        SDP Overlay Grouping framework for immersive telepresence media streams
                       draft-overlay-grouping-00

Abstract

   This document defines semantics that allow for signalling a new SDP
   group "OL" for overlays in an immersive telepresence session. The
   "OL" attribute can be used by the application to relate all the
   overlay media streams enabling them to be added as overlay on top of
   the immersive video. The overlay grouping semantics is required, if
   the media data is seperate and transported via different protocols.

Status of This Memo

   This Internet-Draft is submitted in full conformance with the
   provisions of BCP 78 and BCP 79.

   Internet-Drafts are working documents of the Internet Engineering
   Task Force (IETF). Note that other groups may also distribute
   working documents as Internet-Drafts. The list of current Internet-
   Drafts is at https://datatracker.ietf.org/drafts/current/.

   Internet-Drafts are draft documents valid for a maximum of six months
   and may be updated, replaced, or obsoleted by other documents at any
   time. It is inappropriate to use Internet-Drafts as reference
   material or to cite them other than as "work in progress."

   This Internet-Draft will expire on April 23, 2021.

Copyright Notice

Table of Contents

1. Introduction

Telepresence [RFC7205] can be described as a technology that allows a person the experience of "being present" at a remote location for video as well as audio telepresence sessions, so as to enable the users sense of realism and presence [TS26.223] . SDP [RFC4566] is being predominantly used for describing the format for multimedia communication session for telepresence conferencing. These use open standards such as RTP [RFC3550] and SIP [RFC3261] .

An SDP session may contain more than one media lines with each media line identified by "m"=line. Each line denotes a single media stream. If multiple media lines are present in a session, a receiver needs to identify relationship between those media lines.

Overlay media stream can be defined as a piece of visual media which can be rendered over an immersive video or image or over a viewport [ISO23090] . When an overlay is transmitted, its media stream needs to be uniquely identified across multiple SDP descriptions exchanged with different receivers so that the streams can be identified in terms of its role in the session irrespective of its media type and transport protocol.

In an immersive telepresence session, one media is streamed as an immersive stream whereas other media streams are overlaid on top of the immersive video/image. An end user can stream more than one overlay, subject to its decoding capacity. When multiple overlay streams are transmitted within a session, the end application upon receiving, needs to be able to relate the media streams to each other. This can be achieved by SDP grouping framework by using the

"group" attribute that groups different "m" lines in a session.
However, the current SDP signalling framework does not provide such
grouping semantics for overlays.

This document describes a new SDP group semantics for grouping the
overlays when an immersive media stream is transmitted for
telepresence conferencing. SDP session description consists of one
or multiple media lines know as "m" lines which can be identified by
a token carried in a "mid" attribute. The SDP session describes a
session-level group level attributes that groups different media
lines using a defined group semantics. The semantics defined in this
memo is to be used in conjuction with [RFC5888] titled "The Session
Description Protocol (SDP) Grouping Framework".

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
"SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
document are to be interpreted as described in [RFC2119] .

3. Overview of Operation

A non-normative description of SDP overlay group semantics is
described in this section. An immersive stream for a telepresence
session may consist of one or more conference rooms with a 360-degree
camera and the remote users using head mounted display for streaming.
"Participant cameras" are used to capture the conference participants
whereas "presentation cameras" or "content cameras" can be used for
document display [RFC7205] . The remote participant can stream any
of the available immersive video in the session as background whereas
other available streams such as the presentation stream or 2D video
from any other room or participant can be used as an overlay on top
of the immersive video/image.

A user with a head mounted display may stream more than one overlay
in a single SDP session. These overlay streams are transmitted via
"m" line in SDP session description. Each "m" line in the session
description is identified by a token carried via the "mid" attribute.
When multiple overlay streams are transmitted within a session, the
end application upon receiving, needs to be able to relate the media
streams to each other. This is achieved by using the SDP grouping
framework [RFC5888]. The session descriptions carries session-level
"group" attribute for the overlays which groups different "m" lines
using overlay(OL) group semantics.

4. **Overlay Stream Group Identification Attribute**

   The "overlay media stream identification" attribute is used to
   identify overlay media streams within a session description. In a
   overlay group, the media lines MAY have different media contents.
   Its formatting in SDP [RFC4566] is described by the following
   Augmented Backus-Naur Form (ABNF) [RFC5234] :

   mid-attribute = "a=mid:" identification-tag
   identification-tag = token
                        ; token is defined in RFC4566

   This documents defines a new group semantics "OL" identification
   media attribute, which is used to identify overlay group media
   streams within a session description. It is used for grouping the
   media streams for different overlays together within a session. An
   application that receives a session description that contains "m"
   lines grouped together using "OL" semantics MUST overlay the
   corresponding media streams on top of the immersive media stream.

5. **Use of group and mid**

   All group and mid attributes MUST follow the rules defined in
   [RFC5888]. The "mid" attribute should be used for all "m" lines
   within a session description . If for any "m" lines within a session,
   no "mid" attribute is identified for a session description, the
   application MUST NOT perform any media line grouping. If the
   identification-tags associated with "a=group" lines do not map to any
   "m" lines, it MUST be ignored.

   group-attribute ="a=group:" semantics
                    *(SP identification-tag)
   semantics = "OL" / semantics-extension
   semantics-extension = token
                        ; token is defined in RFC4566

6. **Example of OL**

   The following two examples show a session description for overlays in
   an immersive telepresence conference. The "group" line indicates
   that the "m" lines with tokens 1 and 2 are grouped for the purpose of
   overlays and intended to be overlaid on top of the immersive video.

   In the first example shown below, two overlays are being transmitted.
   The first media stream (mid:1) carries the video stream, and the
   second stream (mid:2) contains an audio stream.

```
v=0
o=Alice 292742730 29277831 IN IP4 233.252.0.74
c=IN IP4 233.252.0.79
t=0 0
a=group:OL 1 2
m=video 30000 RTP/AVP 31
a=mid:1
m=audio 30002 RTP/AVP 31
a=mid:2
```

The second example, below, uses 'content' attribute with the media
streams which are transmitted for overlay purpose.

```
v=0
o=Alice 292742730 29277831 IN IP4 233.252.0.74
c=IN IP4 233.252.0.79
t=0 0
a=group:OL 1 2
m=video 30000 RTP/AVP 31
a= content:slides
a=mid:1
m=video 30002 RTP/AVP 31
a=content:speaker
a=mid:2
```

7. Security Considerations

All security considerations as defined in [RFC5888] apply:

Using the "group" parameter with FID semantics, an entity that
managed to modify the session descriptions exchanged between the
participants to establish a multimedia session could force the
participants to send a copy of the media to any destination of its
choosing.

Integrity mechanisms provided by protocols used to exchange session
descriptions and media encryption can be used to prevent this attack.
In SIP, Secure/Multipurpose Internet Mail Extensions (S/MIME)
[RFC8550] and Transport Layer Security (TLS) [RFC8446] can be used to
protect session description exchanges in an end-to-end and a hop-
byhop fashion, respectively.

8. IANA Considerations

The following contact information shall be used for all registrations
included here:

```
   Contact:          Rohit Abhishek
                     email: rabhishek@rabhishek.com
                     tel  : +1-816-585-7500
```

   This document defines a new SDP group semantics for overlays for a
   immersive telepresence session. This attribute can be used by the
   application to group all the overlays in a session. Semantics values
   to be used with this framework should be registered by the IANA
   following the Standards Action policy [RFC8126]. This document adds
   a new group semantics and follows the registry group defined in
   [RFC5888].

   The following semantics needs to be registered by IANA in Semantics
   for the "group" SDP Attribute under SDP Parameters.

```
   Semantics            Token          Reference
   ----------------------------------------------
   Overlay              OL             RFCXXXX
```

   The "OL" attribute is used to group different media streams to be
   rendered as overlays. Its format is defined in Section 4 .

   The IANA Considerations section of the RFC MUST include the following
   information, which appears in the IANA registry along with the RFC
   number of the publication.

   o A brief description of the semantics.

   o Token to be used within the "group" attribute. This token may be
      of any length, but SHOULD be no more than four characters long.

   o Reference to a standards track RFC.

9. References

9.1. Normative References

   [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
              Requirement Levels", BCP 14, RFC 2119,
              DOI 10.17487/RFC2119, March 1997,
              <https://www.rfc-editor.org/info/rfc2119>.

   [RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston,
              A., Peterson, J., Sparks, R., Handley, M., and E.
              Schooler, "SIP: Session Initiation Protocol", RFC 3261,
              DOI 10.17487/RFC3261, June 2002,
              <https://www.rfc-editor.org/info/rfc3261>.

   [RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V.
              Jacobson, "RTP: A Transport Protocol for Real-Time
              Applications", STD 64, RFC 3550, DOI 10.17487/RFC3550,
              July 2003, <https://www.rfc-editor.org/info/rfc3550>.

   [RFC4566] Handley, M., Jacobson, V., and C. Perkins, "SDP: Session
              Description Protocol", RFC 4566, DOI 10.17487/RFC4566,
              July 2006, <https://www.rfc-editor.org/info/rfc4566>.

   [RFC5234] Crocker, D., Ed. and P. Overell, "Augmented BNF for Syntax
              Specifications: ABNF", STD 68, RFC 5234,
              DOI 10.17487/RFC5234, January 2008,
              <https://www.rfc-editor.org/info/rfc5234>.

   [RFC5888] Camarillo, G. and H. Schulzrinne, "The Session Description
              Protocol (SDP) Grouping Framework", RFC 5888,
              DOI 10.17487/RFC5888, June 2010,
              <https://www.rfc-editor.org/info/rfc5888>.

   [RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for
              Writing an IANA Considerations Section in RFCs", BCP 26,
              RFC 8126, DOI 10.17487/RFC8126, June 2017,
              <https://www.rfc-editor.org/info/rfc8126>.

   [RFC8446] Rescorla, E., "The Transport Layer Security (TLS) Protocol
              Version 1.3", RFC 8446, DOI 10.17487/RFC8446, August 2018,
              <https://www.rfc-editor.org/info/rfc8446>.

   [RFC8550] Schaad, J., Ramsdell, B., and S. Turner, "Secure/
              Multipurpose Internet Mail Extensions (S/MIME) Version 4.0
              Certificate Handling", RFC 8550, DOI 10.17487/RFC8550,
              April 2019, <https://www.rfc-editor.org/info/rfc8550>.

9.2. Informative References

   [ISO23090]
              "Information technology -- Coded representation of
              immersive media -- Part 2: Omnidirectional MediA Format
              (OMAF) 2nd Edition", ISO ISO 23090-2:2020(E), February
              2020.

   [RFC7205] Romanow, A., Botzko, S., Duckworth, M., and R. Even, Ed.,
              "Use Cases for Telepresence Multistreams", RFC 7205,
              DOI 10.17487/RFC7205, April 2014,
              <https://www.rfc-editor.org/info/rfc7205>.

   [TS26.223]
              "3rd Generation Partnership Project; Technical
              Specification Group Services and System Aspects;
              Telepresence using the IP Multimedia Subsystem (IMS);
              Media Handling and Interaction", 3GPP TS26.223, March
              2020.

Author's Address

   Rohit Abhishek
   Tencent
   2747 Park Blvd
   Palo Alto 94588
   USA

   Email: rabhishek@rabhishek.com