**Source:** Nokia Corporation[[1]](#footnote-2)

**Title: [5G\_RTP] Reuse of RTP header extensions in multiple streams**

**Document for** Agreement

**Agenda item:** 10.8

# Introduction

Use cases in 5G\_RTP require sending metadata along with media which can be achieved using RTP header extensions (HE). For example, a RTP HE can be used to send pose together with the rendered frame, as agreed in SA4 #122.

When both video and audio are delivered together, or when either audio or video is delivered using multiple real-time streams (e.g., depth + texture), multiple RTP streams may be associated with the same header extension data, e.g., the same pose may have been used for generating multiple streams. This may lead to sending the same/related header extension data multiple times in different streams.

The required amount of data for pose information can be quantified based on the information from the TR 26.926 (FS\_XRtraffic) and 5G\_RTP PD. According to the 5G\_RTP PD, the HE for rendered pose has a size of 36+2\*n bytes, where n is the number of action identifiers in the header extension. A maximum limit of 10 actions per RTP header extension for rendered pose is allowed. Assuming 5 actions each with 2-byte IDs, a total of 56 bytes are required for a single pose sample.

Clause 6.5.3.1 of TR 26.926 lists the mean and max video frame sizes for the recommended XR split rendering configurations. Among the listed configurations, VR 2-6 and VR 2-8 have the lowest and highest packet sizes, respectively. Assuming that the pose HE (56 bytes) is added to every packet of a rendered frame, 1500 bytes per packet and a frame rate of 60 fps, the mean/max amount of data required for pose is 66-87 kbps for VR 2-6 and 184-235 kbps for VR 2-8, which are non-negligible data volumes for the air interface. Note that these values are given per eye buffer and will be even higher for higher frame rates.

This contribution defines a generic mechanism to reuse the rendered pose RTP HE added to one stream in multiple streams.

# Reusing RTP header extensions

# 2.1. Basic syntax

According to RFC 8285, an RTP header extension (HE) potentially used in a media stream is indicated using the extmap attribute of the form:

a=extmap:<value>[“/”<direction>] <URI> <extensionattributes>

For the rendered pose HE, a new parameter is included in the <extensionattributes>:

media:<mid-list>

The parameter indicates, using the media ID (MID) of media, to which other media the HE with the ID indicated by <value> is applicable to. <mid-list> is a string containing a list of MID values separated by a semi colon.

An example SDP description using this extension attribute is presented below. In the example, the rendered pose RTP HE with URI urn:3gpp:xr-rendered-pose provided in the video stream with MID m1 is also applicable to another video stream with MID m3.

v=0

o=alice 2890844526 2890844526 IN IP4 host.atlanta.example.com

s=SDP Session

c=IN IP4 host.atlanta.example.com

t=0 0

m=application 1001 UDP/DTLS/SCTP webrtc-datachannel

a=sendonly

m=video 23458 RTP/AVP 96

a=mid:m1

a=recvonly

a=rtpmap:96 H264/90000

**a=extmap:1 urn:3gpp:xr-rendered-pose media:m3**

m=audio 23468 RTP/AVP 97

a=mid:m2

a=recvonly

a=rtpmap:97 PCMU/8000

m=video 23478 RTP/AVP 97

a=mid:m3

a=recvonly

a=rtpmap:96 H264/90000

## Extended syntax

In some scenarios, the HE data in one stream may not exactly match the other stream due to different frame rates of two streams (which leads to different time instances of the HE data) or offsets between two streams (e.g., pose difference in two stereo views). In order to address these use cases, additional parameters can be added to the <extensionattributes> of the rendered pose HE as follows:

a=extmap:<value> urn:3gpp:xr-rendered-pose media:<mid-list> interp:<value> offset:<value>

interp indicates the interpolation method when pose in one media (e.g., video) does not match all media samples in the other media (e.g., audio). <value> can be provided as a lookup table for different semantics. Example semantics are given as follows:

* + value=1 – client uses the last x values to interpolate the pose for the current time of the media that is re-using the RTP HE
  + value=2 – client uses the pose of the last received media sample
  + value=3 – client uses the pose to be a constant if the value is missing.

NOTE: It is left to the discretion of the application to select an appropriate value for x.

offset indicates the pose translation between one media (e.g., audio) that provides the RTP HE and other media (e.g., video) that reuses the RTP HE. Another example is when stereoscopic video is used, and the offset values provide the pose difference between the two views. The offset values are given in x, y, z coordinates and indicate the position difference in meters according to the coordinate system associated to the stream containing the RTP HE.

The below line demonstrates the usage of these extensions in the example given above. Interpolation mode 1 is used such that the last x samples are used to interpolate the value of the current sample in the dependent stream. The offset is given as 5cm in x coordinates; this could be, for example, the difference between two stereo views.

...

a=extmap:1 urn:3gpp:xr-rendered-pose media:m3 interp:1 offset:0.05;0

# Proposal

We propose to incorporate the above text into 5G\_RTP PD.

1. Contact: Serhan Gül, Saba Ahsan, Igor Curcio, Nokia Technologies, Finland. Emails: íserhan.guel, saba.ahsan, igor.curcioý@nokia.com. [↑](#footnote-ref-2)