**3GPP TSG-SA WG4 Meeting #131S4-250052r1**

**CH, Geneva, 17 – 21 February 2025**

|  |
| --- |
| *CR-Form-v12.2* |
| **CHANGE REQUEST** |
|  |
|  |  | **CR** |  | **rev** | **3** | **Current version:** |  |  |
|  |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* |
|  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Proposed change affects:*** | UICC apps |  | ME | **X** | Radio Access Network |  | Core Network | **X** |

|  |
| --- |
|  |
| ***Title:***  | 5G\_RTP\_Ph2] PSI Guidelines for HEVC tiles |
|  |  |
| ***Source to WG:*** |  |
| ***Source to TSG:*** |  S4 |
|  |  |
| ***Work item code:*** | 5G\_RTP\_Ph2 |  | ***Date:*** | 1 |
|  |  |  |  |  |
| ***Category:*** | B |  | ***Release:*** |  Rel-19 |
|  | *Use one of the following categories:****F*** *(correction)****A*** *(mirror corresponding to a change in an earlier release)****B*** *(addition of feature),* ***C*** *(functional modification of feature)****D*** *(editorial modification)*Detailed explanations of the above categories canbe found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | *Use one of the following releases:Rel-8 (Release 8)Rel-9 (Release 9)Rel-10 (Release 10)Rel-11 (Release 11)…Rel-16 (Release 16)Rel-17 (Release 17)Rel-18 (Release 18)Rel-19 (Release 19)* |
|  |  |
| ***Reason for change:*** |  In FS\_5G\_RTP\_Ph2 KI#13, the following aspect was concluded for the normative work:* *- Consider extending the PSI guidelines in TS 26.522 [2] for the case when a PDU Set is defined as a tile (as opposed to a video frame or slice).*
* *- Coordinate with SA2 and RAN2 on potential benefits of signaling PDU Set type to the 5G network.*

Based on this, an objective was added to the 5G\_RTP\_Ph2 WID:*9. Conduct normative work on guidelines for marking PDU Sets that are not defined as video frames or slices and potentially signalling PDU Set type to the 5G network.*  *NOTE: This objective requires coordination with SA2.*TS 26.522 generally assumes that video PDU Sets (H.264/AVC, H.265/HEVC) either carry frames or slices. However, the general definition of a PDU Set is not limited to these constructs, and it may be beneficial for the application to define PDU Sets differently.In addition to frames and slices, video PDU Sets can be defined as other sub-picture constructs such as H.265/HEVC tiles. Such PDU Sets are particularly relevant for immersive media use cases, where it could be beneficial to assign different importance to different regions of a picture, for example depending on content saliency and/or user’s viewing direction, as in the examples of 360-degree and volumetric video streaming.This requires a dynamic spatial adaptation of the PDU Set Importance (PSI) based on user’s viewport and/or other content-related metadata. Depending on the selected criteria (e.g., the user’s actual viewport), the sender can dynamically modify the PSI for each tile. Alternatively, the sender may consider a group of tiles as a PDU Set. |
|  |  |
| ***Summary of change:*** | Provide guidelines on setting PSI for HEVC tiles |
|  |  |
| ***Consequences if not approved:*** | No PSI guidelines for HEVC tiles  |
|  |  |
| ***Clauses affected:*** | 4.2.6.2.6 |
|  |  |
|  | **Y** | **N** |  |  |
| ***Other specs*** |  | **x** |  Other core specifications  | TS/TR ... CR ...  |
| ***affected:*** |  | **x** |  Test specifications | TS/TR ... CR ...  |
| ***(show related CRs)*** |  | **x** |  O&M Specifications | TS/TR ... CR ...  |
|  |  |
| ***Other comments:*** |  |
|  |  |
| ***This CR's revision history:*** | Rev4 ….Rev3 [S4-250052]: Addressed the comments received at the RTC SWG Telco on Feb 5 regarding the provided example and the offline comments received after the telco.Rev2 [S4aR250045]: Provided an example and different adaptation criteria.Rev1 [S4aR250031]: Removed the signaling of PDU Set type to the 5G network.  |

\* \* \* \* First change \* \* \* \*

### 4.2.6 Guidelines for PDU Set Marking

#### 4.2.6.1 End of Data Burst Field

NOTE: These detailed guidelines are FFS.

#### 4.2.6.2 PDU Set Importance Field

##### 4.2.6.2.1 General

In general, whenever the RAN needs to discard packets (e.g., under congestion situations), it is better to discard packets of lower importance rather than discarding packets randomly. If a discarded packet is critical for the media stream, the QoE may be severely degraded. For this reason, the PDU Set Importance (PSI) field can be used to mark PDU Sets with their importance level. The PSI field can then be used by the RAN to discard PDU sets. In case of congestion, PDU Sets with higher PSI values are more likely to be discarded.

PDU Sets that contain audio data should be assigned a lower PSI value (i.e., higher importance) compared with PDU Sets that contain other media types.

NOTE 1: PDU Sets that carry immersive audio data are not necessarily assigned a lower PSI value compared with the other media PDU Sets. The PSI value of immersive audio PDU Sets is FFS.

PDU Sets that contains the reference frames present in the video bitstream should be assigned a lower PSI value compared with PDU Sets that contain non-reference frames.

NOTE 2: It is assumed that the video bitstream uses referencing structures that have no coding delay caused by out-of-order output, as typically done for low-delay applications.

The following clauses provides the guidelines for the 3GPP video codecs on setting the PSI field in the RTP HE for PDU Set marking. For specific PSI value ranges, refer to clause 4.2.6.2.5.

##### 4.2.6.2.2 H.264/AVC Codec

In an H.264/AVC bitstream, NAL units with the nal\_unit\_type field assigned the value 5 (refer to Table 7.1 in the H.264/AVC specification [2]) are Instantaneous Decoding Refresh (IDR) pictures. When the Type field value in the NAL Unit header of an RTP packet is 5, then the corresponding PDUs in that PDU Set should be set with higher importance.

The parameter set NAL units such as Sequence Parameter Set (SPS) and Picture Parameter Set (PPS) are important for decoding the bitstream. Therefore, PDU sets with a Type field value equal to 7, 8, 13 or 15 (refer to Table 7.1 in the H.264/AVC specification[2]) in the NAL Unit header of the RTP packet should be assigned a higher importance (lower PSI value) relative to PDU Sets with other Type field values.

+---------------+

|0|1|2|3|4|5|6|7|

+-+-+-+-+-+-+-+-+

|F|NRI| Type |

+---------------+

Figure 4.2.6-1: Format of the H.264/AVC NAL unit header

The NAL unit type octet contains the NRI (nal\_ref\_idc) field highlighted in Figure 4.2.6-1. The NRI field indicate the relative transport priority. A value of b00 indicates that the content of the NAL unit is not used to reconstruct reference pictures for inter picture prediction. Such NAL units can be discarded by the RAN (in case of congestion) without risking the integrity of the reference pictures. Values greater than b00 indicate that the decoding of the NAL unit is required to maintain the integrity of the reference pictures. The highest transport priority is b11, followed by b10, and then by b01; finally, b00 is the lowest. PDU Sets with an NRI value b00 should be set with lower importance relative to the PDU Sets with other NRI values. PDU Sets with an NRI value b11 should be set with higher importance relative to the PDU Sets with other NRI field values.

The Type and NRI fields can be used to set the PSI. The PSI value assignment based on the Type and NRI field values is for further study.

##### 4.2.6.2.3 H.265/HEVC Codec

Different from H.264/AVC, H.265/HEVC NAL unit header (shown in Figure 4.2.6-2) is two bytes, contains a 6-bit Type field, a 5-bit LayerID field, a 3-bit TID field, and no NRI field. The Type and TID fields in the NAL unit header indicate the relative transport priority and can be used to set the PSI.

NAL unit types 0–31 indicate Video Coding Layer (VCL) NAL unit types; 32–40 indicate non-VCL NAL unit types. NAL unit types 41–47 are reserved, and NAL unit types 48–63 are unspecified.

+---------------+---------------+

|0|1|2|3|4|5|6|7|0|1|2|3|4|5|6|7|

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

|F| Type | LayerId | TID |

+-------------+-----------------+

Figure 4.2.6-2: Format of the H.265/HEVC NAL unit header

All VCL NAL units of the same access unit have the same NAL unit type, which defines the type of the access unit and its coded picture. There are three basic classes of pictures in H.265/HEVC: intra random access point (IRAP) pictures, leading pictures, and trailing pictures.

In an H.265/HEVC bitstream, NAL units with the nal\_unit\_type field assigned a value in the range 16 to 23 (inclusive) (refer to Table 7.1 in the H.265/HEVC specification [3]) are Intra Random Access Pictures (IRAP) pictures. This includes IDR, CRA, and BLA picture types as well as types 22 and 23, which currently are reserved for future use. When the Type field value in the NAL Unit header of RTP packet is in the range 16 to 23 (inclusive), then the corresponding PDUs in that PDU Set should be assigned higher importance (i.e., lower PSI value).

The parameter set NAL units such as Sequence Parameter Set (SPS), Picture Parameter Set (PPS), Video Parameter Set (VPS) are important for decoding the bitstream. Therefore, PDU Sets with payload Type field value in the NAL Unit header of RTP packet in the range 32 to 34 (inclusive) should be assigned higher importance (lower PSI value).

RFC 7798 [6] specifies Aggregation Packets (APs) to enable the reduction of packetization overhead for small NAL units, such as most of the non-VCL NAL units, which are often only a few octets in size. An AP aggregates NAL units within one access unit. Each NAL unit to be carried in an AP is encapsulated in an aggregation unit. An AP consists of a payload header (denoted as PayloadHdr) followed by two or more aggregation units. In an AP, the Type field in the PayloadHdr is equal to 48. APs are typically used to aggregate parameters sets (VPS, SPS, PPS) into a single packet. When APs are used, the sender should consider the NAL unit types of the aggregation units while assigning the PSI value. For example, if the aggregation unit contains parameter sets, PDU Sets containing those should be assigned higher importance (lower PSI value).

It could be that there are PDUs with different NAL unit types in a PDU Set. For example, if the first PDU in PDU Set is a prefix SEI message or Access Unit Delimiter (AUD), it would be misleading if the sender looked only at the first PDU of the PDU Set to determine the PSI value. The sender should ignore the NAL units with non-VCL NAL unit types 35 and 39 and instead consider NAL unit types of the subsequent VCL NAL units while determining the PSI value for such PDU Sets.

A leading picture is a picture that follows a particular IRAP picture in decoding order and precedes it in output order. There are two types of leading pictures in H.265/HEVC: Random access decodable leading (RADL) pictures and Random access skipped leading (RASL) pictures. A RADL picture is a leading picture that is guaranteed to be decodable when random access is performed at the associated IRAP picture. Therefore, RADL pictures are only allowed to reference the associated IRAP picture and other RADL pictures of the same IRAP picture. A RASL picture is a leading picture that may not be decodable when random access is performed from the associated IRAP picture. Only other RASL pictures are allowed to be dependent on a RASL picture. Hence, in H.265/HEVC bitstreams, RASL pictures can be discarded during random access. H.265/HEVC provides mechanisms to enable specifying the conformance of a bitstream wherein the originally present RASL pictures have been discarded. Consequently, system components can discard RASL pictures, when needed, without worrying about causing the bitstream to become non-compliant.

PDU Sets with Type field value equal to 6 or 7 (refer to Table 7.1 in H.265/HEVC specification [3]) in the NAL Unit header of RTP packet are RADL pictures. PDU Sets with Type field value equal to 8 or 9 (refer to Table 7.1 in H.265/HEVC specification [3]) in the NAL Unit header of RTP packet are RASL pictures. PDU Sets that contain RADL pictures should be assigned lower importance (higher PSI value) relative to the IRAP pictures and higher importance (lower PSI value) relative to the RASL pictures in the bitstream.

In video coding, temporal scalability is the option to decode only some of the frames in a video stream instead of the whole stream. This enables a media server to reduce the bitrate sent towards viewers that don’t have enough bitrate or CPU to handle the whole stream. In H.265/HEVC, pictures with lowest temporal identifier value (TID) are used as reference pictures in the bitstream and are important for decoding the dependent frames. PDU Sets with TID value 1 (lowest possible value) should be set with higher importance (lower PSI value) relative to PDU Sets that have a higher TID value. The PSI value for such pictures should be lower for IRAP pictures and slightly higher for non-IRAP pictures compared to the pictures with higher TID values. Pictures with highest TID value cannot be used as reference pictures and can be discarded at the network level when the throughput is not good, or network conditions are unstable. PDU Sets with higher TID values should be set with lower importance (higher PSI value) compared with the PDU Sets with lower TID values.

In H.265/HEVC, each leading picture and trailing picture type has two type values. The even picture type numbers indicate sub-layer non-reference pictures and odd picture type numbers indicate sub-layer reference pictures. An encoder can use the sub-layer non-reference picture types for pictures that are not used for reference for prediction of any picture in the same temporal sub-layer. Note that a sub-layer non-reference picture may still be used as a reference picture for prediction of a picture in a higher temporal sub-layer.

PDU Sets that contain sub-layer reference picture types should be assigned a lower PSI value compared with the PDU sets with the corresponding sub-layer non-reference picture types.

##### 4.2.6.2.4 PSI based on affected PDU Sets

When the transport layer is forced to perform immediate dropping/discarding of a PDU Set but has a freedom of selection among the PDU Sets, the PDU Set with smaller degrees of artifact would be the better choice in most cases. Dropping of a PDU Set may corrupt the decoded output of itself and the other PDU Sets though they may already be transmitted perfectly to the receiving end or yet in a queue waiting to be transmitted. The degrees of artifact can be explicitly transferred as the number of affected frames which precedes/follows the PDU Set, or can be implicitly transferred as the importance value where the lower value means the higher PDU Sets are affected while higher values proportionally mean less number of PDU Sets are affected, for example. By considering such a quantization of various affected PDU Sets can be translated into importance field, using 4 bits to represent 16 possible size ranges is recommended.

The information on the size of propagation error which caused by the dropping of each PDU Set may be provided by the application layer. The information may present the size of error propagation implicitly with a proportional mapping of error propagation size to an index such as the importance of the PDU Set in the media stream.

The importance value of a PDU Set in PDU Set Information (PSI) RTP HE is set as follows:

- The error propagation size is mapped to importance field value. The higher the error propagation size of a PDU Set, that PDU Set is more important, and it shall be assigned the lower PSI value. PDU Sets with low error propagation are of less importance and the PSI value for such PDU Sets shall be higher compared to PDU sets with higher error propagation size.

##### 4.2.6.2.5 PSI mapping based on PDU Set dependencies

RTP senders should consider that multiplexed RTP streams are treated as a single Multimedia Session and set the PSI field accordingly, i.e., the PSI field for one bitstream that depends on other RTP stream(s) in the same Multimedia Session may need to be set taking the PSI field for PDU Sets in other multiplexed RTP streams into account. In some cases, dependencies can exist across bitstreams even when they are not multiplexed, particularly for XR services.

In case of such dependencies, it may not be sufficient to set the PSI values based on codecs and media types alone. PSI values shall be set in this case based on the following, which are listed in an increasing order of importance, i.e., decreasing order of PSI values.

- The PDU Set is considered not necessary for the processing of any other PDU Set. Such PDU Sets should be assigned the highest PSI values 14-15. When multiplexing, if a PDU Set is assigned PSI value of 15, similar PDU Sets of all streams should be assigned the PSI value 15 to prevent unfair treatment. If interdependency is known, e.g., in stereo streams (left eye is more important than right eye), then the more important stream can be assigned the PSI value 14.

- In H.264/AVC, these include the PDU Sets with an NRI value equal to b00 in the NAL unit header.

- In H.265/HEVC, the NAL unit header does not contain a field like NRI that indicates the relative transport priority. Hence, it is up to the application to identify such PDU Sets.

- The PDU Set is necessary for the processing of some PDU Sets of the stream to which it belongs. Such PDU Sets should be assigned a PSI value in the range 9-13 (inclusive). The lower end of the range should be used for IDR/IRAP pictures since they are more important for decoding of the bitstream.

- In H.264/AVC, these include:

- IDR pictures with nal\_unit\_type equal to 5

- Non-IDR pictures with nal\_unit\_type in the range 1 to 4 (inclusive)

- In H.265/HEVC, these include:

- IRAP pictures with nal\_unit\_type field assigned a value in the range 16 to 23 (inclusive)

- RADL or RASL pictures with nal\_unit\_type in the range 6 to 9 (inclusive)

- The PDU Set is necessary for the processing of all the other PDU Sets of the stream to which it belongs. Such PDU Sets should be assigned a PSI value in the range 6-8 (inclusive).

- In H.264/AVC, these include:

- SPS, PPS, i.e., NAL units with the nal\_unit\_type field equal to 7, 8, 13 or 15

- In H.265/HEVC, these include:

- SPS, PPS, VPS, i.e., NAL units with the nal\_unit\_type field in the range 32 to 34 (inclusive)

- The PDU Set is necessary for the processing of some PDU Sets of the stream to which it belongs and also necessary for the processing of some PDU Sets of some other streams to which it does not belong. Such PDU Sets should be assigned a PSI value in the range 4-5 (inclusive).

NOTE 1: Values in this and lower range shall be used for assigning PSI values to PDU Sets in multiplexed streams or if dependencies exist across non-multiplexed bitstreams. Use cases for those cases are FFS. In case only a single RTP stream is present, the ranges provided by the previous bullet points shall be used.

NOTE 2: Considerations for multiplexed audio streams are FFS.

- The PDU Set is necessary for the processing of all PDU Sets of the stream to which it belongs and also of some other streams to which it does not belong. Such PDU Sets should be assigned a PSI value in the range 2-3 (inclusive).

- The PDU Set is necessary for the processing of all PDU Sets of all streams. Such PDU Sets should be assigned the lowest PSI value 1.

##### 4.2.6.2.6 PSI Guidelines for H.265/HEVC tiles

When using tiled H.265/HEVC encoding, an RTP sender may define a PDU Set as a collection of one or more HEVC tiles, where each PDU Set comprises all PDUs carrying the data for these tiles. This definition is particularly relevant for immersive media use cases, as it enables assignment of different PSI values to distinct regions of a picture. Moreover, PSI values could be dynamically adjusted based on factors such as content saliency and the user’s viewing direction, which is especially beneficial for applications like 360-degree video streaming and volumetric video streaming.

 HEVC Motion Constrained Tile Set() .

NOTE: In HEVC, intra prediction process for each tile is independen from other tiles within the same picture. However, in terms of inter-picture prediction, motion vectors for each tile can point beyond tile boundaries in the reference pictures. The MCTS technique in HEVC constrains the inter-picture prediction process within a specified set of tiles to reference only regions within the same set of tiles in previous pictures in decoding process. When MCTS is used, the tiling grid of the picture is required to be constant within a coded video sequence. An MCTS can consist of one or more HEVC tiles.

An RTP sender may dynamically modify the PSI for each MCTS over time based on specific criteria. For example, it could assign lower PSI to PDU Sets within the the user’s (actual or predicted) viewport or pose, thereby prioritizing the regions currently viewed by the user and those likely to be based on the viewport prediction. The PSI adaptation criteria may also include eye gaze or other content/scene-related metadata such as content saliency.

For example, a 360-degree video sender could determine which spatial regions are likely to fall into the user’s viewport based on the user’s actual or predicted pose and assign lower PSI values to PDU Sets covering regions that are more likely to fall into the user’s viewport.

An example scenario demonstrating dynamic adaptation of PSI is shown in Figure 4.2.6.2.6-1. In the example, it is assumed that an MCTS comprises a single HEVC tile. However, the operation can be generalized to apply to other cases.

At time t=t1, the sender assigns PSI=3 to the HEVC tiles that are predicted to be fully within the user’s viewport at t=t2, and PSI=5 to the HEVC tiles that are predicted to be partially within the user’s viewport at t=t2. Background tiles (i.e., tiles that are predicted to be outside the viewport at t=t2) are assigned PSI=10.

Note that the PSI values in the example are chosen arbitrarily and the actual configuration depends on the application requirements. The sender could also assign PSI in an even more fine-granular way by considering the percentage of a tile covered by the user’s predicted viewport.



**Figure 4.2.6.2.6-1: Example scenario for dynamic adaptation of PSI**