**Source: Nokia Corporation1**

**Title: [5G\_RTP] Definition of the Pose RTP HE for 3DoF pose**

**Agenda Item: 10.7**

**Document for: Discussion and Agreement**

# 1 Introduction

In SA4#126, an RTP header extension for pose was agreed, which may either be used to indicate the pose used for rendering the media (rendered pose) or the XR pose sent to another UE or to a server. One of the remaining issues is the definition of the header extension for 3DoF.

Although it is possible to use the existing RTP header extension for 3DoF by setting the position fields x,y,z to 0 or an unused value, this would mean that 12 bytes are wasted. As per the comments received at the RTC SWG telco on January 10, this may have considerable effects on audio data where typical bitrates are much lower compared to video.

To address the above concerns, this contribution introduces a 3DoF version of the RTP header extension for pose that does not include the position fields.

# 2 Proposal for TS 26.522

**===================================== Change 1 =====================================**

### 4.4.3 RTP Header Extension for Pose

An RTP sender that uses RTP to deliver pre-rendered video streams to a UE should include an RTP header extension for pose to indicate the XR pose used for rendering the media (rendered pose). The RTP header extension for pose may be used for signaling either a 6DoF XR pose or a 3DoF XR pose. The RTP header extension for pose may also be used with audio streams.

The RTP header extension for pose may also be used by a UE to indicate the XR pose to another UE or to a server.

An RTP client that supports the RTP header extension for pose shall negotiate the use the of the extension using SDP. The signaling of the RTP header extension for pose shall follow the SDP signaling design, the syntax, and semantics of the "extmap" attribute as outlined in RFC8285. The header extension shall be registered with IANA.

For IANA registration, the "reference" field in the registry is 3GPP TS 26.522.

The ABNF syntax for this header extension extends the "extmap" attribute as follows:

*extensionname* = "urn:3gpp:xr-pose"

*extensionattributes* = "3DOF" / "6DOF" ["media:" 1\*(SP token)]The extension attribute “3DOF” indicates that the sender uses the RTP header extension to signal a 3DoF XR pose, i.e., an XR pose that does not include the position fields x, y, z.

An RTP client that supports the RTP header extension for pose and receives an SDP offer with "a=extmap" attribute with the URN "urn:3gpp:xr-pose" and the extension attribute “3DOF” shall include the extension attribute “3DOF” in the SDP answer.

The extension attribute “6DOF” indicates that the sender uses the RTP header extension to signal a 6DoF XR pose, i.e., an XR pose that includes both the position fields x, y, z and the orientation fields rx, ry, rz, rw.

An RTP client that supports the RTP header extension for pose and receives an SDP offer with "a=extmap" attribute with the URN "urn:3gpp:xr-pose" and the extension attribute “6DOF” shall include the extension attribute “6DOF” in the SDP answer.

The extension attribute "media" is followed by a list of tokens for "mid" (as defined in RFC 5888) for media streams that can reuse the pose included in the RTP header extension. Further details on reuse are provided later in the section.

An RTP client that supports the RTP header extension for pose and receives an SDP offer with "a=extmap" attribute with the URN: "urn:3gpp:xr-pose" shall remove the attribute from the answer for any media that will not use the extension, and retain it for any media that will use it.

If the RTP header extension for pose is used by a server, the server should use the RTP header extension for pose to associate the selected pose with the rendered frame. The server delivers the rendered frames using one or more video streams, depending on the view and projection configuration that is selected by the UE.

If negotiated successfully, an RTP sender should add the RTP header extension for pose to the RTP stream. The frequency of RTP header extension for pose shall be at least once in a frame. It may be sent more often but not necessarily in every RTP packet.

The 2-byte (RFC 8285) RTP header extension format shall be used for signalling the RTP header extension. Format of the header extension for 6DoF XR pose is shown below.

0 1 2 3  
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1  
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+  
| 0x100 |appbits| length |  
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+  
| ID | L | rx …

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

| ry …

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+  
 | rz …

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+  
 | rw …

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+  
 | x …

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+  
 | y …

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+  
 | z …

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+  
| | XR timestamp …

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

| XR timestamp continued …

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

| XR timestamp continued | action\_id #1 |

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

| action\_id #2 | ... |

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

If the RTP header extension is used for signaling a 3DoF XR pose, the fields x, y, z shall be omitted. Format of the header extension for 3DoF XR pose is shown below.

0 1 2 3  
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1  
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+  
| 0x100 |appbits| length |  
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+  
| ID | L | rx …

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

| ry …

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+  
 | rz …

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+  
 | rw …

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+  
 | | XR timestamp …

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

| XR timestamp continued …

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

| XR timestamp continued | action\_id #1 |

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

| action\_id #2 | ... |

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

The fields rx, ry, rz, rw, x, y, z are defined in single-precision floating-point format (binary32 as per ISO/IEC 60559:2020).

**rx (32 bits):** x coordinate of the orientation quaternion of the pose.

**ry (32 bits):** y coordinate of the orientation quaternion of the pose.

**rz (32 bits):** z coordinate of the orientation quaternion of the pose.

**rw (32 bits):** w coordinate of the orientation quaternion of the pose.

**x (32 bits):** x coordinate of the position of the pose in meters.

**y (32 bits):** y coordinate of the position of the pose in meters.

**z (32 bits):** z coordinate of the position of the pose in meters.

**[**

**XR timestamp (64 bits)**: Timestamp for the pose. If the header extension is used for rendered pose, this timestamp indicates the predicted XR runtime display time. Otherwise, this timestamp indicates the associated XR runtime display time for the predicted XR pose. XR timestamp uses the XR system clock and is represented in nanoseconds. The timestamp is passed to the XR runtime together with the rendered swapchain images (e.g. as part of the xrEndFrame call in OpenXR).

]

NOTE 1: It is left to the discretion of the application how to use the XR timestamp.

[Editor’s Note: Rendered pose is sent from the SR server to the SR client. If the pose is not rendered pose, it is sent from a UE to a server or to another UE.]

**action\_id (32 bits)**: A list of actions corresponding to the pose x, y, z, rx, ry, rz, rw coordinates. An action\_id uniqely identifies an action and it may be an action identifier as defined in the action format of TS 26.119 Clause 6.2.3. The number of action identifiers in one RTP header extension for pose shall be no more than 10. Hence, the size of the header extension is 36+2\*n, if a 6DoF XR pose is used, or 24+2\*n, if a 3DoF XR pose is used, where n is the number of action identifiers in the header extension.

If the RTP header extension for pose is sent by a server, it should contain an action\_id field as defined above, with the list of action identifiers identifying the processed actions for the rendering of the frame.

If the RTP header extension for pose is sent by a UE, it should contain an action\_id field as defined above, with the list of action identifiers identifying the action for which the pose coordinates apply.

NOTE 2: A peer to a UE XR client should be aware of the UE actions configuration in an action space. Signalling aspects for the UE actions configuration are defined in other specifications such as TS 26.119 and TS 26.565.

NOTE 3: An XR server should be aware of the XR space used by the XR client for the pose fields defined above. Signalling aspects for this XR space are defined in other specifications such as TS 26.119 and TS 26.565.