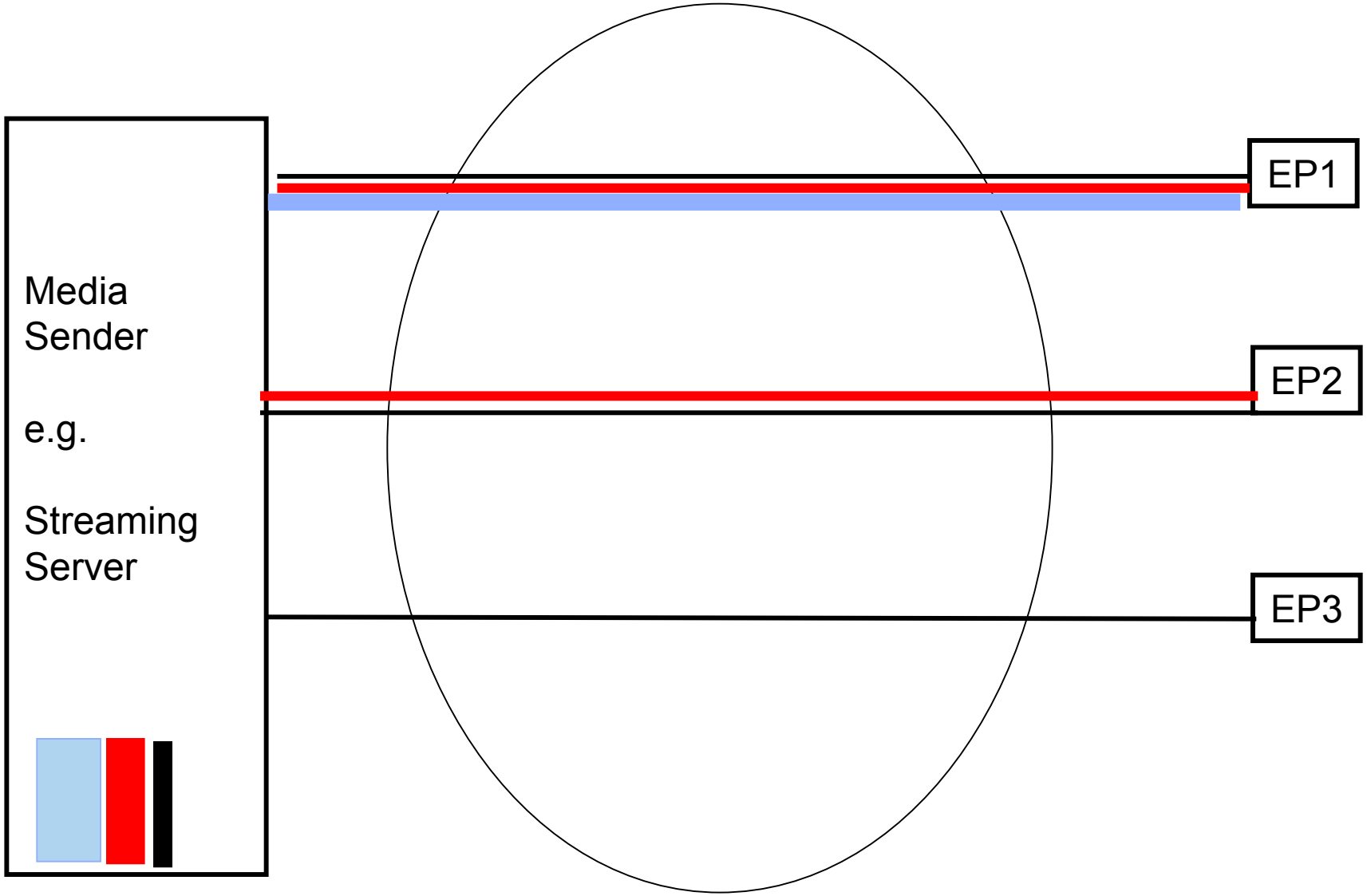
Signaling layered coding structures and the SVC payload format

Stephan Wenger

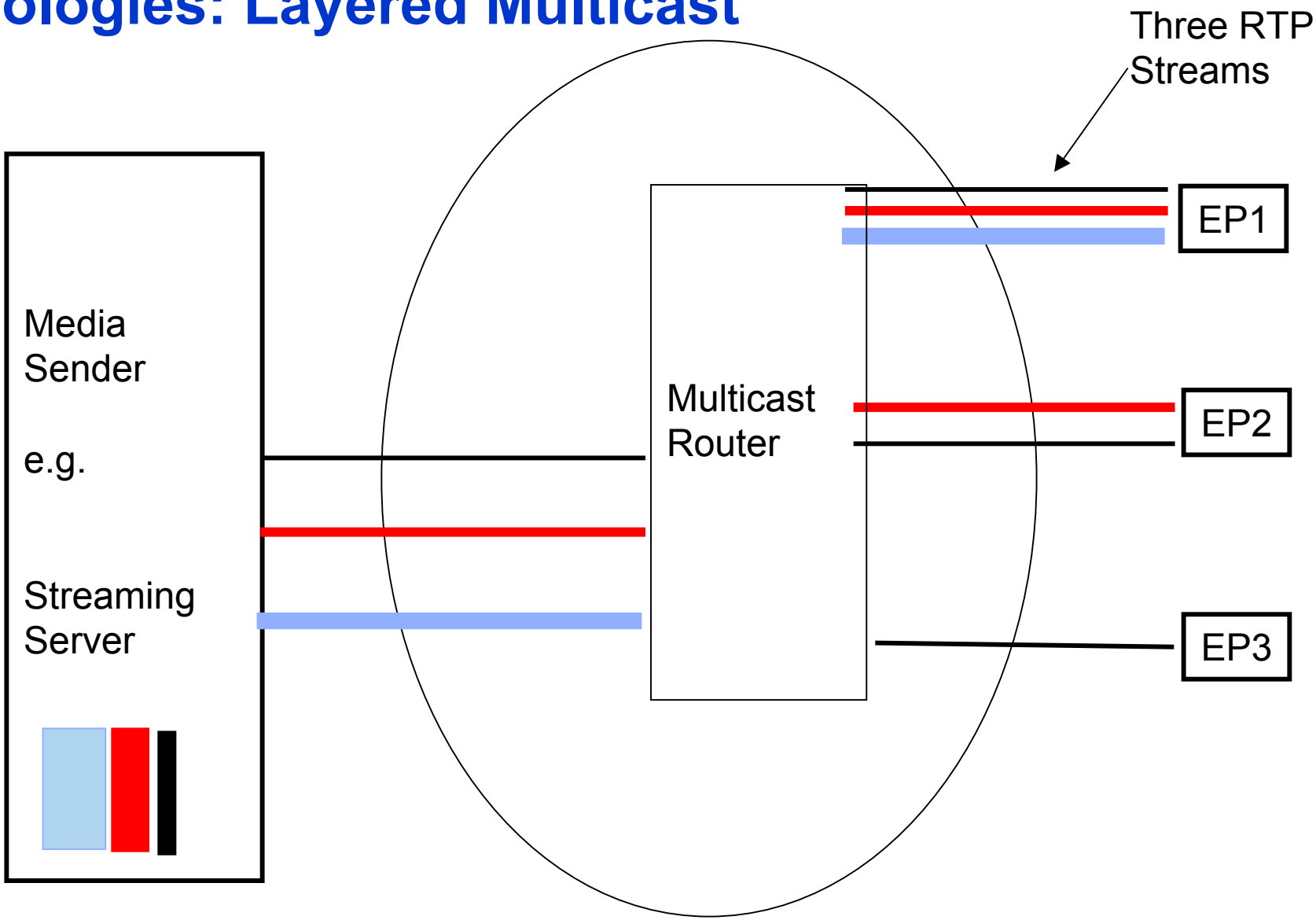
Topologies

- Simulcast, point-to-point
  - Seems to be boring but…
  - layered coding makes sense even here, to avoid real-time transcoding in server and/or an excessive number of available bit streams (file size)
  - May require offer/answer signaling support for a layered bit stream
- (receiver driven) layered multicast
  - Great in theory, but has problems in practice (NAT/Firewall, port # limitations, packet stream overhead, …)
- Layered multicast in core network, “layer combining” in MANEs close to network edge, point-to-point from there
  - Avoids most of the problems of RDLM, but adds the problem of MANEs
    - Inside or outside the security context?
    - Intercept/regenerate RTCP? signaling?
    - Offer/Answer between MANE and endpoint, so to provide MANE with endpoint capabilities so that the MANE can accumulate layers according to capabilities?
  - What are MANEs? Media Aware Network Elements. Are they signaling aware?

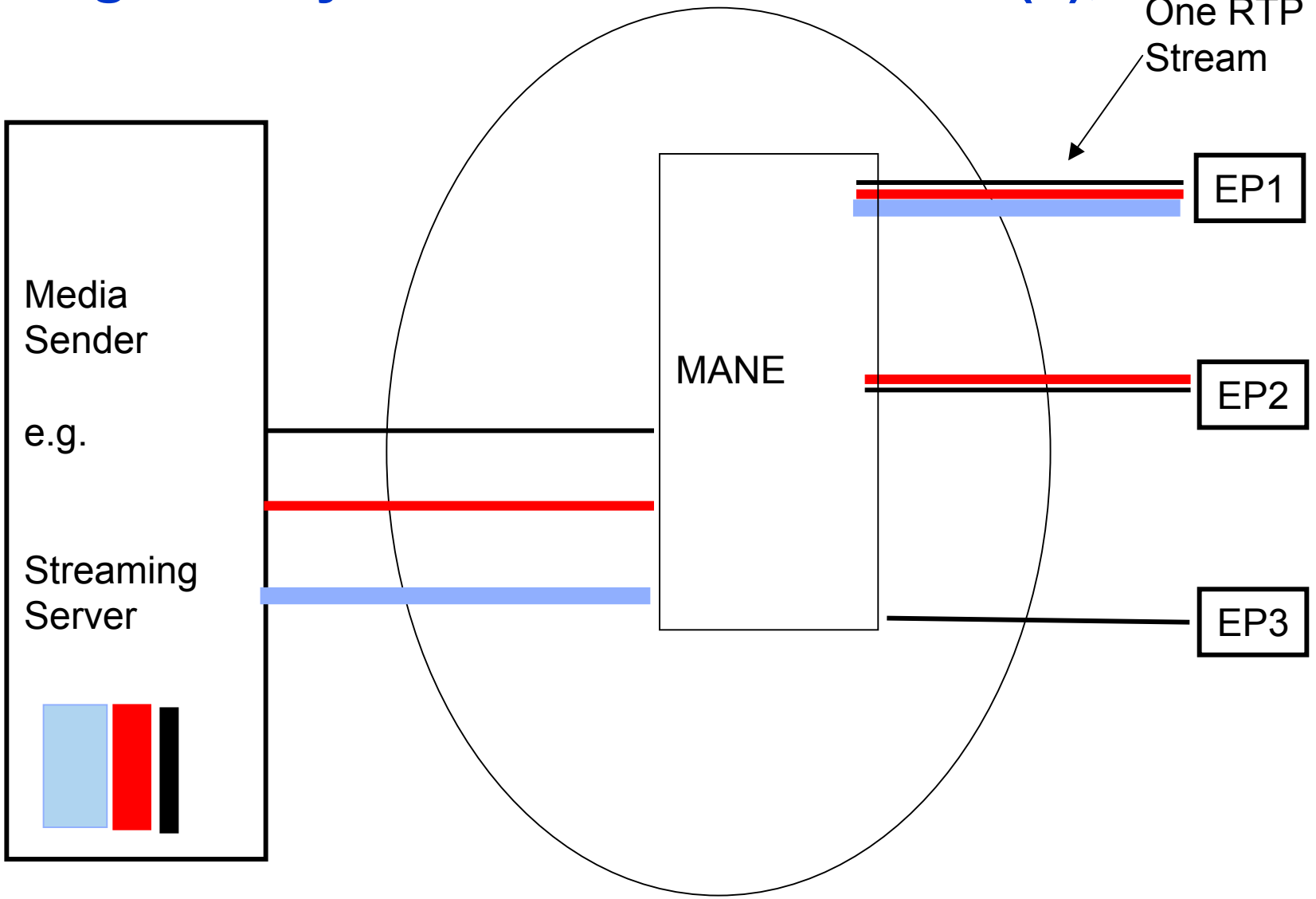
Topologies: Simulcast



Topologies: Layered Multicast



Topologies: Layered Multicast to MANE(s),



Signaling for layered codec streams in AVT – not much

- RFC3550's "view" on layered coding
  - "Progressive layers of a […] signal across multiple RTP sessions each carried on its own multicast group" (section 2.4)
  - "[…] a single SSRC identifier space SHOULD be used across the sessions of all layers and the core (base) layer SHOULD be used for SSRC identifier allocation […]
- RFC3551 does not mention layers except in the context of congestion control
- RFC2429: layers may be sent in their own session, layering structure MUST stay during the lifetime of a session, in-band signaling of layer dependencies (H.263 Annex O's ELNUM, RLNUM, ScalabilityType), no SDP section
- Draft-ietf-avt-rfc2429-bis-06.txt: no additional scalability support, Annex O support cannot be signaled (in-line with deployment situation)
- Draft-ietf-avt-rtp-jpeg2000-08.txt mentions JP2k's scalability, but relies on in-band support only. No multiple RTP sessions
- draft-ietf-avt-rtp-vc1-01.txt does not mention scalability, although technically possible
- RFC2250 and RFC2343 do not support scalability (and do not contain signaling)
- RFC3016 LATM: all RTP packet SHOULD contain data of a single layer only, association in-band (StreamMuxConfig, in its own RTP packet), no discussion of scalable video, no signaling support
- RFC3555, MIME types, no mentioning of layered coding support
- RFc3640 does not mention scalability nor signaling support for it
- RFC3984 does not mention scalability nor signaling support for it (save NRI)

Signaling for layered codec streams in MMUSIC – not much

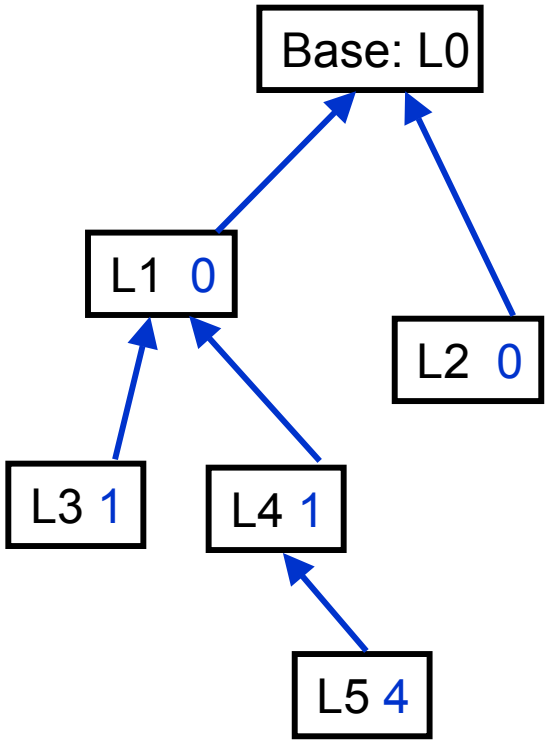
- Draft-ietf-mmusic-sdp-new-25.txt mentions scalable codecs in section 5.7 in conjunction with the connection field. The “c=“ field can either be at session or at media level.
  - Session level: “For applications requiring multiple multicast groups, we allow the following notation […]: **<base multicast address>[/<ttl>]/<number of addresses>**
  - **c=IN IP4 224.2.1.1/127/3**
  - Semantically identical to three media-level c= statements as follows
    - **c=IN IP4 224.2.1.1/127 c=IN IP4 224.2.1.2/127 c=IN IP4 224.2.1.3/127**
- This session layer shortcut notation is nice to have, but does not solve the problem of signaling layer dependencies

Options to signal layer dependencies

- Layer dependency signaled in a media-level attribute
  - Each layer is signaled as its own media in SDP
  - Attribute could reference the basing layer e.g. by transport address, unique identifier, …
  - Also needed is information such as layer types, profile/level of this layer, … all that is necessary for a potential receiver to decide whether it wants to receive this layer
- Layer dependency signaled in a media- or session level attribute, using a bitstring in the native syntax of the payload
  - This is what we currently have in draft-wenger-avt-svc-00.txt
  - Advantage: description can be much more flexible than a simple reference, and would be under control of the media standardization committee.
  - Logical place: media description of the base layer
    - But then we couldn't use plain H.264 any more
- Layers being sent in a single RTP session, all info is hidden in the stream, MANEs need to intercept and interpret in-band config information, keep this context, and forward packets accordingly. This is the RFC3016 solution. Bad ☹
  - Also: no offer-answer possible for point-to-point cases.

Layer dependency on media level w/ layer identifier

- Assumption: No Multiple Description Coding (MDC) Support
  - There is a *single* prediction dependency for each layer (non-degenerated tree)
- Under this assumption, it would be sufficient to add, for each scalable layer, an identifier of the layer it directly depends on, into the media description
  - Identifier could be an integer, a string, or a transport address (any preferences?)
  - Signal on the SDP media level only
  - Receiver has to puzzle out the various options possible, by creating the tree



Layer dependency signaling through bitstring

- Media committee defines a bit string that contains, in a binary compressed format, all possible layer dependencies, and all types of scalability information of each layer
- This bitstring is included, in BASE64 format, in one or all of these
  - Session-level SDP
  - Media level SDP of the base layer; the enhancement layers are described such that they can be identified as enhancement layer, which serves as an indication that they are useless unless being subscribed to something else (unspecified what it is)
  - Media level SDP of all layers (redundant, but perhaps makes the parsing easier?)
- Media receiver can interpret this bit string, if the payload is meaningful for them. If a receiver cannot handle the payload in question, it does not subscribe to this media (or does not accept such an offer in offer/answer scenario).
- Advantage: media committee retains full control over possible layering structures; also, likely, a very compact representation which reduces traffic during session announcement or capability negotiation
- Disadvantage: some odd binary stuff in SDP

Summary: Pros and Cons

- Layer identifier
  - + in-line with SDP principles, simple, straightforward
  - - may not be able to express full functionality; alignment problem with JVT standardization
- Bitstring concept (on whatever SDP-level)
  - + allows for all the flexibility of JVT; JVT would take care of defining sense-making operation points with all their expertise
  - - EXTREMELY payload specific
  - - A binary syntax (but we did that before, see RFC3984 parameter sets attribute)

SVC payload format: other properties

- SVC is an extension (Annex) of H.264/AVC
  - Fixes a number of problems of previous layered codec designs (very limited cycle explosion for multiple layers, reasonable coding efficiency of enhancement layers, error resilience)
  - Parameter set concept – no need for header repetition and such
  - NAL unit concept – transport awareness built into the bit stream
    - NAL unit header is in the process to change
  - It's currently unclear whether NAL unit header changes have direct implications on STAP, MPAP, FU design
    - Tradeoff: when keeping STAP, MTAP, FU design as is, and when aggregating more than one layer into a MANE-incoming bit stream, a MANE needs to parse at least parts of said bit stream to make intelligent decisions. Which adds functionality to MANEs previously unnecessary. However, please remember: MANE == *Media Aware* network element
    - This will be sorted out over time as SVC standardization progresses – no rush here
- Author's goal: keep draft aligned as much as possible with RFC3984
  - Ideally, force every independently transported SVC base layer to use RFC3984, and make it decodable on "legacy" devices
    - Unclear whether this is achievable
  - Limit standardization and implementation efforts

In other news

- Need to reconsider attributes of RFC 3984. Example
  - Bitrate in media SDP: refers to this layer only, but we may want, in addition, the complete bit rate of all layers necessary for this quality (could be puzzled out of the individual media SDPs, though)
  - Profile and Level ID;
  - Others? Need to go through them one-by-one
- SVC also considers a Fine Granularity Scalability mechanism
  - When in need to reduce bit rate by a few per cent, chop off the last n bytes of enhancement layer packets
    - N is signaled in the bit stream on a per packet basis
    - For large bit rate changes, coarse grain scalability (layer dumping) is more appropriate
  - Tricky when security etc. is in the game
  - We need to study whether
    - a) this is useful in an RTP environment?
    - b) do we need support in the RTP payload format, or is this a sender/MANE implementation issue?
- AVPF extension draft (TMMBR / TMMA) can be used to inform sender or MANE to dump layers (or engage FGS) in case of congestion (or to add layers/FGS bits)

Way forward

- Draft is presented here *very early*, to allow coordinated standardization
  - Same idea as H.264 and RFC 3984, which (we think) was successful
  - Suggest to give it WG item status early as well, credibility in other SDOs

- WG item?