

mmusic
Internet-Draft
Intended status: Standards Track
Expires: August 5, 2021

R. Abhishek
S. Wenger
Tencent
February 1, 2021

SDP Superimposition Grouping framework
draft-abhishek-mmusic-superimposition-grouping-01

Abstract

This document defines semantics that allow for signaling a new SDP group "supim" for superimposed media in an SDP session. The "supim" attribute can be used by the application to relate all the superimposed visual media streams enabling them to be added as an overlay on top of any visual media stream. The superimposition grouping semantics is helpful, if the media data is separate and transported via different sessions.

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 August 5, 2021.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of

the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Terminology	4
3. Superimposition Group Identification Attribute	4
4. Use of group and mid	5
5. "superposition" Attribute for Superimposition Group Identification Attribute	5
6. Example of Supim	6
7. Relationship with CLUE (informative)	7
8. Security Considerations	7
9. IANA Considerations	7
10. Acknowledgements	9
11. References	9
11.1. Normative References	9
11.2. Informative References	10
Authors' Addresses	10

1. Introduction

Media superimposition herein is defined to be a visual media (video/image/text) which is superimposed on top of an already existing visual media such that the resulting foreground and background media can be displayed simultaneously. Superimposition can be recursive in that visual media that is superimposed against its background can, in turn, be the background of another superimposed visual media. The superimposed visual media displayed over a background media content may be anywhere between opaque and transparent. Examples of applications for video superimposition include real-time multi-party gaming, where these superimposed media maybe used to provide additional details or stats about each player, or multi-party teleconferencing where visual media from users in the teleconference may be superimposed on a background media or over each other. An example is shown in the figure below, where three foreground media have been superimposed over a background media, with one foreground media being partly superimposed over another foreground media.

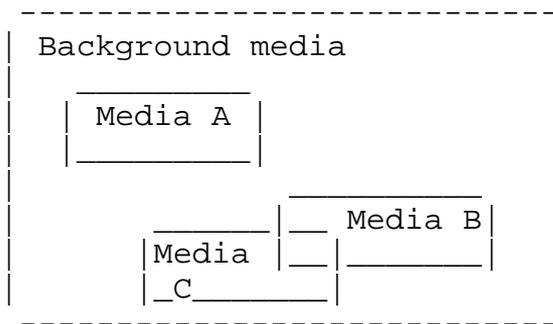


Figure 1: A example of media superimposition

SDP is predominantly used for describing the format for multimedia communication sessions. Many SDP-based systems use open standards such as RTP [RFC3550] for media transport and SIP [RFC3261] for session setup and control. An SDP session may contain more than one media description with each media description identified by "m"=line. Each line denotes a single media stream. If multiple visual media lines are present in a session, at present, their superimposition (foreground/background) relationship at the rendering device is undefined. This memo introduces a mechanism in which certain rendering information becomes available. The rendering information herein is limited to the foreground/background relationship of each grouped media vis-a-vis each other through a layer order value, and optionally a transparency value. Where, spatially, the media is rendered is not covered by this memo, and is in many application scenarios a function of the user interface. The CLUE framework [RFC8845] is available when the application requires defining capture (camera ports), and their geo-spatial relationship to each other is needed. The superimposition grouping as described below enables a compliant receiver/renderer implementation to know the relative relevance of the visual media as coded by the sender(s) and, in a compliant implementation, observed by the renderer through superimposition when needed. Of course, assuming sufficient screen real-estate, a renderer may not have to rely on superimposition mechanisms at all--when there is enough screen real-estate available, a valid display strategy may well be to show all media without overlapping and hence without superimposition. However, when the screen real-estate becomes insufficient, then the information provided by the mechanisms defined in this memo can be used to order (in the sense of foreground to background) the visual media according to a hierarchy chosen by the sender or a middlebox, and based on their application knowledge.

When multiple superimposed streams are transmitted within a session, the receiver needs to be able to relate the media streams to each

other. This is achieved by the SDP grouping framework [RFC5888] by using the "group" attribute that groups different "m" lines in a session. By using a new superimpose group semantic defined in this memo, a group's media streams can be uniquely identified across multiple SDP descriptions exchanged with different receivers, thereby identifying the streams in terms of their role in the session irrespective of its media type and transport protocol. These superimposed streams within the group may be multiplexed based on the guidelines defined in [draft-ietf-avtcore-multiplex-guidelines-12].

This document describes a new SDP group semantics for grouping the superimposition in an SDP session. An SDP session description consists of one or multiple media lines known as "m" lines which can be identified by a token carried in a "mid" attribute. The SDP session describes a session-level group level attribute that groups different media lines using a defined group semantics. The semantics defined in this memo is to be used in conjunction with "The Session Description Protocol (SDP) Grouping Framework"[RFC5888].

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. Superimposition Group Identification Attribute

The "superimposition media stream identification" attribute is used to identify the relationship of superimposed media streams within a session description. In a superimposition group, the media lines MAY have different media formats but, to be meaningful, SHOULD be visual media. There is no defined behavior for the rendering of non-visual media being grouped in a superimposition group. Its formatting follows [RFC5888] in the use of 'mid' attribute to identify the media line to be included in the superimposition.

This document defines a new group semantics "supim" identification media attribute, which is used to identify superimposition group media streams within a session description. It is used for grouping the foreground and the background media streams intended for the purpose of composition with foreground media to be superimposed over the background media stream. An application that chooses to implement the extension, receives a session description that contains "m" lines grouped together using "supim" semantics MUST superimpose the foreground media streams on top of the background media stream in case there is overlap. For non-supporting devices, these media streams are treated as independent media streams.

4. Use of group and mid

All group and mid attributes MUST follow the rules defined in [RFC5888]. The "mid" attribute MUST be used for all "m" lines covering visual media within a session description for which a foreground/background relationship is to be defined. The foreground/background relationship of visual media within a session description that is not covered in a group is undefined. No more than one group MUST be used within one session. If the identification-tags associated with "a=group" lines do not map to any "m" lines, it MUST be ignored.

```
semantics = "supim" /; semantics extension
           as defined in RFC5888
```

5. "superposition" Attribute for Superimposition Group Identification Attribute

This memo defines a new media-level attribute, "superposition", with the following ABNF [RFC5234]. The identification-tag is defined in [RFC5888].

```
superimposition-attribute =
    "a=superposition:" "transparency:" transparency-tag,
                          "layer:" layer-tag
transparency-tag =transparency-value *("," transparency-value) CRLF
transparency-value= alpha
layer-tag =layering-order *("," layering-order) CRLF
layering-order = beta
```

Alpha describes the transparency for the media stream. It is identified by its transparency-tag values in the transparency-attribute. The transparency value must be an ASCII representation of an 8 bit signed integer with values between "-128" and "127", and linear weighting between the two extremes. A value of -128 means media stream is opaque and the highest value of 127 means it is transparent. Beta represents the layering order value for the media stream. The layering order value is an integer value between 0 to n, where the value 0 represents the most background layer. For each k within 0..n, a reconstructed sample of the k-th media is superimposed (while perhaps applying an alpha transparency value) on the 0 to k-th reconstructed samples in the same spatial position. The transparency attribute MUST be omitted for layer with order 0, and the default transparency value for background media stream of -128 is applied.

6. Example of Supim

The following example shows a session description for superimposed media stream in an SDP session. The "group" line indicates that the "m" lines with tokens 1, 2 and 3 are grouped for the purpose of superimposition.

In the example shown below, three media streams are being transmitted for superimposition. The background media stream along with the foreground media streams are grouped together using "supim". All media streams are video with "superposition" attribute. Media stream with layer order value 0 is intended for background.

```
v=0
o=Alice 292742730 29277831 IN IP4 233.252.0.74
c=IN IP4 233.252.0.79
t=0 0
a=group:supim 1 2 3
m=video 30000 RTP/AVP 31
a=mid:1
a= superposition:transparency= -128, layer=0
m=video 30002 RTP/AVP 31
a=mid:2
a= superposition:transparency=35, layer=1
m=video 30003 RTP/AVP 31
a=mid:3
a= superposition:transparency=75, layer=2
```

The transparency value is used for composing the foreground with the background media [Wiki.Alpha-compositing]. The "layer" value is relevant when two or more media streams are to be composed. When the transparency value of the foreground is -128, the composed image will be the foreground image, as it is being displayed as opaque. Similarly, if the transparency value for the foreground media is 127, the resulting image will be the background media, as the foreground media stream is being presented fully transparent, hence invisible. The details of the weighting of foreground and background sample values based on a given alpha value is left undefined herein, beyond the abstract definition that alpha equal to -128 means opaque, and alpha equal to 127 means transparent, and the weighting is to be implemented such that it is visually linear for the values in between. We do not define a weighting formula as these formulae would depend on many factors such as the colorspace and the sampling structure of the media.

7. Relationship with CLUE (informative)

Edt. Note: maybe we remove this section later once there is a general understanding why CLUE in its current form is unsuitable. The CLUE framework [RFC8845] and its associated suite of I-Ds and RFCs describe a telepresence framework that, at the first glance seems to have a lot in common with the technology proposed herein. CLUE defines captures (camera ports), and their geo-spatial relationship to each other. A render can use this information to put the reconstructed samples of the streams from the various captures into a suitable arrangement such that visually pleasant rendering can be achieved. However, CLUE does not describe the relative relevance of the captures. For that reason, CLUE would need to be extended in a spirit very similar to the one described in this memo to achieve the desired functionality. CLUE has not seen wide deployment outside its intended key application (large room, multiple camera telepresence systems). It's not reasonable to assume that small systems would willingly implement the overhead the (comparatively complex) CLUE protocols require when a simple SDP extension can serve the same purpose.

8. 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-by-hop fashion, respectively.

9. IANA Considerations

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

Rohit Abhishek <rabhishek@rabhishek.com>

Stephan Wenger <stewe@stewe.org>

The IETF MMUSIC working group <mmusic@ietf.org> or its successor
as designated by the IESG.

This document defines a new SDP group semantics for media superimposition for a SDP session. This attribute can be used by the application to group the foreground and the background media streams to be superimposed together 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
-----	-----	-----
Superimposition	supim	RFCXXXX

The "supim" attribute is used to group different media streams to be superimposed together with one background media stream and rest foreground streams. Its format is defined in Section 3.

The SDP media-level attribute "superposition" needs to be registered by IANA Semantics for "att-field (media-level only)". The registration procedure in [RFC8866] applies.

SDP Attribute ("att-field (media level only)":

```

Attribute name: superposition: transparency, layer
Long form: superimposition transparency, superimposition layer
Type of name: att-field
Type of attribute: media level only
Subject to charset: no
Purpose: RFC 5583
Reference: RFC 5583
Values: alpha, beta

```

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.

10. Acknowledgements

The authors would like to thank Christer Holmberg and Paul Kyzivat for reviewing the draft and providing key ideas.

11. References

11.1. 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>>.
- [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>>.

[RFC8866] Begen, A., Kyzivat, P., Perkins, C., and M. Handley, "SDP: Session Description Protocol", RFC 8866, DOI 10.17487/RFC8866, January 2021, <<https://www.rfc-editor.org/info/rfc8866>>.

11.2. Informative References

[draft-ietf-avtcore-multiplex-guidelines-12]
Westerlund, M., Burman, B., Perkins, C., Alvestrand, H., and R. Even, "Guidelines for using the Multiplexing Features of RTP to Support Multiple Media Streams", draft-ietf-avtcore-multiplex-guidelines-12 (work in progress), June 2020.

[RFC8845] Duckworth, M., Ed., Pepperell, A., and S. Wenger, "Framework for Telepresence Multi-Streams", RFC 8845, DOI 10.17487/RFC8845, January 2021, <<https://www.rfc-editor.org/info/rfc8845>>.

[Wiki.Alpha-compositing]
"Alpha compositing", <https://en.wikipedia.org/wiki/Alpha_compositing>.

Authors' Addresses

Rohit Abhishek
Tencent
2747 Park Blvd
Palo Alto 94588
USA

Email: rabhishek@rabhishek.com

Stephan Wenger
Tencent
2747 Park Blvd
Palo Alto 94588
USA

Email: stewe@stewe.org