Negotiating Human Language in Real-Time Communications

Abstract

   Users have various human (i.e., natural) language needs, abilities,
   and preferences regarding spoken, written, and signed languages.
   This document defines new Session Description Protocol (SDP) media-
   level attributes so that when establishing interactive communication
   sessions ("calls"), it is possible to negotiate (i.e., communicate
   and match) the caller's language and media needs with the
   capabilities of the called party. This is especially important for
   emergency calls, because it allows for a call to be handled by a call
   taker capable of communicating with the user or for a translator or
   relay operator to be bridged into the call during setup. However,
   this also applies to non-emergency calls (for example, calls to a
   company call center).

   This document describes the need as well as a solution that uses new
   SDP media attributes.

Status of This Memo

   This is an Internet Standards Track document.

   This document is a product of the Internet Engineering Task Force
   (IETF). It represents the consensus of the IETF community. It has
   received public review and has been approved for publication by the
   Internet Engineering Steering Group (IESG). Further information on
   Internet Standards is available in Section 2 of RFC 7841.

   Information about the current status of this document, any errata,
   and how to provide feedback on it may be obtained at
   https://www.rfc-editor.org/info/rfc8373.

Table of Contents

1. Introduction

   A mutually comprehensible language is helpful for human
   communication. This document addresses the negotiation of human
   language and media modality (spoken, signed, or written) in real-time
   communications. A companion document [RFC8255] addresses language
   selection in email.

   Unless the caller and callee know each other or there is contextual
   or out-of-band information from which the language(s) and media
   modalities can be determined, there is a need for spoken, signed, or
   written languages to be negotiated based on the caller's needs and
   the callee's capabilities. This need applies to both emergency and
   non-emergency calls. An example of a non-emergency call is when a
   caller contacts a company call center; an emergency call typically
   involves a caller contacting a Public Safety Answering Point (PSAP).
   In such scenarios, it is helpful for the caller to be able to
   indicate preferred signed, written, and/or spoken languages and for
   the callee to be able to indicate its capabilities; this allows the
   call to proceed using the language(s) and media forms supported by
   both.

   For various reasons, including the ability to establish multiple
   streams using different media (i.e., voice, text, and/or video), it
   makes sense to use a per-stream negotiation mechanism known as the
   Session Description Protocol (SDP). Utilizing SDP [RFC4566] enables
   the solution described in this document to be applied to all
   interactive communications negotiated using SDP, in emergency as well
   as non-emergency scenarios.

   By treating language as another SDP attribute that is negotiated
   along with other aspects of a media stream, it becomes possible to
   accommodate a range of users' needs and called-party facilities. For
   example, some users may be able to speak several languages but have a
   preference. Some called parties may support some of those languages
   internally but require the use of a translation service for others,
   or they may have a limited number of call takers able to use certain
   languages. Another example would be a user who is able to speak but
   is deaf or hard of hearing and desires a voice stream to send spoken
   language plus a text stream to receive written language. Making
   language a media attribute allows standard session negotiation to
   handle this by providing the information and mechanism for the
   endpoints to make appropriate decisions.

   The term "negotiation" is used here rather than "indication" because
   human language (spoken/written/signed) can be negotiated in the same
   manner as media (audio/text/video) and codecs. For example, if we
   think of a user calling an airline reservation center, the user may

be able to use a set of languages, perhaps with preferences for one
or a few, while the airline reservation center may support a fixed
set of languages. Negotiation should select the user's most
preferred language that is supported by the call center. Both sides
should be aware of which language was negotiated.

In the offer/answer model used here, the offer contains a set of
languages per media (and direction) that the offerer is capable of
using, and the answer contains one language per media (and direction)
that the answerer will support. Supporting languages and/or
modalities can require taking extra steps, such as bridging external
translation or relay resources into the call or having a call handled
by an agent who speaks a requested language and/or has the ability to
use a requested modality. The answer indicates the media and
languages that the answerer is committing to support (possibly after
additional steps have been taken). This model also provides
knowledge so both ends know what has been negotiated. Note that
additional steps required to support the indicated languages or
modalities may or may not be in place in time for any early media.

Since this is a protocol mechanism, the user equipment (UE) client
needs to know the user's preferred languages; while this document
does not address how clients determine this, reasonable techniques
could include a configuration mechanism with a default of the
language of the user interface. In some cases, a UE client could tie
language and media preferences, such as a preference for a video
stream using a signed language and/or a text or audio stream using a
written/spoken language.

This document does not address user interface (UI) issues, such as if
or how a UE client informs a user about the result of language and
media negotiation.

1.1. Applicability

Within this document, it is assumed that the negotiating endpoints
have already been determined so that a per-stream negotiation based
on SDP can proceed.

When setting up interactive communication sessions, it is necessary
to route signaling messages to the appropriate endpoint(s). This
document does not address the problem of language-based routing.

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. Desired Semantics

The desired solution is a media attribute (preferably per direction) that may be used within an offer to indicate the preferred language(s) of each (direction of a) media stream and within an answer to indicate the accepted language. When multiple languages are included for a media stream within an offer, the languages are listed in order of preference (most preferred first).

Note that negotiating multiple simultaneous languages within a media stream is out of scope of this document.

4. The Existing 'lang' Attribute

RFC 4566 [RFC4566] specifies an attribute 'lang' that is similar to what is needed here but is not sufficiently specific or flexible for the needs of this document. In addition, 'lang' is not mentioned in [RFC3264], and there are no known implementations in SIP. Further, it is useful to be able to specify language per direction (sending and receiving). This document therefore defines two new attributes.

5. Solution

An SDP attribute (per direction) seems the natural choice to negotiate human language of an interactive media stream, using the language tags of [BCP47].

5.1. The 'hlang-send' and 'hlang-recv' Attributes

This document defines two media-level attributes: 'hlang-send' and 'hlang-recv' (registered in Section 6). Both start with 'hlang', short for "human language". These attributes are used to negotiate which human language is selected for use in (each direction of) each interactive media stream. (Note that not all streams will necessarily be used.) Each can appear for media streams in offers and answers.

In an offer, the 'hlang-send' value is a list of one or more language(s) the offerer is willing to use when sending using the media, and the 'hlang-recv' value is a list of one or more

language(s) the offerer is willing to use when receiving using the media. The list of languages is in preference order (first is most preferred). When a media is intended for interactive communication in only one direction (e.g., a user in France with difficulty speaking but able to hear who indicates a desire to receive French using audio and send French using text), either 'hlang-send' or 'hlang-recv' MAY be omitted. Note that the media can still be useful in both directions. When a media is not primarily intended for language (for example, a video or audio stream intended for background only), both SHOULD be omitted. Otherwise, both SHOULD have the same value. Note that specifying different languages for each direction (as opposed to the same, or essentially the same, language in different modalities) can make it difficult to complete the call (e.g., specifying a desire to send audio in Hungarian and receive audio in Portuguese).

In an answer, 'hlang-send' is the language the answerer will send if using the media for language (which in most cases is one of the languages in the offer's 'hlang-recv'), and 'hlang-recv' is the language the answerer expects to receive if using the media for language (which in most cases is one of the languages in the offer's 'hlang-send').

In an offer, each value MUST be a list of one or more language tags per [BCP47], separated by white space. In an answer, each value MUST be one language tag per [BCP47]. [BCP47] describes mechanisms for matching language tags. Note that Section 4.1 of RFC 5646 [BCP47] advises to "tag content wisely" and not include unnecessary subtags.

When placing an emergency call, and in any other case where the language cannot be inferred from context, each OFFERed media stream primarily intended for human language communication SHOULD specify the 'hlang-send' and/or 'hlang-recv' attributes for the direction(s) intended for interactive communication.

Clients acting on behalf of end users are expected to set one or both of the 'hlang-send' and 'hlang-recv' attributes on each OFFERed media stream primarily intended for human communication when placing an outgoing session, and either ignore or take into consideration the attributes when receiving incoming calls, based on local configuration and capabilities. Systems acting on behalf of call centers and PSAPs are expected to take the attributes into account when processing inbound calls.

Note that media and language negotiation might result in more media streams being accepted than are needed by the users (e.g., if more preferred and less preferred combinations of media and language are all accepted). This is not a problem.

5.2. No Language in Common

A consideration regarding the ability to negotiate language is
whether the call proceeds or fails if the callee does not support any
of the languages requested by the caller. This document does not
mandate either behavior.

When a call is rejected due to lack of any language in common, the
SIP response has SIP response code 488 (Not Acceptable Here) or 606
(Not Acceptable) [RFC3261] and a Warning header field [RFC3261] with
a warning code of 308 and warning text indicating that there are no
mutually supported languages; the warning text SHOULD also contain
the supported languages and media.

Example:

Warning: 308 proxy.example.com "Incompatible language
   specification: Requested languages not supported. Supported
   languages are: es, en; supported media are: audio, text."

5.3. Usage Notes

A sign-language tag with a video media stream is interpreted as an
indication for sign language in the video stream. A non-sign-
language tag with a text media stream is interpreted as an indication
for written language in the text stream. A non-sign-language tag
with an audio media stream is interpreted as an indication for spoken
language in the audio stream.

This document does not define any other use for language tags in
video media (such as how to indicate visible captions in the video
stream).

This document does not define the use of sign-language tags in text
or audio media.

In the IANA registry for language subtags per [BCP47], a language
subtag with a Type field "extlang" combined with a Prefix field value
"sgn" indicates a sign-language tag. The absence of such "sgn"
prefix indicates a non-sign-language tag.

This document does not define the use of language tags in media other
than interactive streams of audio, video, and text (such as "message"
or "application"). Such use could be supported by future work or by
application agreement.

5.4. Examples

   Some examples are shown below. For clarity, only the most directly
   relevant portions of the SDP block are shown.

   An offer or answer indicating spoken English both ways:

      m=audio 49170 RTP/AVP 0
      a=hlang-send:en
      a=hlang-recv:en

   An offer indicating American Sign Language both ways:

      m=video 51372 RTP/AVP 31 32
      a=hlang-send:ase
      a=hlang-recv:ase

   An offer requesting spoken Spanish both ways (most preferred), spoken
   Basque both ways (second preference), or spoken English both ways
   (third preference):

      m=audio 49250 RTP/AVP 20
      a=hlang-send:es eu en
      a=hlang-recv:es eu en

   An answer to the above offer indicating spoken Spanish both ways:

      m=audio 49250 RTP/AVP 20
      a=hlang-send:es
      a=hlang-recv:es

   An alternative answer to the above offer indicating spoken Italian
   both ways (as the callee does not support any of the requested
   languages but chose to proceed with the call):

      m=audio 49250 RTP/AVP 20
      a=hlang-send:it
      a=hlang-recv:it

   An offer or answer indicating written Greek both ways:

      m=text 45020 RTP/AVP 103 104
      a=hlang-send:gr
      a=hlang-recv:gr

   An offer requesting the following media streams: video for the caller
   to send using Argentine Sign Language, text for the caller to send
   using written Spanish (most preferred) or written Portuguese, and

audio for the caller to receive spoken Spanish (most preferred) or
spoken Portuguese:

```
m=video 51372 RTP/AVP 31 32
a=hlang-send:aed

m=text 45020 RTP/AVP 103 104
a=hlang-send:sp pt

m=audio 49250 RTP/AVP 20
a=hlang-recv:sp pt
```

An answer for the above offer, indicating text in which the callee
will receive written Spanish and audio in which the callee will send
spoken Spanish. (The answering party has no video capability):

```
m=video 0 RTP/AVP 31 32
m=text 45020 RTP/AVP 103 104
a=hlang-recv:sp

m=audio 49250 RTP/AVP 20
a=hlang-send:sp
```

An offer requesting the following media streams: text for the caller
to send using written English (most preferred) or written Spanish,
audio for the caller to receive spoken English (most preferred) or
spoken Spanish, and supplemental video:

```
m=text 45020 RTP/AVP 103 104
a=hlang-send:en sp

m=audio 49250 RTP/AVP 20
a=hlang-recv:en sp

m=video 51372 RTP/AVP 31 32
```

An answer for the above offer, indicating text in which the callee
will receive written Spanish, audio in which the callee will send
spoken Spanish, and supplemental video:

```
m=text 45020 RTP/AVP 103 104
a=hlang-recv:sp

m=audio 49250 RTP/AVP 20
a=hlang-send:sp

m=video 51372 RTP/AVP 31 32
```

   Note that, even though the examples show the same (or essentially the
   same) language being used in both directions (even when the modality
   differs), there is no requirement that this be the case. However, in
   practice, doing so is likely to increase the chances of successful
   matching.

6. IANA Considerations

6.1. att-field Subregistry of SDP Parameters

   The syntax in this section uses ABNF per RFC 5234 [RFC5234].

   IANA has added two entries to the "att-field (media level only)"
   subregistry of the "Session Description Protocol (SDP) Parameters"
   registry.

   The first entry is for 'hlang-recv':

   Attribute Name: hlang-recv
   Long-Form English Name: human language receive
   Contact Name: Randall Gellens
   Contact Email Address: rg+ietf@coretechnologyconsulting.com
   Attribute Value: hlang-value
   Attribute Syntax:
      hlang-value = hlang-offv / hlang-ansv
         ; hlang-offv used in offers
         ; hlang-ansv used in answers
      hlang-offv = Language-Tag *(SP Language-Tag)
         ; Language-Tag as defined in [BCP47]
      SP   = 1*" " ; one or more space (%x20) characters
      hlang-ansv = Language-Tag
   Attribute Semantics: Described in Section 5.1 of RFC 8373
   Usage Level: media
   Mux Category: NORMAL
   Charset Dependent: No
   Purpose: See Section 5.1 of RFC 8373
   O/A Procedures: See Section 5.1 of RFC 8373
   Reference: RFC 8373

The second entry is for 'hlang-send':

```
Attribute Name:           hlang-send
Long-Form English Name:   human language send
Contact Name:             Randall Gellens
Contact Email Address:    rg+ietf@coretechnologyconsulting.com
Attribute Value:          hlang-value
Attribute Syntax:
   hlang-value =  hlang-offv / hlang-ansv
Attribute Semantics:      Described in Section 5.1 of RFC 8373
Usage Level:              media
Mux Category:             NORMAL
Charset Dependent:        No
Purpose:                  See Section 5.1 of RFC 8373
O/A Procedures:           See Section 5.1 of RFC 8373
Reference:                RFC 8373
```

6.2. Warning Codes Subregistry of SIP Parameters

IANA has added the value 308 to the "Warning Codes (warn-codes)"
subregistry of the "Session Initiation Protocol (SIP) Parameters"
registry. (The value lies within the range allocated for indicating
problems with keywords in the session description.) The reference is
to this document. The warn text is "Incompatible language
specification: Requested languages not supported. Supported
languages are [list of supported languages]; supported media are:
[list of supported media]."

7. Security Considerations

The Security Considerations of [BCP47] apply here. An attacker with
the ability to modify signaling could prevent a call from succeeding
by altering any of several crucial elements, including the
'hlang-send' or 'hlang-recv' values. RFC 5069 [RFC5069] discusses
such threats. Use of TLS or IPsec can protect against such threats.
Emergency calls are of particular concern; RFC 6881 [RFC6881], which
is specific to emergency calls, mandates use of TLS or IPsec (in
ED-57/SP-30).

8. Privacy Considerations

Language and media information can suggest a user's nationality,
background, abilities, disabilities, etc.

9. References

9.1. Normative References

[BCP47]   Phillips, A., Ed. and M. Davis, Ed., "Matching of Language
            Tags", BCP 47, RFC 4647, DOI 10.17487/RFC4647, September
            2006.

            Phillips, A., Ed. and M. Davis, Ed., "Tags for Identifying
            Languages", BCP 47, RFC 5646, DOI 10.17487/RFC5646,
            September 2009.

[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>.

[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>.

[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>.

9.2. Informative References

[RFC3264]   Rosenberg, J. and H. Schulzrinne, "An Offer/Answer Model
            with Session Description Protocol (SDP)", RFC 3264,
            DOI 10.17487/RFC3264, June 2002,
            <https://www.rfc-editor.org/info/rfc3264>.

[RFC5069]   Taylor, T., Ed., Tschofenig, H., Schulzrinne, H., and M.
            Shanmugam, "Security Threats and Requirements for
            Emergency Call Marking and Mapping", RFC 5069,
            DOI 10.17487/RFC5069, January 2008,
            <https://www.rfc-editor.org/info/rfc5069>.

   [RFC6881] Rosen, B. and J. Polk, "Best Current Practice for
              Communications Services in Support of Emergency Calling",
              BCP 181, RFC 6881, DOI 10.17487/RFC6881, March 2013,
              <https://www.rfc-editor.org/info/rfc6881>.

   [RFC8255] Tomkinson, N. and N. Borenstein, "Multiple Language
              Content Type", RFC 8255, DOI 10.17487/RFC8255, October
              2017, <https://www.rfc-editor.org/info/rfc8255>.

Acknowledgements

Contributors

Author's Address

   Randall Gellens
   Core Technology Consulting

   Email: rg+ietf@coretechnologyconsulting.com
   URI: http://www.coretechnologyconsulting.com