

SDP media alignment in SIP

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC2026](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts. 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."

The list of current Internet-Drafts can be accessed at

<http://www.ietf.org/ietf/1id-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at

<http://www.ietf.org/shadow.html>.

Abstract

This document defines an SDP media attribute. This attribute is intended to be used in conjunction with SIP in order to align different media streams belonging to a session. The use of this attribute allows sending media from a single flow (several media streams), encoded in different formats during the session, to different ports and host interfaces.

1. Introduction

SIP [[1](#)] is an application layer protocol for establishing, terminating and modifying multimedia sessions. SIP carries session descriptions in the bodies of the SIP messages but is independent from the protocol used for describing sessions. SDP [[2](#)] is one of the protocols that can be used for this purpose.

[Appendix B](#) of [[1](#)] describes the usage of SDP in relation to SIP. It states: "The caller and callee align their media description so that the nth media stream ("m=" line) in the caller's session description corresponds to the nth media stream in the callee's description."

SDP media alignment in SIP

This way of performing the media alignment is not efficient when a single flow comprises several media streams. This is a common situation when AP (Application Server) components [3] are employed. It is also common for systems that handle different codecs on different port numbers (or on different interfaces).

2. Media flow definition

The RTSP RFC [4] defines a media stream as "a single media instance, e.g., an audio stream or a video stream as well as a single whiteboard or shared application group. When using RTP, a stream consists of all RTP and RTCP packets created by a source within an RTP session".

This definition assumes that a single audio (or video) stream maps into an RTP session. The RTP RFC [5] defines an RTP session as follows: "For each participant, the session is defined by a particular pair of destination transport addresses (one network address plus a port pair for RTP and RTCP)".

However, there are situations where a single media instance, e.g., an audio stream or a video stream is sent using more than one RTP session. Two examples (among many others) of this kind of situation are cellular systems using SIP and systems receiving DTMF tones on a different host than the voice. Both examples are described in later sections.

We introduce the definition of media flow:

Media flow consists of a single media instance, e.g., an audio stream or a video stream as well as a single whiteboard or shared application group. When using RTP, a media flow comprises one or more RTP sessions.

For instance, in a two party call where the voice exchanged can be encoded using GSM or PCM, the receiver wants to receive GSM on a port number and PCM on a different port number. Two RTP sessions will be established, one carrying GSM and the other carrying PCM.

At any particular moment just one codec is in use. Therefore, at any moment one of the RTP sessions will not transport any voice. Here the systems are dealing with a single flow (one audio stream) and two RTP sessions.

2.1 SIP and cellular access

Systems using a cellular access (such as UMTS or EDGE) and SIP as a signalling protocol need to receive media over the air. During a session the media can be encoded using different codecs. The encoded media has to traverse the radio interface. The radio interface is generally characterized by being bit error prone and associated with

relatively high packet transfer delays. In addition, radio interface resources in a cellular environment are scarce and thus expensive,

SDP media alignment in SIP

which calls for special measures in providing a highly efficient transport [6]. In order to get an appropriate speech quality in combination with an efficient transport, precise knowledge of codec properties are required so that a proper radio bearer for the RTP session can be configured before transferring the media. These radio bearers are dedicated bearers per media type, i.e. codec.

In UMTS, for instance, when the RTP packets shall be delivered over the air interface, a packet filtering function routes the packets to the proper radio bearer towards the UMTS/SIP terminal. The packet filtering function operates using a Traffic Flow Template (TFT) [7], which is established when configuring the radio bearer. The TFT hence specifies the profile of the data that should be carried by the radio bearer. A TFT can contain the following data:

- Source Address and Subnet Mask.
- Protocol Number (IPv4) / Next Header (IPv6).
- Destination Port Range.
- Source Port Range.
- IPSec Security Parameter Index (SPI).
- Type of Service (TOS) (IPv4) / Traffic class (IPv6) and Mask.
- Flow Label (IPv6).

It is worth noticing that just certain combinations of these parameters are allowed.

The media has to have different destination port numbers for the different possible codecs in order to be filtered and routed properly to the correct radio bearer. Therefore, several RTP sessions are used for a single media flow.

2.2 DTMF tones

Some voice sessions include DTMF tones. Sometimes the voice handling is performed by a different host than the DTMF handling (e.g. [section 5.4](#), figures 3 and 4 of [3]). In this situations it is necessary to establish two RTP sessions: one for the voice and the other for the DTMF tones. Both RTP sessions are logically part of the same media flow.

3. Flow identification attribute

A new "flow identification" media attribute is defined. It is used for identifying media flows within a session. It provides a means for aligning a number of flows (rather than a number of media streams) within a session between members participating in the session. Its formatting in SDP is described by the following BNF:

```
fid-attribute      = "a=fid:" identification-tag
identification-tag = token
```

The identification tag is unique within the SDP session description. The following examples illustrate its usage.

Camarillo/Holler/Eriksson

3

SDP media alignment in SIP

4. Examples of flow identification attribute

4.1 UMTS/SIP terminal

In the following example John uses a traditional access such as an ethernet while Laura has a UMTS/SIP terminal. The caller John sends an INVITE with the following session description to the callee Laura.

```
v=0
o=John 289085535 289085535 IN IP4 first.example.com
t=0 0
c=IN IP4 111.111.111.111
m=audio 20000 RTP/AVP 0 8
a=fid:1
```

The callee Laura is on a UMTS/SIP terminal. She configures the necessary radio bearers and implements the TFTs:

All the incoming IP packets with destination port UDP 30000 will be carried by the radio access bearer configured for G-711 u-law (payload type 0).

All the incoming IP packets with destination port UDP 30002 will be carried by the radio access bearer configured for G-711 A-law (payload type 8).

Accordingly, the following SDP is returned to the caller in a 200 OK response:

```
v=0
o=Laura 289083124 289083124 IN IP4 second.example.com
t=0 0
c=IN IP4 222.222.222.222
m=audio 30000 RTP/AVP 0
a=fid:1
m=audio 30002 RTP/AVP 8
a=fid:1
```

With the current way of performing SDP media alignment in SIP the

callee would have accepted the call and immediately after re-INVITED the caller with the new SDP. The fid attribute saves many RTTs.

Besides saving bandwidth and RTTs the fid attribute provides a means for describing a logical relationship between media streams that belong to the same flow.

4.2 Application Server Components

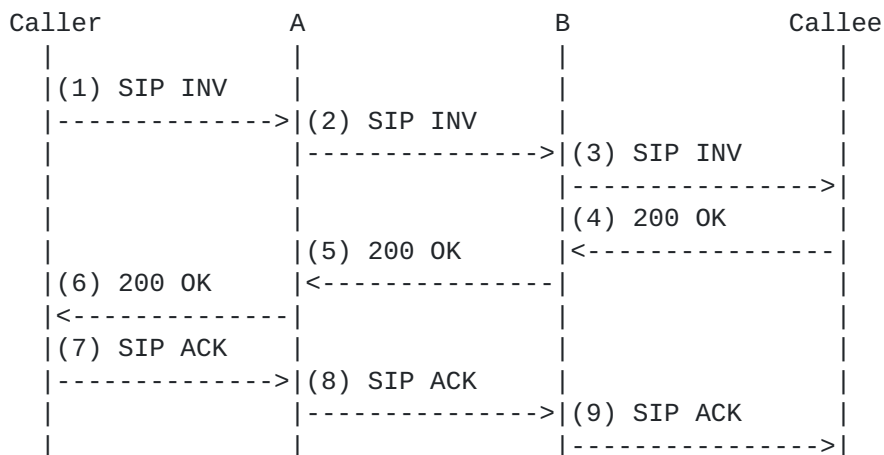
SDP media alignment in SIP

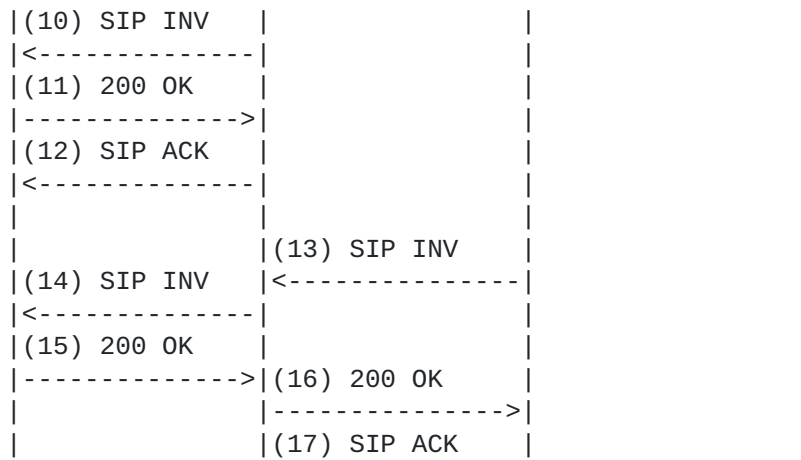
In [section 5.4](#) of "An Application Server Architecture for SIP" [3] contains two examples (figures 3 and 4) where DTMF tones are received by a different host than the voice stream. In both situations using the fid attribute to perform media alignment would save a tremendous amount of messages exchanged and reduce the global session establishment time.

Let us take figure 4. A UAC sends an INVITE with just a voice stream. There are two ASs in the path that want to receive DTMF tones.

Three steps are needed in order to set the session up:

- 1) A session is established between the UAC and the callee. This involves three messages from the caller's point of view (INVITE-200 OK-ACK).
- 2) The session is modified by A (one of the ASs that wants to receive DTMF tones). It adds an "m" line to the session description indicating that it wants to receive DTMF tones. This involves three more messages from the caller's point of view (INVITE-200 OK-ACK)
- 3) The session is modified once more by B (the other AS that also wants to receive DTMF tones). It adds another "m" line indicating that it wants to receive DTMF tones. This involves three more messages from the caller's point of view (INVITE-200 OK-ACK).





SDP media alignment in SIP

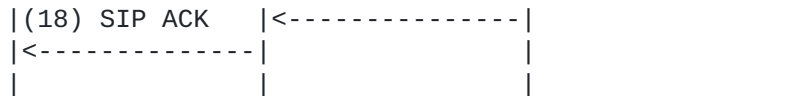


Figure 4 of "An AS Component Architecture for SIP" [3]

The whole session is not correctly set up until the end of this sequence of messages. If the caller is using a low-rate access this can take a long time.

The use of the fid attribute would reduce these nine messages that the caller sees to just three (INVITE-200 OK-ACK). B would add an "m" line to the 200 OK from the callee with the same fid value as the voice stream. Then A would add another "m" line, again with the same fid value than the two previous "m" lines.

As a result, the caller receives a 200 OK indicating that just one flow is established, but also that all the DTMF tones should be sent to A and B. For a low-rate access the establishment time has been reduced a lot.

5. Media-level versus session-level attribute

Syntactically fid is a media-level attribute. It provides information about a media stream defined by an "m" line. Semantically fid would be defined as a session-level attribute since it provides flow hierarchy inside a session description.

6. Backward compatibility

A system that understands the fid attribute MUST add it to any SDP session description that it generates.

If a response to a request that included the fid attribute also includes it media alignment is performed based on the fid attribute

rather than on matching of nth lines.

6.1 Caller does not support fid

This situation does not represent a problem. The SDP in the INVITE will not contain any fid attribute and the callee will use the "nth-line" method to perform media alignment.

The callee will need a re-INVITE in order to receive the proper media encoding on the proper interface.

6.2 Callee does not support fid

The callee will ignore the fid attribute. It will consider that the session comprises several media streams.

Different implementations would behave in different ways.

SDP media alignment in SIP

In the case of audio and different "m" lines for different codecs an implementation might decide to act as a mixer with the different incoming RTP sessions, which is the correct behavior.

If an implementation decides to refuse the request (e.g. 488 Not acceptable here or 606 Not Acceptable) the caller should re-try the request without the fid attribute and only one "m" line per flow. Note that even re-INVITES without the fid attribute adding new "m" lines would probably fail in this situation because the callee does not support multiple "m" lines. Therefore, this problem is related to UAs that do not handle multiple "m" lines rather than to the fid attribute.

7. Acronyms

AP	Application Server
BNF	Backus-Naur Form
DTMF	Dual Tone Multi Frequency
EDGE	Enhanced Data rates for GSM and TDMA/136 Evolution
GSM	Global System for Mobile communication
IP	Internet Protocol
PCM	Pulse Code Modulation
RFC	Request For Comments
RTCP	RTP Control Protocol
RTP	Real-time Transport Protocol
RTSP	Real-Time Streaming Protocol
RTT	Round Trip Time
SDP	Session Description Protocol
SIP	Session Initiation Protocol
TFT	Traffic Flow Template
UA	User Agent

UAC User Agent Client
UMTS Universal Mobile Telecommunication System
WLAN Wireless Local Area Network

8. Acknowledgments

The authors would like to thank Adam Roach for his feedback on this document.

9. References

- [1] M. Handley/H. Schulzrinne/E. Schooler/J. Rosenberg, "SIP: Session Initiation Protocol", [RFC 2543](#), IETF; March 1999.
- [2] M. Handley/V. Jacobson, "SDP: Session Description Protocol", [RFC 2327](#), IETF; April 1998.
- [3] J. Rosenberg/P. Mataga/H. Schulzrinne, "An Application Server Component Architecture for SIP", [draft-rosenberg-sip-app-components-00.txt](#), IETF; November 2000.

Camarillo/Holler/Eriksson

7

SDP media alignment in SIP

- [4] H. Schulzrinne/A. Rao/R. Lanphier, "Real Time Streaming Protocol (RTSP)", [RFC 2326](#), IETF; April 1998.
- [5] H. Schulzrinne/S. Casner/R. Frederick/V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", [RFC 1889](#), IETF; January 1996.
- [6] L. Westberg/M. Lindqvist, "Realtime Traffic over Cellular Access Networks", [draft-westberg-realtime-cellular-02.txt](#), IETF; May 2000. Work in progress.
- [7] 3G TS 23.060 v3.2.1 General Packet Radio Service Description.

10. Authors' Addresses

Gonzalo Camarillo
Ericsson
Advanced Signalling Research Lab.
FIN-02420 Jorvas
Finland
Phone: +358 9 299 3371
Fax: +358 9 299 3052
Email: Gonzalo.Camarillo@ericsson.com

Jan Holler
Ericsson Research
S-16480 Stockholm

Sweden
Phone: +46 8 58532845
Fax: +46 8 4047020
Email: Jan.Holler@era.ericsson.se

Goran AP Eriksson
Ericsson Research
S-16480 Stockholm
Sweden
Phone: +46 8 58531762
Fax: +46 8 4047020
Email: Goran.AP.Eriksson@era.ericsson.se