Interim Meeting, Feb 15 morning – John Leslie

BlueSheet

c1: Paul Kyzivat
l8: Zaheduzzaman Sanker
l7: Bo Burman
l6: John Leslie
l5: Christian Groves
l4: Roni Even
l3: Jonathan Lennox
: Keith Drage on WebEx
l2: Gyubong Oh
l1: Stephen Wenger
r1: Spencer Dawkins
r2: Andy Pepperell
r3: Allyn Romanow
r4: Gerard Fernando
r5: Brian Baldino
r6: Espen Berger
r7: Mark Duckworth
r8: Rob Hansen
: Scott Pennock

CLUE Interim 15 Feb 2012

(trying to get wireless to work)
0906 recording starts
0910 Mary: starting, "read the NoteWell"; agenda-bash, new use-cases, need feedback on list; framework (Mark), break, Roni on framework; lunch, clue data model; rtp usage (discuss again tomorrow)... day two...
Roni: two separate issues under RTP usage...
Mary: breakouts tomorrow
Roni: going to 5:00 p.m. tomorrow?
0916 Mary: Gyubong Oh
Gyu: several comments on-list
(slides mostly illegible due to projector glitching)
Gyu: video streams plus a presentation stream; proposing that presentation stream can be managed by all parties, dynamically
Roni: should questions wait?
Gyu: yes; next slide: multiple devices use cases (illegible)
(interruption, propose to download slides to our computers... projector got better)
Gyu: propose media stream distribution by ?
Roni: not a limitation of BFCP H239 issue: need to write BCP of how to do this with BFCP
Jonathan?: remote control...
Rob: BFCP is about control of floor

Roni: can be more than one control; two persons can have the floor, the rest is up to the application; BFCP allows a process to be floor-control
Espen: CLUE can express technical, but not the user experience, most folks are satisfied switching by audio level
Roni: if you want more than one, you need a BCP document explaining how to use the application doing multiple-control
Espen: interesting use-case, but should be seen as something separate
Jon: do we need a way of associating capture-set
Mark: floor-control is out-of-scope for CLUE
Roni: say we have one capture-set for presentation, everyone has to choose video that goes with it
Espen: no mechanism today... think that's a good thing
Jon: if you have policy not defined in CLUE...
Roni:
Jon: need to understand interaction of CLUE with BFCP
Mary: can we go on... next use-case... (trying to get WebEx to show right slide)
(lunga pausa...)
0951 Mark: issues around framework document...
Mark: version 2 major changes spatial relationships, v3 higher-level reorganization, more concise, address some review comments
Mark: capture set, top level, area of scene, scale, simultaneous sets (media captures), list of capture set entries (each a suggestion of grouping media capture), one could be audio, another video
Mark: attributes for each media capture: area, point, audio format... open issue on composed/switched; possibilities for how it's encoded; each media capture has pointer to encoding groups (may duplicate)
Mark: information from provider to consumer, could include multiple capture set
Roni: simultaneous-set is what's needed... list of what user can ask for; user can say take one with all entries and create his own set; the lists are suggestions, may be subsets of the complete one... needs to be clarified in documents
Andy: zero or more pairs, which can't be supplied simultaneously;
Roni: in each set lines, all MCs in set 1 can be supplied simultaneously
Jon: two ways to encode... tradeoffs... four cameras, each with two modes; you can say pairs that don't work or sets which do work
Roni: limitation is entries have subsets
Andy: one other point: might be a need for simultaneous sets to not be capture set
Espen: doesn't matter... one of the restrictions of what you can send simultaneously
Jon: is the intention to match uplink bitrate
Andy: no
Paul: at one point bandwidth was a constraint
Mark: that was ??
Roni: we just need to clarify... how you can select all or part
Roni: framework starts from capture set... no definition of anything above capture set
r5: in framework we have spatial definition "above"; but that's not hierarchy
Roni: we discussed this in Taiwan
Paul: can you have two endpoints sending capture sets for the same scene?
several: no... we had something... how do you describe the environment
Jon: two different endpoints in same room... they can coordinate or not
1011 (trying to get next slide)
Mark: agree with Roni, we need to get to how individual encodings relate to SDP... encoding group, includes video encodings, audio encodings (ID, bandwidth, height, width, framerate)

Andy: example two sets, cannot sum to more than ?
Jon: consumer request, may be less than provider limit
Roni: other constraints... my view, what you have in video encoding are 6184... same thing; nothing like "group" available in SDP, not a CLUE-specific problem; we're talking about structure inside CLUE data-model, we don't want to duplicate information.
Roni: I understand the concept of encoding groups; I object to duplicating information
(free-for-all)
Mary: this is not the data model, this is framework of information CLUE needs
Jon: stream coming out of transcoding MCU... single transcoding... do you need to advertise the limit
Mark: advertise 1080p30, but provider can't supply a lesser requesti
Roni: provider response, can't do that, but this is close...
1028 break
1036 Mark: spatial relationship, discussed in Taiwan, 3D cartesian
Jon: why area of scene and area of capture separate?
Brian: may be worth making more explicit
Espen: rendering hint...
Mark: area of scene is higher level
Roni: area of scene not applicable in MCU case
Mary: area of scene is not physical, but virtual
Stephan: camera axis could be diagonal through center, can we capture camera axis through ??
Roni: need scaling information
Mark: can consumer construct? yes (including angles)
Stephan: this picture doesn't show... axis doesn't go through lens... we should write this down
Mark: not sure I got everything
Stephan: geometric correction you would need...
Mark: do you think there _is_ enough information?
Roni: can only be deducted if known scale
Mark: "camera axis perpendicular", are you saying that should be required
Stephan: on whiteboard... camera axis, second camera, same field, possibly wide-angle, possible if not sensible; you can deduct camera axis, point is smooth geometric correction won't work without this input; I would say current info is fine, you can calculate, but this isn't written down
Roni: I support... NB case where three cameras are in center
Stephan: I'll provide text on-list
WebEx: optional attribute saying geometric correction already done
Paul: troubled by area-of-scene, defines an area, not a volume, and you see a volume
Stephen: question of definition, you're not really defining...
Spencer: are you asking about somebody passing in front?
Mark: plane of interest
Brian: you might have horizontal for microphone, vertical for camera
Paul: in theatre-of-the-round, how would you do area of scene; seems to me we need volume of scene...
Mark: this simplified version seems good enough... didn't cause problem
Spencer: back-wall is the limit
Jon: you're describing the thing you want life-size
Espen: we don't have use-case where this is needed, we could drop it... I don't know how to use it as a receiver.
Paul: if you're only offered one camera, you'd want to know if it was center or right...
Mark: you can figure out where it fits into the overall scene, receiver would always use differences
Christian: you're only sending one (without sending area of scene)

Brian: maybe we need to be more explicit about area not listed
Mary: think we need to take this to list
1101 Allyn: previous topic, are we comfortable with different terms, there has been confusion, we wanted to make these constructs are clear
Stephan: document is greatly improved
Zahed: difficult to read (area of capture?)
Mary: ASCII-art would help
Gerard: every capture...
Paul: encoding group can span multiple captures
Zahed: some not instantiated
Allyn: consumer sends back request, then they happen
Roni: text is still missing something
Espen: scenes have name (speaker, audience), high-level examples would help
Brian: idea of capture scene is useful
Jon: are scene and capture-set synonomous? if so we should pick one name
Brian: scene is conceptual, capture is representation
Bo: "entries" unclear in text: when you define entries, they're supposed to be mutually exclusive
Several: no
Andy: not the job of CLUE to say you should choose just one; two captures _could_ be viewed as alternatives
Brian: how many you select is based on other constraints
Andy: first determinant is whether you're an endpoint
Mary: examples could help
Andy: they are alternative representations of the scene; every entry is an alternative representation; the coordinates may not give you enough information
Brian: it's a suggestion from the provider side
Roni: what the provider thinks will make sense to the consumer; the consume will not normally select more than more
Brian: provider isn't trying to guess what consumer wants: he's describing himself
Mark: provider listing alternatives because some consumer will find that useful; one consumer will look for something "easy"; I agree with Roni that the consumer doesn't need the info
Jon: "which do I trust" is something I'd rather avoid
Rob: entries become useful when you start thinking virtual rooms
Zahed: entries are required, recommendation aren't?
Andy: simultaneous sets may span multiple captures; example middlebox supplying two swithched... capture sets are always usable
Roni: currently we have physical units and unscaled units, plus no-physical units; I say you also need "no order at all"
Brian: you could provide no information at all; we can explicitly say "if no spatial info"...
Brian: I think we just need to say "if no spatial info" to make explicit that option
Christian: if you have area of capture, you need to include a scale
Mary: we need to move on
Mark: ticket 5 & 7, we should combine; composed/switched... instead of boolean, list of alternatives
John: what is the benefit of "composed"
Rob: MCU is likely to be doing the composing for resource-limited receivers
Espen: some choices have to be done early -- e.g. camera-pointing
Roni: consumer will be able to select from multiple choices, provider shouldn't excessively limit
Andy: consumer-capability message

Andy: start with consumer advertising capabilities, then provider sends options, then consumer chooses... provider must not send before seeing consumer advertising: which media type, capture and capture-set attributes it understands: forward-compatibility
Paul: would this include constraints -- attributes, media types, would you want bandwidth? combinatorial... which ten of a million choices.
Andy: other things could be in there; any new attribute, even proprietary, can be added
Allyn: do people think this idea is good?
Stephan: WG needs to nail-down sooner rather than later what happens after connection establishment
Espen: extensibility -- maybe not the only way
Roni: no need for three messages to have extensibility
Jon: "I understand CLUE" -- could this be part of that
Stephan: that's probably a detail
Jon: depends on scope... customizing...
Mark: voice activity detection -- more specific proposal to list
Mark: ticket 4 sharing metadata -- not ready yet
Jon: WebRTC
Mark: other review comments not in issue tracker, maybe talk about later in Interim
1142 break for lunch, to resume at 1300

1302 Mary: CLUE Framework
Roni: some things I think more important to discuss (slide, 6 major issues)
Roni: consumer capabilities, is there a need for three messages, looks redundant to me, could be optional; we'll have separate discussion about this
Roni: CLUE and SDP media description,
Jon: do you mean m-line?
Roni: terminology not very well specified in those documents
Roni: draft-even-clue-rtp-mapping... left media wide, left media zoom can map into same media capture, one payload type into more than one media capture... combination of ssrc and payload type... depends on what media capture is...
Jon: not clear how this works with encoding groups
Roni: separate issue; number of m-lines is bounded by number of captures?
Jon: captures with different values can't share same m-line
Andy: captures and m-lines are very different, no 1-1 mapping, might have simulcast
Jon: an m-line that says this will carry up to...
Andy: doesn't follow that you need an m-line for every ??
Mary: could we go on to your encoding?
Roni: any mapping can be done using...
Roni: Individual encodingq 5 variables, some codec-specific, all specified also in SDP; propose to map SDP info to CLUE and not duplicate
Gerard: if 264 comes up...
Roni: different things for different codecs
Andy: how do you express that mappings can be split... total encoding power...
Roni: some systems limited by computation, not bandwidth, OTOH this is not CLUE-specific
Andy: this will impact CLUE more, need to come up with some way of modeling
Roni: encoding is unrelated to semantics of stream, not necessarily part of CLUE
Andy: idea of coming up with some number which represents ability
Jon: need to understand kind of encoding semantics we need for CLUE, that present a huge bucket of requirements to MMUSIC

Andy: you might say within an encoding group you can't have multiple codecs
Roni: here you bound it... some information I want to have... the way you bound it together is not right
Andy: are you saying it's wrong to bind every media capture to exactly one encoding group?
Roni: constraint of multiple, not one... what I'm saying is this assumes you're using one codec
Espen: very useful to support 264 as a start
Roni: if you don't limit to one codec you have no way to support more than one
Espen: try to define a metric like how many pixels you can send
Mary: would you (Roni) be satisfied if we specify only-one?
Roni: we had this discussion on ?? list; H323 designed by video folks, SIP by audio folks
Andy: this was intentionally simplistic
Stephan: mark this as needing further study -- doesn't matter for audio at all; matters on video -- limited number of those (presently)
Roni: resource management is always a critical issue
Roni: other issues, switching/mixing; CLUE message rate; CLUE message size
Mary: we could open issue-tracker items
Stephan: general discussion isn't something an issue-tracker can manage
Mary: we can't talk about message size, e.g., yet
Jon: message size and rate depend on so many other issues...
1338 Espen: MCU behaviour
Espen: use-case one, Alice wants MCU to determine use of two screens
Espen: policy use-cases, segment/site switching; hollywood-squares; select who is on the screen
Espen: typical three-screen room, composed carries no info how it's composed, swithched shows any one of three cameras
Espen: MCU - example; you can't know what is on the screen: A is 3-camera, B, C, D, individual cameras (B with five people)
Espen: 3x stream offer, 3 video captures, three audio captures
Mark: names of streams represent what?
Espen: you cannot assume what they contain
Paul: are these names intended to be shown to the end-user?
Mark: I think this came from an example where the names were not conveyed to end-user
Paul: these things can be recursively composed... get too small
Espen: in this case, no relation screen-to-screen; could be useful for MCU to have same policy for all three captures
Mark: what does "no spatial information required except" mean?
Mark: one intent of area-of-capture is to match audio to video... if you want to match audio to video, you would have spatial info
Roni: specific rule for lip-sync... you want location sync, not lip-sync... lip-sync is synchronizing timing of video and audio (not the person speaking)... it's not for CLUE to do the synchronization
Andy: which streams you sync will change during the call; if streams are swiched, the synchronizations will necessarily change
Jon: if they're both switched...
Mark: Andy, lip-sync example, current doc talks of sync...
Andy: I think receiver should do syncing wherever relevant
Jon: you want all streams from same room sync'd... probably corresponds to a capture set
Espen: Switch policy-Site; when site A speaks, three screens replaced
Jon: how much is MCU communicating "what's important"
Espen: MCU with <switched-policy>
Mark: looks like policy associated with single capture; vs. capture-set...

Espen: in this case, I think captures should have the same policy
Espen: MEC with <video-layout>; e.g. hollywood-squares
Espen: Participant-lock; if MCU advertises option, receiver can request
Espen: Correlate;
Espen: Issues; 25-streams...
Paul: policy has to be combined with set of input captures to be useful
Espen: MCU should be capable of understanding how to combine all inputs... choose... don't think you need to know (all) individual inputs
Paul: two ways to represent -- each screen independent, pick any three (different policies); or not that level of freedom: must pick policy first
Jon: use case there is source-selection, everything else switching
WebEx: does this mean participant is always locked?
Andy: somebody might be providing zero appropriate capture; up to provider what to do in absence of locked participant
WebEx: not convinced that conference should be locked to a particular capture
Jon: my model is I can choose a capture set which is just that one person
Espen: CLUE has never talked of understanding individual participants... policy on top which may be available
Paul: didn't we discuss the distinction between captures and people, and decide that people is out-of-scope... deal with it down to the resolution of a capture
Jon: I'm interested in the left camera at that site... other somewhat related issue, we have no way of doing life-size-displays, geometric captures... but available for all other sites...
Allyn: was there some kind of decision about participant-lock
Espen: CLUE should be extensible
Andy: if we could do participant-lock, at which level should it happen?
1431 Mary: data model (3:20 item on agenda), we just finished the morning stuff, want to finish review of slides
Paul: started drawing this, wasn't thinking of presenting -- for my own understanding; this shows relationships more explicitly;
Paul: didn't elaborate audio yet... capture video comes from a camera; switched has algorithm for switching... set of inputs to a mix; for mixed you need to show how they're positioned.
Paul: still divorced from encoding; doesn't need to all be the same encoding
Gerard: is audio going to be similar to video
Paul: didn't know what to put in there, so I put in nothing
Stephan: up-arrow shows?
Paul: that's inheritance; all other lines are relationships, if no arrow, bi-directional
Paul: rendering belongs where? what is endpoint going to do with it?
Mark: this is modeling something different from provider advertisement
Paul: closer to a model of a provider... replace room with endpoint, this represents data you need in an endpoint, we may or may not want to model receiver as well
Mark: what a provider needs to fill in... don't think render device belongs here... maybe capture-set here is same thing as media-capture
Paul: you could have single device capturing audio and video; if there's nothing interesting there, we could excise it
1446 Andy: hierarchy of elements, not XML exactly... intended to express structure, which things are mandatory
Gerard: significance of * and !
Andy: * means zero or more; ! means at least 1

Mary: if we did UML, elements should have same names; do people want UML?
Mary: this is defining a message
Jon: defining state after message
Mary: message may be a misnomer...
Roni: supported-media-capture-attribute... usually we say we support RFC NNNN
Andy: listing every attribute would be the cleanest way
Stephan: profiling based on document is a bad idea
Jon: feature-tags is the usual model
Rob: support-RFC can define things which are not attributes
Andy: after receiving this list, you should know exactly which attributes are supported
Mary: take decision on how-to-model to list
Christian: what is the relationship of this to the framework document?
Allyn: this updates framework
WebEx: which gets us to automatically-generating the protocol?
1500 break
1534 Jon: RTP Usage for CLUE
Jon: RTP Requirements; easily dozens of media sources at a time; potentially choosing from hundreds; asymmetric, dynamic
Roni: hundreds of inputs? or participants?... signaling, not necessarily media? No problem
Jon: Architectural constraints; don't want to confuse endpoints, middleboxes; don't want more than about three UDP streams
Roni: what do you mean, confuse?
Jon: Source multiplexing; all of same media type over single RTP session; if you want to do them separately, you can... e.g. you need different quality for them; could have splices model going to different IP addresses; consumer might request capture X on transport stream Y
Roni: if I have one-stream, can ask for camera or presentation...
Jon: transports with different characteristics, want to match to streams; didn't want to require bundle for this (bundle is proposal that different m-lines want to use same transport)
Jon: Complications; when you receive, need to know how it corresponds to your requests
Jon: UseCase1; sources constant over complete session
Jon: UseCase2; receiver has only two screens, source might move between captures; provider notes something interesting happens on other side of room, both screens switch
Jon: UseCase3; MCU chooses which site to relay
Jon: Mapping sources; SSRC is a random number (can't encode anything into it)
Jon: Need for accurate mappings; which hardware is required, e.g.; have to throw packets away; some models where you learn this 500 msec later
Jon: Need for flexible mappings; moving among captures; don't want to force an i-frame; case where decoding needs no change (though rendering does need to change); avoid unnecessary consequences
Gerard: why are you switching
Jon: sending i-frame has negative consequences, can avoid it, e.g. swiching among 16 particpants, nth-most-recent; there will be cases that need i-frame, but this doesn't
Roni: that breaks some switching cases, where parameters overlap each other
Jon: if every source is a different SSRC, not an issue
Jon: Sending mapping outside media; CLUE messaging reliable, but not time-guaranteed, might be heavyweight, RCTP SDES unreliable, etc.
Andy: does this break on two instances of same capture?
Jon: maybe capture-ID is the wrong choice
Roni: for switched capture, why not provide every SSRC that could happen

Jon: could be an enormous conference, thousands; every distinct encoder state has its own SSRC
Espen: unnecessary to send every SSRC up-front
Jon: different SSRC
Roni: what if he's providing his own SSRC
Jon: that would break moving-capture... every distinct low-level encoding is own SSRC
Rob: need to re-encrypt the entire message... referencing back to things that no longer exist... decoder doesn't know what changed
Roni: if I'm receiving streams, and one changes... header extension to say same
Jon: other thing: lip-sync bound to CSRC? which audio to sync to
Roni: in most cases, the audio would be mixed (not switched)
Espen: quite common to use switched audio
Jon: Sending in media stream; no latency, no confusion; processing load for MCU, could push you over MTU limit
Jon: RTP Header Extension; could be costly, have to re-authenticate, "MUST only be used for data that can safely be ignored" -- wouldn't be in-scope for CLUE
Jon: When to send CaptureID; every-packet, as-needed, or for some period after any change
Roni: how does captureID get created? how link to what codec, etc.
Jon: that's orthogonal, payload type; if I have an SSRC, encoder could switch among allowed payload types; my model of switched capture is SSRC represents low-level capture being switched, metadata allows you to understand...
Roni: that's signaling... different users map payload-type differently
Jon: who I got it from, I know they are different things to encode
Bo: if you define SSRC being between receiver and mixer, how would it change the meaning in RTP?
Andy: it doesn't change that at all
Roni: requirement you have, if you switch stream currently being rendered, it will continue to have same SSRC... now rendered to a different place... OK...
Roni: would like to have some requirement to reference when I write anything
1626 Bo: Multi-stream Media
Mary: not necessarily something CLUE would do, not the document in DISPATCH
Bo: informational, presenting nuts&bolts view that could be useful; ties together how endpoints with different capabilities cooperate in conference
Bo: Overview; tries to avoid trans-coding
Bo: Assumptions; different "quality categories", high-quality endpoint can receive lower-quality stream
Bo: Low Quality Sender; where to render if you have spare real-estate
Bo: Medium Quality Sender; "dual high" means two screens
Roni: what is distinction two 30" vs one 60"
Bo: dual can present two different things; one sender, single camera, receiver with two screens, choose which
Roni: why "dual" instead of "multi"
Bo: low receiver cannot receive "out-of-box"; could introduce transcoder; could send two qualities; also scalable encoding
Bo: Dual Channel Sender; same but reverse, receiving chooses which to render; policies as we discussed earlier
Bo: Multiple Channel Sender; announce how many sources you can send, mixer could announce "many", receivers limit how many actually sent
Bo: Multi-quality Local Composition; mixer sends many, receiver does composition of those
Roni: what entity is announcing this, how does this relate to RTP mixer?

Bo: Mixer Stream Roles; SSRC is constant from mixer, CSRC indicates original, need new idr, decoder needs to be refreshed
Paul: what's the maximum-receive SSRC?
Bo: could be announcing e.g. three decoders low-quality... maximum concurrent
Ba: two slides showing where CSRC channges... M1 switched, M2 composed
Bo: Receiver Stream Selection; receiving A participant in M1, not happy, "give me participant D instead"
Jon: why are you rewriting SSRC
Bo: mixer is sending its own SSRC, describes channel
Paul: you're trying to use SSRC in order to demux
Bo: Avoid Unused Streams; many senders, some streams not presented, and are paused
Bo: Expected Outcome: "take into accont" only
Paul: to some extent an alternative to what Jonathan presented
Jon: any time anything changes... slide 13, he was receiving A on M3, switches to M1
Mark: are there assumptions about how screens would map to SDP?
Bo: whatever low-quality would be different payload types, but could be sent in one M-line
Mark: number is negotiated?
Roni: mapping very simple, content change, still 1-1 relation; similar to what I proposed; whether we use RTP Mixer or RTP translator... the mixer can switch, using own SSRC
Espen: keep c-name from original source?
Bo: if you're forwarding carried as CSRCs
Espen: sounds like two suggestions for doing exactly the same thing
Jon: more than one source for same capture (selected source), could receive stream once marked for different captures
(discussion)
1658 Gerard: Use Cases for RTP Multiplexing; meant to be presented by colleague
Gerard: slide 6; 5-6 use-cases of multiplexing; six use-cases (on slides)
Gerard: slide 7; start traditional all-separate, as we go along merge
Gerard: Use case 3; one transport flow, same SSRC
Jon: coming out of end-point? or MCU? several sources in same packet, defitnitely AVT material
Gerard: use case 4; multiple streams one flow; "ordinary"
Gerard: use case 5; just like usecase2 but both audio and video; the common quality of service;
Gerard: I can't explain case 6; no action item, just presenting; colleague had visa issue re Switzerland
Roni: this isn't very good in presence of packet loss
Mary: Roni?
1705 Roni: SDP description of RTP stream... encoding/decoding capability of... RTP stream can have different media capture, can map to same RTP stream... VCn... 0-6 from example in framework, need six RTP streams... mapping is fixed, new RTP ID... group capture ID... 6 m-lines, each describe codec
Jon: not using bundle
Roni: could use bundle on top of that... need for new media capture... RTP-ID 1... coming from left-camera... more interesting case is loudest-speaker
Jon: require a way to move a source...
Roni: encoding... you can know individual encoding... defined for m-line... depends on number of encodings
Paul: demux by payload type? recall Magnus having a rant...
Roni: not demux by payload
Jon: possible in framework model to request more than 1 encoding type of same capture (360p and 720p)
Roni: without describing, you can't do that

Andy: capture definition needs to map to encoding
Roni: must include both... don't know it can send them
Espen: don't think there's a requirement to say every single possibility must be defined
Allyn: two different approaches, one says "how to solve" not constrained by current, as if it were a green field; other way looks within the environment we currently have, craft a solution within this domain
Mary: I look at this as top-down vs. bottom-up
Roni: need for equipment that cannot-fail
Mary: we're not picking one thing or the other
Roni: OK, but I'm saying can't break the infrastructure where this must work
Espen: still perfectly valid to do out-of-band negotiation; framework rather than protocol draft
Roni: we need to start from the top
Paul: in my experience, neither top-down nor bottom-up works if done blindly
Andy: too big? 30+ audio-visual pairs
Espen: as practical, you shouldn't expect more than 6 m-lines concurrently
Roni: we can solve that using media capabilities... you have to show... if we don't say how we start the call... don't support large SDP, and can't send before defining... if you need to send 5-7 streams...
Jon: probably are SDPs out there that will break on too many media types?
Paul: definitely ones that will break on more than one video and one audio
Roni: changing too often
Andy: simulcast problem
Mary: time to start wrapping up... two different proposals for tomorrow; talk about signaling first... second option some breakouts until 1500
Roni: hoping to stay that long... need to start to look at how this holds together staring from SIP initiation
Mary: this puts more between RTP stuff and framework
Allyn: love to see work on issues we haven't solved on mailing-list
Mary: Issues for further discussion; I suggest brainstorm about requirements
Allyn: think we're beyond that
Paul: need to be some decisions
Mary: I don't see people on the same page
Mary: first alternative, issue identification at 1300, we'll do this one; do we want to start issue identification at 1020
Spencer: RTP usage is worth 1 hour 20 minutes
Espen: I'd be surprised if we had new approach by tomorrow
Mary: typing
Roni: first item, no document
Mary: slides have details; we have presentation, draft-hansen
Stephan: I prefer updated draft
Mary: they decided it wouldn't fly as written
Allyn: not a proposal
Mary: quick vote: revisit RTP (0),
ticket 5&7 compose/switched (),
consumer capabilities (),
CLUE & SDP (),
hard-coded 264 (),
spatial relationships (),
data model ()
Rob: how spatial relationships interact with switched/composed ()

Gyu: second use-case ()
Allyn: source-selection ()
Roni: how we do SIP-CLUE startup ()
Stephan: +1
Espen: relationship between capture-scene and capture-set ()
Mary: please rank from most-important to least-important; I'll give you paper, collect them and announce in the morning
Jon: announcing tonight would enable us to prepare slides
(lengthy discussion of how to count votes)
1800 lots of folks leaving

Interim Meeting, Feb 15 afternoon – Allyn Romanow

Roni CLUE framework comments and issues see slides

- Can we use just 2 messages, not 3? Better for offer answer
- CLUE and SDP media description – draft-even-clue-rtp-mapping-00

How many m-lines are needed? Roni argues many, Jonathan and Andy argue one for video, one for audio, as now
Conf event pkg covers conference management

- Individual encoding

Params are in SDP and some are codec specific, need to know which codec
Map SDP to CLUE and not duplicate info
CLUE is not just for H.264
But within an encoding group can there be multiple codecs? Should the framework say something about this?
Roni – the CLUE constraints are too simplistic. It is only for h.264 - this isn't general enough
Is CLUE extensible for another codec?
They had this problem with H.239 in SIP. This is a problem in SIP
Something more sophisticated is needed
They tried to address this with media capabilities
Stephan agrees this is an oversimplified solution and therefore harmful
This is a self-contained issue
CLUE was intentionally simple in this respect
Doesn't matter at all for audio
Matters on video
Number of video codecs is smaller than 5
We can define a multiplier on video complexity. We could find something. Not impossible to nail, without complexity says Stephan.
Do 1 H.264, 2H.263, it's a simplification
System should include audio as well as video

Slide 4 What is message rate?
What is clue message size?

Espen Use Cases of MCU, see slides

- Discussion about what the slides meant
- Mark – distinguish between directional rendering vs lip-sync, not the same.

Location synchronization, not lip-sync

Lip sync is time synchronization
Time synchronization – there is work being done, says Roni.
VC1 and AC1 may need synchronization at some times and not at others
Which streams you lip sync together change during the conference

- A lot of info that we statically know up front. Slide.

Different case if switched or MCU. Mixed??
Syncing things with the same CNAME
Needs clarification in the framework. Switching case requires considering

- MCU policies
  Segment, site, round robin, etc.

What is a policy associated with? Capture set? Or capture set entry
Wants to announce policy
Configure ask for one policy or skip and get a default

- Participant lock- need out of clue messaging for who is in the conference. Participant selection, segment selection

- Issues- how does clue handle 25 streams

Participant locking discussion-capability on part of provider, choice of consumer to decide
Should this be in CLUE?
Interesting use case, no consensus on whether should be in CLUE
If you are doing it, should be on a capture level or a participant level?

Data Model Paul K, UML
Diagram, not sent out
Relationship between things
Valuable to see relationships in a systematic form

Data Model Andy
Not quite in XML, followed CCMP format
Discussion whether we use messages or not
Maybe would be better
Take decision on how to model to the mailing list

Jonathan RTP for CLUE, see slides

- Requirements – many media sources, e.g., continuous presence
- Architectural constraints.. want few UDP flows
- Don't want SIP mid boxes to reject out of hand
- Discussion as to whether media type is right designation
- Doesn't want to require BUNDLE for this

Bundle describes how indicate 2 indpt mlines go over the same transport

- Use cases, requirements

avoid unnecessary i-frames

- Looks at different possible solutions for mapping

Send mapping outside the media
Lip sync – needs to be bound to SSRC, another reason for not doing it outside media

Sending capture ids in media stream – pros and cons

- Strawman proposal- header extension

Re-auth necessary with SRTP

RFC 5285 safely ignored. Details not in scope for CLUE, AVTEXT

- Recommendation must support send capture id always, may negotiate option 2 – send as needed
- The capture ID – how it's created. In advertisement
- SDP signaling unchanged

- Important idea. A switched capture is the SSRC.. source being switched, come and go, need mapping to know how to match.

The benefit of this approach is it adds multi-stream capability without changing how SSRCs are used.

- Roni wants to have the requirements more explicit.

Bo Burman multi-stream Media Conferencing

- Mary – this is not necessarily CLUE work

Informational. Presenting the nuts and bolts

- Use cases, avoid transcoding, proposed methods
- Assumptions- EPs different qualities, as high quality as possible, can always render a lower quality
- Use case- single stream provider - Dual means the receiver has to choose where to put the one stream that is sent
- Different strategies
- Dual channel case
- Multiple channel

- Rewrite SSRCs. Overwrite active speaker

Use SSRC same function as Jonathan's CaptureID

- Stream selection by receiver

- Expected outcome- CLUE take the use cases and proposed solutions into account.
- This is an alternate proposal to Jonathan's
- Has rewriting. Every switch is an I-frame, everytime something changes.
- Are there assumptions on how map to SDP? No
- High, medium, low- would be 3 PTs. But could be in one mline, he thinks.

- Difference if using an RTP mixer and an RTP translator. Mixer can switch streams. Provides its own SSRC. Translator uses original SSRC

Gerard Fernando RTP multiplexing see slides

Roni Mapping RTP mapping, from his doc

- Observation that some of the proposed solutions seem to ignore SDP and others solutions want to craft a solution that uses SDP.

- SDP big- 30 plus audio video pairs, does it have its own issues? Practically, 6 mlines can go thru. Vendors not implementing support for more than 6 mlines.
- Can use media capabilities draft

- Won't allow moving media that's not in SDP vs not move media that has too many mlines.

- How do you really send this info, what is the transport?

- SDP – don't want to send often

- Roni's scheme doesn't allow simulcast

WRAP UP Mary
2 proposals for tomorrow
Signaling
Then – RTP
Vs breakouts, no breakouts
No more on RTP usage
Go over list of issues
Took a poll

Agenda for tomorrow – Mary will send

Interim Meeting, Feb 16 – John Leslie

BlueSheet (clockwise)

l1: Paul Kyzivat

l8: Zaheduzzaman Sanker

l7: Bo Burman

l6: John Leslie

l5: Christian Groves

l4: Roni Even

l3: Jonathan Lennox

: Keith Drage on WebEx

: Gyubong Oh (arrived later Thurs)

l2: Stephan Wenger

r1: Spencer Dawkins

r2: Andy Pepperell

r3: Allyn Romanow

r4: Gerard Fernando

r5: Brian Baldino

r6: Espen Berger

r7: Mark Duckworth

r8: Rob Hansen

: Scott Pennock

0905 Mary: starting WebEx... being recorded... agenda...

0911 Allyn: defer my presentation until Steve arrives

0912 Mary: #3 (relationship, capture-scene/set)

Jon: does a capture scene have one scene, or many?

(Keith joined WebEx)

Mary: we rearranged agenda

Spencer: did we say scene has one capture set

Paul: "area of scene" isn't part of any one capture

Mark: intent capture scene was conceptual, not part-of... capture set would be concrete representation ...

Roni: provider advertisement, does it include many capture sets?

Mark: yes

Roni: how do I know what each represents; give me an example of two capture set

Mark: one video/audio, one presentation

Roni: you should be able to tell from the outside

Allyn: original framework document didn't have scene; notion of scene arose, but now it's not used; anyone who sees the need for a capture scene, explain

Jon: IMHO more evocative name for capture set

Christian: endpoint saying this room would be the scene

Brian: gets a little tricky with scene=set... each entry in the set is an alternate view of the same scene

Roni: section 4... need to address this issue there in document; when I look at capture set... we need to define... in document we need structure drawn

Espen: concept, capture scene, one or more capture sets

Paul: haven't heard why you have more than one set

Andy: our original intent, used "scene" as alternative... 1:1 mappinng

Mary: that text needs to be clarified

Roni: makes sense to me to have one capture set to a scene... when I get N capture sets I should know what they mean without looking inside them

Paul: difference between scene and room... scene is coordinate space embodied in an area of capture, though area is 2D and coordinate space is 3D

Roni: what information do we want to have in description...

Paul: presentation is a different coordinate space, thus room may have multiple scenes

Andy: saw no reason to tie capture sets together (physically close together)

Christian: I'm getting confused... yesterday we talked of presentation at capture-set level

Andy: we never said capture-set tagged as ??

Roni: example, room with presenter and rest-of-room; there's no relation between them; if I have two "main"s, I have no way to tell them apart; we need some info

(Steve arrived 0931)

Espen: if two capture sets available,

WebEx: definitions in framework doc, from receiver's point-of-view, one item from each capture set in the scene

Steve: think you select a capture-set; idea is capture-set is a scene

Roni: text in 6.2... "scene" where you mean "capture scene"

Paul: explore this difference, if presentation is attribute of capture

Andy: transcoding into a single capture, tagged as both presentation & media

Roni: for units, "unknown" scale same for every capture in capture-set

Brian: MCU pulling from lots of sources will use "no scale" with or without presentation

Steve: presentation from document camera, capture is spatially related to rest of room -- different from slide case; might be two scenes in a room which are unrelated to each other

Roni: we raised most of the points

Christian: term to consider: "spatial relationship"; do we say anything in a capture set _is_ spatially related

Mark: don't think we want to say that

Christian: if they're not spatially related, why do I put them together

Andy: ...

Roni: what I think I heard: if we have a scene, one set per scene... not yet sure what information should be captured

Mark: no way to relate coordinates

Paul: presentations coming from room, but no spatial relationship

Mary: think we agreed 1:1 mapping

Allyn: still unclear what "scene" intends to represent

Andy: we liked "set" because of the presentation

0946 Mary: are we going to stick with capture-set, or scene

Steve: I like "scene" ... benefit of distinguishing scene vs. ?

Mary: take to mailing-list

Christian: want to recored difference of capture-sets... spatial relationship... what can I use capture-sets for?

Allyn: don't think this is in any way resolved, leave open the reason for capture scene in the first place

Mary: we do have more related issues

Stephan: unclear to her that capture scene and capture set are the same think

Roni: we're spending too much time on this topic

Mary: we've gone as far as we will today; quick break now (0954) until 1005

1009 Allyn: if we agree on the criteria we will use, that's progress

Allyn: pieces of CLUE where we need to talk about signaling; consumer capabilities (not clear it will be there; advertisement of capabilities (some are encodings); selection process (back to provider); media being sent... each side may play both roles

Spencer: does this roughly match offer/answer

(Gyubong arrived 1014)

Allyn: Parameters;

Allyn: Dynamic Messages; not just setup at beginning

Allyn: Using SDP; helpful to discuss SDP-for-which-parts

Roni: also need to clarify what they represent, e.g. what-I-can-send, what-I-can-receive

Allyn: whether we want SDP for session-level or media-level

Paul: that's what SDP means

Steve: one idea is no relationship of CLUE to SDP; this is looking for middle-ground: what fits in SDP and what doesn't

Allyn: "current SDP usage"

Mary: is this saying whether we want CLUE stuff in SIP messages?

Allyn: do we want to use SDP to set up the cas

Stephan: SDP is a language, not how to use that language; do we want go through effort to describe CLUE stuff in SDP syntax? ?

Allyn: that's the general question, I was trying to ask in terms of what is/isn't CLUE-specific... four specific parts

Roni: we need to agree what type of information we provide: what is content-semantic; also description of codec...

Allyn: is it good idea to use SDP for the four items -- we possibly will come to different conclusions... different parts/aspects...

Mary: I would almost think... is this from people outside CLUE

Allyn: we would use SDP in describing m-lines

Jon: one choice, single SDP m-line "application/CLUE"

Allyn: "SDP-as-typically-used"

Roni: it's not SDP; it's offer/answer; we need to ask whether signaling should be offer/answer

Allyn: not all as one

Mary: charter says we use SIP

Stephan: basic call would consist of audio plus RTP channel... different from... offer/answer is one way of using SDP language

Rob: doesn't require replacing offer/answer, additional info outside SDP

Stephan: that's not the discussion... understanding...

Mark: offer/answer implies one-way-to-use SDP; we do not have any offer/answer that doesn't use SDP

Paul: I'm hearing question which CLUE stuff can be mapped onto existing SDP, handled through offer/answer mapping onto existing SDP parameters, which on new SDP parameters, which can't

Stephan at whiteboard: (1), (2) fits into SDP syntax, not offer/answer

Paul: (2) is empty set

Stephan: (3) new SDP syntax, offer/answer ideas; (4) new SDP & something-else-not-offer/answer; (5) XML; (6) SDP+XML existing offer/answer

Rob: additional... separate-channel

Stephan: (7) two-stage

Keith: exactly same as ??

Stephan: #6 media-control work

Rob: we may want to divide the information

Allyn: should be considered for different aspects of CLUE... not all one... things I think would help us decide... latency requirements, three different issues

Stephan: not with SDP...

Allyn: second thing, what do intermediaries need-to-know about CLUE info, three dimensions...

Jon: conversely, what info do intermediaries not want to know

Allyn: SDP questions; is there a concern about bloat

Paul: if you put it within SDP, then things which parse SDP will have to parse it

Jon: limit to maximum SDP before networks stop working

Allyn: CLUE doesn't fit within offer/answer, different aspects may fit...

Christian: one aspect related to latency/bloat: at what point do we need the info?

Jon: "can it fit" is wrong question... engineering trade-off

Roni: about latency... latency of what

Allyn: do you mean encoding-info needs real-time, nothing to do with SDP

Roni: in advertisement, you say "I can do up to HD"; change... what are the dynamics of change... perhaps change once-per-minute, latency would be crucial

Steve: one example, capture attributes have to be at receiver before first packet arrives, otherwise rendering won't be done... late-joiner problem

Rob: case which concerns me, virtual-room (100 things behind you), you need to know spatial information to do transforms

Allyn: end-point needs to know?

Roni: you can get it at the beginning of call

Steve: eventually...

Allyn: not with dynamic during call

Roni: room doesn't change in middle of call... that's how event package works

Christian: different problems with different latency requirements

Allyn: do we have requirements less than half-a-second?

Espen: for lip-sync...

Allyn: do we agree CLUE needs to work for both mixers and translators?

Roni: no... case where it provides its own timing, question whether it's feasible to implement; it's OK to say this piece works under this architecture

Jon: mixer can be a switch, translator has to be a switch

Allyn: are there architectures that should be left out of CLUE because they don't work?

Mary at whiteboard...

(I left room)

1310 resume after lunch, WebEx restart

Mary: time to move back to issues; how do we start call

Paul at whiteboard: point-to-point... ground-rule... initial invite shouldn't look too disconcerting to unaware

Jon: could have m-lines that unaware could ignore... two parts, compatibility mode...

Roni: here we want to provide multiple m-lines; no way to do that in SDP... could be different behaviors, select first, reject all, maybe support one which will be OK for you. bundle is _a_ way to multiplex, my proposal is grouping for CLUE

Andy: 1:1 mapping captures to encoding...

Roni: talking about SDP; just having labels doesn't help; need to figure out whether other side understands the offer.

Jon: if he thinks you're offering 18 m-lines... likely to reject them all

Roni: you cannot predict the action (unless offering just one audio); send second offer when I know what he understood

Paul: sounds like you're assuming can be done in one offer/answer... initial invite includes sufficient m-line info to include all encodings; m-lines send-only?

Roni: can offer send-receive

Paul on whiteboard: ...

Jon: proposing one audio, one video, indication ability for multiple

Paul: after initial offer/answer, you have CLUE session up? where is callee's advertisement?

Jon: an option: 3 m-lines: audio, video, CLUE-signaling

Paul: are those m-lines suitable for the full... ? maybe non-CLUEful accepts audio, video, and that's it... maybe CLUEful rejects audio, video, accepts CLUE-signaling

Paul: option B...

(long discussion)

1414 Mary: we've done what we can today, time for some drafts

(discussion continued)

1449 (after break) Mary: Consumer Capability continued; theoretically should simplify

Paul: which is the primary reason? give provider mechanism to filter down to more manageable size.

Andy: might never advertise the complete list of capabilities... worry about difficulty adding new attributes (captures that differ only in an attribute receiver doesn't know)

Steve: presumably provider can still send full list if it wants to? or does provider need to adjust to capabilities of receiver? you can't differentiate them anyway, so I pick...

(discussion continued)

1513 Mary: Source Selection;

Jon: I know which particpant I want to see... which camera

Andy: different granularities...

(discussion)

Mary: should this be in use-cases, or round-2? put in appendix of use-cases

1531 Mary: Policies;

(discussion)

1552 Mary: don't need solution in CLUE; glue between CLUE and XCON, e.g., not in charter; short break

(discussion continued)

1633 Mary: Issues Discussed;

Mark: what we discussed was _some_ policies; selecting what sources go into a composed capture (not in scope now); site-switching rather an exception...

Mary: issues we didn't get to...

1651 Mary: I think we are done

Interim Meeting, Feb 15/16– Spencer Dawkins

2 Wednesday

Working group chairs bashed the agenda – no changes.

2.1 Use Cases 02 – comments and proposal

Gyubong Oh

This is a continuation of comments made on the mailing list.

Presentation use case:

• Audio/Video stream, presentation stream OR Video presentation control OR single presentation stream

• View is that BFCP limitations prevent satisfying REQMT-15 for multiple sources, seen by all, and variation in placement, number and size of presentations.

• Proposal is that presentations may be dynamic, and managed by all participants

• Roni – single presenter limitation is not BFCP, it's H.239 policy, but there's no document that tells how to handle multiple tokens. No problem; just need a BCP document that describes how to do this in an interoperable way (don't even need CLUE to do this). H.323 DOES define a single presenter …

• Jonathan – are you asking that any participant could advance the slides? Interesting, but probably more like remote input, and probably not in scope for CLUE.

• We could have more than one floor active at a single time.

• We don't have a standard protocol for collaboration – that's why we're still doing presentations using video.

• Do we need to add anything to the use case document? Use cases are examples, not exhaustive set.

• Do we need to associate capture sets with a floor (if we have multiple floors).

• Floor control is out of scope for CLUE, but we're just talking about how to do multiple floors at once.

Multiple devices use case:

• Previously, per-device media rendering based on resolution/bandwidth

• Did not meet REQMTs

We did not have time to discuss the multiple devices use case during the meeting.

2.2 Framework

Two revisions since IETF 82 (-02 and -03).

2.2.1 Information hierarchy

We discussed the relationship between Providers, Capture Sets (with Attributes, Simultaneous Sets, and Entries which point to Media Captures), and Media Captures.

We're concerned about Simultaneous Sets becoming combinatorial (for large numbers of participants with multiple complex capture sets).

We don't have a use case for these types of environments yet. Do we need them?

We're still trying to figure out whether specific requirements are met by existing protocols – we're focused on the data requirements, not the protocol requirements, for now.

We have been talking about ceilings on resolution/bandwidth, but we need

to talk about floors, as well – otherwise people may end up with media streams that are unacceptable.

2.2.2 Spatial Relationship Coordinates

This topic is from the framework document, and hasn't changed since Taiwan.

Point of capture and four corners of the area of capture, all with X/Y/Z coordinates.

Coordinates can be physical or virtual. We can use the coordinates to reconstruct the angles of cameras within the capture set, to allow geometric correction – and we'll include this in the framework.

Would we assume that the camera angle is perpendicular to the four corners of the area of capture? Christian assumes so …

Paul points out that we're not defining a plane, we're defining a four-sided pyramid. There may be participants who aren't on the base of the pyramid – multiple rows of tables covered by the same camera is a reasonable example, and there are other, less reasonable, examples! What we're trying to describe is the part of the pyramid where people are displayed life-size.

Do we have a use case that needs this? Do we know how receivers would use the information?

Are we talking about a "scene"? Are they synonyms? We think the scene is a concept and the capture set is the description of one representation of the scene.

The document says entries are supposed to be mutually exclusive, but that's not correct. We expect that endpoints will be using multiple audio and video streams in normal operation.

We have physical and virtual units of measure – and we have “unknown scale”. Roni thinks we should have a coordinate system that says there’s no relationship between captures – that’s different from “unknown scale, but they’re all the same”.

2.2.3 Tickets #5 and #7 – Composed and switched captures

These tickets may be talking about the same thing.

• Attributes as list of alternatives, consumer chooses one.

• Provider advertises <video-layouts>, consumer chooses one.

• Provider advertises <switch-policies>, consumer chooses one.

• Provider advertises list of sources that go into composed/switched source (if that’s known).

How much are we going to specify about what middleboxes do?

2.2.4 Consumer Capability Message

Proposal to add this detail to the framework.

Three-way messaging – consumer capability is the first message sent by the consumer, telling provider what attributes and media types it understands. Consumer can be simpler and implement only the subset of features they are interested in – this also allows forward compatibility with new features, and there are scenarios where multiple captures differ only in attributes that the consumer doesn’t understand.

This isn’t just offer-answer (matching SDP announcements) – it really is three-way.

2.2.5 Voice Activity Detection

There’s a proposal on the way, but not here yet.

Ticket #4 – share metadata with W3C (WebRTC).

Not ready for this yet (just not stable enough).

2.2.6 Roni's issue list for -03

Message flow:

• Is there a need for three messages?

• Let producers advertise full information and let consumers choose –

two messages would map better onto SDP offer/answer.

SDP media description

• Do we need to map a media capture to an RTP stream in CLUE?

• Is this part of CLUE transport work?

• Roni thinks the number of simultaneous media captures bounds the

number of m-lines, but Jonathan does not – and if it does not, that

would be a problem!

Individual encoding

• We have seven attributes in CLUE that are also specified in SDP, and

some are codec-specific ("maxH264Mbps") – so support for a new codec

could require new attributes in both CLUE and in SDP.

• Can we map the SDP information to CLUE, so not duplicate information

in both places?

• If we're assuming H.264, no problem. If we're not, we may have a problem.

• Would we limit a encoding group to H.264? Would we limit an encoding

group to one codec (whatever codec is chosen dynamically)?

• Roni is concerned that what we're doing isn't extensible, but if we do

something extensible, that may require MMUSIC to do things in SDP (and

that may not be a small amount of work, that would not be under our

control).

• Roni's point is that H.320/H.323 was done by video people who thought in terms of alternative encodings, while SIP/SDP was done by audio people who didn't see the value …

Other issues:

• Who decides what will be sent – producers or consumers?

• What is the CLUE message rate?

• What is the CLUE message size?

2.3 CLUE MCU Use Cases

We don't have any use cases that include MCUs, and we need that.

MCUs could select media streams on behalf of users, and this could be based on user policy.

We had a fairly confusing conversation about "virtual" – apparently this is what we were talking about earlier with video sources that don't have the same scaling.

Paul pointed out that this is (potentially) recursive – compositions of captures of compositions of captures …

This proposal assumes MCUs optimize by having the same policy for all capture streams and enable lip-sync by matching RTCP SDES CNAME (this matching also syncs the video streams coming out of a single site).

No spatial information is required, except for "next active speaker" matching.

MCU with <switched-policy> - examples would be segment, site, round-robin switching every 10 seconds, TEXT+, etc.

Participant lock – a named participant (from RFC 4575 or XCON), can be requested by conference identifier.

Correlate information:

• From RTCP: CNAME (RFC 3350), SRCNAME to label individual sources.

• Lip-sync using CNAME and SRCNAME

Paul doesn't think the policy name is sufficient to identify streams – that could work if that's the only thing we support, but not clear how it interacts with someone deciding to pin one screen to one input while letting the MCU handle everything else – that trips over "participant lock" pretty quickly.

Participant-lock is actually a capability – ("participant-lockable"?).

Didn't we decide that CLUE stopped decomposing when we got down to a capture? Not to individual participants in a capture, right?

All the spatial magic is gone when you're getting a stream from an MCU, right? So you need the MCU to handle that, if anything is going to?

2.4 Data Model

Paul was drawing UML for the data model, and this is where he ended up … does anyone else understand UML? It's an object-oriented language with no methods – similar to Entity-Relationship diagrams.

This is very early work, and already needs to change based on discussion earlier in the meeting, but it already allowed us to talk about cardinality in the meeting (and if two things are 1:1, are they really 1:1, are they really the same thing, and do we need both of them?).

We also reviewed an overview of CLUE messages and attributes. Again, this is early work, but helpful.

We'll decide on the mailing list how we actually represent these concepts in the documents …

2.5 RTP Usage for CLUE

Jonathan, Allyn and Paull Witty …

CLUE allows SIP to have:

• Dozens of media sources,

• Choosing among hundreds of inputs (for switched or composed captures),

and

• The number of media sources can be asymmetric and dynamic.

MCUs do this today, but not in a standardized way.

We have architectural constraints:

• Don't confuse non-CLUE endpoints with initial offer/answer,

• Don't confuse SIP middleboxes with signaling messages, and

• Transport flow usage (for NATs, firewalls, port consumption on MCUs)

should be restrained – unless you need multiple flows, don't use

multiple flows! Send all sources of the same media type over a single

RTP session, even if they come from multiple captures.

Three categories of use cases:

• Static streams,

• Dynamic switching between a small number of streams, and

• Dynamic switching between a large number of streams and sources.

(Noticing that as you choose a different set of video streams, what was

once the "left" video stream is now the "right" video stream)

Some equipment needs to know source-to-capture mappings before decoding

(dedicated decoder hardware, decoder hardware associated with specific

displays/speakers).

Jonathan is trying to allow stream movement without requiring

unnecessary i-frames or duplicate transmission, especially on the high-bandwidth streams we're expecting with telepresence.

Sending mapping information – where?

• In CLUE signaling is reliable but not time-guaranteed, and might be very heavyweight.

• In RTCP SDES signaling is unreliable and subject to RTCP timing/bandwidth rules.

Not sure Capture-ID is the right index – could be receiving multiple instantiations of RTP streams for the same capture at once (not recommended, but possible in some scenarios).

Roni thinks all audio is switched (with its own timestamps, etc.) but not everyone does it that way.

Sending capture IDs in media streams – no latency, and you know how to interpret what you got, but you add to MTU size and MCU processing requirements.

RTP header extension? Modification could be costly (especially for SRTP – authentication required) and tripping over "RTP header extensions MUST be safe to ignore" – and isn't in scope for CLUE anyway.

We'll start by asking Jonathan to add high-level requirements at the beginning of his document …

2.6 Multi-stream Media Conferencing

These presentations are informational for CLUE, but aren't in scope for CLUE.

The goal here is to allow endpoints with different capabilities to participate in the same conference, with the best experience possible

("highest-quality media possible").

Draft-westerlund-avtcore-max-ssrc allows selection of multiple media streams of different quality.

Mixer SSRCs become "roles", need not change often, and can be tied to a certain decoder resource.

Draft-westerlund-dispatch-stream-selection allows a receiver to "reach through" a mixer, selecting a specific source SSRC to a specific mixer SSRC.

Draft-westerlund-avtext-rtp-stream-pause allows senders to notice that no one is receiving streams, and to pause streams with no receivers.

This isn't CLUE work, but CLUE might keep these proposals in mind when designing CLUE.

2.7 Use Cases for RTP Multiplexing

Gerard Fernando presenting for a Chinese colleague with visa problems L.

Proposal to put multiple streams in a single RTP packet (which means they MUST have the same SSRC!). This is more useful for audio because the per-packet payloads for a single source are short.

Proposal to multiplex all media streams in a single RTP session.

Fernando said this is the way MPEG works, but Roni challenged the MPEG analogy because MPEG-2 is usually used with retransmission – it's not good for the conversational applications we're thinking about.

2.8 RTP Mapping in CLUE

This is Roni presenting his draft (no slides – the draft was submitted after the cutoff, so Roni chose not to prepare slides for this meeting).

Some of these proposals are top-down, others bottom-up – and the

bottom-up proposals are closer to SDP today.

We have some concerns about the number of m-lines in the SDP, and we have some concerns about what happens when CLUE hits unmodified network infrastructure (it's common to drop media streams that haven't appeared in an SDP m-line, for example).

Paul said we'll probably have problems with anything more than one audio m-line and one video m-line. Jonathan said we'll probably have problems with anything that's not G.711 for audio. J

3 Thursday

At end-of-day Wednesday, we decided we would spend Thursday focusing on the issues we haven't closed on, even after discussion on the mailing list. Mary and Paul polled the working group on priorities, and that's what we used for our agenda on Thursday.

3.1 Signaling with CLUE

We did include this topic in the Thursday discussions (from the original agenda).

Allyn/Stephen/Robert are just trying to frame the discussion of the existing document, not make a proposal.

Data flow:

• Not clear whether we would advertise consumer capabilities (see later discussion).

• Advertisement and selection (which roughly map to offer/answer?)

Parameters:

• Capture attributes plus encoding attributes.

Dynamic Messages

• Advertisement and selection can happen at any point throughout the call.

Using SDP for CLUE signaling

• What goes into SDP? Session level, media level, encoding/encoding groups, capture attributes, selection? (This is actually a series of questions, and we should probably start at the end of the list and work backwards. J)

• Latency requirements? If they're too stringent, SDP won't work.

• What do intermediaries need to know about CLUE?

• Are we advertising what we can send? Yes – that matches SDP practice.

• "CLUE signaling doesn't have a relationship to SDP at all" – this is trying to take a middle path.

• SDP isn't a call control protocol, it's a language. Do we want to go through the effort of describing CLUE stuff in SDP syntax? That's the general question we need to answer.

• Two types of information – what's the content of the stream, and what's the codec information.

• Roni - we aren't talking about SDP, we're talking about offer/answer. That's the first decision.

• We have to use SDP for session-level/media-level media parameters – that's in our charter. But we could be using "something else" for any other purpose – SDP is one way, not the only way.

• Paul – what CLUE stuff can be mapped onto EXISTING SDP syntax and offer/answer model, what can be mapped onto NEW SDP syntax and offer/answer model, and what should be mapped onto something else?

• Stephen – we actually have more than two cases – depending on whether

you can use the SDP syntax required in the offer/answer model or not, and whether you include XML (either as a replacement for SDP syntax, or in addition to SDP syntax).

• The two-stage/three-stage model discussion fits into this conversation.

• Keith – MEDIACTRL MRB uses a combination of XML and SDP for its consumer interface – that's not an unknown model.

• Allyn – these are the right questions, but there's not one answer for the framework – the list of questions above may have different answers in each case.

• Allyn – the question about what intermediaries need to know also applies here.

• Jonathan – there may be things that intermediaries don't want to be bothered with (just based on frequency of updates, etc).

Is there a concern about SDP bloat and the need to preserve backwards compatibility?

• Elements that parse SDP will have to parse SDP that describes CLUE.

• The point in call setup where information is needed, also applies here.

Do we agree that CLUE doesn't map into the offer/answer model?

• We think that everyone in the room agrees …

• Jonathan – not "is it possible", but "is it good engineering" (and Paul agreed).

• Roni – what is described in SDP is maximums.

• Stephen – if we're sending capture attributes in SDP, we need to look at when we need the information – this is even true in the point-to-point case, with late joiners.

• Roni – there's information that doesn't change (like, the attributes of your capture set), and information that does. That's part of our latency requirements discussion.

• Real-time requirements? Probably just-in-time requirements …

Do we agree that CLUE needs to work with mixers and with translators?

• Roni's not there yet …

• Allyn thinks this is in the requirements (or was at some point).

• Are there architectures that should be out of scope for CLUE?

Are we conflating latency requirements with frequency of updates?

• There can be infrequent updates with very low latency requirements.

Mary did the table for each of the questions against "frequent updates", "realtime", "intermediates need information", and "size of SDP". Here's how far we got …

Type of information Frequent Updates Realtime Intermediary need info Size

CLUE encodings Medium/low N Y (policy) Large

CLUE encoding groups

CLUE capture attributes

CLUE capture selection

Realistically, we need strawman examples, to complete this, and we need strawman data models, to work on strawman examples!

3.2 Issue Discussion

3.2.1 Framework structure – Captures

Capture sets are a well-formed concept, and map to an "area of a scene".

"Capture scene", that's 1:1?

What does the provider advertisement actually provide? Many capture sets (potentially). What do we call what's in the provider advertisement?

We didn't have capture scenes in the original concept. Now it's in the framework document (in Section 4), but it's not used anywhere else. What is it really?

"This room is the scene, with more than one capture set in the scene"?

"Alternate views of the same scene"?

Need to explore whether we have more than one capture set for a scene – not clear what that means, but a 1:1 relationship between capture set and capture scene where the document was headed – it's just not what the document says.

Roni thinks it's important to understand what a capture set means without rooting around inside it, and Paul agrees.

Is there a difference between a scene and a room? Paul thinks the scene is the 3-D coordinate space. If you have two capture sets with the same coordinates, what does that mean?

Perhaps a room has multiple scenes – the audio/video, and the presentation, for example.

Do capture sets have the presentation attribute, or do individual captures? Need to figure that out …

Presenter view plus rest of the room? That would be different coordinate sets – no relation between the coordinates at all.

Document authors didn't think we could come up with a finite list of capture set types.

Could have free text descriptions – but machines can't base decisions on

free text descriptions.

Keith – is the scene one capture from each capture set? Steven thinks we select capture sets, not captures.

Roni – probably need to write “capture scene” where we have “scene”, if we keep this concept.

Paul – presentation would be its own scene, but if that’s an attribute of a capture, there could presentation and non-presentation captures in a capture set.

Roni – “scale is the same for every capture in the capture set” – but that isn’t true for presentations.

Stephen – document camera is still used, and that’s different from presentation slides.

Christian – need to consider spatial relationships – anything in the same capture set is spatially related; two things that aren’t in the same capture set, don’t have to be. Is that what the captures in a capture set have in common?

Andy – multiple capture scenes in provider advertisement, each with its own scaling?

Roni – if we have a scene, there’s one capture set per scene (need to fix document). We aren’t sure what information to include at the scene level yet. If we decide the scene has information about itself, spatial scaling might be part of that information.

Allyn – still not clear what a scene really means.

Jonathan thinks “capture set” is too abstract – no one can guess what that means. Paul thinks it’s a set of captures J

Andy – scene was people sitting at a table, not what comes out of a VGA connection.

Concluding:

• We think that capture sets are capture scenes – we're just not sure which name we prefe, but we need to pick one and update the documents.

• If we allow multiple capture sets, we need to understand the difference between them (why send two, and not one?).

• We're distinguishing between two captures (three video streams vs. one stream that represents all three streams) and two captures (three video streams associated with one stream of presentation).

Allyn - we're still trying to figure out what a scene is (what it was in the first place)! That's why we don't know whether a capture set is a capture scene.

3.2.2 CLUE Call Flows

We're trying to avoid CLUE confusing endpoints that aren't CLUE-ful.

We have to offer/answer in each direction (using our model).

I have some architecture questions here:

• Would we require bundling? "No intermediaries" is like "land of unicorns". We would always bundle, and that means we'd already have ports open at intermediaries. Roni's proposal actually uses the bundling mechanism, while Jonathan's multiplexes within a single channel.

• Is anything good going to happen when we send a really rich INVITE to someone who's not CLUE-ful? We're talking about sending one audio/one video blindly – that would work in most cases, and if it doesn't, nothing more complex would work, either.

Previous practice on dual video was to send one video line plus support for BFCP, and then adding a second video stream if you support BFCP.

There's a low upper bound on what we can do to accommodate CLUE-less intermediaries … working through unmodified intermediaries is really limiting.

All the proposals we discussed would require more than one offer/answer exchange, and all the proposals we discussed would support (but not require) consumer capabilities.

Keith – we don't have to choose the same transport for every kind of information – MEDIACTRL MRB transports some information via XML/HTTP, and other information via a MEDIACTRL control channel.

Keith – we need a mixer interface control channel (for MRF).

The pushback CLUE got at Taipei from the larger RAI community was that everything that COULD be in SDP, SHOULD be in SDP. Allyn's got a proposal that is SDP-only, but people reacted in horror to it (maybe we need to distribute it more widely!)

There is a new SDP directorate – we should show that proposal to them …

We made progress on discussing INVITE J but there's still lots of work to do.

3.2.3 Consumer capability message

I am concerned that this may end up as a way to generate a multitude of subset consumer profiles, and significantly INCREASE provider complexity.

Andy pointed out (correctly in my view) that there are two questions – does it help for producers to know what consumers can do early in session setup, and is this the right way to tell producers what

consumers can do?

Not that the IAB loves protocol profiles, but having arbitrary lists of protocol components that are supported seems like a particularly bad way to achieve this goal. I pointed this out, and there was general agreement in the room with my point.

We need more investigation here.

3.2.4 Source selection

What are we really selecting? Capture level, or person-level, or …

Is this supposed to work the same as the site-level switching today, or better, or … ?

We need to figure out how XCON roster lists and CLUE interact, but that's not in our charter now.

We'll put source selection in an appendix of the use case document, so we don't lose track of it.

Some CLUE stuff will be fixed in XCON ("these are the six people in these three captures" in CCMP, for example).

3.2.5 MCU with <switch-policy>

We can think of other policies, but "loudest speaker" covers a lot of ground.

How do we do switching in a conference with deaf people? Jonathan says they do formal floor control (there's not much else to suggest).

The axes are how, why, and when …

We may want to do layouts on some capture sets, but not others …

This isn't all orthogonal. You could round-robin through segments or sites.

Andy doesn't think switching policies and layouts are CLUE-specific –

this should be worked on, but outside CLUE. It's needed for general videoconferencing, and is useful, but it's not in our charter.

We can handle site vs. segment switching as an attribute for now.

3.2.6 Is "Composed" a Boolean, or a data structure?

Based on previous issue – Boolean is as far as CLUE should go. Beyond that, is manipulating based on policy …

Would someone send two composed captures, and need to explain what the difference is? Is the reason you would care because it affects the layout, or are there other reasons?

XCON punted on policy, too ("it will always be done by the implementer in a proprietary way").

We need to think about what goes into a composed capture.