CLUE
Internet-Draft
Intended status: Standards Track
Expires: March 15, 2013

A. Romanow
F. Andreasen
A. Krishna
Cisco Systems

September 11, 2012

Investigation of Session Description Protocol (SDP) Usage for ControLling mUltiple streams for tElepresence (CLUE) draft-romanow-clue-sdp-usage-02

Abstract

This draft investigates how SDP offer/answer can be used to communicate CLUE encoding and encoding group parameters.

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 http://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 March 15, 2013.

Copyright Notice

Copyright (c) 2012 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 (http://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 3
2. Terminology 3
3. Background 3
4. Assumptions and Design Principles 4
5. Encoding 5
6. Encoding group 5
7. Solution overview 8
8. Provider advertisement in the initial offer/answer 9
8.1. Bidirectional SDP messages 13
8.2. Unidirectional SDP messages 13
9. Multiple offer/answer exchanges 15
10. Representation of Encoding and Encoding group in SDP 19
11. Concerns about conveying CLUE parameters via SDP 21
12. Security Considerations 22
13. Acknowledgements 22
14. IANA Considerations 22
15. References 22
15.1. Normative References 22
15.2. Informative References 22
Appendix A. Changes From Earlier Versions 23
A.1. Changes From Draft -01 23
Authors' Addresses 23

1. Introduction

This draft investigates how SDP could be used to carry CLUE, ControLling mUltiple streams for tElepresence, encoding and encoding group provider advertisements. CLUE specifies how multiple streams can be communicated in a telepresence conference in a standardized manner allowing interoperability. A familiarity is assumed with the CLUE framework document and the concepts defined therein [CLUEFRWK]. The media provider describes to the media consumer both media captures and encodings that it is capable of sending, and the consumer "configures" the provider to send the captures using the potential streams it wants to receive. In order for an interactive telepresence session to be established between A and B, each party needs to configure its role as both a media provider and a media consumer.

Encoding parameters are maxBandwidth, maxMbps, maxFps, maxWidth, maxHeight. Encoding group parameters are maxGroupBandwidth and maxGroupVideoMbps.

There are 2 reasons we investigated using SDP for CLUE provider advertisements of encoding and encoding groups. It was suggested that:

1. By carrying the encoding and encoding group parameters in SDP, the call-control agents, proxies, and other intermediaries can monitor and potentially modify these values thereby allowing administrators to easily configure and enforce either static limitations or more dynamic call-shaping approaches.
2. SDP is the usual protocol for communicating the encoding and encoding group related parameters in SIP-based systems

Media captures, capture sets, and simultaneous transmission information are all communicated in a new CLUE protocol, whereas encoding and encoding group parameters are communicated in SDP.

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 RFC 2119 [RFC2119] and indicate requirement levels for compliant implementations.

3. Background

The following abstractions defined in the CLUE framework document

need to be understood in considering this investigation.

- Media capture - fundamental representation of the content, audio and video so far defined. Each media capture is associated with one or more encoding groups.
- Capture set - A capture set is most usefully thought of as being a set of table rows, with each row consisting of a set of media captures. Media captures within a row can always be used simultaneously, whereas media captures in different sets may or may not be used simultaneously.. By including multiple media captures in a row within a capture set, the provider is signaling that those captures together form a representation of that capture set's scene.
- Simultaneous transmission set - there are often physical constraints that determine whether captures can be sent at the same time or not. For example, a camera may be able to create a video capture which is close up and another which is zoomed out. These 2 views cannot be done simultaneously. The simultaneous transmission set expresses the physical constraints on simultaneity.
- Encoding - Represents a way to encode a media capture in terms of its associated media characteristics, specifically maxBandwidth, maxFps, maxMbps, maxHeight and maxWidth. Each encoding is uniquely identified within an encoding group. [Practically it's desirable to have a unique ID among all encodings in the conference]
- Encoding group - An encoding group includes set(s) of encodings, plus parameters that apply to the group as a whole, specifically maxGroupBandwidth and maxGroupMpbs. Each encoding group has a unique identifier in the SDP.

4. Assumptions and Design Principles

- For each CLUE type ("purpose") of media capture (audio, video-main, or video-presentation), there will be at most one corresponding SDP media stream ("m=" line). Note: video-main and video-presentation are both using "video" media type.
- Multiple media captures of the same CLUE type ("purpose"), or different encodings of a given media capture will all use the same SDP media stream (e.g. via RTP multiplexing with a different SSRC for each).
- It is acceptable to use more than one offer/answer exchanges to setup the whole Telepresence session.
- The Telepresence session is established by setting up sets of unidirectional streams (not bidirectional streams).

5. Encoding

The CLUE framework calls an RTP stream that the provider is capable of sending, an Encoding (note that an RTP session may contain multiple RTP streams, with each stream having a unique SSRC). The parameters of an encoding are: encodingID, maxBandwidth, maxMbps, maxWidth, maxHeight, maxFrameRate, as described below in the following Figures: Figure 1 shows video encoding parameters and Figure 2. shows audio encoding parameters.

Name	Description
encodingID	A unique identifier for the individual encoding
maxBandwidth	Maximum number of bits per second
maxMbps	Maximum number of macroblocks per second: ((width + 15) / 16) * ((height + 15) / 16) * framesPerSecond
maxWidth	Video resolution's max supported width in pixels
maxHeight	Video resolution's max supported height in pixels
maxFrameRate	Maximum supported frame rate

Figure 1: Video encoding parameters

Name	Description
encodingID	A unique id for the encoding
maxBandwidth	Max # of bits per second

Figure 2: Audio encoding parameters

6. Encoding group

In addition to encodings, the Framework defines an Encoding Group as a set of individual encodings plus parameters that apply to the whole set, as shown in Figure 3, Encoding group. The two parameters for the group of encodings are maxGroupBandwidth and maxGroupVideoMbps. Max group bandwidth refers to both audio and video, and is the

maximum value of the sum of all of the configured bitrates. The max group video macroblocks per second refers only to video and is the maximum value for the sum of all the macroblocks per second in the encoding group.

Name	Description
maxGroupBandwidth	Max # bits per sec for all encodings combined
maxGroupVideoMbps	Max # macroblks/sec all video encodings combined
videoEncodings[]	Set of available encodings (list of encodingIDs)
audioEncodings[]	Set of available encodings (list of encodingIDs)

Figure 3: Encoding group parameters

Additional characteristics of encoding groups include:

1. A media provider advertises one or more encoding groups.
2. Each encoding group includes one or more individual encodings.
3. Each individual encoding can represent a different way of encoding media. For example one individual encoding may be 1080p60 video, another could be 720p30, with a third being CIF.
4. While a typical 3 codec/display system might have one encoding group per "codec box", there are many possibilities for the number of encoding groups a provider may be able to offer and for the encoding values in each encoding group. The consumer uses the media capture attribute EncGroupID to know which set of encodings to consider using for that capture and what the constraints are.
5. In a provider advertisement, the sum of the maxBandwidths of the encodings advertised in an encoding group may total more than the maxGroupBandwidth, as these are potential encodings. Similarly for maxGroupVideoMbps. The group restriction however must be respected by the provider in terms of actual set of encodings used at any given point in time.
6. In a consumer request, the sum of the maxBandwidths of the encodings selected in an encoding group may total more than the maxGroupBandwidth. However the provider will send values under the max of the individual encodings and which sum to a total below or equal to that of the group max.

The following diagram Figure 4, Association of captures and encodings, depicts the association between the encodings (expressed in SDP) and the media captures, expressed in elsewhere (in a CLUE protocol perhaps). In this diagram the consumer can match Capture 1

with any of the encodings in Encoding group 1.

The consumer can match capture 2 with any of the encodings in encoding group 2, and match capture 3 wiht any of the encodings in encoding group 1.

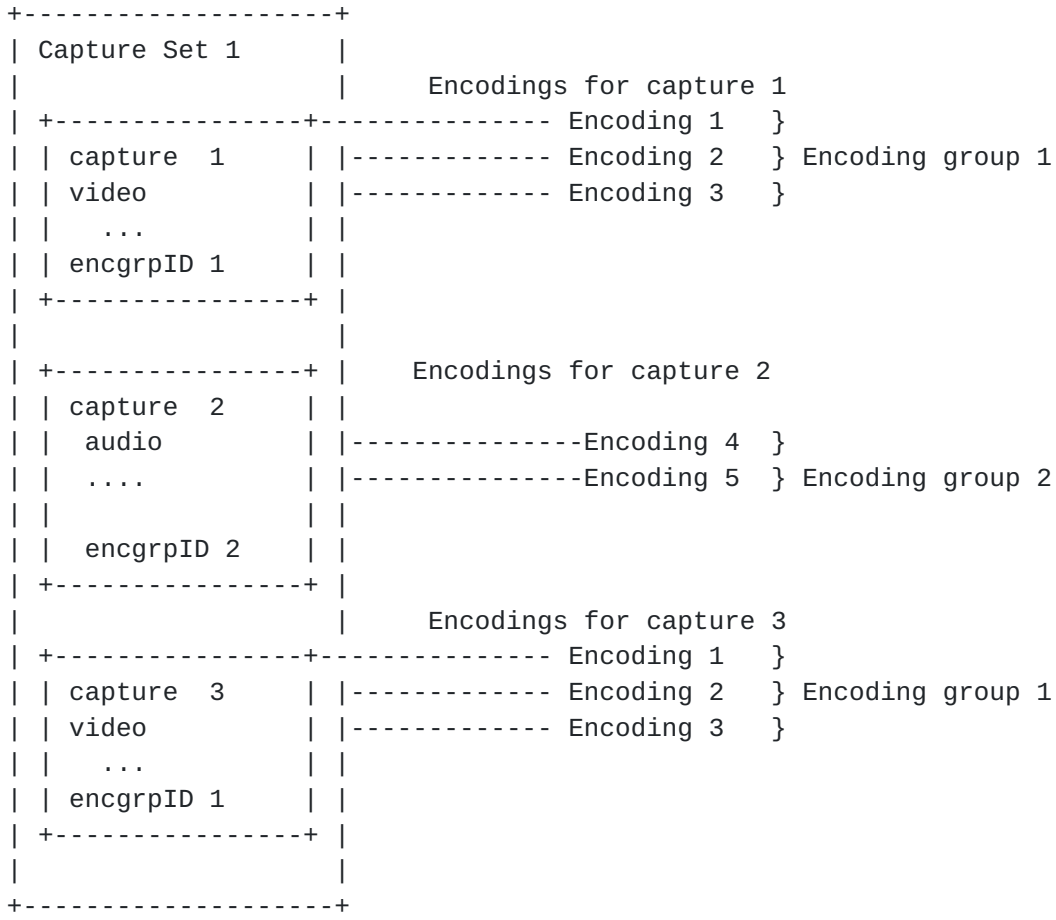


Figure 4: Association of captures and encodings

In this example showing a capture set containing 2 captures, the first captureID 1 is video and is mapped to encoding group 1, encgrpID 1. The consumer can map capture 1 to any or all of the potential RTP streams within encoding group 1. The use case for mapping a single video capture to more than one encoding at a given point in time is simulcast. The information contained in the encoding group parameters helps the consumer choose how to do this mapping by communicating the resource limitations.

7. Solution overview

Today telepresence endpoints actively negotiate audio and video SDP media streams using the SDP offer/answer model during call establishment and during midcall features, such as hold resume. However, in the CLUE framework, each party in the conference is usually both a provider and a consumer whether point-to-point or multipoint. The parameter values for one provider may well not be the same values for the other provider. Thus each party needs to advertise its provider encoding parameter values to the other side. This is not the typical communication model for SDP offer/answer, which is more often a bidirectional agreement on parameter values.

As specified in the CLUE framework, parties in a telepresence conference do not negotiate a single value between them; therefore offer/answer [RFC3264] is not used for the full negotiation of all encoding related values. Rather, we propose, sending CLUE provider advertisements for encodings and encoding group in SDP offer/answer, if it valuable for intermediaries to know these values. Then the consumer uses another protocol, such as a new CLUE protocol, to select which captures it wants at a given point in time and how it wants them encoded, subject to the encoding and encoding group parameters specified in the SDP. Provider advertisements not related to encoding, that is, capture attributes such as spatial relations would also be communicated in the CLUE protocol, rather than in SDP, as they are not provided as part of the SDP media descriptions for the streams. In this approach, the CLUE protocol carries the information related to media captures, as well as implements the control functions described in the CLUE framework including configuring the actual encodings used.

For example, if A and B are parties (either endpoints or MCUs) in a telepresence conference, A is a provider for B, and B is a provider for A. We propose that A and B send to each other provider advertisements for encoding and encoding groups through SDP; and A and B send to each other provider advertisements for Captures and consumer requests through a CLUE protocol.

The sending of advertisements of capture information and encoding information and the responding consumer request happens not only at call setup but also throughout the conference whenever there is a change in what the provider is capable of sending, or a change in which captures and streams the consumer wants to receive.

We will now explore two approaches for sending provider encoding advertisements in SDP. In the first approach, the provider encoding advertisements are accomplished in the initial offer/answer. A sends its provider encoding advertisements in the offer and B sends its

provider encoding advertisements in the answer. There are two message options here: bidirectional or unidirectional.

In the second approach, a series of offer/answer messages are sent from each participant. Both approaches have pluses and minuses. How each approach handles the situation where the parties advertise different types (SDP media capabilities) is of interest. For example one party may offer video-presentation and the other may not offer it.

8. Provider advertisement in the initial offer/answer

One possibility for exchange of CLUE parameters between 2 endpoints or MCUs is in the initial SDP offer/answer exchange. One party sends provider encoding advertisements as the SDP offer and the other party answers with its provider encoding advertisements, as shown in Figure 5, Initial offer/answer exchange. A then sends its capture information to B through the CLUE protocol (and vice versa).

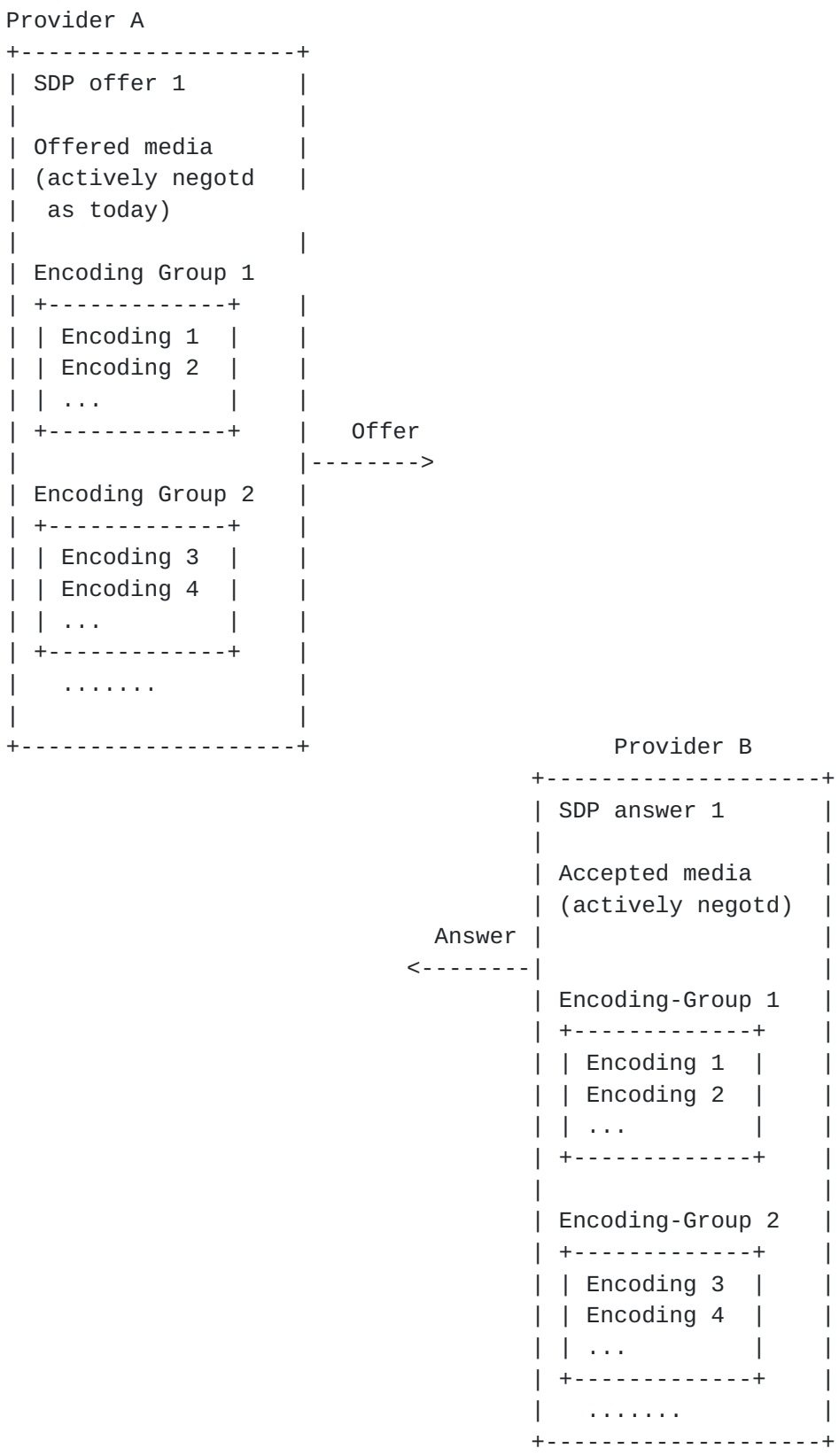
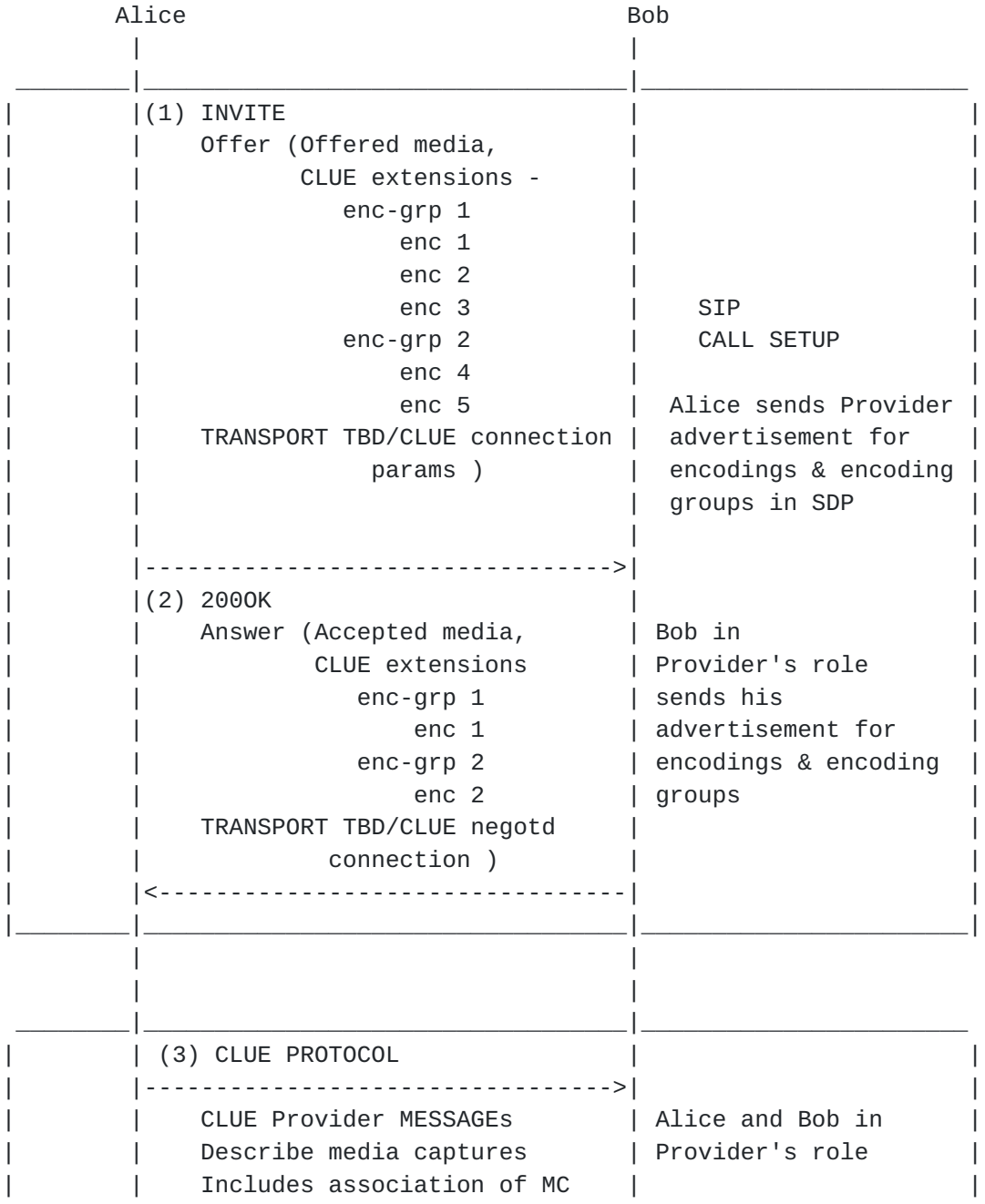

Provider A
SDP offer 1
Offered media
(actively negotd
as today)
Encoding Group 1
Encoding 1
Encoding 2
...
Offer
Encoding Group 2
Encoding 3
Encoding 4
...
....
Provider B
SDP answer 1
Accepted media
(actively negotd)
Answer
Encoding-Group 1
Encoding 1
Encoding 2
...
Encoding-Group 2
Encoding 3
Encoding 4
...
....

Figure 5: Initial offer answer exchange

There are two possibilities: birdirectional or unidirectional SDP messages.

Figure 6, Call message sequence illustrates an end-end call with the message sequence where provider advertisements are in initial offer/ answer.



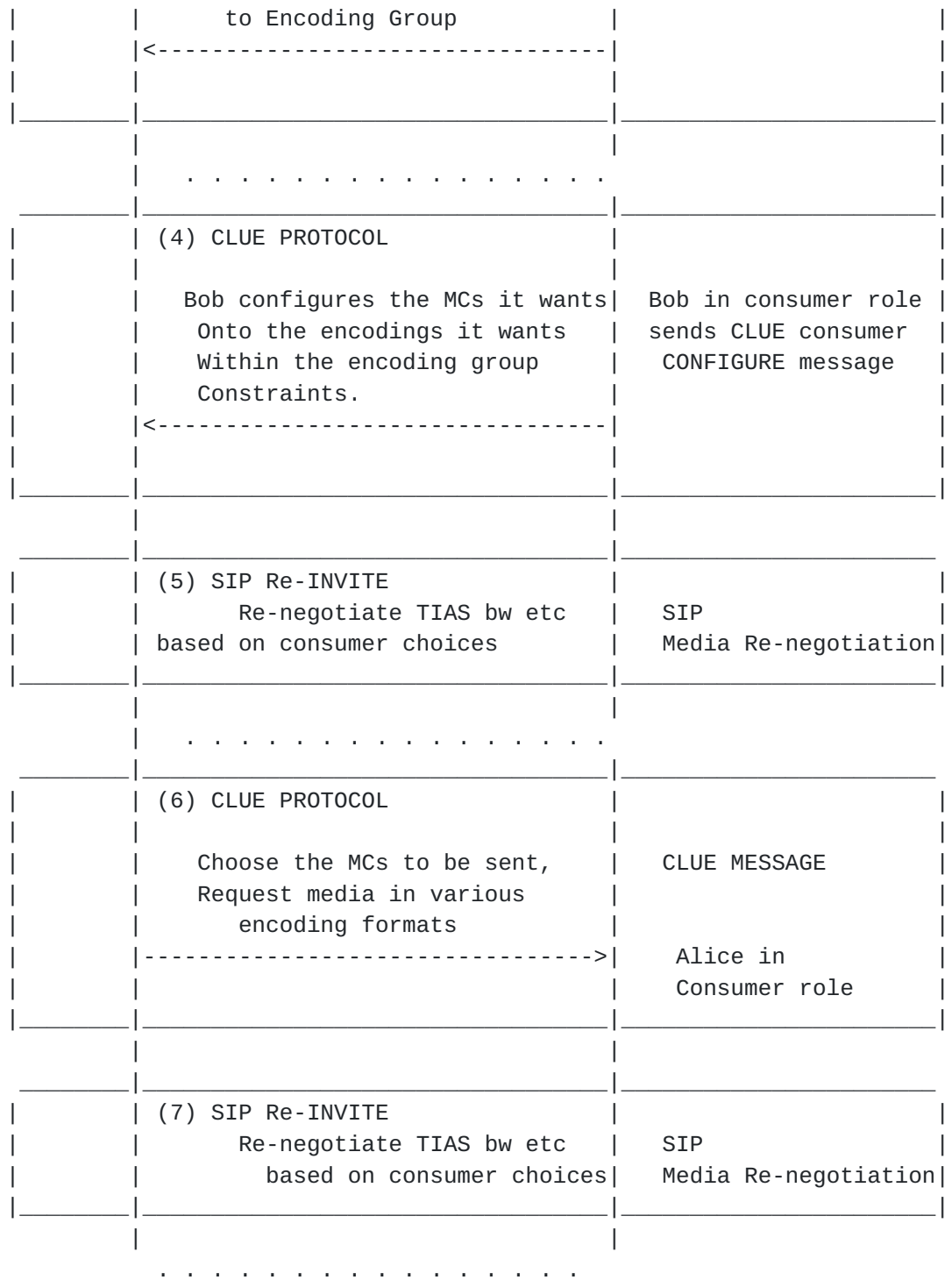


Figure 6: Call message sequence for CLUE advertisements in initial offer/answer

8.1. Bidirectional SDP messages

The challenge with a bidirectional approach is that the answerer must not respond with a type that was not in the offer. Suppose A offers two SDP media streams video-main and audio. B however wants to offer 3 SDP media streams: video-main, audio, and video-presentation. B cannot add a new stream type in the answer.

One way around this is for the offerer to always offer all 3 CLUE media types ("purpose") whether it will send them or not, i.e., be a provider. This allows both sides to negotiate use of each "purpose" as a provider and/or a consumer. The SDP direction attributes ("sendrecv", "sendonly", "recvonly", and "inactive") are used to indicate the real intent for a stream; a provider sends whereas a consumer receives, and hence a "sendrecv" stream covers both provider and consumer role. Streams can of course be rejected as per normal offer/answer procedures (e.gg., if neither side wants to use a particular stream such as video-presentation. An adjustment to reflect the real situation can also bemade in a subsequent offer/ answer exchange.

Using the example above, A will only send video main and audio, however, following the specification, A offers all 3 audio, video main, video presentation. B is going to send presentation, B accepts all 3 audio, video-main and video-presentation and hence generates an answer with them. The CLUE protocol negotiations take place. Then there is an updated offer in which A does not offer video presentation, thus making the SDP media streams accurate.

8.2. Unidirectional SDP messages

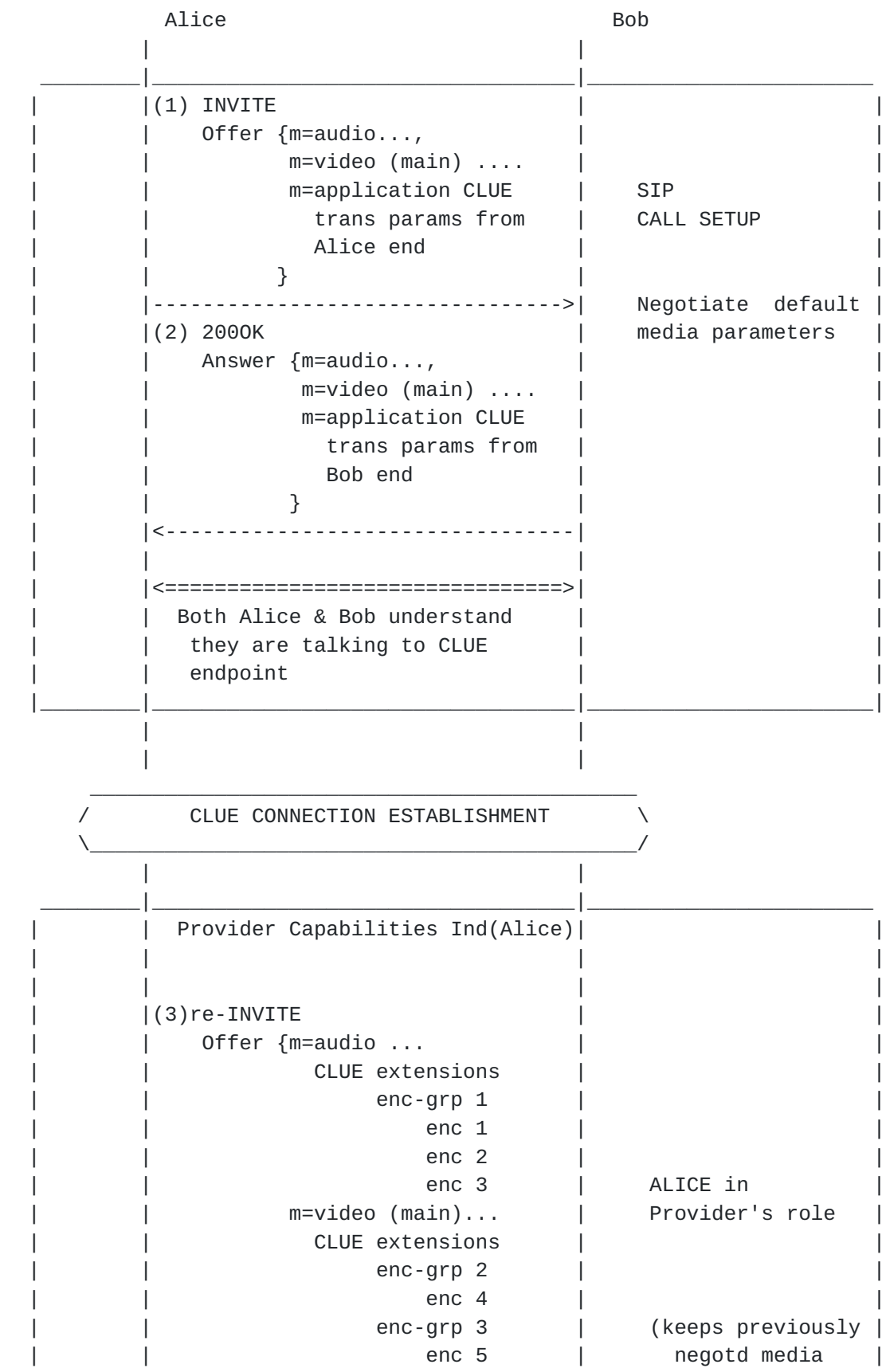
```
1) Offer (INVITE) from Alice to Bob
(Alice provider=sendonly, Alice consumer=recvonly)
================================================
m=audio 10000 RTP/AVP 96 ;Alice provider audio
a=sendonly
a=rtpmap:96 PCMU
...
m=video 10002 RTP/AVP 96 97;Alice provider video-main
a=sendonly
a=rtpmap:96 H264
a=rtpmap:97 H265
...
m=video 10004 RTP/AVP 96 ;Alice provider video-presentation
a=sendonly
a=rtpmap:96 H264
...
```

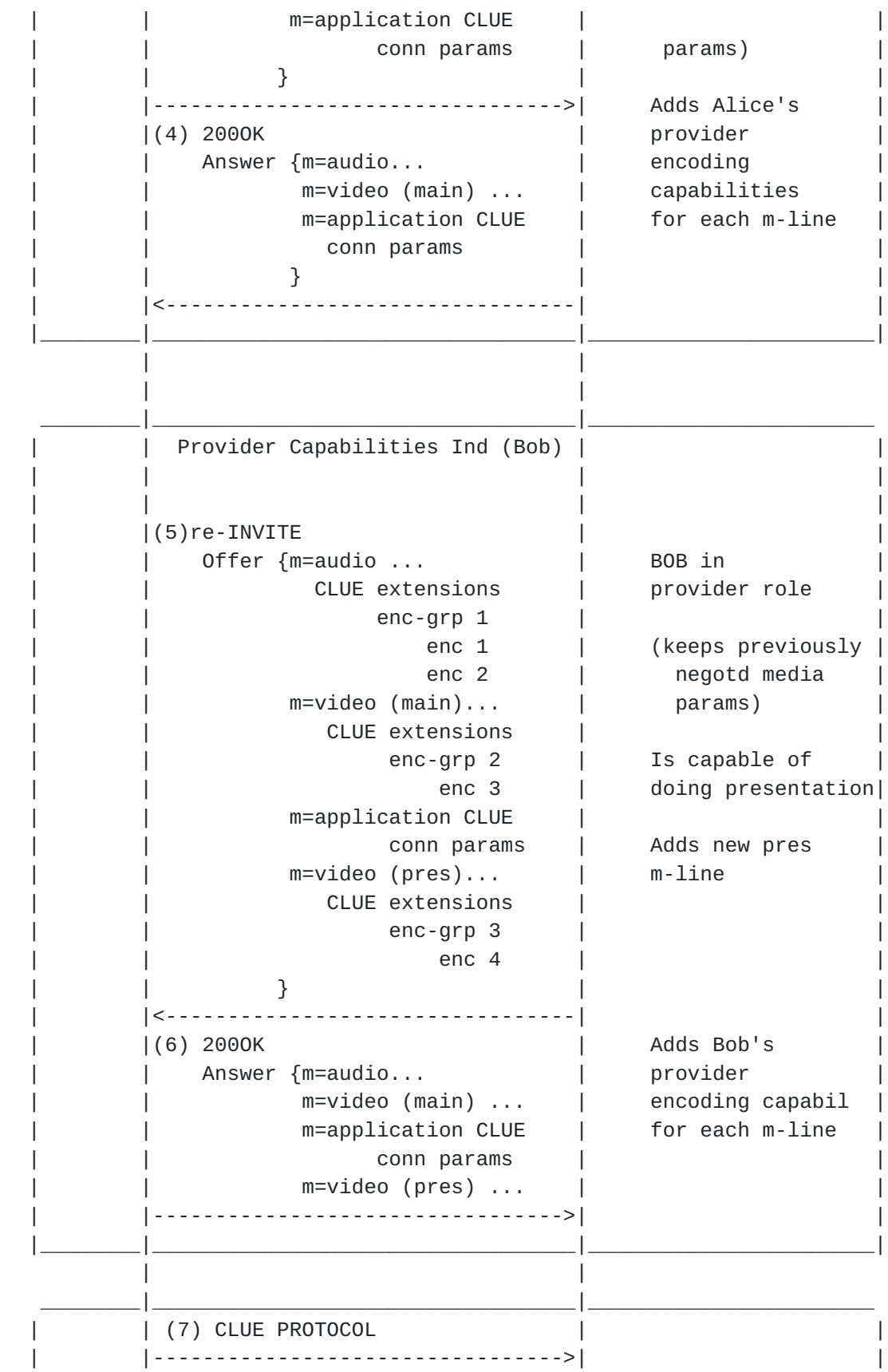
```
   m=audio 20000 RTP/AVP 96 ;Alice consumer audio
   a=recvonly
   a=rtpmap:96 PCMU
   ...
   m=video 20002 RTP/AVP 96 97 ;Alice consumer video-main
   a=recvonly
   a=rtpmap:96 H264
   a=rtpmap:97 H265
   ...
   m=video 20004 RTP/AVP 96 ;Alice consumer video-presentation
   a=recvonly
   a=rtpmap:96 H264
   ...

   2) Answer from Bob to Alice
   (Bob provider=sendonly, Bob consumer=recvonly)
   ==============================================
   m=audio 30000 RTP/AVP 96 ;Bob consumer audio
   a=recvonly
   a=rtpmap:96 PCMU
   ...
   m=video 30002 RTP/AVP 96 97 ;Bob consumer video-main
   a=recvonly
   a=rtpmap:96 H264
   a=rtpmap:97 H265
   ...
   m=video 30004 RTP/AVP 96 ;Bob consumer video-presentation
   a=recvonly
   a=rtpmap:96 H264
   ...
   m=audio 40000 RTP/AVP 96 ;Bob provider audio
   a=sendonly
   a=rtpmap:96 PCMU
   ...
   m=video 40002 RTP/AVP 96 97;Bob provider video-main
   a=sendonly
   a=rtpmap:96 H264
   a=rtpmap:97 H265
   ...
   m=video 0 RTP/AVP 96 ;Bob does not want to be a provider
        for video-presentation, so rejected this stream (port=0)
   a=inactive
   a=rtpmap:96 H264
   ...
```

9. Multiple offer/answer exchanges

An alternate approach to using the intial offer answer for the provider encoding advertisements is to do multiple offer/answer messages in which both sides make provider advertisements. The initial offer/answer establishes that the parties are CLUE capable and subsequent offer/answer messages exchange provider encoding advertisements. The disadvantage of this approach is that it requires a number of offer/answer messages.

In this method, the call is initially established with configurable default media values. The call flow is shown in detail in Figure 7


Alice
Bob
(1) INVITE
Offer {m=audio...,
m=video (main) ....
m=application CLUE
trans params from
Alice end
}
SIP
CALL SETUP
Negotiate default
media parameters
(2) 200OK
Answer {m=audio...,
m=video (main) ....
m=application CLUE
trans params from
Bob end
}
Both Alice & Bob understand
they are talking to CLUE
endpoint
CLUE CONNECTION ESTABLISHMENT
Provider Capabilities Ind(Alice)
(3)re-INVITE
Offer {m=audio ...
CLUE extensions
enc-grp 1
enc 1
enc 2
enc 3
m=video (main)...
CLUE extensions
enc-grp 2
enc 4
enc-grp 3
enc 5
ALICE in
Provider's role
(keeps previously
negotd media


m=application CLUE
conn params
}
params)
(4) 200OK
Answer {m=audio...
m=video (main) ...
m=application CLUE
conn params
}
Adds Alice's
provider
encoding
capabilities
for each m-line
Provider Capabilities Ind (Bob)
(5)re-INVITE
Offer {m=audio ...
CLUE extensions
enc-grp 1
enc 1
enc 2
m=video (main)...
CLUE extensions
enc-grp 2
enc 3
m=application CLUE
conn params
m=video (pres)...
CLUE extensions
enc-grp 3
enc 4
}
BOB in
provider role
(keeps previously
negotd media
params)
Is capable of
doing presentation
Adds new pres
m-line
(6) 200OK
Answer {m=audio...
m=video (main) ...
m=application CLUE
conn params
m=video (pres) ...
Adds Bob's
provider
encoding capabil
for each m-line
(7) CLUE PROTOCOL

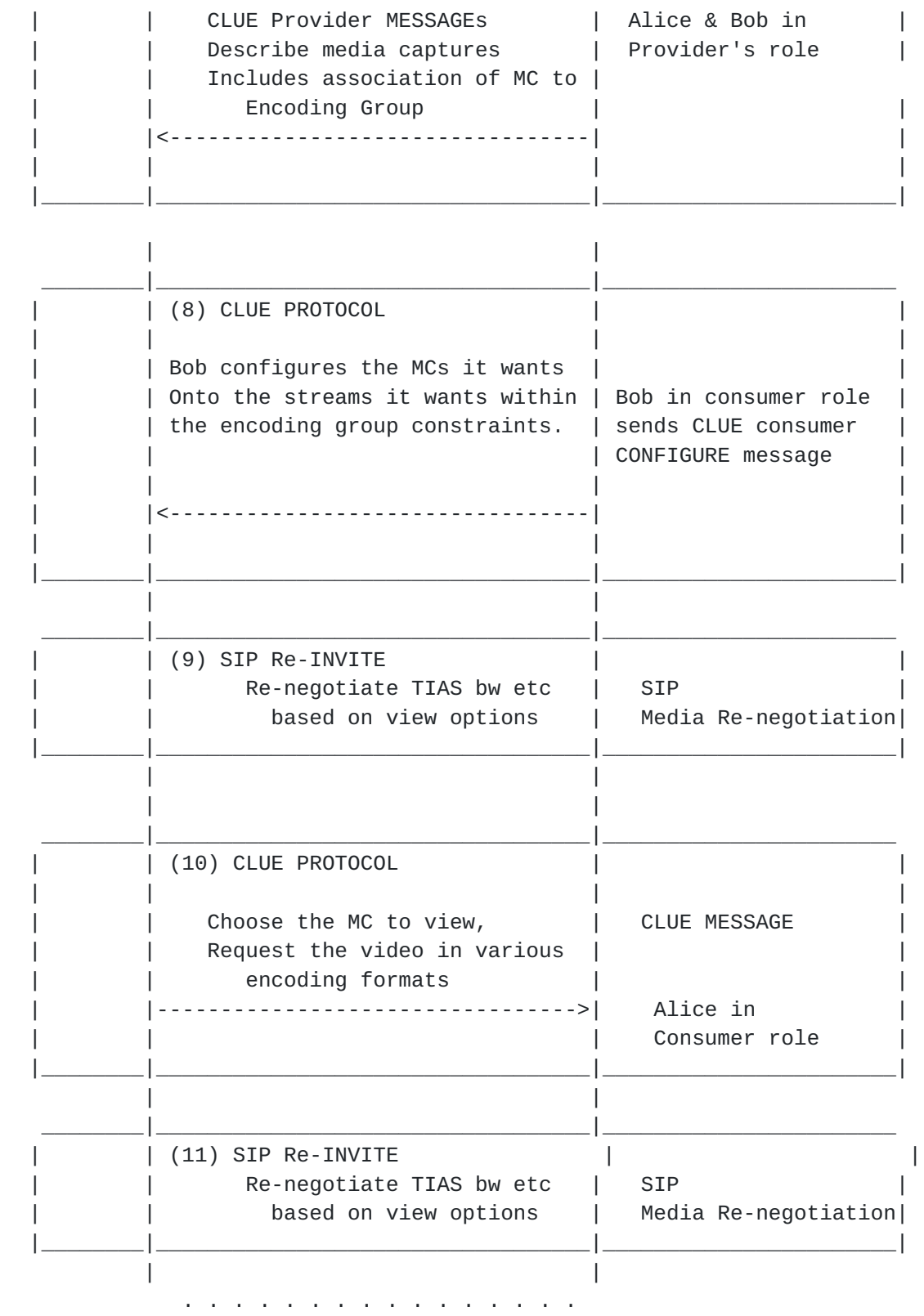


Figure 7: Call message sequence for CLUE advertisements multiple

offer/answers

The offerer establishes a CLUE application with transport connection. The answerer acknowledges, thus indidcating its willingness to participate in CLUE signaling. No encoding advertisements are exchanged at this point.

The CLUE connection is established

Alice, in the provider's role, indicates ecnoding and encoding group advertisement parameters using re-INVITE (new offer/answer). In this example, Alice does not indicate support for video-presentation. Bob acks back and does not provide any of his provider encoding advertisements at this point

Bob now signals his provider encoding advertisements using a fresh re-INVITE (third offer/answer). Bob is capable of video-presentation. It adds a video-presentation m=line and indicates its provider encoding advertisements. Alice may decide to not actively do presentiaton at this time, and may stop the media (port 0) in 200OK. But Alice is aware of the CLUE advertisement for the video-presentation to be used in the future.

Alice and Bob use CLUE messages exchange provide capture advertisements, including bindings to encoding groups.

Bob in consumer role uses the consumer configure CLUE message to select which captures and encodings he wants.

Bob sends a re-INVITE (4th offer/answer) to negotiatie TIAS bandwidth based on the consumer requests. This enables call control and intermediaries to manage bandwidth resources.

Alice in consumer role uses the consumer configure CLUE message to slect which captures and encodings she wants.

Alice sends a re=INVITE to negotiate new TIAS bandwidth based on the consumer requests, updating previous parameter values.

This process continues as options change at either end.

10. Representation of Encoding and Encoding group in SDP

We want a mechanism that will allow the provider to advertise encodings and encoding groups. We can define three new SDP attributes: clue, enc, encgrp.

Encgrp is the encoding group. It consists of an encoding group ID, a list of encodings, and a list of encoding group parameters, maxGroupBandwidth, maxGroupVideoMbps.

a=encgrp:<encgrp-number> <list of enc nums> <list of group params>

Enc is the encoding. It consists of an encoding id and a list of clue parameters and their values for the encoding.

a=enc: <enc-number> <list of clue nums>

Clue consists of CLUE encoding parameters: max-fps, max-mbps, max-bitrate, imageattr. The parameter list can be extended in the future.

a=clue:<clue-number> <clue attribute=value>

where clue attributes in our context would be "imageattr[x= ,y=]" [draft-ietf-mmusic-image-attributes],"max-mbps", "max-br", max-fps

Example of 2 video encodings:

```
a=clue:1 imageattr[w=1920, y=1088]
a=clue:2 max-fps=60
a=clue:3 max-mbps=244800
a=clue:4 max-br=4000
a=clue:5 imageattr[x=960, y=554]
a=clue:6 max-fps=30
a=clue:7 max-mbps=61200

a=enc:1 clue=1,2,3,4          Encoding 1 and its clue attributes
a=enc:2 clue=5,6,7,4          Encoding  2 and its clue attributes
Example of 2 video encoding groups
a=encgrp: 1 enc=1,2,3 grpparm=3,4,5
a=encgrp:2 enc=4,5,6,7 grpparm=2,4,5
```

Figure 8: Example 2 video encodings

11. Concerns about conveying CLUE parameters via SDP

There are some potential concerns in using SDP for CLUE provider capability announcements. Passing information to two separate protocols and processing it from them complicates the protocol and implementation. Secondly, if the two SIP peers do not communicate directly (e.g. due to SIP Record-Route), then reINVITEs/UPDATEs may be delay-prone compared to an independent and direct end-to-end protocol; this is relatively unimportant so long as any parameters expected to change repeatedly, or that must be sent when a stream switches (i.e., near real-time) are kept out of the SDP.

The encoding and encoding group will potentially be large and significantly expand the SDP size. A compact representation is suggested in the above to try and alleviate this. Further work and testing will be required to see if it is a problem in practice, and whether it is inherent to CLUE to begin with.

The presence of the CLUE protocol negotiation provides a way to ameliorate this problem for systems operating in a mixed CLUE/ non-CLUE environment that don't wish to send the information to non-CLUE endpoints: the initial offer contains the CLUE protocol m-line, but not the CLUE media descriptions. The answerer sees that the caller is CLUE-capable and includes the media description in their answer, and the offerer can then reINVITE once negotiation is complete with a media description.

The value to the intermediaries of the encoding parameters needs to be considered carefully. The intermediaries will have the TIAS value which is updated after CLUE protocol activity. What about the encoding parameters? It is unclear they are of much use to intermediaries by themselves - they are the potential maximum values of potential RTP streams. How are intermediaries expected to use these?

It seems the encoding group parameter values, especially the maxGroupBandwidth could be of interest to intermediaries augmenting the TIAS information. An alternative to including provider advertisements in SDP might be to include an attribute in SDP for the sum of all the Encoding group maxGroupBandwidths, which seems more valuable than all the individual encoding group maxGroupBandwidths.

Another factor to consider is that suppose an intermediary changes the maxGroupBandwidth value - that change would be seen by the Consumer, not by the provider, and it is the provider who is the one who needs to see a change in encoding group available resources. This could require actual negotiation of these parameters, instead of the currently declarative unidirectional use specified in the draft

currently.

Note on real time requirements: The provider advertisements and Consumer selection should be as quick as possible, but do not have hard real time requirements. They can be accomplished within the time scale of signaling.

There is however one scenario that has timing requirements more stringent than signaling typically provides. Consider a switch=true audio capture, the information on the spatial location of the audio capture needs to happen in real time. The speaker changes frequently and the location of the speaker must be known quickly (in near real-time) in order to assign the incoming stream to a decoder.

Editor's Note: add reference to draft-clue-hansen

12. Security Considerations

TBD

13. Acknowledgements

Thanks to Rob Hansen for discussions on the advantages and disadvantages of using SDP.

14. IANA Considerations

TBD

15. References

15.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

15.2. Informative References

[I-D.ietf-clue-framework]
Romanow, A., Duckworth, M., Pepperell, A., and B. Baldino, "Framework for Telepresence Multi-Streams", draft-ietf-clue-framework-06 (work in progress), July 2012.

[RFC3264] Rosenberg, J. and H. Schulzrinne, "An Offer/Answer Model with Session Description Protocol (SDP)", RFC 3264, June 2002.

Appendix A. Changes From Earlier Versions

Note to the RFC-Editor: please remove this section prior to publication as an RFC.

A.1. Changes From Draft -01

- Version -02 is being submitted to keep the document available for discussion until decisions are made.

Authors' Addresses

Allyn Romanow
Cisco Systems
San Jose, CA 95134
USA

Email: allyn@cisco.com

Flemming Andreasen
Cisco Systems
Iselin, NJ
USA

Email: fandreas@cisco.com

Arun Krishna
Cisco Systems
San Jose, CA
USA

Email: arukrish@cisco.com