. According to SA input Sa4-V600040, the ratio between STD and mean value as well as that between MAX and mean value is shown in Table 1   Table 1   |  |  |  |  | | --- | --- | --- | --- | | Birate | Configuration | STD/Mean(Percent) | Max Packet Size/Mean(Percent) | | 30Mbps | VR2-3, VR2-4 | 2 | 106 | | VR 2-1, VR2-2, VR2-6 | 8 | 124 | | VR 2-5 | 13 | 139 | | 45Mbps | VR2-7, VR 2-8 | 15 | 145 |   In Table 1, the working assumption in last meeting could only cover the 45Mbps cases with a modification of the max value from 1.5 to 1.45.   1. We notice that some companies used numerical characteristics of raw data to determine STD/Mean and Max/Mean. We are fine with this method as we introduced in R1-2103278. However, if we directly use the ratios presented in Table 1, on the one hand, different ratios would be considered in different bit rate cases, i.e., 30Mbps and 45Mbps, or CBR and VBR. On the other hand, non-negligible bias could be found in the CDF curves ranging from 5%-95%. 2. We utilized a reasonable fitting method as illstrated in R1-2103278 and found that there is an unified ratio(around 3%) between standard deviation and mean vaule for VR2 with different bit rate cases, i.e., 30Mbps and 45Mbps, or CBR and VBR, according to SA4 input Sa4V200640.   Table 2   |  |  |  |  | | --- | --- | --- | --- | | **Configuration** | **Mean**  **(Byte) x 10^4** | **STD**  **(Byte)** | **STD / Mean**  **(%)** | | **VR2-1** | 5.992 | 1884 | 3.14 | | **VR2-2** | 5.853 | 1605 | 2.74 | | **VR2-3** | 6.11 | 1496 | 2.20 | | **VR2-4** | 5.73 | 1463 | 2.31 | | **VR2-5** | 5.478 | 1245 | 2.27 | | **VR2-6** | 5.991 | 1602 | 2.67 | | **VR2-7** | 8.942 | 2637 | 2.95 | | **VR2-8** | 8.732 | 2614 | 2.99 |  1. For minimum packet size, our understanding is that this is not a necessary variable. Only if error case of packet size less than 0 is generated, which is in essence highly unlikely to take place, we can re-generate the packet size by using the original distribution function. |
| LG | We are fine with the approach the Moderator took for the proposal. In addition to that, the proposal to make the min/max symmetrical seems to make sense to us. If we take that proposal, then we don’t see much difference from the majority view :--) |
| QC | We are ok with the FL’s proposal. We think it is reasonable to have symmetrical truncation to keep the mean value unchanged. |
| InterDigital | We are OK with FL’s proposal, provided the majority of companies with fine with the proposed parameters. |
| Samsung | OK with the FL proposal. Also OK (but not critical) to have a symmetric truncation around the mean. |
| AT&T | Ok with the FL proposal, but agree with Ericsson and others that a symmetric min/max values around the mean seems to be better. |
| Intel | we believe the statistical model and modeling parameters should be firmly rooted in the analysis of SA4 P-trace models (or similar raw data) – an overly simplistic model (based on avg. for example) with no case-to-case variation can also lead to specification work that is not practically useful while showing gains on paper.  We observe CBR and VBR having quite different frame-size characteristics, with CBR having almost constant frame-size with very little variation  We also observe that usage of slicing is also helping to minimize variation of frame-sizes.  We did not observe symmetric distribution for many cases  Proposal:   * A single set of values is not sufficient to model all use-cases, encoding and delivery models. * This should be work in progress in RAN1 and further work on this can be done as other P-traces and use-cases become available from SA4. Confirmation with SA4 is also a good point. * Model should be rooted in the analysis of SA4 P-trace models (or similar data) |
| DOCOMO | We are fine with the FL proposal. |

1. **DL Jitter Model**

RAN1#104-e agreement on Jitter for DL video stream for a single UE

|  |
| --- |
| * Per the agreed statistical traffic model, arrival time of packet k is k/X1000 [ms] + J [ms], where X is the given fps value and J is a random variable.  * J is drawn from a truncated Gaussian distribution:   + Mean: [0]   + STD: [2 ms]   + Range: [[-4, 4]ms]     - Note: The values ensure that packet arrivals are in order (i.e., arrival time of a next packet is always larger than that of the previous packet)   + Note: The above values for mean, STD and Range are working assumption for initial simulations, and is to be revisited potentially with more inputs from companies in RAN1#104-bis-e |

Companies’ views in RAN1#104bis-e tdocs are presented in the table below.

|  |  |
| --- | --- |
| Huawei | Mean: 0; STD: 2 ms; Range: [-4, 4]ms |
| OPPO | Mean: 0; STD: 3 ms; Range: [-5, 5]ms |
| vivo | Mean: 0; STD: 2 ms; Range: [-4, 4]ms |
| CATT | Mean: 20 ms; STD: 6.35 ms; Range: [9, 31] ms |
| MTK | Mean: 0; STD: 2 ms; Range: [-4, 4]ms |
| Nokia | Mean: 0; STD: 3 ms; Range: [-6, 6]ms |
| Ericsson | Mean: 0; STD: 2 ms; Range: [-4, 4]ms |
| Qualcomm | Mean: 0; STD: 2 ms; Range: [-4, 4]ms |
| ZTE | Presented different values for different VR2 configurations. |
|  |  |

**Summary**:

* clear majority of companies propose to confirm the RAN1#104-e Working Assumption, Mean: 0; STD: 2 ms; Range: [-4, 4]ms.

1. **Based on the discussions and proposals in tdocs for RAN1#104bis-e, the moderator makes the following proposal. Please share your comments.**

**Moderator proposal:** Confirm the following WA.

* Jitter for DL video stream for a single UE
  + Per the agreed statistical traffic model, arrival time of packet k is k/X x 1000 [ms] + J [ms], where X is the given fps value and J is a random variable.
  + J is drawn from a truncated Gaussian distribution:
    - Mean: 0 ms
    - STD: 2 ms
    - Range: [4, 4] ms
      * Note: The values ensure that packet arrivals are in order (i.e., arrival time of a next packet is always larger than that of the previous packet)
    - Other values can be optionally evaluated

|  |  |
| --- | --- |
| **Company** | **Comment** |
| FUTUREWEI | OK with proposal |
| CATT | The formula for k-th packet arrival at the gNB buffer by the formula (k/X)\*1000 + J [ms]. For X= 60 fps, the 1st, 2nd, 3rd, and k-th packets generated from the XR source are 16.67, 33.34, 50.00, …, (k/X)\*1000 [ms]. The k-th packet arrival at the gNB buffer with the delay jitter modelling by random variable J in [ms] uses the formula (k/X)\*1000 + J [ms] as follows,  1st packet 16.67 + J [ms]  2nd packet 33.34 + J [ms]  3rd packet 50.00 + J [ms]  ……  k-th packet (k/X)\*1000 + J [ms]  Random variable J is considered the network delay jitter, which is the transport delay with variation. The value J would never be negative. If J is negative, it implies that the transport delay is negative and the packet arrived at the gNB is faster than the packet generated from XR source.  We don’t think the current formula is correct. |
| OPPO | According to SA4 input (copied as below for reference), encoder pre-delay is varying between 10 to 20ms. Thus, the range should be [-5, 5], rather than [-4, 4] |
| Ericsson | Support. Looks like there is a typo in the expression: range should be [-4,4]  We would also be ok with the values   * + - Mean: 4 ms     - STD: 2 ms     - Range: [0, 8] ms   The simulation results would be identicial. |
| Xiaomi | We are fine with FL proposal. |
| vivo | Fine with the proposal. |
| MTK | We observed larger jitter values in real field than [-4, 4] (in Google Stadia), where the jitter range can be up to 32ms ([-16,16]). However, we can accept current FL proposal to assist progress with the following modification to clarify how the [-4, 4] range is set:   * + Range: [4, 4] ms     - Note: The values are set to ensure that packet arrivals are in order (i.e., arrival time of a next packet is always larger than that of the previous packet), rather than real measurement   Other values can be optionally evaluated |
| Huawei, HiSilicon | We are fine with the proposal. |
| Nokia, NSB | If we consider the variation of inter-arrival time between two consecutive frames as per SA4 traces (please, refer to R1-2102827), we observe the jitter is larger than the proposed values. Therefore, by analysing the traces, we suggested the range [-6, 6] ms. However, for the sake of progress, we can accept the values proposed by the moderator if that is the majority view. |
| ZTE | OK for evaluation but not for potential enhancement. Jittering stats should be derived according to the SA raw data or application information. We believe a larger range than the current WA std and range needs to be evaluated. Any value of the max min of absolute values 16-30 ms is acceptable for us. For the STD, any value within the range of (4,10) should reflect the data the application. In this sense, MTK’ s suggestion on the note should be captured. |
| Sony | Fine with FL proposal and with the correction on jitter range [-4,4]. |
| LG | Even if the range suggested by OPPO seems to make more sense as we agreed to derive our model based on SA4 input, but we are okay with the Moderator proposal if it is a majority view. |
| QC | We are fine with confirming with jitter range [-4,4]. |
| InterDigital | We are ok with FL’s proposal with jitter range [-4,4] |
| Samsung | Fine with the proposal to confirm the values from RAN1#104-e |
| AT&T | Agree with Ericsson’s proposal. We prefer a non-negative jitter value to be modeled. |
| Intel | We should not rush into this. We believe stats should be derived from SA4 data (not from handwaving arguments) and would be better if more time is avaialble untill SA4 data is available for various use-cases. |
| DOCOMO | We think OPPO has a valid point so that jitter range should be [-5,5]. However, we can accept [-4, 4] for the sake of progress. |

1. **DL Per UE KPI (Baseline): Definition of whether each UE is satisfied or not in case of single DL stream per UE.**

RAN1#104-e agreement

|  |
| --- |
| * Baseline: A UE is declared a satisfied UE if more than X (%) of packets are successfully transmitted within a given air interface PDB.   + The exact value of X is FFS, e.g., 99, 95     - FFS different values for I-frame and P-frame if evaluation of them is agreed.   + Other values can be optionally evaluated |

Companies’ views in RAN1#104bis-e tdocs are presented in the table below.

|  |  |
| --- | --- |
| Huawei | X=99, 95, <95 |
| OPPO | For each identified traffic/service, whether a UE is satisfied or not is not determined based on the following tuples, where the detailed values is to be determined based on the traffic models.  DL: {Data rate, Packet Delay Budget, Packet Error Rate}  UL: {Data rate, Packet Delay Budget, Packet Error Rate} |
| vivo | 99 |
| MTK | 99 |
| Nokia | 99 |
| Ericsson | 99 |
| Xiaomi | 99.9 |
| Qualcomm | 99 |
| Samsung | 99 |
| ZTE | 99 |
| LGE | 95 |
| InterDigital | For a given XR/CG applications, employ link performance metrics such as user DL/UL throughput to determine a user’s experience satisfaction by the percentage (70%, 90%) of achieved maximum throughput |
|  |  |

**Summary**:

* X = 99 (7 companies)
* X = 95 (1 company)
* X = 99.9 (1 company)
* There are some other views.

1. **Based on the discussions and proposals in tdocs for RAN1#104bis-e, the moderator makes the following proposal. Please share your comments.**

**Moderator proposal:**

* In case of single stream per UE in DL, a UE is declared a satisfied UE if more than X (%) of packets are successfully delivered within a given air interface PDB.
  + The baseline X value is 99.
    - Other values can be optionally evaluated
    - FFS different values for I-frame and P-frame if evaluation of them is agreed.

|  |  |
| --- | --- |
| **Company** | **Comment** |
| FUTUREWEI | Agree with proposal |
| CATT | We are OK with the assumption of X=99 to achieve 1% frame error rate. However, the higher layer error control protocol, such as RLC AM and TCP, could achieve 10-4 packet error rate with MAC layer frame error rate higher than 1%. Thus, we should allow X <=95 being evaluated as optional parameters. |
| OPPO | Support |
| Ericsson | Support |
| Xiaomi | Although we think higher layer error control is not effective considering the short delay constraints, we can accept FL proposal considering the potential evaluation complexity. |
| vivo | Agree with the proposal |
| MTK | We can accept FL proposal. In the meantime, after the PDB/PER and file size statistical values are finalized in RAN1, we suggest to send an LS to SA4 so they can progress their work (Ex. Develop a quality evaluation model based on statistical models and PDB/PER setting) and provide further feedback to RAN1. |
| Huawei, HiSilicon | The user experience is a key characteristic of XR/CG services compared with URLLC services. In realistic XR/CG services, there are multiple user experience levels, depending on the network transmission quality, etc. This is also reflected in SA4 outcome. For example, in the table under Issue 7, the E2E Latency requirement includes multiple values (e.g., 100ms, 200ms) instead of one single value.  Higher packet success rate X% and smaller PDB lead to better user experience. Therefore, RAN1 needs to evaluate multiple combinations of (PSR, PDB) to reflect multiple user experience levels so that the SI’s outcome is close to real applications and more informative. In addition, according to our initial simulation results (R1-2102322 section 3.1.1.1), the user experience levels (i.e., different PSR/PDB values) will heavily impact the network capacity.  If RAN1 just agrees X=99 as a baseline, and let companies optionally choose other values, RAN1 may face the following issues:   * The physical meaning of X=99 is unclear, e.g., why RAN1 chooses such value, what’s the corresponding user experience level. * If there is no principle/guideline on choosing optional (PSR, PDB) values, there could be too many combinations, resulting in large simulation workload and companies’ results may not be comparable. And the user experience level of each (PSR, PDB) combination is still unclear.   So we suggest RAN1 to pick a small, limited number of (PSR, PDB) values, to reflect different user experience levels.  For example, it seems most companies think PSR X=99% can be one value. And RAN1 also agreed PDB=10ms (VR/AR) as a baseline. Thus, it seems RAN1 implicitly assume (X=99, PDB=10ms) can lead to an acceptable user experience level. Based on this, we suggest to further evaluate the combinations of (PSR, PDB) in Index#1, #3 of the following proposal to reflect increased and decreased user experience levels, respectively.  **Revised proposal:**   * **In case of single stream per UE in DL, a UE is declared a satisfied UE if more than X (%) of packets are successfully delivered within a given air interface PDB.**    + **~~The baseline X value is 99.~~**   + **The following (X, PDB) combinations are prioritized for evaluation:**     - **Index#1 (optional): X=99, PDB=7 ms (VR/AR), PDB=12 ms (CG)**     - **Index#2 (baseline): X=99, PDB=10 ms (VR/AR), PDB=15 ms (CG)**     - **Index#3 (optional): X=95, PDB=13 ms (VR/AR), PDB=18 ms (CG)**     - **RAN1 assumes the corresponding XR quality index #1, #2, #3 reflect user experience levels in decreasing order**     - **Other values can be optionally evaluated**     - **FFS different values for I-frame and P-frame if evaluation of them is agreed.** |
| Nokia, NSB | We agree with the proposal |
| ZTE | ok |
| Sony | Support |
| LG | We are okay with the Moderator proposal especially given the SA4 input on the new 5QIs. |
| QC | We agree the proposal. |
| InterDigital | We are Ok with FL’s proposal to use X=99% as baseline |
| Samsung | OK with the 99% value. Somewhat smaller values (e.g. 95%) should also be considered. |
| AT&T | We can accept 99% as a compromise, however for certain applications (e.g. AR) we believe 99.9% may better reflect the actual requirements and could be evaluated optionally. |
| Intel | This is okay, we should clarify this is for a single video stream with no separation of I-frame, P-frame etc. in the main bullet. |
| DOCOMO | Support |

|  |
| --- |
| **DL Per UE KPI (Baseline): Definition of whether each UE is satisfied or not in case of single DL stream per UE**  **Moderator proposal for 1st round of email discussion:**   * In case of single stream per UE in DL, a UE is declared a satisfied UE if more than X (%) of packets are successfully delivered within a given air interface PDB.   + The baseline X value is 99.     - Other values can be optionally evaluated     - FFS different values for I-frame and P-frame if evaluation of them is agreed   **Observation from 1st round of email discussion**  The moderator proposal is supported by all companies w/ some comments: FUTUREWEI, CATT (in addition, optionally evaluate X <= 95), OPPO, Support, Xiaomi, vivo, MTK (Send LS to SA4), HW (in addition, optionally evaluate additional (X, PDB) values, (99, 7) for VR/AR, (95, 13) for VR/AR, (99, 12) for CG, (95, 18) for CG), Nokia, ZTE, Sony, LG, QC, InterDigital, Samsung (in addition, optionally evaluate smaller X values, e.g., 95), AT&T (in addition, optionally evaluate X=99.9), Intel, DCM  **New moderator proposal**:  In case of single stream per UE in DL, a UE is declared a satisfied UE if more than X (%) of packets are successfully delivered within a given air interface PDB.   * The baseline X value is 99. * Other values of X can be optionally evaluated, e.g., X < = 95, 99.9. * Additional combinations of (X, PDB) values can be optionally evaluated, e.g.,   + (99, 7), (95, 13) for VR/AR   + (99, 12), (95, 18) for CG * FFS different values for I-frame and P-frame if evaluation of them is agreed. |

1. **DL Per UE KPI (Optional): Definition of whether each UE is satisfied or not in case of single DL stream per UE.**

RAN1#104-e agreement

|  |
| --- |
| * Baseline: A UE is declared a satisfied UE if more than X (%) of packets are successfully delivered within a given air interface PDB.   + The exact value of X is FFS, e.g., 99, 95     - FFS different values for I-frame and P-frame if evaluation of them is agreed.   + Other values can be optionally evaluated * Per UE KPI   + FFS: In addition to the baseline, the following additional method is FFS     - When determining a XR/CG user is satisfied or not, the following factors are considered. FFS how to use those factors.       * Packet loss information       * Packet delay information       * Some XR/CG source related information if they can be available within RAN, e.g. the mapping between packet and slices or frames and the packet importance   + Multiple data streams traffic model |

Companies’ views in RAN1#104bis-e tdocs are presented in the table below.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Huawei | Proposal 2: XR Quality Index (XQI) is defined to reflect the impact of network transmission on user experience in XR and CG services.  Observation 1: In real XR/CG applications, there could be multiple user experience levels, depending on the network transmission quality, etc. Therefore, evaluating a single combination of (PSR, PDB) is not enough since RAN1 does not clearly know its physical meaning and the user experience level it corresponds to.  Proposal 3: RAN1 evaluates multiple combinations of (PSR, PDB) to reflect multiple user experience levels, so that the SI’s outcome is close to real applications and more informative.  Observation 2: If there is no principle/guideline on choosing (PSR, PDB) values, there could be too many combinations and face the following issues:   * Too many combinations will result in large simulation workload * If companies choose quite different values to evaluate, their results are not comparable * The physical meaning and user experience level of each (PSR, PDB) combination is still unclear   Proposal 4: RAN1 discusses and agrees on multiple typical combinations of (PSR, PDB), wherein each combination represents one user experience level. Such combinations of (PSR, PDB) are prioritized in RAN1 evaluations.  Proposal 5: RAN1 agrees on the following table for evaluating multiple combinations of (PSR, PDB):   * The detailed values of packet success rate and PDB of each level will be separately discussed  |  |  |  | | --- | --- | --- | | XR Quality Index (XQI) | Description | (Packet success rate X%, PDB (ms)) | | 5 | Excellent | (X1, T1) | | 4 | Good | (X2, T2) | | 3 | Fair | (X3, T3) | | 2 | Poor | (X4, T4) | | 1 | Bad | (X5, T5) |   Proposal 6: For VR/AR DL video, RAN1 agrees on the following Table 3 for evaluating multiple combinations of (PSR, PDB);  Table 3. XQI table for VR/AR DL video   |  |  |  |  | | --- | --- | --- | --- | | XR Quality Index (XQI) | Description | (Packet success rate X%, PDB (ms)) | | | Single-stream | Multi-stream  {I-stream, P-stream} | | 5 | Excellent | (99, 7) | { (99.5, 7), (95, 7) } | | 4 | Good | (99, 10) | { (99.5, 10), (95, 10) } | | 3 | Fair | (95, 13) | { (95.5, 13), (90, 13) } | | 2 | Poor | (95, 20) | { (95.5, 20), (90, 20) } | | 1 | Bad | (X <95, or PDB>20) | { (X <95.5, or PDB>20),  or (X <90, or PDB>20) } |   Proposal 7: For CG DL video, RAN1 agrees on the following Table 5 for evaluating multiple combinations of (PSR, PDB):  Table 4. XQI table for CG DL video   |  |  |  |  | | --- | --- | --- | --- | | XR Quality Index (XQI) | Description | (Packet success rate X%, PDB (ms)) | | | Single-stream | Multi-stream  {I-stream, P-stream} | | 5 | Excellent | (99, 12) | { (99.5, 12), (95, 12) } | | 4 | Good | (99, 15) | { (99.5, 15), (95, 15) } | | 3 | Fair | (95, 18) | { (95.5, 18), (90, 18) } | | 2 | Poor | (95, 25) | { (95.5, 25), (90, 25) } | | 1 | Bad | (X <95, or PDB>25) | { (X <95.5, or PDB>25), or  (X <90, or PDB>25) } | |

1. **Please share your comments on additional per UE KPI, XR Quality Index (XQI) proposed by Huawei.**

|  |  |
| --- | --- |
| **Company** | **Comment** |
| CATT | Company could provide optional KPI for evaluation of technologies |
| Ericsson | Do not support as an optional KPI. Companies are still free to contribute, of course. |
| Xiaomi | We think SA4 input is necessary on this issue. |
| vivo | We suggest to discuss single-stream PER and PDB requirements firstly, then extend to multi-stream case. Regarding to multiple user experience levels, we are wondering whether we can change those requirements or not for a single QoS flow from RAN perspective. Consulting SA2/4 for more information by sending LS might be needed. |
| MTK | We are fine with Huawei’s suggestion to add additional optional KPIs. In the meantime, we suggest to send an LS to SA4 so they can progress their work (Ex. Develop a quality evaluation model based on statistical models and PDB/PER setting) and provide further feedback to RAN1. |
| Huawei, HiSilicon | The key point we think is needed for the SI is that RAN1 should evaluate multiple combinations of (PSR, PDB) to reflect different user experience levels, so that RAN1’s evaluation results can be more informative in demonstrating the feasibility of supporting XR. A single, spot value provides a limited investigation.  Such evaluations give a clear view of multiple combinations of (PSR, PDB) and their corresponding user experience levels, and can facilitate RAN1’s discussion in the future.  It is worth noting that the XQI table is proposed based on ideas similar to ITU mean opinion score (MOS) method, which uses a rating scale (value 1~5) to reflect different user experience levels as below. More details can be found in our Tdoc R1-2102321 section 4.2.2.  Table 1. Different user experience levels of video and audio in ITU MOS   |  |  | | --- | --- | | Absolute category rating scale value | Description | | 5 | Excellent | | 4 | Good | | 3 | Fair | | 2 | Poor | | 1 | Bad | |
| Nokia, NSB | We do not support it as an optional |
| ZTE | Better to involve SA for this XQI definition and categorization. |
| Sony | Do not support.  The XR Quality Index suggested by Huawei is problematic. It infers that user experience has been properly evaluated. The table is a list of different levels of quality of service based on two QoS parameters but there should not be any suggestion that the quality of experience QoE has been validated. The risk of such misunderstanding is evident due to writing "Excellent, Good, Fair, Poor, Bad" and index from 5 to 1 as this happens to be the scale defined by ITU-T in several QoE evaluation methodologies. The risk of misunderstanding is very high. We suggest to do as suggested by SA4; give the simulation results to SA4 and let them evaluate for what scenarios the user experience is adequate. It is within the mandate of SA4. |
| LG | Our suggestion would be to trigger the discussion in the SA4 as it involves the new QIs relevant for XR applications. For our discussion on the values of PER and PDB, we basically rely on the input from SA4. We don’t prefer to create the QIs in RAN1 based on our needs. |
| QC | We share the similar view with Sony. The mapping between scores and (PER, PDB) are a bit arbitrary; not based on actual measurements / survey. Thus, from RAN1 point of view, it is not clear what we can get from there. It is not different from evaluating with multiple different X values and PDB values, which can be done as optional evaluation. |
| InterDigital | We do not see much benefit in introducing the XQI as a UE KPI in addition to the previously discussed per-UE KPI for single stream of X=99% and PDB=10ms (VR/AR) or PDB=15ms (CG). For minimizing the number of evaluation combinations for the X%:PDB values, we think the previously discussed values (i.e. X=99 and PDB=10/15 ms) can be retained as baseline for capturing per-UE performance. However, companies can still show the per-UE performance using other values. |
| Samsung | Agree with Sony/Qualcomm. Also, the more the variables, the less likely it would be to converge or to have sufficient diversity/number for the evaluations when all parameters are concluded. |
| AT&T | We don’t believe this is a necessary metric as an input to the SI and can instead be developed as an outcome or recommendation once evaluations have been made. |
| Intel | we don’t mind additional KPIs but we have similar concerns as Sony, unless we do a thorough job, it may provide a wrong perception that user experience is properly evaluated. It is also difficult to see how RAN1 can do a good job coming up with a perceptual metric for video. |
| DOCOMO | We don’t think it should be optional, while company can still provide additional KPIs. |

|  |
| --- |
| **Question for 1st round of email discussion**   * Please share your comments on additional per UE KPI, XR Quality Index (XQI) proposed by Huawei   **Summary of comments on the moderator proposal for 1st round of email discussion**   * Support: HW, MTK * Not supportive: CATT, Ericsson, Xiaomi, vivo, Nokia, ZTE, Sony, LG, QC, InterDigital, Samsung, AT&T, Intel, Docomo   **New moderator proposal**  We do not discuss this further. Instead, this has been reflected with modifications in the following RAN1 agreement on per UE KPI.  RAN1 agreement  In case of single stream per UE in DL, a UE is declared a satisfied UE if more than X (%) of packets are successfully delivered within a given air interface PDB.   * The baseline X value is 99. * Other values of X can be optionally evaluated, e.g., X < = 95, X=99.9. * Additional combinations of (X, PDB) values can be optionally evaluated, e.g.,   + (99, 7), (95, 13) for VR/AR   + (99, 12), (95, 18) for CG * FFS: Different values for I-frame and P-frame if evaluation of them is agreed. |

## DL Traffic Model: Multiple Streams

1. **DL multiple streams per UE.**

RAN1#104-e agreement

|  |
| --- |
| * FFS if there are multiple streams (if adopted) * FFS the following in RAN1#104-bis-e   + Whether/how to model and evaluate I-frame and P-frame for both DL and UL, e.g., separate definition of fps, packet size, QoS requirements (e.g., PER, PDB), etc.   + Whether/how to separately model and evaluate two streams of video and audio/data for both DL and UL   + Whether/how to model and evaluate FOV (high-resolution) and non-FOV (lower-resolution omnidirectional) streams, e.g., separate definition of fps, packet size, QoS requirements (e.g., PER, PDB), etc |

Companies’ views in RAN1#104bis-e tdocs are presented in the table below.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Huawei | Proposal 4: For DL video of AR/VR/CG, adopt M1=2 for modelling I-frame and P-frame separately, and adopt the multi-stream traffic model in following Table 6.  Table 6. Multi-stream model for DL video   |  |  |  | | --- | --- | --- | | Application | AR/VR/CG | | | Two data streams, i.e. M1 = 2 | * Stream #1: I-stream * Stream #2: P-stream | | | Option 1: slice-based | Option 2: frame-based | | Packet modelling | Slice-level | Frame-level | | Traffic arrival pattern | Both streams are periodic with the same FPS. | Follow the GOP structure, e.g. GOP size K = 8. | | Number of packets per stream at a time | * Stream #1: 1 * Stream #2: N-1 * N is the number of slice per frame, e.g. N = 8. | * Stream #1: 1 * Stream #2: 1 | | Average data rate | Stream #1: Stream #2 = : (N-1) | Stream #1: Stream #2 = : (K-1) | | is average size ratio between one I-frame/slice and one P-frame/slice, e.g. = 2.   * Other values can be optionally evaluated. | |   Note: the QoS requirement for each stream is separately discussed in the KPI part |
| OPPO | Not support to model multiple streams |
| vivo | Proposal 5: For multiple streams XR traffic model in DL, GOP-based/slice-based multiple streams traffic model in Table 2/Table 3 can be considered.  Table 2. GOP-based multiple streams traffic model (FPS=60)   |  |  |  |  | | --- | --- | --- | --- | | **Traffic model** | I-frame | P-frame | Note | | **Date rate (Mbps)** |  |  | GOP length = 1 second | | **Packet size distribution** | Truncated Gaussian distribution | |  | | **Mean packet size (Bytes)** | 272177 | 90725 | The average ratio of I-frame size and P-frame size is around 3:1 | | **STD of packet sizes (Bytes)** | 40826 | 13608 | 15% of Mean packet size | | **Maximum packet size (Bytes)** | 408265 | 136087 | 1.5 \* Mean packet size | | **Minimum packet size (Bytes)** | 900 | 300 |  | | **Packet arrival interval (ms)** | 1000 |  | 1 I-frame and 59 P-frames in one second | | **Packet delay budget (ms)** | 10 | |  |   Table 3. Slice-based multiple streams traffic model (FPS=60)   |  |  |  |  | | --- | --- | --- | --- | | **Traffic model** | I-frame | P-frame | Note | | **Date rate (Mbps)** |  |  |  | | **Packet size distribution** | Truncated Gaussian distribution | |  | | **Mean packet size (Bytes)** | 28125 | 65625 | The average ratio of I-slice size and P-slice size is around 3:1, and each encoded video frame contains 1 I-slice and 7 P-slices | | **STD of packet sizes (Bytes)** | 4218 | 9843 | 15% of Mean packet size | | **Maximum packet size (Bytes)** | 42187 | 98437 | 1.5 \* Mean packet size | | **Minimum packet size (Bytes)** | 93 | 217 |  | | **Packet arrival interval (ms)** |  |  | each encoded video frame contains 1 I-slice and 7 P-slices | | **Packet delay budget (ms)** | 10 | |  |   Proposal 2: For a given data rate, single stream with two-eye buffers can be modelled as:   * Model 1: each packet representing both eyes buffers arrives at the same time at X FPS and the sum of packet size for both eyes is equal to the size of a packet in simulation. * Model 2: packet representing left or right eye buffer arrives at 2\*X FPS and the packet size of left or right eye is the size of a packet in simulation. |
| MTK | Adopt the IDR refresh model for both UL/DL videos for RAN1 evaluation considering its low encoding complexity and wide usage in current industry (e.g. Google Stadia, Nvidia Geforce Now)  Proposal 8: Discuss two possible options:   1. FoV vs. non-FoV 2. FoV vs. low resolution Omnidirectional stream  * FFS: Need for different QoS requirements for the two streams.   FFS: co-existence with the QoS requirements for I/P-frames  Adopt Tp as PDB for P-Frames and Ti as PDB for I-frames with Tp < Ti.  FFS Tp = 8 ms and Ti = 12 ms |
| Nokia | Proposal 5: Consider a single stream in downlink and single stream in uplink for VR1 and VR2 applications as a baseline.  Proposal 6: Consider a signle stream in downlink and a single stream in uplink for CG application as a baseline.  Proposal 8: Following SA4 input, consider no differentiation between the types of packets/frames as well as FOV/non-FOV as the baseline evaluation of XR/CG applications. |
| Ericsson | [Proposal 5 RAN1 should not model and evaluate I-frame and P-frame separately which will require introducing new traffic parameters.](#_Toc68631143)  [Proposal 6 RAN1 should avoid including multiple streams caused by a frame type, voice traffic, and non-FoV which will increase traffic modelling complexity and evaluation options.](#_Toc68631144) |
| Apple | Two streams (scene/video + audio/data) for downlink for evaluation on AR2. The audio/data flow is modeled as:   * Periodic:   + 10 milliseconds for framing (SA4 input: 10 ms for data stream and 20 ms for audio) * Data rate   + 0.756 Mbps/s or 1.12 Mbps (SA4 input: 256/512 Kbps for audio, 0.5 Mbps for data) * Packet size: constant packet size calculated from periodicity and data rate   End-to-end (mouth-to-ear) latency: 100 ms (SA4 input: 100 ms for both data and audio stream), air interface latency: 30 ms |
| Xiaomi | In XR evaluation, do not differentiate I-frame stream and P-frame stream in DL video stream. |
| Samsung | If the P-frame needs to have smaller PER or PDB that the I-frame, separate models can be defined; otherwise, a single model for the I-frame is used. |
| LGE | Multi-stream is not further considered in DL for VR1/VR2/CR/AR1/AR2 applications |
| AT&T | In addition to M1=1 and M2=1 streams, support M1=2 and M2=2, where in both the DL and UL a user has one video stream based on a Truncated Gaussian packet size distribution and one data/control stream based on a fixed packet size and inter-arrival time |

**Summary**:

* Five companies propose not to evaluate multiple streams per UE in DL.
* Five companies are open to evaluate multiple streams

1. **Based on the discussions and proposals in tdocs for RAN1#104bis-e, the moderator makes the following proposal. Please share your comments.**

**Moderator proposal:**

* In addition to single stream per UE in DL which is baseline, two streams are optionally evaluated.
* FFS details of traffic model, KPI per stream, and per UE KPI.

|  |  |
| --- | --- |
| **Company** | **Comment** |
| FUTUREWEI | Suggest that the single stream is baseline while the two streams evaluation in DL is Optional while adding that the FFS details are specifically for the optional scheme as an example  Baseline: Single stream per UE in DL  Optional: Two streams are evaluated   * + FFS details for Traffic model, KPIs etc. |
| CATT | We support moderator’s proposal. Two stream evaluation models would only be useful when two streams do not multiplex for transport between the XR server to the gNB. Two stream model also require CORE network to set the quality index and gNB to set different priority index for scheduling since gNB scheduler would not open the packet to be scheduled for identifying different streams. |
| OPPO | We prefer only single stream. However, for the sake of progress, we can support two streams as optional |
| Xiaomi | We prefer only single video stream for DL. We are not sure whether I-frame and P-frame can be differentiated from RAN aspect. Also it is not clear on how to set different QoS requirements for them. Note that in P-trace file from SA4, the “importance” parameter of P-frames are quite diverse, which can be from 1 to 7. We can accept FL proposal to leave it as FFS. |
| vivo | Fine with the proposal. |
| MTK | We support moderator’s proposal. We also suggest to first agree the statistical values of packet size derived by vivo in Table 2 and Table 3 which are based on SA4 inputs. The PDB and PER determination of multi-streams can be further discussed. |
| Huawei, HiSilicon | In summary, modelling multi-stream not only reflects realistic XR/CG applications, but also helps the TR better report the network transmission performance and provide insights on further enhancements. If RAN1 only evaluates single-stream model, it seems RAN1 is only evaluating URLLC-like traffic with different data rate and QoS requirements. So multi-stream model is a key differentiator between XR/CG traffic and URLLC traffic, and can make the SI’s outcome more informative since some key characteristics of XR/CG traffic are considered.  In realistic XR/CG applications, there are multiple data streams (e.g., I/P frame, FOV/omnidirectional, video/audio, etc.). This is also reflected in the SA4 outcome. For example, SA4 already provided the traces for VR2-1~VR2-6 [1] (see S4aV200634 clause 4.2.1), which includes I/P frame/slice. ([1] <http://dash.akamaized.net/WAVE/3GPP/XRTraffic/Traces/Candidate/VR2>).  Therefore, our 1st priority is to make multi-stream models as mandatory. However, if some companies think this may increase simulation workload and have strong concerns here, for the sake of progress, we might ok to make multi-stream models as optional.  In addition, several companies including Huawei, HiSilicon have already proposed detailed models for I/P frame based on SA4 outcomes. So we also suggest the following Proposal#2 to give more details on “Option 1: I-frame + P-frame” of multi-stream model.  Suggested proposals are:  **Proposal#1:**  **In addition to single stream per UE in DL which is baseline, two streams are optionally evaluated with the following options.**   * **Option 1: I-frame + P-frame** * **Option 2:** **video + audio/data** * **Option 3:** **FOV + omnidirectional stream** * **Note: Other options are not precluded** * **Note:** **For each option above, RAN1 strives to agree on the details of traffic model, KPIs, etc., during RAN1#104b-e.**   **Proposal#2:**  **For “Option 1: I-frame + P-frame” of multi-stream model, adopt the following table for modelling I-frame and P-frame separately.**  Table. Two-stream model for video   |  |  |  | | --- | --- | --- | | **Application** | **AR/VR/CG** | | | **Two data streams, i.e. M1 = 2** | * **Stream #1: I-stream** * **Stream #2: P-stream** | | | **Option 1A: slice-based** | **Option 1B: frame-based** | | **Packet modelling** | **Slice-level** | **Frame-level** | | **Traffic arrival pattern** | **Both streams are periodic with the same FPS.** | **Follow the GOP structure, e.g. GOP size K = 8, 60** | | **Number of packets per stream at a time** | * **Stream #1: 1** * **Stream #2: N-1 (N is the number of slice per frame, e.g. N = 8)** | * **Stream #1: 1** * **Stream #2: 1** | | **Average data rate** | **Stream #1: Stream #2 = : (N-1)** | **Stream #1: Stream #2 = : (K-1)** | | **is average size ratio between one I-frame/slice and one P-frame/slice, e.g. = 2. Other values can be optionally evaluated.** | |   **Note: the QoS requirement for each stream is separately discussed in the KPI part** |
| Nokia, NSB | We support single stream per UE in DL as a baseline. We do not support any additional streams per UE as optional for DL. From SA4 input, it is stated that single stream per DL is assumed (at least for VR and CG). Please, find the copied text as a reference (VR2 from SA4):   1. Content Model:    1. Rendered scene output with 2 eye buffers at 2Kx2K at 60 fps, 8bit.    2. Content and Trace Preview is here: <http://dash.akamaized.net/WAVE/3GPP/XRTraffic/Traces/Qualcomm-VR2>   No audio (considered small, could be added) according to [S4aV200626](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/3GPP_SA4_AHOC_MTGs/SA4_VIDEO/Docs/S4aV200626.zip), clause 7.2.12. |
| ZTE | We are fine with moderator’s proposal. Two views we would like to share as follow:   1. From the simulation’s perspective, I/P frame modelling for video stream use cases are regarded as the first priority when considering two streams evaluations in DL, since other two-stream models, i.e. FoV and non-FoV, are similar with the I/P frame modelling. 2. From the product implementation’s perspective, it is hard for PHY to distinguish two different streams with different QoS requirements, i.e. PER, PDB and etc, according to current 5QI. Therefore, we prefer SA first settled the 5QI values corresponding to streams with different QoS requirements, i.e. whether the differentiation lies in one or both of PER and PDB. |
| LG | We support the Moderator proposal. |
| QC | We prefer single stream evaluation. But, if companies want to evaluate multiple streams, then, they can do it optionally. We think those companies should provide rational for different quality requirements for different streams with their results. |
| InterDigital | We think evaluating 2 streams in DL, where different streams (e.g. control traffic and video) use separate traffic models and have different KPIs/QoS requirements, is beneficial for showing realistic scenarios for CR/VR/AR. However, for progress sake we are ok with FL’s proposal. |
| Samsung | It is ACKed that there are cases for two streams. However, we support the moderator’s proposal as we don’t expect materially different conclusions vs. the single-stream case (even when the most stringent requirements are mixed between two streams) and the benefit of the proposal is obvious. |
| AT&T | We support the optional modelling of two streams for the DL. We believe I-frame + P-frame can be the starting point for defining the traffic model characteristics and QoS requirements for each stream |
| Intel | We are fine to model multi-stream. |
| Apple | We continue to support the study of multiple flows. In R1-2103833, we studied 3 traffic models for DL:   * + - 1. 2 flows (video stream + audio/data stream)       2. 2 flows (video stream + audio/data stream) with merged traffic for two flows       3. A single flow (video)   We saw difference among them, using a single flow won’t reveal complications for MCS selection and scheduling, etc.  For the data flows, besides the video stream, data/audio stream can be modelled as:   * Periodic:   + 10 milliseconds for framing (SA4 input: 10 ms for data stream and 20 ms for audio) * Data rate   + 0.756 Mbps/s or 1.12 Mbps (SA4 input: 256/512 Kbps for audio, 0.5 Mbps for data) * Packet size: constant packet size calculated from periodicity and data rate   End-to-end (mouth-to-ear) latency: 100 ms (SA4 input: 100 ms for both data and audio stream), air interface latency: 30 ms |
| DOCOMO | We are fine with the FL proposal. |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DL multi-streams per UE**  **Moderator proposal for 1st round of email discussion**   * In addition to single stream per UE in DL which is baseline, two streams are optionally evaluated. * FFS details of traffic model, KPI per stream, and per UE KPI   **Summary of comments on the moderator proposal for 1st round of email discussion**   * Support optional evaluation of two streams: FUTUREWEI, CATT, OPPO, vivo, MTK, HW, ZTE, LG, QC, InterDigital, Samsung, AT&T, Intel, Apple, DCM * No evaluation of two streams: Xiaomi, Nokia   **New moderator proposal**  Given the clear majority view, RAN1 confirm that two streams in DL are optionally evaluated.  **Q1 for round 2 email discussion**  Assuming that two streams is optionally evaluated for DL, please provide your comments on the following options. All of the options may not have the same level of interest/support/priority.   * Option 1: I-frame + P-frame * Option 2: video + audio/data * Option 3: FOV + omnidirectional stream   **Q2 for round 2 email discussion**  If two streams of I-frame and P-frame are optionally evaluated for DL, please provide your view on how to evaluate them, i.e.,  traffic model, KPI per stream, and per UE KPI. Below is one proposal on the table. Please provide your comment on this proposal if any.  Table. Two-stream model for video   |  |  |  | | --- | --- | --- | | Application | AR/VR/CG | | | Two data streams, i.e. M1 = 2 | * Stream #1: I-stream * Stream #2: P-stream | | | Option 1A: slice-based | Option 1B: frame-based | | Packet modelling | Slice-level | Frame-level | | Traffic arrival pattern | Both streams are periodic with the same FPS. | Follow the GOP structure, e.g. GOP size K = 8, 60 | | Number of packets per stream at a time | * Stream #1: 1 * Stream #2: N-1 (N is the number of slice per frame, e.g. N = 8) | * Stream #1: 1 * Stream #2: 1 | | Average data rate | Stream #1: Stream #2 = : (N-1) | Stream #1: Stream #2 = : (K-1) | | is average size ratio between one I-frame/slice and one P-frame/slice, e.g. = 2. Other values can be optionally evaluated. | |   Note: the QoS requirement for each stream is separately discussed in the KPI part |

## UL CG/VR

1. **WA for UL CG/VR**

RAN1#104-e agreement

|  |
| --- |
| Working assumption: On UL Traffic model and QoS parameters   * CG/VR: single stream (pose/control) * Traffic model for Pose/control   + Periodic: 4ms (no jitter)     - Other values can be optionally evaluated.   + Fixed: 100 bytes (SA4 input)   + PDB: 10 ms |

Companies’ views in RAN1#104bis-e tdocs are presented in the table below.

|  |  |
| --- | --- |
| Huawei | Confirm the WA. |
| OPPO | Confirm the WA. |
| Vivo | Confirm the WA. |
| Xiaomi | Confirm the WA. |
| Qualcomm | Confirm the WA. |
| ZTE | Confirm the WA. |
| DOCOMO | Confirm the WA. |
| MTK | Confirm the WA. |

**Summary**:

* Seven companies propose to confirm the above WA.

1. **Based on the discussions and proposals in tdocs for RAN1#104bis-e, the moderator makes the following proposal. Please share your comments.**

**Moderator proposal:** On UL Traffic model and QoS parameters

* CG/VR: single stream (pose/control)
* Traffic model for Pose/control
  + Periodic: 4ms (no jitter)
    - Other values can be optionally evaluated.
  + Fixed: 100 bytes (SA4 input)
    - PDB: 10 ms.
* A UE is declared a satisfied UE if more than X (%) of packets are successfully delivered within the given air interface PDB.
  + The baseline X value is 99.
  + Other values can be optionally evaluated

|  |  |
| --- | --- |
| **Company** | **Comment** |
| FUTURWEI | OK with proposal. One suggestion may be that the subbullet *Other values can be optionally* *evaluated* may be removed. This may help narrow down the results for calibration of results between companies. |
| CATT | OK with moderator’s proposal |
| OPPO | Support |
| Ericsson | Support |
| Xiaomi | We are fine with FL proposal. |
| vivo | Agree with the proposal. |
| MTK | Support. We think the subbullet can be kept since it is optional. |
| Huawei, HiSilicon | We are ok with the first two main bullets.  We suggest to postpone the discussion on the 3rd main bullet since it’s related to Issue 3, 4 (i.e., X for DL). For similar issues, we suggest to first discuss on DL, and once agreements are made, we can adapt them to UL easily. |
| Nokia, NSB | Support |
| ZTE | OK |
| Sony | Support |
| LG | We are okay with the proposal. |
| QC | We support the FL proposal with following comments.  We think, in practice, X value for pose (i.e., the % of successfully transmitted packets) could be lower than 99. The reason is as follows.  In edge server, when a FL video frame is generated, the latest received pose info is more important than others since it includes latest pose/motion info which is critical for rendering DL video frame. Stale pose info (which have been received before the latest pose) might be also used, but it is not as important as the latest one. Thus, given that we have 4ms periodicity which is much shorter than 16.67ms, the X could be lower than 99. |
| InterDigital | We are ok with FL’s proposal |
| Samsung | OK to confirm the WA but as mentioned previously and in our Tdoc, a 99% target of satisfied UEs may lead to different conclusions that a somewhat more relaxed value such as 95%. It will be useful to consider additional optional values. |
| AT&T | We can accept 99% as a compromise, however for certain applications (e.g. AR) we believe 99.9% may better reflect the actual requirements and could be evaluated optionally. |
| Intel | OK, comment from HW also makes sense |
| DOCOMO | Support |

|  |
| --- |
| **On CG/VR in UL**  **Moderator proposal for 1st round of email discussion**   * CG/VR: single stream (pose/control) * Traffic model for Pose/control   + Periodic: 4ms (no jitter)     - Other values can be optionally evaluated.   + Fixed: 100 bytes (SA4 input)     - PDB: 10 ms. * A UE is declared a satisfied UE if more than X (%) of packets are successfully delivered within the given air interface PDB.   + The baseline X value is 99.   + Other values can be optionally evaluated   **Observation from 1st round of email discussion**   * Supported by all companies w/ some comments: FUTUREWEI (remove the bullet, Other values can be optionally evaluated), CATT, OPPO, Ericsson, Xiaomi, vivo, MTK (keep the bullet pointed by FUTUREWEI), HW (third main bullet should be further discussed once X value for DL is agreed), Nokia, ZTE, Sony, LG, QC (on 3rd main bullet, evaluate also smaller X values), InterDigital, Samsung (on 3rd main bullet, optionally evaluate other X values, e.g., X = 95), AT&T (on 3rd main bullet, optionally evaluate other X values, e.g., X = 99.9), Intel (same comments as HW), DCM   **New moderator proposal**:  On UL Traffic model and QoS parameters   * CG/VR: single stream (pose/control) * Traffic model for Pose/control   + Periodic: 4ms (no jitter)     - Other values can be optionally evaluated.   + Fixed: 100 bytes     - PDB: 10 ms.   + A UE is declared a satisfied UE if more than X (%) of packets are successfully delivered within the given air interface PDB.     - The baseline X value is 99.     - Other X values can be optionally evaluated, e.g., <99 including 95, 99.9. |

## UL: AR

1. **UL AR traffic model**

SA4 has discussed the multiple streams for AR applications. As shown in the following table, there are a variety of services in uplink transmission, such as pose, video, audio, data, etc., where video, camera scene, audio and data streams have similar E2E latency requirement. The required bit rate of audio and data is less than the video stream.

Uplink multiple streams for AR application

|  |  |  |
| --- | --- | --- |
| **Media** | **Format and Model** | **E2E Latency requirement** |
| **3/6DOF Pose** | **Same as for split rendering** | **UL: 5-10 ms** |
| **Video + Depth** | **1080p, Capped VBR 10/20 Mbit/s for UL** | **Conversational 100ms, 200ms** |
| **2D Video is split rendering** | **1080p or 4K (2 eyes)** | **60ms, 100ms** |
| **Front Facing Camera\*** | **720p, CBR 3 Mbit/s for UL** | **Conversational 100ms, 200ms** |
| **Audio (MPEG-H)** | **256/512 kbps for both UL/DL** | **Conversational 100ms, 200ms** |
| **Data Stream** | **0.5 Mbps for both UL/DL** | **Conversational 100ms, 200ms** |

Companies’ views in RAN1#104bis-e tdocs on UL AR traffic model are presented in the table below.

|  |  |
| --- | --- |
| Huawei | Proposal 6: There are M2=3 streams in UL traffic model of AR, where   * One stream for UL pose/control, * The model for UL pose/control of VR/CG can be reused. * Two streams for UL video to model I-stream and P-stream separately, * The multi-stream model for DL video of VR/AR/CG can be reused. |
| OPPO | One stream for video and another stream for audio or control/pose   * Pose/Control (Same as CG/VR)   + Periodic: 4ms (no jitter)   + Fixed: 100 bytes (SA4 input)   + PDB: 10 ms * UL Video stream   + 10Mbps @60fps (baseline)   + 20Mbps@60fps (optional)   + For packet size, reuse the truncated Gaussian distribution for DL packet size except the mean value is adjusted according to the data rates of UL video. Other parameters are kept the same   + For jitter, reuse the truncated Gaussian distribution for DL packet arrival jitter. All parameters are kept the same   + 10ms (baseline)     - Other values can be evaluated optionally   A UE is declared as satisfied only when all streams meets their corresponding requirements  Not support to model and evaluate I-frame and P-frame for the evaluation of XR/CG on NR |
| vivo | * Option 1: single pose stream. * Option 2: single video stream. * Option 3: two streams with pose/control and video streams   Table 5. Single stream traffic model of video in UL   |  |  |  | | --- | --- | --- | | **Traffic model** | Video | Note | | **Data rate (Mbps)** | 20 |  | | **Frame per second** | 60 |  | | **Packet size distribution** | Truncated Gaussian distribution |  | | **Mean packet size (Bytes)** | 41667 | Average data rate / FPS / 8 [Bytes] | | **STD of packet sizes (Bytes)** | 6250 | 15% of Mean packet size | | **Maximum packet size (Bytes)** | 62500 | 1.5 \* Mean packet size | | **Minimum packet size (Bytes)** | 4167 | 0.1 \* Mean packet size | | **Packet arrival interval (ms)** | 16.67 |  | | **Packet delay budget (ms)** | 60 | 60ms for AR | |
| MTK | * M2=2 or 3 for AR to model video and control/pose separately   No need to model the audio stream separately |
| Futurewei | * A single video stream for a UE: periodic with 60 fps, no jitter * Average data rate: 20 Mbps @ 60 fps (baseline) * Truncated Gaussian distribution is used for the packet size distribution of video stream for UL AR with the following parameters:   + Mean: derived from fps and average data rate   + STD: 15% of Mean packet size   + Max packet size: 1.5 x Mean packet size   + Min packet size: 0.5 x Mean packet size * PDB: 60 ms (baseline) |
| Nokia | Adopt a single stream of video in UL for AR2: XR Conversational as a baseline. The average data rate is 10 Mbit/s (1080p) and the frame rate is 60 fps. The PDB is 10 ms.  Proposal 4: No jitter is assumed for the UL video stream.  Proposal 7: Consider a single stream in downlink and a single stream in uplink for AR application as a baseline. Any additional streams consider as optional. |
| Ericsson | [Proposal 1 The bit rates for AR UL scene can be the range of 2Mbps to 20Mbps and the latency requirement is similar as DL AR/VR video, i.e., 5ms to 20ms.](#_Toc68631139)  [Proposal 5 RAN1 should not model and evaluate I-frame and P-frame separately which will require introducing new traffic parameters.](#_Toc68631143)  [Proposal 6 RAN1 should avoid including multiple streams caused by a frame type, voice traffic, and non-FoV which will increase traffic modelling complexity and evaluation options.](#_Toc68631144) |
| Xiaomi | An UL pose stream and a single UL video data stream are used as UL traffic model for AR2 use case |
| Apple | 3 streams (scene/video + audio/data + pose/control) for UL and two streams (scene/video + audio/data) for downlink can be used for evaluation on AR2. The audio/data flow is modeled as:   * Periodic:   + 10 milliseconds for framing (SA4 input: 10 ms for data stream and 20 ms for audio) * Data rate   + 0.756 Mbps/s or 1.12 Mbps (SA4 input: 256/512 Kbps for audio, 0.5 Mbps for data) * Packet size: constant packet size calculated from periodicity and data rate   End-to-end (mouth-to-ear) latency: 100 ms (SA4 input: 100 ms for both data and audio stream), air interface latency: 30 ms |
| Qualcomm | * Stream 1: pose/control (same as VR/CG) * Stream 2: aggregated stream for scene, video, data, and audio.   + Truncated Gaussian distribution for Packet size with same parameter values |
| Samsung | 2 UL streams (pose and scene upload) |
| Sony | 3 streams for UL (pose and two cameras) |
| LGE | Multi-stream is not further considered in UL for VR1/VR2/CG applications   * In the case where two streams are used for UL traffic modelling, one for control/pose and the other for video stream/scene update/audio/data, a UE is declared a satisfied UE if more than X1 (%) of packets for control/pose are successfully transmitted within a given air interface PDB1, AND more than X2 (%) of packets for video stream/scene update/audio/data are successfully transmitted within a given air interface PDB2.   + X1 is [99] and X2 is [95]   + PDB1 is 10ms and PDB2 is [100]ms   Proposal 5: UL Traffic model for video stream/scene update/audio/data   * Frame per second (fps)   + 60 fps (SA4 input) – no jitter * Average data rate for UL video stream:   + 20 Mbps (~half of the average of VR/AR average data rate for DL video stream) * Truncated Gaussian distribution is used for the packet size distribution   + Parameters of Truncated Gaussian distribution for Packet size (note: these parameter values are those before the truncation)     - Mean: Derived from average data rate and fps as follows.       * (average data rate) / (fps for video stream, i.e., # packets per second in our statistical model) / 8 [bytes]     - STD: same as in DL     - Max packet size: same as in DL     - Min packet size: FFS (need input from SA4) * PDB   + [100] ms (based on SA4 input) |
| InterDigital | 2 different traffic flows in UL as mandatory for AR evaluations. FFS for using more than 2 dfferent traffic flows in UL for AR |
| AT&T | support M2=2, where a user has one video stream based on a Truncated Gaussian packet size distribution and one data/control stream based on a fixed packet size and inter-arrival time |
| DOCOMO | * *Adopt two streams for UL for AR applications*   + *Traffic model for pose/control information can be same as CG/VR.*   + *Traffic model for scene update/video/audio data,*     - *Periodicity: 60 fps*     - *Data rate: 20 Mbps*     - *PDB: 60 ms* |

**Summary**:

* Ten companies support evaluation of two streams, one for pose/control and the other for scene update/video/audio/data
* Four companies propose to evaluate only one stream for UL AR.
* One company propose 3 streams (scene/video + audio/data + pose/control).
* One company propose 3 streams for UL (pose and two cameras)

1. **Based on the discussions and proposals in tdocs for RAN1#104bis-e, the moderator makes the following proposal. Please share your comments.**

**Moderator proposal:**

* For UL AR,
  + Baseline: two streams are defined as follows.
    - Stream 1: pose/control
      * Traffic model and QoS parameters are same as those for pose/control for UL CG/VR.
    - Stream 2: aggregated stream for scene, video, data, and audio.
      * Packet size: Truncated Gaussian distribution with the same parameter values with DL (Mean, STD, Max, Min)
      * Periodicity: 60 fps
      * Data rate: 10 Mbps (baseline), 20 Mbps (optional)
      * PDB: 60 ms
  + Single stream (Stream 2 above) and/or more than two streams can be optionally evaluated.
  + A UE is declared as satisfied only when each stream meets the following requirement.
    - In case of the above baseline (2 streams)
      * X value for Stream 1 follows that for pose/control of UL CG/VR
      * X value for Stream 2 follows that for DL video streaming for single stream case.
      * Other values can be optionally evaluated
    - In case of single stream (Stream 2 above) and/or more than two streams
      * FFS

|  |  |
| --- | --- |
| **Company** | **Comment** |
| FUTUREWEI | First in our views, the second bullet   * + Single stream (Stream 2 above) and/or more than two streams can be optionally evaluated.   may not be very clear, does that mean that the baseline is subset of this subbullet? Or stream 1 may not be included in such? The statement seems to be broad and may need further clarification. Also having such may diversify the results a lot from companies making it difficult to align the results and draw conclusions.  Second, in our views, two options may be considered based on companies views.   * + - Option 1: Stream 1: aggregated stream for scene, video, data, and audio     - Option 2: Stream 1: aggregated stream for scene, video, data, and audio and Stream 2 modelling pose/control       * FFS whether Option 1 and Option 2 is mandatory or optional   The specific TM parameters for both streams may be as FL proposed   * + - Stream 1: pose/control       * Traffic model and QoS parameters are same as those for pose/control for UL CG/VR.     - Stream 2: aggregated stream for scene, video, data, and audio.       * Packet size: Truncated Gaussian distribution with the same parameter values with DL (Mean, STD, Max, Min)       * Periodicity: 60 fps       * Data rate: 10 Mbps (baseline), 20 Mbps (optional)       * PDB: 60 ms |
| CATT | We are generally OK with the proposal. We would like to clarify how two streams are multiplexed and transmitted by PUSCH in the evaluation. |
| OPPO | The PDB should be 10 ms or 15 ms. The current value of 60 ms seem the E2E latency, rather than the latency of air interface. The following table summarizes the E2E and air-interface PDB for each services.   |  |  |  | | --- | --- | --- | |  | Maximum latency for slice (SA4) | PDB (RAN1) | | VR/AR DL | 60ms | 10ms | | CG | 80ms | 15ms | | AR UL (Video) | 80ms | ?? | |
| Ericsson | We think 2 UL streams will complicate the evaluations without any benefit. The 10Mbps stream will limit performance. Suggest evaluating only one stream. As CATT indicates, there will be more conditions to agree on, or the results will not be comparable.  We are OK with the parameters for stream 2: although it is not clear to us why the PDB for AR is so much larger than the PDD for CG, we believe that performance will be similar: there is little benefit in increasing the PDB beyond the frame arrival interval.  We believe that we should have jitter stream 2: a large part of the jitter comes from the encoding delay, which is similar in UL and DL. Propose to reuse the same jitter model as for DL. |
| Xiaomi | We are fine with FL proposal |
| vivo | Fine with the proposal.  From the perspective of capacity evaluation, stream 2 also can be evaluated as baseline for single-stream case, since stream 1 has negligible impact to capacity. But for power consumption evaluation, above baseline is fine, since stream 1 could have significant impact on power consumption. |
| MTK | We are generally fine with FL proposal while we share the same question with OPPO. The current value of 60 ms from SA4 seems to be the E2E latency. Or is there a reason that UL video latency can be much larger than DL?  For the multiple streams in UL, we think the structure of DL multiple streams can be reused (Ex. I/P frame) if agreed.  For UL video jitter, we think that can be optionally evaluated if some companies deem necessary. |
| Huawei, HiSilicon | This issue is tightly related to Issues 5, 3, 4 (i.e., multi-stream for DL, (PSR, PDB) values for DL video).  For similar issues, we suggest to first discuss on DL, and once agreements are made, we can adapt them to UL easily. For example, the I/P frame model for DL video and UL video could be very similar. So we suggest to postpone the discussion of Issue 7. |
| Nokia, NSB | We do not support the proposal. We suggest to consider single video stream for UL AR. The evaluation methodology for single stream is more or less clear among all companies. Even with that, the numerical results shown by companies demonstrate a large deviation in terms of supported UEs. If we continue to complicate the evaluation methodology the comparison among the companies will be even more challenging. Therefore, we propose to evaluate a single video stream per UL AR as a baseline. The evaluation of two streams can be considered as optional. |
| ZTE | OK |
| Sony | Similar view as OPPO. Why the PDB for UL AR is 60 ms? This is far larger than PDB for DL at 10 ms. Note: There is no definition of UL PDB. In the last meeting, we defined DL PDB. |
| LG | Support the Moderator’s proposal in general. We need a clarification on whether the 60ms for stream 2 is intended for E2E delay. And also we think the 10Mbps is rather small as a baseline for stream 2 considering that it was 30 and 45 @ 60 fps for DL video streaming. And in the third main bullet, with the understanding that “each stream meets the following requirement” means both of them shall meet their own requirements, we are okay with the third bullet. |
| QC | We support the FL proposal. The value of PDB of 60ms is because this traffic is **conversational** information which is different from **motion/tactile** information. Thus, longer PDB is fine for such type of traffic. |
| InterDigital | We are OK with FL’s proposal to evaluate 2 streams in UL for AR as baseline. For the PDB we prefer using either 10ms or 15ms as baseline for AR video. Conversational AR can be considered as optional with PDB of 60ms.  We are ok with per-UE KPI for 2 streams where UE is declared satisfied only when each stream meets its corresponding requirement. |
| Samsung | Although initially supportive of 2 streams, we currently prefer single stream for similar reasons as outlined by Ericsson and Nokia. |
| AT&T | We believe that aligning the DL and UL assumptions for data rate, jitter, and PDB is important to evaluate certain use cases (at least as one possible combination) |
| Intel | We support the proposal in general (PDB needs more discussion as pointed out), jitter due to encoding delay can be further discussed, 1 stream is an oversimplification (again our concern is unhelpful specification impact) |
| Apple | * On “Single stream (Stream 2 above) and/or more than two streams can be optionally evaluated.”, we believe it is beneficial to spell out the modelling details for 3 streams which can be optionally evaluated.   More specifically, the audio/data flow is modelled as:   * Periodic:   + 10 milliseconds for framing (SA4 input: 10 ms for data stream and 20 ms for audio) * Data rate   + 0.756 Mbps/s or 1.12 Mbps (SA4 input: 256/512 Kbps for audio, 0.5 Mbps for data) * Packet size: constant packet size calculated from periodicity and data rate   End-to-end (mouth-to-ear) latency: 100 ms (SA4 input: 100 ms for both data and audio stream), air interface latency: 30 ms  In R1-2103833, we studied 3 traffic models for UL:   * + - 1. 3 flows (video stream + audio/data+pose/control)       2. 3 flows (video stream + audio/data+pose/control), but audio/data packet is delayed to be aligned with video packet       3. 2 single flow (video + pose/control)   We witness difference among them, using two single flows won’t reveal complications for MCS selection and scheduling, etc.    RAN1 also needs to reach an agreement to convert E2E latency to air interface latency. |
| DOCOMO | We are fine with the proposal in general but prefer 10 ms or 15 ms should as 60 ms seems E2E latency. |

|  |
| --- |
| **On AR in UL**  **Moderator proposal for 1st round of email discussion**   * For UL AR,   + Baseline: two streams are defined as follows.     - Stream 1: pose/control       * Traffic model and QoS parameters are same as those for pose/control for UL CG/VR.     - Stream 2: aggregated stream for scene, video, data, and audio.       * Packet size: Truncated Gaussian distribution with the same parameter values with DL (Mean, STD, Max, Min)       * Periodicity: 60 fps       * Data rate: 10 Mbps (baseline), 20 Mbps (optional)       * PDB: 60 ms   + Single stream (Stream 2 above) and/or more than two streams can be optionally evaluated.   + A UE is declared as satisfied only when each stream meets the following requirement.     - In case of the above baseline (2 streams)       * X value for Stream 1 follows that for pose/control of UL CG/VR       * X value for Stream 2 follows that for DL video streaming for single stream case.       * Other values can be optionally evaluated     - In case of single stream (Stream 2 above) and/or more than two streams       * FFS   **Observation from 1st round of email discussion**   * # streams   + Single stream:     - (3) Ericsson, Nokia, Samsung   + Two streams as proposed by moderator     - (11): CATT, OPPO, Xiaomi, vivo, MTK, ZTE, LG, QC, InterDigital, Intel, DCM   + Three streams (1): Apple * PDB for different streams   + Control/pose: 10ms     - This is the same as for pose/control for CG/VR (agreed by all companies)   + A stream aggregating scene, video, data, and audio     - 10ms or 15ms (4) : OPPO, Ericsson, InterDigital, AT&T     - 60ms (10): FUTUREWEI, CATT, Ericsson, Xiaomi, vivo, MTK, ZTE, LG, QC, Samsung,   + Separate audio/video     - 30ms: Apple       * Note: audio/data flow proposed by Apple: 10ms periodicity, 0.756 Mbps/s or 1.12 Mbps (packet size is determined by periodicity and data rate) * Jitter   + Same jitter model as for DL video stream: Ericsson, MTK (optional), AT&T, Intel (further discussion)   **New moderator proposal:**  On UL AR,   * Option 1: Two streams as defined below (baseline)   + Stream 1: pose/control     - Traffic model and QoS parameters are same as for pose/control for UL CG/VR.   + Stream 2: A stream aggregating streams of scene, video, data, and audio.     - Packet size: Truncated Gaussian distribution with the parameter values same as for DL     - Periodicity: 60 fps       * Jitter (optional): same model as for DL     - Data rate: 10 Mbps (baseline), 20 Mbps (optional)     - PDB: 60 ms (baseline), 10/15 ms (optional)     - PER requirements:     - FFS separate streams for I-frame and P-frame * Option 2: Single stream as defined below (optional)   + Packet size: Truncated Gaussian distribution with the parameter values same as for DL   + Periodicity: 60 fps     - Jitter (optional): same model as for DL   + Data rate: 10 Mbps (baseline), 20 Mbps (optional)   + PDB: 60 ms (baseline), 10/15 ms (optional)   + FFS separate streams for I-frame and P-frame * Option 3: Three streams as defined below (optional)   + Stream 1: pose/control     - Traffic model and QoS parameters are same as for pose/control for UL CG/VR.   + Stream 2: A stream aggregating streams of scene and video     - Packet size: Truncated Gaussian distribution with the parameter values same as for DL     - Periodicity: 60 fps       * Jitter (optional): same model as for DL     - Data rate: 10 Mbps (baseline), 20 Mbps (optional)     - PDB: 60 ms (baseline), 10/15 ms (optional)     - FFS separate streams for I-frame and P-frame   + Stream 3: A stream aggregating streams of audio and data     - Periodicity: 10ms     - Data rate: 0.756 Mbps/s or 1.12 Mbps     - Packet size: determined by periodicity and data rate     - PDB: 30 ms * In case multiple steams are evaluated for UL AR, a UE is declared as satisfied only when each stream meets the requirement that X (%) of packets are successfully delivered within a given air interface PDB.   + X value for pose/control: follow X for pose/control for CG/VR   + X value for Stream 2 in Option 1: follow X for DL video stream   + X value for Option 1: follow X for DL video stream   + X value for Stream 2 in Option 3: follow X for DL video stream   + X value for Stream 3 in Option 3: follow X for DL video stream |

## Others

1. **Other issues**
2. **Please discuss other issues that are not discussed above.**

|  |  |
| --- | --- |
| **Company** | **Comment** |
| OPPO | The truncated Gaussian modelling for DL packet size and jitter is reused for the AR UL Stream 2 |
| Ericsson | UL video traffic should also have jitter. |
| Vivo | We would like to further discuss the following issues.  **Issue 9: Association between jitter and PDB**  From the perspective of DL transmission validity, the data packets need to be transmitted within PDB. Whether air interface PDB can be affected by jitter or not should be considered. The following two options are identified for handling air interface PDB when jitter is considered.  • Option 1: air interface PDB is affected by jitter, and actual PDB = (ideal PDB – jitter) for each packet.  • Option 2: air interface PDB is not affected by jitter, and actual PDB = ideal PDB.  If the transmission of a packet before the air interface is delayed due to network jitter, assuming the E2E PDB is fixed, the corresponding time left for downlink transmission over the air interface will be shorter which may result in a higher probability of packet loss thus have an impact on user XR experience. In our opinion, for the association between jitter and air interface PDB, actual PDB = (ideal PDB – jitter) for each packet.  **Issue 10: Two eyes odelling**  According to the outcome of XR work from SA4, the following two different types of XR video traffic are proposed regarding the frame arrival time in the case of X FPS.  - Traffic source type 1: every 1/X s, the packets of both eyes arrive at the same time for each frame.  - Traffic source type 2: every 1/(2\*X) s, the packet of left eye and right eye arrive in turn, e.g. the packet of left eye arrives at odd frames, while the packet of right eye arrives at even frames.  In our opinion, the following proposal can be considered.  Proposal: For a given data rate, single stream with two-eye buffers can be modelled as:   * Model 1: each packet representing both eyes buffers arrives at the same time at X FPS and the sum of packet size for both eyes is equal to the size of a packet in simulation. * Model 2: packet representing left or right eye buffer arrives at 2\*X FPS and the packet size of left or right eye is the size of a packet in simulation. |
| MTK | For both DL and UL video, we see the need to at least list 2 streams with I/P frame differentiation as optional evaluation since both **Google Stadia** (<https://passthroughpo.st/stadias-hidden-limitation-video-encoding/>) and **Nvidia Geforce Now**  (<https://docs.nvidia.com/drive/drive_os_5.1.6.1L/nvvib_docs/index.html#page/DRIVE_OS_Linux_SDK_Development_Guide/NvMedia/nvmedia_nvmvid_enc.html>) uses the IDR (Instantaneous Decoder Refresh) refresh model for video encoding, where in this kind of encoding, I-frame has a much larger size than P-frame. Therefore, to conduct realistic evaluations of capacity in RAN1, it seems necessary. |
| Intel | The current non-IP packet based statistical model lacks the following:   * different distributions for IP packet sizes (limited, unlimited, etc.) * variation in the number of packets per burst, even for a given frame-size * variation in burst length, even for a given frame-size * assymmetry in frame-size distribution below and above mean * variation due to use-cases (CG video is envisioned to be more interactive than VR)   Proposal: Consider trace-based traffic model leveraging the SA4 work as an alternative model for RAN1 for accuracy purposes. |

# Summary

# List of contributions in RAN1 #104b-e

1. [R1-2102320](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2102320.zip) Traffic model for XR and Cloud Gaming Huawei, HiSilicon
2. [R1-2102418](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2102418.zip) Discussion on the XR traffic models for evaluation OPPO
3. [R1-2102546](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2102546.zip) Discussion on traffic models of XR vivo
4. [R1-2102616](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2102616.zip) XR traffic model CATT
5. [R1-2102686](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2102686.zip) Traffic Model for XR and CG MediaTek Inc.
6. [R1-2102769](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2102769.zip) XR traffic model FUTUREWEI
7. [R1-2102827](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2102827.zip) On Traffic Model for XR study Nokia, Nokia Shanghai Bell
8. [R1-2102955](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2102955.zip) Traffic model for XR Ericsson
9. [R1-2102969](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2102969.zip) Discussion on Traffic Model for XR services Xiaomi
10. [R1-2103054](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2103054.zip) Traffic Model for XR Intel Corporation
11. [R1-2103128](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2103128.zip) Views on XR traffic model Apple
12. [R1-2103192](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2103192.zip) Remaining Issues on XR Traffic Models Qualcomm Incorporated
13. [R1-2103264](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2103264.zip) Traffic model for XR Samsung
14. [R1-2103278](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2103278.zip) Further Discussion on Traffic Model for XR Evaluations ZTE, Sanechips
15. [R1-2103317](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2103317.zip) Considerations on XR traffic model Sony
16. [R1-2103360](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2103360.zip) Discussion on traffic models for XR evaluation LG Electronics
17. [R1-2103429](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2103429.zip) UL traffic flows for XR applications InterDigital, Inc.
18. [R1-2103437](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2103437.zip) XR Traffic Model Considerations AT&T
19. [R1-2103598](file:///C:\Users\wanshic\OneDrive%20-%20Qualcomm\Documents\Standards\3GPP%20Standards\Meeting%20Documents\TSGR1_104b\Docs\R1-2103598.zip) Discussion on traffic model for XR NTT DOCOMO, INC.

# Appendix-A (proposals in RAN1#104bis-e tdocs)

**Huawei, HiSilicon**

*Observation 1: For a given VR video, the parameters of the video packet size distribution are related to video encoding configurations, e.g. error resilience, rate control, etc.*

*Observation 2: In the frame-based I/P-stream model for AR/VR/CG, the packet arrival of I-stream and P-stream has a Group-Of-Pictures (GOP) structure.*

*Observation 3: In the slice-based I/P-stream model for AR/VR/CG, both streams have periodic traffic with packet arrival interval 1/FPS.*

*Proposal 1: The following parameters for truncated Gaussian distribution for packet size can be a starting point (note: these parameter values are those before the truncation)*

* *STD: 15% of Mean packet size*
* *Max packet size: 2 \* Mean packet size*
* *Min packet size : 25% of Mean packet size*

*Proposal 2: The following parameters for mean, STD and range of jitter for DL video stream can be a starting point for initial evaluation.*

* *Mean: 0*
* *STD: 2 ms*
* *Range: [-4, 4]ms*

*Proposal 3: If jitter is considered, the remaining scheduling time of a packet is affected by jitter, i.e. remaining scheduling time = air interface PDB – jitter.*

*Proposal 4: For DL video of AR/VR/CG, adopt M1=2 for modelling I-frame and P-frame separately, and adopt the multi-stream traffic model in following Table 6.*

*Table 6. Multi-stream model for DL video*

|  |  |  |
| --- | --- | --- |
| *Application* | *AR/VR/CG* | |
| *Two data streams, i.e. M1 = 2* | * *Stream #1: I-stream* * *Stream #2: P-stream* | |
| *Option 1: slice-based* | *Option 2: frame-based* |
| *Packet modelling* | *Slice-level* | *Frame-level* |
| *Traffic arrival pattern* | *Both streams are periodic with the same FPS.* | *Follow the GOP structure, e.g. GOP size K = 8.* |
| *Number of packets per stream at a time* | * *Stream #1: 1* * *Stream #2: N-1* * *N is the number of slice per frame, e.g. N = 8.* | * *Stream #1: 1* * *Stream #2: 1* |
| *Average data rate* | *Stream #1: Stream #2 = : (N-1)* | *Stream #1: Stream #2 = : (K-1)* |
| *is average size ratio between one I-frame/slice and one P-frame/slice, e.g. = 2.*   * *Other values can be optionally evaluated.* | |

*Note: the QoS requirement for each stream is separately discussed in the KPI part.*

*Proposal 5: Confirm the following working assumption on traffic model for UL pose/control of CG/VR.*

* *On UL Traffic model and QoS parameters*
  + *CG/VR: single stream (pose/control)*
  + *Traffic model for Pose/control* 
    - *Periodic: 4ms (no jitter)* 
      * *Other values can be optionally evaluated.*
    - *Fixed: 100 bytes (SA4 input)*
    - *PDB: 10 ms*

*Proposal 6: There are M2=3 streams in UL traffic model of AR, where*

* *One stream for UL pose/control,*
* *The model for UL pose/control of VR/CG can be reused.*
* *Two streams for UL video to model I-stream and P-stream separately,*
* *The multi-stream model for DL video of VR/AR/CG can be reused.*

**OPPO**

Proposal 1: For the evaluation of XR/CG on NR, the case where M1=1 and M1=2 stream in DL and UL is mandatory

* + Not support to model multiple streams
  + If RAN1 agrees to support M1>1 or M2>1 for evaluation, companies can also evaluate M1=2 and/or M2=2 optionally
    - One stream for video
    - Another stream for audio or control/pose

Proposal 2: If multiple streams are used in DL and/or UL, a UE is declared as satisfied only when all streams meets their corresponding requirements.

Proposal 3: Not support to model and evaluate I-frame and P-frame for the evaluation of XR/CG on NR.

Proposal 4: For the case of 120fps, companies can select one or more values for evaluation

* VR/AR: 60, 90, 120 Mbps
* CG: 16, 60, 90 Mbps

Proposal 5: For the distribution of packet sizes, confirm the working assumption of truncated Gaussian distribution by removing the brackets, i.e.,

* STD: 15% or 20% of Mean packet size
* Max packet size: 1.5 x Mean packet size
* Min packet size: 0.5 x Mean packet size
  + Companies can report whether this parameter is used or not

Proposal 6: For the distribution of inter-packet arrival jitter, adopt the truncated Gaussian distribution with the following parameters:

* Mean: 0
* STD: 3ms
* Range: (-5ms, 5ms)

Proposal 7: For air interface PDB for DL video stream, no more mandatory value is needed.

Proposal 8: For the UL traffic of AR conversational, support the following two different models:

* Pose/Control (Same as CG/VR)
  + Periodic: 4ms (no jitter)
  + Fixed: 100 bytes (SA4 input)
  + PDB: 10 ms
* Data rate for UL Video stream
  + 10Mbps @60fps (baseline)
  + 20Mbps@60fps (optional)

Proposal 9: For UL video stream of AR conversational:

* For packet size, reuse the truncated Gaussian distribution for DL packet size except the mean value is adjusted according to the data rates of UL video. Other parameters are kept the same
* For jitter, reuse the truncated Gaussian distribution for DL packet arrival jitter. All parameters are kept the same

Proposal 10: For UL video stream of AR conversational, the air interface PDB

* 10ms (baseline)
* Other values can be evaluated optionally

**vivo**

Observation 1: For multiple streams modelling, it is necessary to consider I-frame stream and P-frame stream in XR traffic model.

Observation 2: There is no need to model audio stream separately in XR traffic model.

Observation 3: FOV and non-FOV streams based XR traffic model are similar to I-frame and P-frame streams based XR traffic model.

Proposal 1: For the association between jitter and PDB, actual PDB = (ideal PDB – jitter) for each packet.

Proposal 2: For a given data rate, single stream with two-eye buffers can be modelled as:

* Model 1: each packet representing both eyes buffers arrives at the same time at X FPS and the sum of packet size for both eyes is equal to the size of a packet in simulation.
* Model 2: packet representing left or right eye buffer arrives at 2\*X FPS and the packet size of left or right eye is the size of a packet in simulation.

Proposal 3: For XR traffic model in DL, the two traffic models in Table 1 are considered as the starting point for XR evaluation.

Proposal 4: Confirm the working assumptions on the truncated Gaussian distribution for packet size and jitter modelling.

* Parameters of Truncated Gaussian distribution for Packet size (note: these parameter values are those before the truncation)
  + Mean: Derived from average data rate and fps as follows.
    - (average data rate) / (fps for video stream, i.e., # packets per second in our statistical model) / 8 [bytes]
  + STD
    - 15% of Mean packet size derived above
  + Max packet size
    - 1.5 x Mean packet size derived above
  + Min packet size
    - 0.1 \* Mean packet size derived above
* Jitter for DL video stream for a single UE
  + Per the agreed statistical traffic model, arrival time of packet k is k/X \* 1000 [ms] + J [ms], where X is the given fps value and J is a random variable.
  + J is drawn from a truncated Gaussian distribution:
    - Mean: 0
    - STD: 2 ms
    - Range: [-4, 4]ms

Proposal 5: For multiple streams XR traffic model in DL, GOP-based/slice-based multiple streams traffic model in Table 2/Table 3 can be considered.

Proposal 6: Confirm the working assumption of UL single stream traffic model for pose/control.

Proposal 7: For UL single stream traffic model for video, the traffic model in Table 5 is supported at least for AR.

Proposal 8: UL multiple streams with both pose/control and video streams are supported for UE power consumption evaluation.

Proposal 9: For XR DL traffic model, consider the following options:

* Option 1: single video stream.
* Option 2: two streams with I-frame and P-frame.

Proposal 10: For XR UL traffic model, consider the following options:

* Option 1: single pose stream.
* Option 2: single video stream.
* Option 3: two streams with pose/control and video streams.

Proposal 11: A UE with multiple streams is declared as a satisfied UE if each stream from the multiple streams has been satisfied, i.e. for each stream more than X (%) of packets are successfully transmitted within a given air interface PDB, where the X value and the given air interface PDB can be set per stream.

Proposal 12: When two streams are modeled for a user in UL, the stream for pose/control information may have an X value of 99.9 and a given air interface PDB of 10ms, and the stream for scene information may have an X value of 99 and a given air interface PDB of 60ms.

Proposal 13: When the DL video traffic is divided into two streams, e.g. one stream for I-frames and the other for P-frames, the stream for I-frames may have an X value of 99, and the stream for P-frames may have an X value of 95, but the two streams may share the same given air interface PDB of 10ms.

**CATT**

Observation 1: The truncated Gaussian distribution can be used for modelling the packet size for XR and parameters are those of Gaussian distribution before truncation.

Observation 2: It observes that

* + The average data rates based on the mean values are 13.68~14.19 Mbps for VR2.
  + The STD of Packet size is 3.32%~8.33% for VR2, which would be no more than 10% of Mean packet size.
  + The maximum packet sizes are limited by the transmission characteristics, such as “30Mbit/s capped VBR with window 200ms” and “30Mbit/s CBR with window 1 frame”, which is about 1.06~1.14 x Mean packet size.
  + The minimum packet size would be at least larger than the minimum IP packet size, i.e. 46Bytes.

Observation 3: There are two alternatives for modeling the jitter from the different aspects:

* + Opt1-Frame Delay (J): The arrival time of packet k is k/X×1000 [ms] + J [ms], where X is the given FPS value and J is a random variable.
  + Opt2-Inter Arrival Time Jitter (JJ): The inter arrival time between the packet k and the packet k+1 is 1/X×1000 [ms] + JJ [ms], where X is the given FPS value and JJ is a random variable.

Observation 4: For Opt1-Frame Delay (J), the absolute arrival time of packet k is k/X×1000 [ms] + J [ms], where X is the given FPS value and J is a random variable. It observes that

* + The statistic distribution for Opt1-Frame Delay (J) is close to the uniform distribution.
  + The value of Opt1-Frame Delay (J) would be always positive.
  + The mean value is 19.8 ms, which is not equal to 0.
  + The STD is 5.71-5.84 ms.
  + The range of Opt1-Frame Delay (J) is [9.09, 30.68] ms.

Observation 5: For Opt2-Inter Arrival Time Jitter (JJ), the inter arrival time between the packet k and the packet k+1 is 1/X×1000 [ms] + JJ [ms], where X is the given FPS value and JJ is a random variable. It observes that

* + Opt2-Inter Arrival Time Jitter (JJ) is characterized by the truncated Gaussian distribution.
  + The value of Opt2-Inter Arrival Time Jitter (JJ) could be either positive or negative.
  + The mean value is 0 ms.
  + The STD is 8.10~8.19 ms.
  + The range of Opt2-Inter Arrival Time Jitter (JJ) is [-19.54, 19.87] ms.
  + The percentage of packet arrival out of order, i.e. the Opt2-Inter Arrival Time Jitter (JJ) less than -1/X×1000 [ms], is 1.06%~1.56%.

Proposal 1:

* + The truncated Gaussian distribution should be used for modelling the packet size of video stream with VBR and the uniform distribution should be used for modelling the packet size of video stream with CBR.

Proposal 2: For the parameters of the statistical distribution for Packet size

* + The mean value of packet size could be derived by the statistical traffic modelling that (average data rate) / (FPS for video stream, i.e., # packets per second in our statistical model) / 8 [bytes].
  + The STD of Packet size could be derived by 5% and 10% of Mean packet size for the video stream with CBR and VBR, respectively.
  + The maximum packet size could be 1.2 time of and equal to Mean packet size for the video stream with VBR and CBR, respectively.
  + The minimum packet size could be limited by the minimum IP packet size, i.e. 46Bytes.

Proposal 3: Either two alternatives can be used for jitter modeling.

* + Opt1-Frame Delay (J): The arrival time of packet k is k/X×1000 [ms] + J [ms], where X is the given FPS value and J is a random variable.
  + Opt2-Inter Arrival Time Jitter (JJ): The inter arrival time between the packet k and the packet k+1 is 1/X×1000 [ms] + JJ [ms], where X is the given FPS value and JJ is a random variable.

Proposal 4: If jitter is modeled as Opt1-Frame Delay (J), in which the arrival time of packet k is k/X×1000 [ms] + J [ms] under the given FPS value, the following parameters could be considered.

* + The uniform distribution is used for modelling the random variable J.
  + Mean: 20 ms
  + STD: 6.35 ms
  + Range: [9, 31] ms

Proposal 5: If jitter is modeled as Opt2-Inter Arrival Time Jitter (JJ), in which the inter arrival time between the packet k and the packet k+1 is 1/X×1000 [ms] + JJ [ms] under the given FPS value, the following parameters could be considered.

* + The truncated Gaussian distribution is used for modelling the random variable JJ.
  + Mean: 0
  + STD: 8 ms
  + Range: [-1/X×1000, 20] ms

**MediaTek Inc.**

Observation 1: CG and XR display different traffic types within the same application, in both UL and DL directions

Observation 2: 5GS system awareness of differentiated frames may be beneficial

Proposal 1: Adopt the IDR refresh model for both UL/DL videos for RAN1 evaluation.

Proposal 2: traffic model shall take into account different traffic types and possibly differentiated frames within the same application, in both UL and DL directions

Proposal 3: In terms of the values of M1 & M2 for evaluation of DL/UL

* For DL:
  + M1=2 to model I-frame and P-frame separately with different QoS requirements for VR/AR/CG
* For UL:
  + M2=1 for VR/CG (agreed in RAN1 #104e)
  + M2=2 or 3 for AR to model video and control/pose separately

Proposal 4: No need to model the audio stream separately

Proposal 5: Adopt the same PER requirements for I-frames and P-frames.

* FFS 99%

Proposal 6: Adopt Tp as PDB for P-Frames and Ti as PDB for I-frames with Tp < Ti.

* FFS Tp = 8 ms and Ti = 12 ms.

Proposal 7: Coordinate and cooperate with SA4 to construct a video quality evaluation block to use in both RAN1 and SA4 to evaluate the proposed QoS requirements and the proposed enhancement.

Proposal 8: Discuss two possible options:

1. FoV vs. non-FoV
2. FoV vs. low resolution Omnidirectional stream

* FFS: Need for different QoS requirements for the two streams.
* FFS: co-existence with the QoS requirements for I/P-frames.

Proposal 9: Confirm Jitter working assumptions.

* + J is drawn from a truncated Gaussian distribution:
    - Mean: 0
    - STD: 2 ms
    - Range: [-4, 4]ms
      * Note: The values ensure that packet arrivals are in order (i.e., arrival time of a next packet is always larger than that of the previous packet)

Proposal 10: Distinguish Jitter parameters depending on XR/CG server location (Edge, Cloud)

Proposal 11: Confirm Packet size working assumptions.

* + Packet Size is drawn from a truncated Gaussian distribution:
    - Mean: (average data rate) / (fps for video stream, i.e., # packets per second in our statistical model) / 8 [bytes]
    - STD: 15% of Mean packet size derived above
    - Range: 1.5 × Mean packet size derived above

**FUTUREWEI**

Proposal 1: FS\_NR\_XR\_eval adopts the following regarding the parameters of truncated Gaussian distribution for packet size:

* STD: 15% of Mean packet size
* Max packet size: 1.5 x Mean packet size
* Min packet size: 0.5 x Mean packet size

Proposal 2: FS\_NR\_XR\_eval supports modeling single stream/flow on each direction as a baseline.

Proposal 3: FS\_NR\_XR\_eval adopts the following regarding the UL traffic model for AR:

* A single video stream for a UE: periodic with 60 fps, no jitter
* Average data rate: 20 Mbps @ 60 fps (baseline)
* Truncated Gaussian distribution is used for the packet size distribution of video stream for UL AR with the following parameters:
  + Mean: derived from fps and average data rate
  + STD: 15% of Mean packet size
  + Max packet size: 1.5 x Mean packet size
  + Min packet size: 0.5 x Mean packet size
* PDB: 60 ms (baseline)

**Nokia, Nokia Shanghai Bell**

Proposal 1: Adopt the following parameters for the packet (frame) size distribution:

* STD: 2% of mean packet (frame) size
* Max packet (frame) size: 1.1 x mean packet (frame) size
* Min packet (frame) size: 0.9 x mean packet (frame) size

Other values for the packet (frame) size distribution are optional.

Proposal 2: Adopt the following parameters for jitter:

* Mean: 0
* STD: 3 ms
* Range: (-6, 6) ms

Other values for the jitter distribution are optional.

Proposal 3: Adopt a single stream of video in UL for AR2: XR Conversational as a baseline. The average data rate is 10 Mbit/s (1080p) and the frame rate is 60 fps. The PDB is 10 ms.

Proposal 4: No jitter is assumed for the UL video stream.

Proposal 5: Consider a single stream in downlink and single stream in uplink for VR1 and VR2 applications as a baseline.

Proposal 6: Consider a signle stream in downlink and a single stream in uplink for CG application as a baseline.

Proposal 7: Consider a single stream in downlink and a single stream in uplink for AR application as a baseline. Any additional streams consider as optional.

Proposal 8: Following SA4 input, consider no differentiation between the types of packets/frames as well as FOV/non-FOV as the baseline evaluation of XR/CG applications

**Ericsson**

[Observation 1 The bit rates requirement of AR UL scene can be lower than VR/AR DL video while the latency requirement of it is similar as VR/AR DL video.](#_Toc68631137)

[Observation 2 Differentiating and evaluating I-frame and P-frame separately is not essential from a XR traffic characteristics and requirement perspective.](#_Toc68631138)

Based on the discussion in the previous sections we propose the following:

[Proposal 1 The bit rates for AR UL scene can be the range of 2Mbps to 20Mbps and the latency requirement is similar as DL AR/VR video, i.e., 5ms to 20ms.](#_Toc68631139)

[Proposal 2 The min value of packet size should be the 50% of mean packet size in order to make a symmetric distribution. STD and the max value of packet size are 15% and 150% of mean packet size, respectively.](#_Toc68631140)

[Proposal 3 Confirm the proposed values for Mean, STD, range of the jitter distribution.](#_Toc68631141)

[Proposal 4 A UE is satisfied if more than 99% of packets are successfully received within a given air interface PDB.](#_Toc68631142)

[Proposal 5 RAN1 should not model and evaluate I-frame and P-frame separately which will require introducing new traffic parameters.](#_Toc68631143)

[Proposal 6 RAN1 should avoid including multiple streams caused by a frame type, voice traffic, and non-FoV which will increase traffic modelling complexity and evaluation options.](#_Toc68631144)

**Xiaomi**

Proposal 1: Audio/Data stream is not considered in DL video stream for VR2, CG and AR2 services.

Proposal 2: In XR evaluation, do not differentiate I-frame stream and P-frame stream in DL video stream.

Proposal 3: For XR DL evaluation, a single DL video stream including frames for both eye buffers is assumed.

- Interleaved eye buffer model can be optionally considered.

Proposal 4: Send LS to SA4 to confirm on the working assumption of packet size & jitter distribution.

Proposal 5: The initial frame generation time should be randomized among different UEs.

Proposal 6: Confirm the working assumption on UL traffic model and QoS parameters for CG/VR and Pose/control.

Proposal 7: An UL pose stream and a single UL video data stream are used as UL traffic model for AR2 use case.

Proposal 8: For per UE KPI, the exact value of X is set to be 99.9

**Intel Corporation**

Observations-1:

* for CBR configurations the frame-size variations are quite small
  + the max/mean frame-size ratio is ~ 1.06
  + the min/mean frame-size ratio is ~0.93
  + the std/mean frame-size ratio is ~0.02
* for cVBR configurations the frame-size variations are larger
  + the max/mean frame-size ratio is ~ 1.18 – 1.94 with smaller ratios corresponding to 8 slice/eye buffer case while large ratio corresponding to 1 slice case
  + the min/mean frame-size ratio is ~ 0.24 – 0.48 with larger ratios corresponding to 8 slice/eye buffer case while small ratio corresponding to 1 slice case
  + the std/mean frame-size ratio is ~ 0.07 – 0.14

Observations-2:

We observe that

* there is significant variation in traffic observed at L2 (P-trace) for a given content model (V-trace) and use-case example: VR2 30 Mbps 2 eye buffers at 2Kx2K at 60 fps, 8bit
* this variation is due to encoding model (slices, encoding delay, etc.) and content delivery model (packetization etc.)
* distribution of frame-size based on P-trace is asymmetric (heavy tail below mean, light tail above mean)

Observations-3:

Comparing V-trace and P-trace frame-sizes we can observe that:

* distribution of frame-size based on P-trace is asymmetric (heavy tail below mean, light tail above mean) while frame-size based on V-trace is more symmetric in shape
* the max/mean frame-size calculated based on V-trace is much larger than that calculated based on P-trace. The encoding and the content delivery model clearly affects the frame-size distribution.

Observations-4:

The current frame-based statistical model lacks the following:

* different distributions for IP packet sizes (limited, unlimited, etc.)
* variation in the number of packets per burst, even for a given frame-size
* variation in burst length, even for a given frame-size
* assymmetry in frame-size distribution below and above mean
* variation due to use-cases (CG video is envisioned to be more interactive than VR)

Proposal-1: Consider more accurate trace-based traffic model leveraging the SA4 work in RAN1 XR simulations

**Apple**

Observation 1: From SA4 traffic model on XR conversational, it is clear that uplink traffic is with substantial throughput requirements.

Observation 2: SA4 study on AR2 indicates multiple data flows are present in both downlink and uplink.

Proposal 1: It is key to include uplink traffic with substantial throughputs in the study of AR2.

Proposal 2: In RAN1 study, data flows with different QoS requirements in XR study should be modeled separately.

Proposal 3: 3 streams (scene/video + audio/data + pose/control) for uplink and two streams (scene/video + audio/data) for downlink can be used for evaluation on AR2. The audio/data flow is modeled as:

* Periodic:
  + 10 milliseconds for framing (SA4 input: 10 ms for data stream and 20 ms for audio)
* Data rate
  + 0.756 Mbps/s or 1.12 Mbps (SA4 input: 256/512 Kbps for audio, 0.5 Mbps for data)
* Packet size: constant packet size calculated from periodicity and data rate

End-to-end (mouth-to-ear) latency: 100 ms (SA4 input: 100 ms for both data and audio stream), air interface latency: 30 ms

**Qualcomm Incorporated**

**Proposal 1**: Adopt the following for DL video streaming

* Parameters of Truncated Gaussian distribution for Packet size (note: these parameter values are those before the truncation)
  + Mean: Derived from average data rate and fps as follows.
    - (average data rate) / (fps for video stream, i.e., # packets per second in our statistical model) / 8 [bytes]
  + STD
    - 7.5% of Mean
  + Max packet size
    - 135% of Mean
  + Min packet size
    - 54.5% of Mean

**Proposal 2**: Confirm the following WA.

* Jitter for DL video stream for a single UE
  + Per the agreed statistical traffic model, arrival time of packet k is k/X x 1000 [ms] + J [ms], where X is the given fps value and J is a random variable.
  + J is drawn from a truncated Gaussian distribution:
    - Mean: 0 ms
    - STD: 2 ms
    - Range: [4, 4] ms
    - Other values can be optionally evaluated

**Proposal 3**: Evaluate two streams in UL for AR as follows.

* Stream 1: pose/control (same as VR/CG)
  + Periodic: 4ms (no jitter)
    - Other values can be optionally evaluated.
  + Fixed: 100 bytes (SA4 input)
  + PDB: 10 ms
* Stream 2: aggregated stream for scene, video, data, and audio.
  + Traffic model is same as DL video stream, i.e.,
    - Truncated Gaussian distribution for Packet size (note: these parameter values are those before the truncation)
      * Mean: Derived from average data rate and fps as follows.
        + (average data rate) / (fps for video stream, i.e., # packets per second in our statistical model) / 8 [bytes]
      * STD
        + 7.5% of Mean
      * Max packet size
        + 135% of Mean
      * Min packet size
        + 54.5% of Mean

**Proposal 4**: Adopt X = 99 in the following except for the case when I-frames and P-frames are separately evaluated.

* Baseline: A UE is declared a satisfied UE if more than X (%) of packets are successfully transmitted within a given air interface PDB. The exact value of X is FFS, e.g., 99, 95
  + FFS different values for I-frame and P-frame if evaluation of them is agreed.

Other values can be optionally evaluated

**Samsung**

Proposal 1: XR traffic models consider 1 DL stream (video) and 2 UL streams (pose and scene upload).

Proposal 2: Confirm the frame-level modeling for packet arrivals.

Proposal 3: If the P-frame needs to have smaller PER or PDB that the I-frame, separate models can be defined; otherwise, a single model for the I-frame is used.

Proposal 4: Confirm the standard deviation and the maximum packet size for the truncated Gaussian distribution as 15% and 1.5x of the mean packet size, respectively.

Proposal 5: Jitter can be evaluated based on a truncated Gaussian distribution or can be abstracted from evaluations that can be instead without jitter for few PDB values - scaling by a jitter distribution can then apply.

Proposal 6: For KPIs, the percentage of UEs that can achieve a target PER and a target PDB suffices for data packets. For PDCCH-based scheduling, whether and how the PDCCH BLER does not impact the target PDB needs to also be considered

**ZTE, Sanechips**

[Observation 1: With Alt 1, the ratio between standard deviation and mean value is 0.08 under the configuration of VR2-1, VR2-2 and VR2-6, while the ratio is around 0.13 under the configuration of VR2-5, when bit rate is 30Mbps.](#_Toc68641007)

[Observation 2: With Alt 1, the ratio between standard deviation and mean value is 0.15 under the configuration of VR2-7, VR2-8, when bit rate is 45Mbps.](#_Toc68641008)

[Observation 3: With Alt1, the ratio between the maximal value and mean value is 1.24 under the configuration of VR2-1, VR2-2 and VR2-6, while the ratio is around 1.39 under the configuration of VR2-5, when bit rate is 30Mbps.](#_Toc68641009)

[Observation 4: With Alt 1, the ratio between standard deviation and mean value is 1.45 under the configuration of VR2-7, VR2-8, when bit rate is 45Mbps.](#_Toc68641010)

[Observation 5: Non-negligible bias could be observed between the CDF curves of the distribution and that of the data samples in the range of 5%-95%.](#_Toc68641011)

[Observation 6: The ratio between standard deviation and mean value is ranging from 4.14% to 4.66% in Gaussian distribution of single eye packet size.](#_Toc68641012)

[Observation 7: The ratio between standard deviation and mean value is ranging from 2.27% to 3.14% in Gaussian distribution of double eyes packet size.](#_Toc68641013)

[Observation 8: To attain jittering information as defined in the CSV files available from [2] , subtraction could be performed between the time\_stamp\_in\_micro\_s value of a representative, e..g, the last fraction to the corresponding rendering time.](#_Toc68641014)

[Observation 9: The values in the WA do not comply with the numerical evaluations](#_Toc68641015)

[Observation 10: Packet loss information and packet delay information cannot provide additional information.](#_Toc68641016)

[Observation 11: If multiple data streams are adopted for DL traffic, the difference of XR/CG source related information may influence the scheduling/collision handling of the different streams](#_Toc68641017)

[Observation 12: Frame based and slice based intra refreshing have an impact on the traffic model aspects such as jittering modelling and inter frame arrival time.](#_Toc68641018)

[Proposal 1: Standard deviation and maximal packet size for DL video streaming traffic are determined as follows:](#_Toc68618182)

[ Single eye packet size](#_Toc68618183)

[- STD = 4% \* mean, MAX = 112% \* mean](#_Toc68618184)

[ Dual eye packet size](#_Toc68618185)

[- STD = 3% \* mean, MAX = 109% \* mean.](#_Toc68618186)

[Note: Minimum file size is not considered](#_Toc68618187)

[Proposal 2: Further discuss in RAN1 the jittering related information for DL video streaming including mean/variance/maximal value using the statistics as starting point.](#_Toc68618188)

[Table 5 Summary of VR2 Jitter Statistics](#_Toc68618189)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Mean (ms)** | **STD (ms)** | **Range [ms, ms]** |
| **VR2-1** | -1.358 | 8.16 | [-25.0740, 18.2330] |
| **VR2-2** | 4.231 | 7.0279 | [-16.0640, 19.8290] |
| **VR2-5** | -4.149 | 7.6014 | [-29.2700, 21.0710] |
| **VR2-6** | 6.697 | 7.0095 | [-18.0640, 22.1750] |
| **VR2-7** | 1.955 | 6.9989 | [-19.0440, 18.3690] |
| **VR2-8** | 0.0489 | 7.8489 | [-23.0550, 18.1960] |

[Proposal 3: Consider the reliability requirement as 95%, i.e. the baseline for per UE KPI is updated as](#_Toc68618191)

[A UE is declared a satisfied UE if more than 99 (%) of packets are successfully transmitted within a given air interface PDB.](#_Toc68618192)

[Proposal 4: When determining a XR/CG user is satisfied or not, the following factors are not considered.](#_Toc68618193)

[ Packet loss information](#_Toc68618194)

[ Packet delay information](#_Toc68618195)

[Proposal 5: Confirm the WA on UL traffic of 100Byte packet size, 4ms periodicity as well as 100ms PDB](#_Toc68618196)

[Proposal 6: Standard deviation and maximal packet size for UL video streaming traffic are determined as follows:](#_Toc68618197)

[ Single eye packet size](#_Toc68618198)

[- STD = 4% \* mean, MAX = 112% \* mean](#_Toc68618199)

[ Dual eye packet size](#_Toc68618200)

[- STD = 3% \* mean, MAX = 109% \* mean.](#_Toc68618201)

[Note: Minimum file size is not considered](#_Toc68618202)

[Proposal 7: Further discuss in RAN1 the jittering related information for UL video streaming including mean/variance/maximal value using the statistics as starting point.](#_Toc68618203)

[Table 5 Summary of VR2 Jitter Statistics](#_Toc68618204)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Mean (ms)** | **STD (ms)** | **Range [ms, ms]** |
| **VR2-1** | -1.358 | 8.16 | [-25.0740, 18.2330] |
| **VR2-2** | 4.231 | 7.0279 | [-16.0640, 19.8290] |
| **VR2-5** | -4.149 | 7.6014 | [-29.2700, 21.0710] |
| **VR2-6** | 6.697 | 7.0095 | [-18.0640, 22.1750] |
| **VR2-7** | 1.955 | 6.9989 | [-19.0440, 18.3690] |
| **VR2-8** | 0.0489 | 7.8489 | [-23.0550, 18.1960] |

[Proposal 8: It's expected from SA that the 5QI values shall be finalized before RAN1 could start the discussion regarding the differentiation of the multiple streams.](#_Toc68618205)

**Sony**

Proposal 1: Support AR2 application as communicated from SA4 with 5 streams for UL and 3 for DL. If simplications are desired, we could have an option with 3 streams for UL (pose and two cameras) and 1 stream for DL (only video).

Proposal 2: Define air interface delay for uplink that is measured from the point when a packet is transmitted by the UE to the point when it is successfully delivered to gNB. FFS: the reference transmission point at the UE side (e.g. TX antenna connector, etc).

Proposal 3: Air interface PDB UL for VR/AR is 10 ms.

Proposal 4: Use the media characteristics / simulation configurations that SA4 already defined in LS R1-2101765.

Observation 1: Separate modelling of I-frame and P-frame is desirable to provide accurate modelling with the cost of increasing modelling complexity.

Proposal 5: Consider the entire video stream (I-frames, P-frames etc) to be transported on a bearer with a single associated QoS class.

Proposal 6: RAN1 should set up different models for different kinds of data streams in AR2 UL.

Proposal 7: RAN1 to study layer-1 aspects of large packet transmission with better reliability than eMBB and/or with low packet delay

**LG Electronics**

Proposal 1: Not to include additional XR applications for RAN1 study than XR applications agreed in RAN1#103-e, e.g., VR1/2, AR1/2 and CG.

Observation 1: AR1 and AR2 are essential applications for XR in 5G ecosystem while VR1, VR2 and CG applications can be considered as extension of traditional multi-media services.

Proposal 2: If prioritization for study among XR applications is necessary, AR1/2 should be prioritized over other XR applications.

Proposal 3: Not to have more than one mandatory values per each application for air interface PDB for DL video stream.

* 10 ms for VR/AR and 15 ms for CG

Proposal 4: If a single stream is used for both I-frame and P-frame for DL traffic modelling, a UE is declared a satisfied UE if more than X (%) of packets are successfully transmitted within a given air interface PDB.

* X is [95] which is a single value

Proposal 5: UL Traffic model for video stream/scene update/audio/data

* Frame per second (fps)
  + 60 fps (SA4 input) – no jitter
* Average data rate for UL video stream:
  + 20 Mbps (~half of the average of VR/AR average data rate for DL video stream)
* Truncated Gaussian distribution is used for the packet size distribution
  + Parameters of Truncated Gaussian distribution for Packet size (note: these parameter values are those before the truncation)
    - Mean: Derived from average data rate and fps as follows.
      * (average data rate) / (fps for video stream, i.e., # packets per second in our statistical model) / 8 [bytes]
    - STD: same as in DL
    - Max packet size: same as in DL
    - Min packet size: FFS (need input from SA4)
* PDB
  + [100] ms (based on SA4 input)

Proposal 6:

* In the case where two streams are used for UL traffic modelling, one for control/pose and the other for video stream/scene update/audio/data, a UE is declared a satisfied UE if more than X1 (%) of packets for control/pose are successfully transmitted within a given air interface PDB1, AND more than X2 (%) of packets for video stream/scene update/audio/data are successfully transmitted within a given air interface PDB2.
  + X1 is [99] and X2 is [95]
  + PDB1 is 10ms and PDB2 is [100]ms

Proposal 7:

* Multi-stream is not further considered in DL for VR1/VR2/CR/AR1/AR2 applications
* Multi-stream is not further considered in UL for VR1/VR2/CG applications
  + FFS whether to support dual-stream for AR application in UL

**InterDigital, Inc.**

Observation 1: For CG there can be more than one traffic flow in UL, consisting of the following traffic types:

* User actions (e.g. gamepad controller, HMD)
* Control data (e.g. protocol flow control, keep-alive messages)

Observation 2: For CG, the UL traffic can be represented by 2 different traffic flows with the following characteristics:

* User Actions
  + Packet arrival is aperiodic and correlated with user activity, where inter-packet arrival follows exponential distribution
  + Packet size can vary depending on the CG platform (i.e. typically Gaussian distribution)
* Control Data
  + Packet arrival can be periodic (e.g. 1/100Hz, 1/10Hz) or correlated with user activity depending on the CG platform. Dedicated periodic flows may exist for indicating lifecycle status.
  + Packet size can vary depending on the CG platform (i.e. typically Gaussian distribution)

Observation 3: For CG, different transport protocols are used for carrying user actions and control data in UL

Observation 4: The sensitivity of QoE to changing QoS is significantly different between the 2 traffic flows

Observation 5: Similar to CG, for VR the UL traffic can be represented by 2 traffic flows

Observation 6: For AR, the UL traffic includes encoded video/media, in addition to user actions and control data

Observation 7: For AR, the traffic characteristics of encoded video/media is significantly different than the other UL traffic flows

Based on these observations, the following conclusions were made:

Proposal 1: RAN1 uses 2 different traffic flows in UL as mandatory for CG evaluations

Proposal 2: The same number of traffic flows in UL used for CG evaluations can also be used for VR evaluations

Proposal 3: RAN1 uses at least 2 different traffic flows in UL as mandatory for AR evaluations. FFS for using more than 2 dfferent traffic flows in UL for AR

It can be further discussed on whether a third traffic flow is considered for AR evaluations.

**AT&T**

Proposal 1: In addition to M1=1 and M2=1 streams, support M1=2 and M2=2, where in both the DL and UL a user has one video stream based on a Truncated Gaussian packet size distribution and one data/control stream based on a fixed packet size and inter-arrival time.

Proposal 2: For both DL and UL consider mixed traffic scenarios with different ratios of UEs with XR and eMBB traffic (e.g. based on FTP Model 3).

**NTT DOCOMO, INC.**

Proposal 1:

* *Consider to study XR conference as optional.*

Proposal 2:

* *Adopt two streams for UL for AR applications*
  + *Traffic model for pose/control information can be same as CG/VR.*
  + *Traffic model for scene update/video/audio data,*
    - *Periodicity: 60 fps*
    - *Data rate: 20 Mbps*
    - *PDB: 60 ms*

Proposal 3:

*RAN1 continues to discuss the statistical models for VR1 and AR1 considering corresponding traffic model discussion in SA4*

# Appendix-B (previous agreements)

## RAN1 #103-e

Agreement: **XR applications**

RAN1 confirms that diverse applications of VR1/2, AR1/2,~~(XR conference FFS),~~ CG are of interest for study. Potential prioritization/down selection of these applications for evaluation is to be discussed after detailed traffic models and relevant evaluation assumptions are stable.

* FFS: other applications, e.g., XR conferencing

Agreement: **Traffic model**

Traffic model for DL and UL should reflect various aspects, e.g., various bit rates, variable frame/packet (definition of frame/packet to be clarified with traffic model as necessary) size, and periodicity (how to model jitter is FFS).  RAN1 will strive to conclude on detailed traffic models in the next RAN1 meeting (104-e) where SA4 outcome on traffic model is expected to be available.

* Statistical model is preferred.
* It is preferred traffic model for both UL and DL have a certain degree of variability so that~~and~~ the total number of traffic models can be reduced.
* Note: Taking into account the fact that the decision on traffic models may hold many other crucial decisions, discussion on traffic model in the next RAN1 meeting is prioritized from the beginning.

Agreement:

Adopt the following deployment for XR/CG evaluations

* Indoor hotspot: FR1 and FR2
  + Detailed definition of Indoor hotspot refers to TR 38.913.
  + Channel model: InH. Detailed definition of InH refers to TR 38.901.
* Dense urban: FR1 and FR2
  + Detailed deployment refers to TR 38.913, where single layer with Marco layer is assumed.
  + Channel model: UMi. Detailed definition of UMi refers to TR 38.901.

FFS: Whether to prioritize FR1 for evaluation.

Note 1: When selecting the deployment and evaluation assumptions for XR/CG evaluations, it is up to company to evaluate FR1 or FR2 or both for the frequency range.

Note 2: It does not mean that all applications are evaluated for all the deployment scenarios.

Agreement:

Urban Macro can be ~~optionally~~ reported for XR/CG evaluations only for FR1.

* FFS: whether Uma is optional or not
* Following parameters can be assumed.

|  |  |
| --- | --- |
| **Parameter** | **Proposed value** |
| **Urban Macro (FR1)** |
| Layout | 21cells with wraparound ISD = 500 m |
| BS Tx power | FR1: 49 dBm/20 MHz |

Agreement:

It is to be further discussed how to prioritize the combinations of deployment scenarios and applications after traffic models for each application are stable.

Agreement:

System capacity is defined as the maximum number of users per cell with at least X % of UEs being satisfied.

* X=90 (baseline) or 95 (optional)
* Other values of X can also be evaluated optionally

Note: The exact ‘satisfied’ requirements will be discussed separately

FFS: how to calculate the percentage of satisfied users across multiple drops of simulations

Agreement:

* Adopt the simulation assumptions in table 1 as below

Table 1: Simulation assumptions for XR evaluation (Part 1) (updated)

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Proposed value** | |
| **Indoor hotspot FR1/FR2** | **Dense urban FR1/FR2** |
| Layout | 120m x 50m ISD: 20m TRP numbers: 12 | 21cells with wraparound ISD: 200m |
| Carrier frequency | FR1: 4 GHz  FR2: 30 GHz | |
| Subcarrier spacing | FR1: 30 kHz  FR2: 120 kHz | |
| BS height | 3m | 25m |
| UE height | hUT=1.5 m | |
| BS noise figure | FR1: 5 dB  FR2: 7 dB | |
| UE noise figure | FR1: 9 dB  FR2: 13 dB | |
| BS receiver | MMSE-IRC | |
| UE receiver | MMSE-IRC | |
| Channel estimation | Realistic  FFS:Ideal(optional) | |
| UE speed | 3 km/h | |
| MCS | Up to 256QAM | |
| BS antenna pattern | Ceiling-mount antenna radiation pattern, 5 dBi | 3-sector antenna radiation pattern, 8 dBi |
| UE antenna pattern | FR1: Omni-directional, 0 dBi,  FR2: UE antenna radiation pattern model 1, 5dBi | |

Agreement:

Adopt the following UE distribution for XR/CG evaluation for outdoor scenario

* For outdoor scenario:
  + FR1: 80% indoor, 20% outdoor
  + FR2: 100% outdoor

Other UE distribution can be evaluated optionally.

Agreement:

Adopt the following TDD configuration for XR/CG evaluation

* FR1:
  + Option 1: DDDSU
  + Option 2: DDDUU
* FR2:
  + Option 1: DDDSU

FFS detailed S slot format

Note: Other TDD configuration or FDD can be optionally evaluated.

Agreement:

Adopt the following BS antenna parameters for indoor scenario for XR/CG evaluation

* FR1:
  + 32 TxRU, (M, N, P, Mg, Ng; Mp, Np) = (4,4,2,1,1;4,4)
  + (dH, dV) = (0.5, 0.5)λ
* FR2:
  + Option 2: 2 TxRU, (M, N, P, Mg, Ng; Mp, Np) = (16, 8, 2,1,1;1,1)
  + (dH, dV) = (0.5, 0.5)λ

Other BS antenna parameters can be optionally evaluated

Agreement:

For XR/CG evaluation, adopt the following assumptions for downtilt

* Dense Urban
  + FFS: 6 or 12 degree
  + ~~Other downtilt can be optionally evaluated.~~
* Indoor hotspot
  + 90° (pointing to the ground)

Other downtilt can be optionally evaluated

Agreement:

* Adopt the simulation assumptions in table 3 as below

Table 3: Simulation assumptions for XR evaluation (Part 3)

|  |  |
| --- | --- |
| **Power control parameter** | Companies should report |
| **Transmission scheme** | Companies should report~~, such as Type I/II codebook, rank assumption~~ |
| **Scheduler** | SU/MU-MIMO PF scheduler (company to report SU or MU),  other scheduler (e.g., delay aware scheduler) is up to companies report |
| **CSI acquisition** | Realistic  Both CSI feedback and SRS are considered  Companies should report  CSI feedback delay, CSI report periodicity, whether using CSI quantization, CSI error model or not,  Assumptions on SRS: periodicity, processing gain, processing delay, etc  and etc. |
| **PHY processing delay** | Baseline: UE PDSCH processing Capability #1  Optional: UE PDSCH processing Capability #2  Companies should report gNB processing delay, e.g. DL NACK to retransmission delay, UL previous transmission to current transmission delay and etc. |
| **PDCCH overhead** | Companies should report |
| **DMRS overhead** | Companies should report |
| **Target BLER** | Companies should report |
| **Max HARQ transmission** | Companies should report |

Agreement:

The following aspects are to be discussed after traffic model is stable.

* For the system capacity definition, how to determine whether a UE is satisfied or not is to be deferred until the exact traffic model along with how to measure E2E user experience is available. Additional metrics to be collected will be further discussed after traffic model is stable.
* Various options for traffic arrival offset among UEs per cell were proposed by companies, e.g., even offset, random offset, no offset. It will be discussed after traffic model is determined.

Agreement:

System bandwidth for XR/CG evaluations are as follows.

* For FR1,
  + Baseline: 100 MHz
  + Optional: 20/40 MHz (FFS: 200 MHz)
* FFS FR2

Agreement:

For outdoor scenarios, the ~~baseline~~ BS antenna parameters are as follows.

* FFS FR1,
  + Option 1: 64 TxRU, (M, N, P, Mg, Ng; Mp, Np) = (8,8,2,1,1;4,8)
  + Option 2: 32 TxRU, (M, N, P, Mg, Ng; Mp, Np) = (8,2,2,1,1,8,2)
  + Option 3: 32TxRUs (M, N, P, Mg, Ng; Mp, Np) = (4,4,2,1,1,4,4)

(dH, dV) = (0.5λ, 0.~~8~~5λ)

* FR2:
  + TxRU, (M, N, P, Mg, Ng; Mp, Np) = (4,8,2,2,2;1,1)

(dH, dV) = (0.5λ, 0.5λ)

Other configurations can be optionally evaluated.

Agreement:

UE antenna parameters for XR/CG evaluations are as follows

* FR1:
  + Baseline: 2T/4R, (M, N, P, Mg, Ng; Mp, Np) = (1,2,2,1,1;1,2), (dH, dV) = (0.5, N/A)λ
  + Optional: 4T/4R, 1T/2R, 2T2R
* FFS FR2: down-selection between the next two options. Please indicate if you have preference.
  + Option 1 (Follow Rel-17 evaluation methodology for FeMIMO in [R1-2007151](file:///E:\Workspace\3GPP%20related\3GPP%20meeting\2021\2021.Q2\RAN1%23104b-e\Summary\Docs\R1-2007151.zip))
    - (M, N, P)=(1, 4, 2), 3 panels (left, right, top)
    - (Mp, Np) is up to company. Need to be reported with simulation result.
  + Option 2 (from TR 38.802 – developed in Rel-14)
    - 4Tx/4Rx: (M, N, P, Mg, Ng; Mp, Np) = (2,4,2,1,2;1,2), (dH,dV) = (0.5, 0.5)λ, the polarization angles are 0° and 90°

Agreement:

BS Tx power for XR/CG evaluations are as follows

* For Indoor hotspot:
  + FR1:
    - 24 dBm per 20 MHz
  + FR2:
    - 23 dBm per 80 MHz. EIRP should not exceed 58 dBm
* For Dense urban:
  + FR1:
    - 44 dBm per 20 MHz
  + FR2:
    - 40 dBm per 80 MHz. EIRP should not exceed 73 dBm

For system BW larger than above, Tx power scales up accordingly.

Agreement:

UE max Tx power for XR/CG evaluations are as follows

* FR1: 23 dBm
* FR2: 23 dBm, maximum EIRP 43 dBm

Agreement: **Baseline power evaluation methodology**

~~If UE power consumption is agreed as a KPI for evaluation of XR performance over NR,~~TR38.840 is the baseline methodology potentially with some modifications if necessary.  RAN1 aim to minimize modeling effort. ~~For example, the following aspects can be considered for further discussion but not limited to.~~

* ~~FFS whether/how to model UE power consumption for UE tx power other than 0dBm and 23dBm,~~
* ~~FFS whether/how to model UE power consumption for UL slots that are not defined in TR38.840~~
* ~~FFS whether/how to model UE power consumption for ‘S’ slot~~
* ~~FFS whether/how to model UE power consumption for 400MHz in FR2 including scaling rule for FR2 BWP adaption.~~
* ~~FFS whether/how to model UE consumption for the corresponding number of Tx antennas~~
* ~~FFS whether/how to model the UE power consumption for UE tx power under FR2~~

Agreement:

* RAN1 continues to discuss evaluation methodologies for UE power consumption and system capacity.
* RAN1 is to discuss whether/how to study/evaluate mobility and coverage at a later stage, e.g., starting from Q1 2021.

## RAN1 #104-e

Agreements**:** RAN1 adopts a parameterized statistical traffic model for evaluation of XR and CG, and KPI with details as shown below (RAN1 strives to agree on the remaining details during RAN1 #104e, based on SA4 input):

* There are M1 and M2 streams in DL and UL respectively
  + At least adopt the case where M1=1 & M2=1
  + FFS the values of M1 and M2, including the possibility of being application-dependent
* DL
  + Air interface Packet Delay budget (PDB)
    - Air interface delay is measured from the point when a packet arrives at gNB to the point when it is successfully delivered to UE
    - Air interface PDB for video streaming
      * VR/AR: [10ms (mandatory), 20ms (optional)]
      * CG: [15ms (mandatory), 30ms (optional)]
        + FFS: other optional values
* Per UE KPI
  + Baseline: A UE is declared a satisfied UE if more than X (%) of packets are successfully transmitted within a given air interface PDB. The exact value of X is FFS.
  + FFS: In addition to the baseline, the following additional method is FFS
    - When determining a XR/CG user is satisfied or not, the following factors are considered. FFS how to use those factors.
      * Packet loss information
      * Packet delay information
      * Some XR/CG source related information if they can be available within RAN, e.g. the mapping between packet and slices or frames and the packet importance
      * Multiple data streams traffic model
  + FFS if there are multiple streams (if adopted)
* FFS additional aspects not addressed above.
* Note 1: Companies are encouraged to provide details such as parameters (e.g., mean, STD, etc.), distributions, etc., by analyzing SA4 input, e.g., V/S/P traces
* Note 2: All FFS points above are to be further discussed in RAN1 #104e

Agreements

* Statistical traffic model for a single DL video stream for a single UE
  + The statistical traffic model for a single UE for a single DL video stream in Figure 1 is adopted, where a packet is assumed to represent multiple IP packets corresponding to a single video frame for modelling/evaluation purposes, e.g., traffic arrival, packet size, evaluation of latency and reliability.



* Frame per second (fps) for DL video stream for a single UE
  + 60 fps (baseline)
  + 120 fps (optional)
  + Other values, e.g., 30, 90 fps can be also optionally evaluated.
* Average data rate for DL video stream:
  + VR/AR: 30, 45 Mbps @60fps (baseline)
    - ~~30,~~ 60 Mbps @60fps (optional)
    - Note: this is the aggregated data rate when applicable
  + CG: 8, 30 Mbps @60fps (baseline)
    - ~~8,~~ 45 Mbps @60fps (optional)
  + Other values (in combination with fps) can be also optionally evaluated.
* Truncated Gaussian distribution is used for the packet size distribution of video stream for AR/VR/CG.
  + Other distribution is not precluded.
* (Working assumption) Parameters of Truncated Gaussian distribution for Packet size (note: these parameter values are those before the truncation)
  + Mean: Derived from average data rate and fps as follows.
    - (average data rate) / (fps for video stream, i.e., # packets per second in our statistical model) / 8 [bytes]
  + STD
    - TBD
  + Max packet size
    - TBD
  + Min packet size
    - TBD
    - FFS whether or not to use this parameter
* Per UE KPI
  + Baseline: A UE is declared a satisfied UE if more than X (%) of packets are successfully transmitted within a given air interface PDB.
    - The exact value of X is FFS, e.g., 99, 95
      * FFS different values for I-frame and P-frame if evaluation of them is agreed.
      * Other values can be optionally evaluated
* DL traffic model: video stream
* (Working assumption) Parameters of Truncated Gaussian distribution for Packet size (note: these parameter values are those before the truncation)
  + Mean: Derived from average data rate and fps as follows.
    - (average data rate) / (fps for video stream, i.e., # packets per second in our statistical model) / 8 [bytes]
  + STD
    - [15% of Mean packet size derived above]
    - Note: The above value is an example for further investigation, and is to be revisited potentially with more inputs from companies in RAN1#104-bis-e
  + Max packet size
    - [1.5 x Mean packet size derived above]
    - Note: The above value is an example for further investigation, and is to be revisited potentially with more inputs from companies in RAN1#104-bis-e
  + Min packet size
    - TBD
    - FFS whether or not to use this parameter
    - Note: This is to be revisited potentially with more inputs from companies in RAN1#104-bis-e.
* Jitter for DL video stream for a single UE
  + (Already agreed) Per the agreed statistical traffic model, arrival time of packet k is k/X1000 [ms] + J [ms], where X is the given fps value and J is a random variable.



* + (Newly proposed agreement) J is drawn from a truncated Gaussian distribution:
    - Mean: [0]
    - STD: [2 ms]
    - Range: [[-4, 4]ms]
      * Note: The values ensure that packet arrivals are in order (i.e., arrival time of a next packet is always larger than that of the previous packet)
    - Note: The above values for mean, STD and Range are working assumption for initial simulations, and is to be revisited potentially with more inputs from companies in RAN1#104-bis-e
* Air interface PDB for DL video stream
  + VR/AR:
    - 10ms
    - Other values, e.g., 5ms, 20 ms can be optionally evaluated.
  + CG:
    - 15ms
    - Other values, e.g., 10ms, 30ms can be optionally evaluated.
  + FFS whether or not to have more than one mandatory value

Working assumption: On UL Traffic model and QoS parameters

* CG/VR: single stream (pose/control)
* Traffic model for Pose/control
  + Periodic: 4ms (no jitter)
    - Other values can be optionally evaluated.
  + Fixed: 100 bytes (SA4 input)
  + PDB: 10 ms
* AR
  + FFS

Agreements: On evaluation of multiple streams/flows:

* FFS the following in RAN1#104-bis-e
  + Whether/how to model and evaluate I-frame and P-frame for both DL and UL, e.g., separate definition of fps, packet size, QoS requirements (e.g., PER, PDB), etc.
  + Whether/how to separately model and evaluate two streams of video and audio/data for both DL and UL
  + Whether/how to model and evaluate FOV (high-resolution) and non-FOV (lower-resolution omnidirectional) streams, e.g., separate definition of fps, packet size, QoS requirements (e.g., PER, PDB), etc

Agreement: Adopt following update for TDD configuration for XR/CG evaluation

* FR1:
  + Option 1: DDDSU
  + Option 2: DDDUU
* FR2:
  + Option 1: DDDSU
  + Option 2: DDDUU

Detailed S slot format is 10D:2F:2U. Other S slot format(s) can also be optionally evaluated.

Further clarify that for option 2 for FR1/FR2, there is [2]-symbol gap at the end of third “D” slot of  DDDUU.

FFS whether or not to differentiate the two options (e.g., mandatory vs. optional)

Agreement**:** For XR evaluation, ideal channel estimation can be optionally evaluated.

Agreements**:** System bandwidth for XR/CG evaluations are as follows.

* For FR1,
  + Baseline: 100 MHz
  + Optional: 20/40 MHz, 2\*100 MHz with CA
* FR2
  + Option 1: 100 MHz
  + Option 2: 400 MHz

Companies should report the CA setting if CA is adopted.

Other system bandwidth can also be optionally evaluated.

Agreements**:**For outdoor scenarios, the BS antenna parameters are as

* Option 1: 64 TxRU, (M, N, P, Mg, Ng; Mp, Np) = (8,8,2,1,1;4,8)
* Option 2: 32 TxRU, (M, N, P, Mg, Ng; Mp, Np) = (8,2,2,1,1,8,2)

Company to report the BS antenna parameters for XR/CG evaluation.

Other BS antenna parameters can also be optionally evaluated.

Agreements**:**For FR2, UE antenna parameters for XR/CG evaluations are as follows.

* Option 1 (Follow Rel-17 evaluation methodology for FeMIMO in R1-2007151)
  + (M, N, P)=(1, 4, 2), 3 panels (left, right, top)
* Option 2 (from TR 38.802 – developed in Rel-14)
  + 4Tx/4Rx: (M, N, P, Mg, Ng; Mp, Np) = (2,4,2,1,2;1,2), (dH,dV) = (0.5, 0.5)λ, the polarization angles are 0° and 90°

Company to report the UE antenna parameters for XR/CG evaluation.

Other UE antenna parameters can also be optionally evaluated.

Agreements**:** For XR/CG evaluation, adopt following assumptions for BS height for Urban Macro

|  |  |
| --- | --- |
| **Parameter** | **Proposed value** |
| **Urban Macro (FR1)** |
| BS height | 25m |

Agreements**:** For Dense urban and Urban Macro, the UE height for indoor UEs is updated as following based on Table 6-1 in TR 36.873.

|  |  |  |
| --- | --- | --- |
|  |  | Urban Micro/Macro cell  with high UE density  (3D-UMi) /(3D-UMa) |
| UE height (*hUT*) in meters | general equation | *hUT*=3(*nfl* – 1) + 1.5 |
| *nfl* for outdoor UEs | 1 |
| *nfl* for indoor UEs | *nfl* ~ uniform(1,*Nfl*) where  *Nfl* ~ uniform(4,8) |

Agreements: At least for XR/CG capacity evaluation, for DL and UL

* Baseline: DL and UL performances are evaluated independently
* Optional: DL and UL performance are evaluated together
* FFS details both the baseline and the optional evaluations

Agreements: For Dense urban for XR/CG evaluation, update the agreement in RAN1 #103e for channel model as follows.

* Dense urban: FR1 and FR2
  + Channel model: ~~UMi~~ UMa. Detailed definition of ~~UMi~~ UMa refers to TR 38.901.

Agreements: For XR/CG evaluation, adopt 12 degree for downtilt for Dense Urban in FR1.

* Other downtilt value can also be optionally evaluated

Agreements: To facilitate further discussion on evaluation of power saving effect of different power saving schemes, the following references are defined.

* Case 1 (baseline): UE power consumption assuming UE is always ON, i.e., UE is always available for gNB scheduling.
* Case 2 (FFS optional or baseline): UE power consumption assuming Rel-15/16 CDRX configuration
  + FFS CDRX configuration details
* Company can also optionally evaluate ~~for~~ other cases, e.g.
  + Genie: UE power consumption assuming that UE is in a sleep state (e.g., micro/light/deep sleep as defined in TR38.840) whenever there is neither DL data reception nor UL transmission. From the gNB scheduling perspective, UE is always available for scheduling, i.e., there is no difference from Baseline in gNB scheduling and corresponding UE Tx/Rx. ~~It is noted that Genie is not a power saving scheme but the result may serve as an upper bound of power saving gain of power saving techniques, which may potentially motivate development of new power saving techniques that can approach the Genie performance.~~
  + R15/16/17 power saving techniques for connected mode, e.g., BWP, PDCCH skipping, search space switching, etc.

**Decision:** As per email posted on Feb 5th,

Agreements:

UE power consumption (i.e., power saving gain of the evaluated scheme) for XR is evaluated in conjunction with impact on latency, user experience, and capacity.  In this regard, the following table is used to collect results for system level simulation from companies as a starting point.

* FFS all UEs or only satisfied UEs are included for obtaining the PS gain

Table 1 Evaluation of UE power saving schemes for e.g., {dense urban, AR, FR1}

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Power Saving Scheme | Power Saving Gain (PSG) compared to Case 1 | | | | #satisfied UEs per cell2 / #UEs per cell3 |
| Baseline | Optional | | |
| Mean PS gain | PS gain of 5%-tile UE in PSG CDF1 | PS gain of 50%-tile UE in PSG CDF1 | PS gain of 95%-tile UE in PSG CDF1 |
| Case 1 | - | - | - | - | K1 / N |
| Case 2 | X1 % | Y1 % | Z1 % | U1% | K2/ N |
| Case X | X2 % | Y2 % | Z2 % | U2% | K3 / N |
|  |  |  |  |  |  |

Note 1: CDF of power saving gains of ~~each~~ UE

Note 2: # of satisfied UEs per cell among # of UEs per cell (=N).

Note 3: # of dropped UEs per cell (=N) that needs to be the same for all power saving schemes to be evaluated.

Note 4: company to provide the detailed simulation assumptions including parameter values for each case, e.g. CDRX parameters

~~Note 5: company can report one or more power saving gain metrics (i.e. mean PS gain or PS gain of 5%/50%/95%/-tile UE in PSG CDF) for each power saving scheme~~

Agreements: For UL UE power consumption evaluation for UE with transmit power X [0,23] dBm, adopt the following

* Option 1 ~~(Baseline)~~: Consider only two Tx power values as defined in TR 38.840
  + Power number is given as **A** for X= [0, M)dBm and **B** for X =[M, 23]dBm, where **A** and **B** (defined in 38.840) correspond to power consumption numbers for a given uplink slot for 0dBm and 23dBm respectively.
    - M = [20]
    - Other value(s) of M can be optionally evaluated
  + ~~Companies to provide detailed assumptions on UE power consumption for Tx power values other than 0 and 23 dBm~~ 
    - ~~E.g. Power number is given as~~ **~~A~~** ~~for X= [0, 20)dBm and~~ **~~B~~** ~~for X =[20, 23]dBm, where~~ **~~A~~** ~~and~~ **~~B~~** ~~(defined in 38.840) correspond to power consumption numbers for a given uplink slot for 0dBm and 23dBm respectively.~~
* Option 2 ~~(FFS mandatory or optional)~~: Linear interpolation method in linear scale for Tx power values other than 0 dBm and 23 dBm
* FFS whether or not to differentiate the two options (e.g., mandatory vs. optional)
* FFS whether or not to consider UE with transmit power less than 0 dBm