

mmusic
Internet-Draft
Intended status: Standards Track
Expires: December 3, 2021

R. Abhishek
S. Wenger
Tencent
June 1, 2021

SDP Superimposition Grouping framework
draft-abhishek-mmusic-superimposition-grouping-02

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 fully or partly superimposed visual media streams enabling them to be added as an overlay on top of any one or more background visual media streams. The superimposition grouping semantics is helpful if the media stream 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 December 3, 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	3
3. Media Superimposition in SDP	3
4. Superimposition Group Identification Attribute	4
5. Use of group and mid	5
6. "superimposition" Attribute for Superimposition Group Identification Attribute	5
7. Example of Supim	6
8. Relationship with Existing Specifications (informative)	7
9. Security Considerations	8
10. IANA Considerations	8
11. Acknowledgements	9
12. References	9
12.1. Normative References	9
12.2. Informative References	10
Authors' Addresses	11

1. Introduction

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 fully or partly superimposed visual media streams enabling them to be added as an overlay on top of any one or more background visual media streams. The superimposition grouping semantics is helpful if the media stream data is separate and transported via different sessions.

Media superimposition herein is defined to be a visual media stream (video/image/text) that is fully or partly superimposed on top of an already existing visual media stream 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 may be 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 over a background media or over each other.

This document describes 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 are to be used in conjunction with "The Session Description Protocol (SDP) Grouping Framework" [RFC5888].

We have studied the existing specifications, including the CLUE framework [RFC8845] and work in MPEG, and found that such work is not covering our intended application space; please refer to Section 8 for details. 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.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Media Superimposition in SDP

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, rendering aspects, including their possible 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 to other media streams 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. An example is shown in Figure 1, where three foreground media streams have been superimposed over a background media stream, with Media B being partly superimposed over Media C.

```

=====
==== Media A ====
=====
=====
                                     ++++++
                                     +---- Media B ----+
#####+++++
#####+++++
#### Media C ####
#####

```

Figure 1: A example of media superimposition

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 MANE (media-aware network element), 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 their 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].

4. Superimposition Group Identification Attribute

The "superimposition media stream identification" attribute, "supim", 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. There is no defined behavior for the rendering of non-visual media being grouped in a superimposition group. It is assumed that all the media streams are that need to be time-synchronized are time-synchronized. Its

formatting follows [RFC5888] in the use of the 'mid' attribute to identify the media line to be included in the superimposition.

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. A media player that chooses to implement the extension and receives a session description that contains "m" lines grouped together using "supim" semantics is able to superimpose the foreground media streams on top of the background media stream in cases where there is overlap. For non-supporting devices, these media streams are treated as independent media streams.

5. 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. Multiple groups MUST not be used within one session. If the identification-tags associated with "a=group" lines do not map to any "m" lines, the identification-tags MUST be ignored.

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

6. "superimposition" Attribute for Superimposition Group Identification Attribute

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

```
superimposition-attribute =
    "superimposition:" super-opt *(SP super-opt)
super-opt = super-trans / super-layer
super-trans = "transparency:" super-trans-val
super-layer = "layer:" super-layer-val
super-trans-val = signed-integer ; range [-128, 127]
super-layer-val = signed-integer ; range [0, 255]

signed-integer =
    <zero-based-integer defined in RFC8866>
    / "-" <integer defined in RFC8866>
attribute = <attribute defined in RFC4566>
attribute =/ superimposition-attribute
```

The transparency for the media stream is identified by its `super-trans-val` values in the `super-trans` attribute. The 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 the media stream is opaque, and the highest value of 127 means it is transparent. Further details of interpretation is to be left open to the implementer. The layering order value for the media stream is identified by `super-layer-val`. It MUST be an integer value between 0 and n, where the value 0 represents the deepest background layer. For each k within 0..n, a reconstructed sample of the k-th media is superimposed (while perhaps applying an `super-trans-val` value) on the 0 to k-th reconstructed samples in the same spatial position. Each "m" line in a session MUST NOT contain more than one instance of `super-opt` attribute.

7. Example of Supim

The following example shows a session description for superimposed media streams 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 videos with "superimposition" attribute. The 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= superimposition:transparency= -128, layer=0
m=video 30002 RTP/AVP 31
a=mid:2
a= superimposition:transparency=35, layer=1
m=video 30003 RTP/AVP 31
a=mid:3
a= superimposition:transparency=75, layer=2
```

The transparency value is used for composing the foreground with the background media [Wiki.Alpha-compositing]. This value itself does not define the transparency of each pixel but is applied to each pixel within a frame and defines the factor by which the transparency of each pixel within a frame is to be increased or decreased. 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 super-trans value is left to the implementation, beyond the abstract definition that value equal to -128 means opaque, and value 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 in this specification as these formulae would depend on many factors such as the colorspace and the sampling structure of the media.

8. Relationship with Existing Specifications (informative)

Edt. Note: maybe we remove this section later once there is a general understanding why the existing specifications in its current form is unsuitable. The CLUE framework [RFC8845] is the IETF's chosen technology for the applications requiring defining multiple "captures" (camera views), and their geo-spatial relationship to each. However, information pertaining to display/rendering is outside of CLUE's scope. While many CLUE-capable receivers infer appropriate rendering strategies from the information offered by CLUE, the CLUE framework has generally assumed non-overlapped rendering of transmitted and reconstructed video streams from the

multiple captures, often on different physical rendering devices. Insofar, we concluded that the CLUE framework neither supports the application we contemplate in this memo, nor would it be sensible to enhance the CLUE specifications with rendering-related mechanisms. There are certain technologies from standards bodies such as MPEG [MPEG-4], often described as "scene descriptions", that to a certain extent can address the applications we contemplate. We evaluated the technologies we are aware of and concluded that something different is required. We base our assumption on a) the complexity of these mechanisms, and b) their design as a metadata media stream, which in the IETF context would be conveyed in RTP sessions or similar, rather than a static or semi-static stream description that is best conveyed at session setup or renegotiation using SDP.

9. 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.

10. 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 value 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 value to the sdp-paramters registry group defined in [RFC5888] [RFC8859].

IANA is requested to register the following semantics value in the "sdp-parameters" in the registry.

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 the rest foreground streams. Its format is defined in Section 4.

IANA is requested to register the semantics value for SDP media-level attribute "superimposition" for "sdp-attributes(media-level only)". The registration procedure in [RFC8866] applies.

SDP Attribute ("sdp-attributes(media level only)"):

```

Attribute name: superimposition: 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: super-trans-val, super-layer-val

```

11. Acknowledgements

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

12. References

12.1. 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>>.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.
- [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>>.
- [RFC8859] Nandakumar, S., "A Framework for Session Description Protocol (SDP) Attributes When Multiplexing", RFC 8859, DOI 10.17487/RFC8859, January 2021, <<https://www.rfc-editor.org/info/rfc8859>>.
- [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>>.

12.2. 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.

[MPEG-4] "MPEG-4 Scene Description and Application Engine",
<[https://mpeg.chiariglione.org/standards/mpeg-4/scene-
description-and-application-engine](https://mpeg.chiariglione.org/standards/mpeg-4/scene-description-and-application-engine)>.

[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