Network Working Group   Stephan Wenger
INTERNET-DRAFT   Umesh Chandra
Expires: November 2006   Nokia
                                                            Magnus Westerlund
                                                                    Bo Burman
                                                                     Ericsson
                                                                 May 24, 2006

Codec Control Messages in the Audio-Visual Profile with Feedback (AVPF)
draft-wenger-avt-avpf-ccm-04.txt>

Status of this Memo

Copyright Notice

Abstract

This document specifies a few extensions to the messages defined in the Audio-Visual Profile with Feedback (AVPF). They are helpful primarily in conversational multimedia scenarios where centralized multipoint functionalities are in use. However some are also usable in smaller multicast environments and point-to-point calls. The

extensions discussed are Full Intra Request, Temporary Maximum Media Bit-rate and Temporal Spatial Trade-off.

TABLE OF CONTENTS

1. Introduction

When the Audio-Visual Profile with Feedback (AVPF) [AVPF] was developed, the main emphasis lied in the efficient support of point-to-point and small multipoint scenarios without centralized multipoint control. However, in practice, many small multipoint conferences operate utilizing devices known as Multipoint Control Units (MCUs). Long standing experience of the conversational video conferencing industry suggests that there is a need for a few additional feedback messages, to efficiently support MCU-based multipoint conferencing. Some of the messages have applications beyond centralized multipoint, and this is indicated in the description of the message. This is especially true for the message intended to carry ITU-T Rec. H.271 [H.271] bitstrings for video back channel messages.

In RTP [RFC3550] terminology, MCUs comprise mixers and translators. Most MCUs also include signalling support. During the development of this memo, it was noticed that there is considerable confusion in the community related to the use of terms such as "mixer", "translator", and "MCU". In response to these concerns, a number of topologies have been identified that are of practical relevance to the industry, but were not envisioned (or at least not documented in sufficient detail) in RTP. These topologies are documented in [Topologies], and this memo frequently refers to sections in that document.

Some of the messages defined here are forward only, in that they do not require an explicit acknowledgement. Other messages require acknowledgement, leading to a two way communication model that could suggest to some to be useful for control purposes. It is not the intention of this memo to open up RTCP to a generalized control protocol. All mentioned messages have relatively strict real-time constraints - in the sense that their value diminishes with increased delay. This makes the use of more traditional control protocol means, such as SIP re-invites, undesirable. Furthermore, all messages are of a very simple format that can be easily processed by an RTP/RTCP sender/receiver. Finally, all messages infer only to the RTP stream they are related to, and not to any other property of a communication system.

The Full Intra Request (FIR) Command requires the receiver of the message (and sender of the stream) to immediately insert a decoder refresh point. In video coding, one commonly used form of a decoder refresh point is an IDR or Intra picture. Other codecs may have

other forms of decoder refresh points. In order to fulfil congestion control constraints, sending a decoder refresh point may imply a significant drop in frame rate, as they are commonly much larger than regular predicted content. The use of this message is restricted to cases where no other means of decoder refresh can be employed, e.g. during the join-phase of a new participant in a multipoint conference. It is explicitly disallowed to use the FIR command for error resilience purposes, and instead it is referred to AVPF's PLI message, which reports lost pictures and has been included in AVPF for precisely that purpose. The message does not require an acknowledgement, as the presence of a decoder refresh point can be easily derived from the media bit stream. Today, the FIR message appears to be useful primarily with video streams, but in the future it may become helpful also in conjunction with other media codecs that support prediction across RTP packets.

The Temporary Maximum Media Bandwidth Request (TMMBR) Message allows to signal, from media receiver to media sender, the current maximum supported media bit-rate for a given media stream. Once a bandwidth limitation is established by the media sender, that sender notifies the initiator of the request, and all other session participants, by sending a TMMBN notification message. One usage scenarios can be seen as limiting media senders in multiparty conferencing to the slowest receiver's maximum media bandwidth reception/handling capability. Such a use is helpful, for example, because the receiver's situation may have changed due to computational load, or because the receiver has just joined the conference and it is helpful to inform media sender(s) about its constraints, without waiting for congestion induced bandwidth reduction. Another application involves graceful bandwidth adaptation in scenarios where the upper limit connection bandwidth to a receiver changes, but is known in the interval between these dynamic changes. The TMMBR message is useful for all media types that are not inherently of constant bit rate.

The Video back channel message (VBCM) allows conveying bit streams conforming to ITU-T Rec. H.271 [H.271], from a video receiver to video sender. This ITU-T Recommendation defines codepoints for a number of video-specific feedback messages. Examples include messages to signal

- the corruption of reference pictures or parts thereof,
- the corruption of decoder state information, e.g. parameter sets,
- the suggestion of using a reference picture other than the one typically used, e.g. to support the NEWPRED algorithm [NEWPRED].

The ITU-T plans to add codepoints to H.271 every time a need arises, e.g. with the introduction of new video codecs or new tools into existing video codecs.

There exists some overlap between H.271 messages and "native" messages specified in this memo and in AVPF. Examples include the

PLI message of [AVPF] and the FIR message specified herein. As a general rule, the "native" messages should be prefered over the sending of VBCM messages when all senders and receivers implement this memo. However, if gateways are in the picture, it may be more advisable to utilize VBCM. Similarly, for feedback message types that exist in H.271 but do not exist in this memo or AVPF, there is no other choice but using VBCM.
Video feedback channel messages according to H.271 do not require acknowledgements on a protocol level, because the appropriate reaction of the video encoder and sender can be derived from the forward video bit stream.

Finally, the Temporal-Spatial Trade-off Request (TSTR) Message enables a video receiver to signal to the video sender its preference for spatial quality or high temporal resolution (frame rate). The receiver of the video stream generates this signal typically based on input from its user interface, so to react to explicit requests of the user. However, some implicit use forms are also known. For example, the trade-offs commonly used for live video and document camera content are different. Obviously, this indication is relevant only with respect to video transmission. The message is acknowledged by an announcement message indicating the newly chosen tradeoff, so to allow immediate user feedback.

2. Definitions

2.1. Glossary

```
ASM    - Asynchronous Multicast
AVPF   - The Extended RTP Profile for RTCP-based Feedback
FEC    - Forward Error Correction
FIR    - Full Intra Request
MCU    - Multipoint Control Unit
MPEG   - Moving Picture Experts Group
PtM    - Point to Multipoint
PtP    - Point to Point
TMMBN  - Temporary Maximum Media Bit-rate Notification
TMMBR  - Temporary Maximum Media Bit-rate Request
PLI    - Picture Loss Indication
TSTA   - Temporal Spatial Trade-off Announcement
TSTR   - Temporal Spatial Trade-off Request
VBCM   - Video Back Channel Message indication.
```

2.2. 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].

Message:
Codepoint defined by this specification, of one of the following types:

Request:
Message that requires Acknowledgement

Acknowledgment:
Message that answers a Request

Command:
Message that forces the receiver to an action

Indication:
Message that reports a situation

Notification:
See Indication.

Note that, with the exception of "Notification", this terminology is in alignment with ITU-T Rec. H.245.

Decoder Refresh Point:
A bit string, packetised in one or more RTP packets, which completely resets the decoder to a known state. Typical examples of Decoder Refresh Points are H.261 Intra pictures and H.264 IDR pictures. However, there are also much more complex decoder refresh points.

Typical examples for "hard" decoder refresh points are Intra pictures in H.261, H.263, MPEG 1, MPEG 2, and MPEG-4 part 2, and IDR pictures in H.264. "Gradual" decoder refresh points may also be used; see for example [AVC]. While both "hard" and "gradual" decoder refresh points are acceptable in the scope of this specification, in most cases the user experience will benefit from using a "hard" decoder refresh point.

A decoder refresh point also contains all header information above the picture layer (or equivalent, depending on the video compression standard) that is conveyed in-band. In H.264, for example, a decoder refresh point contains parameter set NAL units that generate parameter sets

necessary for the decoding of the following slice/data partition NAL units (and that are not conveyed out of band). To the best of the author's knowledge, the term "Decoder Refresh Point" has been formally defined only in H.264; hence we are referring here to this video compression standard.

Decoding:
The operation of reconstructing the media stream.

Rendering:
The operation of presenting (parts of) the reconstructed media stream to the user.

Stream thinning:
The operation of removing some of the packets from a media stream. Stream thinning, preferably, is performed media aware, implying that media packets are removed in the order of their relevance to the reproductive quality. However even when employing media-aware stream thinning, most media streams quickly lose quality when subject to increasing levels of thinning. Media-unaware stream thinning leads to even worse quality degradation.

2.3. Topologies

Please refer to [Topologies] for an in depth discussion.

3. Motivation (Informative)

This section discusses the motivation and usage of the different video and media control messages. The video control messages have been under discussion for a long time, and a requirement draft was drawn up [Basso]. This draft has expired; however we do quote relevant sections of it to provide motivation and requirements.

3.1. Use Cases

There are a number of possible usages for the proposed feedback messages. Let's begin with looking through the use cases Basso et al. [Basso] proposed. Some of the use cases have been reformulated and commented:

1. An RTP video mixer composes multiple encoded video sources into a single encoded video stream. Each time a video source is added,

the RTP mixer needs to request a decoder refresh point from the video source, so as to start an uncorrupted prediction chain on the spatial area of the mixed picture occupied by the data from the new video source.

2. An RTP video mixer that receives multiple encoded RTP video streams from conference participants, and dynamically selects one of the streams to be included in its output RTP stream. At the time of a bit stream change (determined through means such as voice activation or the user interface), the mixer requests a decoder refresh point from the remote source, in order to avoid using unrelated content as reference data for inter picture prediction. After requesting the decoder refresh point, the video mixer stops the delivery of the current RTP stream and monitors the RTP stream from the new source until it detects data belonging to the decoder refresh point. At that time, the RTP mixer starts forwarding the newly selected stream to the receiver(s).

3. An application needs to signal to the remote encoder a request of change of the desired trade-off in temporal/spatial resolution. For example, one user may prefer a higher frame rate and a lower spatial quality, and another use may prefer the opposite. This choice is also highly content dependent. Many current video conferencing systems offer in the user interface a mechanism to make this selection, usually in the form of a slider. The mechanism is helpful in point-to-point, centralized multipoint and non-centralized multipoint uses.

4. Use case 4 of the Basso draft applies only to AVPF's PLI and is not reproduced here.

5. Use case 5 of the Basso draft relates to a mechanism known as "freeze picture request". Sending freeze picture requests over a non-reliable forward RTCP channel has been identified as problematic. Therefore, no freeze picture request has been included in this memo, and the use case discussion is not reproduced here.

6. A video mixer dynamically selects one of the received video streams to be sent out to participants and tries to provide the highest bit rate possible to all participants, while minimizing stream transrating. One way of achieving this is to setup sessions with endpoints using the maximum bit rate accepted by that endpoint, and by the call admission method used by the mixer. By means of commands that allow reducing the maximum media bitrate beyond what has been negotiated during session setup, the mixer can then reduce the maximum bit rate sent by endpoints to the lowest common denominator of all received streams. As the lowest

common denominator changes due to endpoints joining, leaving, or network congestion, the mixer can adjust the limits to which endpoints can send their streams to match the new limit. The mixer then would request a new maximum bit rate, which is equal or less than the maximum bit-rate negotiated at session setup, for a specific media stream, and the remote endpoint can respond with the actual bit-rate that it can support.

The picture Basso, et al draws up covers most applications we foresee. However we would like to extend the list with two additional use cases:

7. The used congestion control algorithms (AMID and TFRC) probe for more bandwidth as long as there is something to send. With congestion control using packet-loss as the indication for congestion, this probing does generally result in reduced media quality (often to a point where the distortion is large enough to make the media unusable), due to packet loss and increased delay. In a number of deployment scenarios, especially cellular ones, the bottleneck link is often the last hop link. That cellular link also commonly has some type of QoS negotiation enabling the cellular device to learn the maximal bit-rate available over this last hop. Thus indicating the maximum available bit-rate to the transmitting part can be beneficial to prevent it from even trying to exceed the known hard limit that exists. For cellular or other mobile devices the available known bit-rate can also quickly change due to handover to another transmission technology, QoS renegotiation due to congestion, etc. To enable minimal disruption of service a possibility for quick convergence, especially in cases of reduced bandwidth, a media path signalling method is desired.

8. The use of reference picture selection as an error resilience tool has been introduced in 1997 as NEWPRED [NEWPRED], and is now widely deployed. It operates the receiver sending a feedback message to the sender, indicating a reference picture that should be used for future prediction. AVPF contains a mechanism for conveying such a message, but did not specify for which codec and according to which syntax the message conforms to. Recently, the ITU-T finalized Rec. H.271 which (among other message types) also includes a feedback message. It is expected that this feedback message will enjoy wide support and fairly quickly. Therefore, a mechanisms to convey feedback messages according to H.271 appears to be desirable.

3.2. Using the Media Path

There are multiple reasons why we propose to use the media path for the codec control messages. First, systems employing MCUs are often separating the control and media processing parts. As these messages are intended or generated by the media part rather than the signalling part of the MCU, having them on the media path avoids interfaces and unnecessary control traffic between signalling and processing. If the MCU is physically decomposite, the use of the media path avoids the need for media control protocol extensions (e.g. in MEGACO [RFC3525]).

Secondly, the signalling path quite commonly contains several signalling entities, e.g. SIP-proxies and application servers. Avoiding signalling entities avoids delay for several reasons. Proxies have less stringent delay requirements than media processing and due to their complex and more generic nature may result in significant processing delay. The topological locations of the signalling entities are also commonly not optimized for minimal delay, rather other architectural goals. Thus the signalling path can be significantly longer in both geographical and delay sense.

3.3. Using AVPF

The AVPF feedback message framework provides a simple way of implementing the new messages. Furthermore, AVPF implements rules controlling the timing of feedback messages so to avoid congestion through network flooding. We re-use these rules by referencing to AVPF.

The signalling setup for AVPF allows each individual type of function to be configured or negotiated on a RTP session basis.

3.3.1. Reliability

The use of RTCP messages implies that each message transfer is unreliable, unless the lower layer transport provides reliability. The different messages proposed in this specification have different requirements in terms of reliability. However, in all cases, the reaction to an (occasional) loss of a feedback message is specified.

3.4. Multicast

The media related requests might be used with multicast. The RTCP timing rules specified in [RFC3550] and [AVPF] ensure that the messages do not cause overload of the RTCP connection. The use of multicast may result in the reception of messages with inconsistent

semantics. The reaction to inconsistencies depends on the message type, and is discussed for each message type separately.

3.5. Feedback Messages

This section describes the semantics of the different feedback messages and how they apply to the different use cases.

3.5.1. Full Intra Request Command

A Full Intra Request (FIR) command, when received by the designated media sender, requires that the media sender sends a "decoder refresh point" (see 2.2) at the earliest opportunity. The evaluation of such opportunity includes the current encoder coding strategy and the current available network resources.

FIR is also known as an "instantaneous decoder refresh request" or "video fast update request".

Using a decoder refresh point implies refraining from using any picture sent prior to that point as a reference for the encoding process of any subsequent picture sent in the stream. For predictive media types that are not video, the analogue applies. For example, if in MPEG-4 systems scene updates are used, the decoder refresh point consists of the full representation of the scene and is not delta-coded relative to previous updates.

Decoder Refresh points, especially Intra or IDR pictures, are in general several times larger in size than predicted pictures. Thus, in scenarios in which the available bandwidth is small, the use of a decoder refresh point implies a delay that is significantly longer than the typical picture duration.

Usage in multicast is possible; however aggregation of the commands is recommended. A receiver that receives a request closely (within 2 times the longest Round Trip Time (RTT) known) after sending a decoder refresh point should await a second request message to ensure that the media receiver has not been served by the previously delivered decoder refresh point. The reason for delaying 2 times the longest known RTT is to avoid sending unnecessary decoder refresh points. A session participant may have sent its own request while another participants request was in-flight to them. Thus suppressing those requests that may have been sent without knowledge about the other request avoids this issue.

Full Intra Request is applicable in use-case 1, 2, and 5.

3.5.1.1. Reliability

The FIR message results in the delivery of a decoder refresh point, unless the message is lost. Decoder refresh points are easily identifiable from the bit stream. Therefore, there is no need for protocol-level acknowledgement, and a simple command repetition mechanism is sufficient for ensuring the level of reliability required. However, the potential use of repetition does require a mechanism to prevent the recipient from responding to messages already received and responded to.

To ensure the best possible reliability, a sender of FIR may repeat the FIR request until a response has been received. The repetition interval is determined by the RTCP timing rules the session operates under. Upon reception of a complete decoder refresh point or the detection of an attempt to send a decoder refresh point (which got damaged due to a packet loss) the repetition of the FIR must stop. If another FIR is necessary, the request sequence number must be increased. To combat loss of the decoder refresh points sent, the sender that receives repetitions of the FIR 2*RTT after the transmission of the decoder refresh point shall send a new decoder refresh point. Two round trip times allow time for the request to arrive at the media sender and the decoder refresh point to arrive back to the requestor. A FIR sender shall not have more than one FIR request (different request sequence number) outstanding at any time per media sender in the session.

An RTP Mixer that receives an FIR from a media receiver is responsible to ensure that a decoder refresh point is delivered to the requesting receiver. It may be necessary to generate FIR commands by the MCU. The two legs (FIR-requesting endpoint to MCU, and MCU to decoder refresh point generating MCU) are handled independently from each other from a reliability perspective.

3.5.2. Temporal Spatial Trade-off Request and Announcement

The Temporal Spatial Trade-off Request (TSTR) instructs the video encoder to change its trade-off between temporal and spatial resolution. Index values from 0 to 31 indicate monotonically a desire for higher frame rate. In general the encoder reaction time may be significantly longer than the typical picture duration. See use case 3 for an example. The encoder decides if the request results in a change of the trade off. An acknowledgement process has been defined to provide feedback of the trade-off that is used henceforth.

Informative note: TSTR and TSTA have been introduced primarily because it is believed that control protocol mechanisms, e.g. a SIP re-invite, are too heavyweight, and too slow to allow for a reasonable user experience. Consider, for example, a user interface where the remote user selects the temporal/spatial trade-off with a slider (as it is common in state-of-the-art video conferencing systems). An immediate feedback to any slider movement is required for a reasonable user experience. A SIP re-invite would require at least 2 round-trips more (compared to the TSTR/TSTA mechanism) and may involve proxies and other complex mechanisms. Even in a well-designed system, it may take a second or so until finally the new trade-off is selected.
Furthermore the use of RTCP solves very efficiently the multicast use case.

The use of TSTR and TSTA in multipoint scenarios is a non-trivial subject, and can be solved in many implementation specific ways. Problems are stemming from the fact that TSTRs will typically arrive unsynchronized, and may request different trade-off values for the same stream and/or endpoint encoder. This memo does not specify a MCU's or endpoint's reaction to the reception of a suggested trade-off as conveyed in the TSTR -- we only require the receiver of a TSTR message to reply to it by sending a TSTA, carrying the new trade-off chosen by its own criteria (which may or may not be based on the trade-off conveyed by TSTR). In other words, the trade-off sent in TSTR is a non-binding recommendation; nothing more.

With respect to TSTR/TSTA, four scenarios based on the topologies described in [Topologies] need to be distinguished. The scenarios are described in the following sub-clauses.

3.5.2.1. Point-to-point

In this most trivial case, the media sender typically adjusts its temporal/spatial trade-off based on the requested value in TSTR, and within its capabilities. The TSTA message conveys back the new trade-off value (which may be identical to the old one if, for example, the sender is not capable to adjust its trade-off).

3.5.2.2. Point-to-Multipoint using Multicast or Translators

RTCP Multicast is used either with media multicast according to Section 2.3.2 of [Topologies], or following RFC 3550's translator model according to Section 2.3.3 of [Topologies]. In these cases, TSTR messages from different receivers may be received

unsynchronized, and possibly with different requested trade-offs (because of different user preferences). This memo does not specify how the media sender tunes its trade-off. Possible strategies include selecting the mean, or median, of all trade-off requests received, prioritize certain participants, or continue using the previously selected trade-off (e.g. when the sender is not capable of adjusting it). Again, all TSTR messages need to be acknowledged by TSTA, and the value conveyed back has to reflect the decision made.

3.5.2.3. Point-to-Multipoint using RTP Mixer

In this scenario the RTP Mixer receives all TSTR messages, and has the opportunity to act on them based on its own criteria. In most cases, the MCU should form a "consensus" of potentially conflicting TSTR messages arriving from different participants, and initiate its own TSTR message(s) to the media sender(s). The strategy of forming this "consensus" is open for the implementation, and can, for example, encompass averaging the participant's request values, prioritizing certain participants, or use session default values. If the Mixer changes its trade-off, it needs to request from the media sender(s) the use of the new value, by creating a TSTR of its own. Upon reaching a decision on the used trade-off it includes that value in the acknowledgement.

Even if a Mixer or Translator performs transcoding, it is very difficult to deliver media with the requested trade-off, unless the content the MCU receives is already close to that trade-off. Only in cases where the original source has substantially higher quality (and bit-rate), it is likely that transcoding can result in the requested trade-off.

3.5.2.4. Reliability

A request and reception acknowledgement mechanism is specified. The Temporal Spatial Trade-off Announcement (TSTA) message informs the request-sender that its request has been received, and what trade-off is used henceforth. This acknowledgment mechanism is desirable for at least the following reasons:

- A change in the trade-off cannot be directly identified from the media bit stream,
- User feedback cannot be implemented without information of the chosen trade-off value, according to the media sender's constraints,
- Repetitive sending of messages requesting an unimplementable trade-off can be avoided.

3.5.3. H.271 Video Back Channel Message

ITU-T Rec. H.271 defines syntax, semantics, and suggested encoder reaction to a video back channel message. The codepoint defined in this memo is used to convey such a message from media receiver to media sender.

We refrain from an in-depth discussion of the available codepoints within H.271 in this memo for a number of reason. The perhaps most important reason is that we expect backward-compatible additions of codepoints to H.271 outside the update/maturity cycle of this memo. The situation is similar to RTP payload format specs - the data carried within the spec is normally not described in any significant detail.

However, we note that some H.271 messages bear similarities with "native" messages of AVPF and this memo, which are known to require caution in multicast environments. One example is the reference picture feedback message, which appears to be critical to contradicting information. While it would perhaps be possible to specify an algorithm to resolve eventual contradictions, this would require an amount of awareness to the details of H.271 and the video codec employed which we would like to avoid in this memo. Therefore, we err on the side of caution and discourage the use of VBCM in topologies other than point-to-point (section 2.3.1 of [Topologies]) and point-to-multipoint utilizing a mixer (section 2.3.4. of [Topologies]). In the former case, obviously, no inconsistency problem exists. In the latter case, it is the mixer's responsibility to resolve the inconsistencies, and the mixer is media aware and can do so.

3.5.3.1. Reliability

H.271 video back channel messages do not require reliable transmission, and the reception of a message can be derived from the forward video bit stream. Therefore, no specific reception acknowledgement is specified.

With respect to re-sending rules, clause 3.5.1.1. applies.

3.5.4. Temporary Maximum Media Bit-rate Request

A receiver, translator or mixer uses the Temporary Maximum Media Bit-rate Request (TMMBR, "timber") to request a sender to limit the maximum bit-rate for a media stream to, or below, the provided value. The primary usage for this is a scenario with MCU (use case 6),

corresponding to topologies in 2.3.3 of [Topologies] (translator) and 2.3.4 of [Topologies] (mixer), but also .2.3.1 of Topologies (point-to-point).

The temporary maximum media bit-rate messages are generic messages that can be applied to any media.

The reasoning below assumes that the participants have negotiated a session maximum bit-rate, using the signalling protocol. This value can be global, for example in case of point-to-point, multicast, or translators. It may also be local between the participant and the peer or mixer. In both cases, the bit-rate negotiated in signalling is the one that the participant guarantees to be able to handle (encode and decode). In practice, the connectivity of the participant also bears an influence to the negotiated value -- it does not necessarily make much sense to negotiate a media bit rate that one's network interface does not support.

An already established temporary bit-rate value may be changed at any time (subject to the timing rules of the feedback message sending), and to any value between zero and the session maximum, as negotiated during signalling. Even if a sender has received a TMMBR message increasing the bit-rate, all increases must be governed by a congestion control algorithm. TMMBR only indicates known limitations, usually in the local environment, and does not provide any guarantees.

If it is likely that the new bit-rate indicated by TMMBR will be valid for the remainder of the session, the TMMBR sender can perform a renegotiation of the session upper limit using the session signalling protocol.

3.5.4.1. MCU based Multi-point operation

Assume a small multiparty conference is ongoing, as depicted in Section 2.3.4 of [Topologies]. All participants (A-D) have negotiated a common maximum bit-rate that this session can use. The conference operates over a number of unicast links between the participants and the MCU. The congestion situation on each of these links can easily be monitored by the participant in question and by the MCU, utilizing, for example, RTCP Receiver Reports. However, any given participant has no knowledge of the congestion situation of the connections to the other participants. Worse, without mechanisms similar to the ones discussed in this draft, the MCU (who is aware of the congestion situation on all connections it manages) has no standardized means to inform participants to slow down, short of forging its own receiver reports (which is undesirable). In

principle, an MCU confronted with such a situation is obliged to thin or transcode streams intended for connections that detected congestion.

In practice, stream thinning - if performed media aware - is unfortunately a very difficult and cumbersome operation and adds undesirable delay. If done media unaware, it leads very quickly to unacceptable reproduced media quality. Hence, means to slow down senders even in the absence of congestion on their connections to the MCU are desirable.

To allow the MCU to perform congestion control on the individual links, without performing transcoding, there is a need for a mechanism that enables the MCU to request the participant's media encoders to limit their maximum media bit-rate currently used. The MCU handles the detection of a congestion state between itself and a participant as follows:

1. Start thinning the media traffic to the supported bit-rate.
2. Use the TMMBR to request the media sender(s) to reduce the media bit-rate sent by them to the MCU, to a value that is in compliance with congestion control principles for the slowest link. Slow refers here to the available bandwidth and packet rate after congestion control.
3. As soon as the bit-rate has been reduced by the sending part, the MCU stops stream thinning implicitly, because there is no need for it any more as the stream is in compliance with congestion control.

Above algorithms may suggest to some that there is no need for the TMMBR - it should be sufficient to solely rely on stream thinning. As much as this is desirable from a network protocol designer's viewpoint, it has the disadvantage that it doesn't work very well - the reproduced media quality quickly becomes unusable.

It appears to be a reasonable compromise to rely on stream thinning as an immediate reaction tool to combat congestions, and have a quick control mechanism that instructs the original sender to reduce its bitrate.

Note also that the standard RTCP receiver report cannot serve for the purpose mentioned. In an environment with RTP Mixers, the RTCP RR is being sent between the RTP receiver in the endpoint and the RTP sender in the Mixer only - as there is no multicast transmission. The stream that needs to be bandwidth-reduced, however, is the one between the original sending endpoint and the Mixer. This endpoint doesn't see the aforementioned RTCP RRs, and hence needs explicitly informed about desired bandwidth adjustments.

In this topology it is the Mixer's responsibility to collect, and consider jointly, the different bit-rates which the different links may support, into the bit rate requested. This aggregation may also take into account that the Mixer may contain certain transcoding capabilities (as discussed in section 2.3.4 of [Topologies]), which can be employed for those few of the session participants that have the lowest available bit-rates.

3.5.4.2. Point-to-Multipoint using Multicast or Translators

In this topology, RTCP RRs are transmitted globally which allows for the detection of transmission problems such as congestion, on a medium timescale. As all media senders are aware of the congestion situation of all media receivers, the rationale of the use of TMMBR of section 3.5.4.1 does not apply. However, even in this case the congestion control response can be improved when the unicast links are employing congestion controlled transport protocols (such as TCP or DCCP). A peer may also report local limitation to the media sender.

3.5.4.3. Point-to-point operation

In use case 7 it is possible to use TMMBR to improve the performance at times of changes in the known upper limit of the bit-rate. In this use case the signalling protocol has established an upper limit for the session and media bit-rates. However at the time of transport link bit-rate reduction, a receiver could avoid serious congestion by sending a TMMBR to the sending side.

3.5.4.4. Reliability

The reaction of a media sender to the reception of a TMMBR message is not immediately identifiable through inspection of the media stream. Therefore a more explicit mechanism is needed to avoid unnecessary re-sending of TMMBR messages. Using a statistically based retransmission scheme would only provide statistical guarantees of the request being received. It would also not avoid the retransmission of already received messages. In addition it does not allow for easy suppression of other participants requests. For the reasons mentioned, a mechanism based on explicit notification is used.

Upon the reception of a request a media sender sends a notification containing the current applicable limitation of the bit-rate, and which session participants that own that limit. That allows all other

participants to suppress any request they may have, with limitation value equal or higher to the current one. The identity of the owner allows for small message sizes and media sender states. A media sender only keeps state for the SSRC of the current owner of the limitation; all other requests and their sources are not saved. Only the participant with the lowest value is allowed to remove or change its limitation. Otherwise anyone that ever set a limitation would need to remove it to allow the maximum bit-rate to be raised beyond that value.

4. RTCP Receiver Report Extensions

This memo specifies six new feedback messages. The Full Intra Request (FIR), Temporal-Spatial Trade-off Request (TSTR), Temporal-Spatial Trade-off Announcement (TSTA), and Video Back Channel Message (VBCM) are "Payload Specific Feedback Messages" in the sense of section 6.3 of AVPF [AVPF]. The Temporary Maximum Media Bit-rate Request (TMMBR) and Temporary Maximum Media Bit-rate Notification (TMMBN) are "Transport Layer Feedback Messages" in the sense of section 6.2 of AVPF.

In the following subsections, the new feedback messages are defined, following a similar structure as in the AVPF specification's sections 6.2 and 6.3, respectively.

4.1. Design Principles of the Extension Mechanism

RTCP was originally introduced as a channel to convey presence, reception quality statistics and hints on the desired media coding. A limited set of media control mechanisms have been introduced in early RTP payload formats for video formats, for example in RFC 2032 [RFC2032]. However, this specification, for the first time, suggests a two-way handshake for one of its messages. There is danger that this introduction could be misunderstood as the precedence for the use of RTCP as an RTP session control protocol. In order to prevent these misunderstandings, this subsection attempts to clarify the scope of the extensions specified in this memo, and strongly suggests that future extensions follow the rationale spelled out here, or compellingly explain why they divert from the rationale.

In this memo, and in AVPF [AVPF], only such messages have been included which

a) have comparatively strict real-time constraints, which prevent the use of mechanisms such as a SIP re-invite in most application scenarios. The real-time constraints are explained separately for each message where necessary
b) are multicast-safe in that the reaction to potentially contradicting feedback messages is specified, as necessary for each message
c) are directly related to activities of a certain media codec, class of media codecs (e.g. video codecs), or the given media stream.

In this memo, a two-way handshake is only introduced for such messages that

a) require a notification or acknowledgement due to their nature, which is motivated separately for each message
b) the notification or acknowledgement cannot be easily derived from the media bit stream.

All messages in AVPF [AVPF] and in this memo follow a number of common design principles. In particular:

a) Media receivers are not always implementing higher control protocol functionalities (SDP, XML parsers and such) in their media path. Therefore, simple binary representations are used in the feedback messages and not an (otherwise desirable) flexible format such as, for example, XML.

4.2. Transport Layer Feedback Messages

Transport Layer FB messages are identified by the value RTPFB (205) as RTCP packet type.

In AVPF, one message of this category had been defined. This memo specifies two more messages for a total of three messages of this type. They are identified by means of the FMT parameter as follows:

```
0:    unassigned
1:    Generic NACK (as per AVPF)
2:    Maximum Media Bit-rate Request
3:    Maximum Media Bit-rate Notification
4-30: unassigned
31:   reserved for future expansion of the identifier number space
```

The following subsection defines the formats of the FCI field for this type of FB message.

4.2.1. Temporary Maximum Media Bit-rate Request (TMMBR)

The FCI field of a TMMBR Feedback message SHALL contain one or more FCI entries.

4.2.1.1. Semantics

The TMMBR is used to indicate the highest bit-rate per sender of a media, which the receiver currently supports in this RTP session. The media sender MAY use any lower bit-rate, as it may need to address a congestion situation or other limiting factors. See section 5 (congestion control) for more discussion.

The "SSRC of the packet sender" field indicates the source of the request, and the "SSRC of media source" is not used and SHALL be set to 0. The SSRC of media sender in the FCI field denotes the media sender the message applies to. This is useful in the multicast or translator topologies where each media sender may be addressed in a single TMMBR message using multiple FCIs.

A TMMBR FCI MAY be repeated in subsequent TMMBR messages if no applicable TMMBN FCI has been received at the time of transmission of the next RTCP packet. The bit-rate value of a TMMBR FCI MAY be changed from a previous TMMBR message and the next, regardless of the eventual reception of an applicable TMMBN FCI.

Please note that a TMMBN message is sent by the media sender at the earliest possible point in time, as a result of any TMMBR messages received since the last sending of TMMBN. The TMMBN message indicates the limit and the owner of that limit at the time of the transmission of the message. The limit is the lowest of all values received since the last TMMBN was transmitted.

A media receiver who is not the owner of the bandwidth limit when sending a TMMBR, MUST request a bandwidth lower than their knowledge of currently established bandwidth limit for this media sender. Therefore, all received requests for bandwidth limits greater or equal to the one currently established are ignored. A media receiver who is the owner of the current bandwidth limit, MAY lower the value further, raise the value or remove the restriction completely by setting the bandwidth limit equal to the session limit.

Once a session participant receives the TMMBN in response to its TMMBR, with its own SSRC, it knows that it "owns" the bandwidth limitation. Only the "owner" of a bandwidth limitation can raise it or reset it to the session limit.

Note that, due to the unreliable nature of transport of TMMBR and TMMBN, the above rules may lead to the sending of TMMBR messages disobeying the rules above. Furthermore, in multicast scenarios it can happen that more than one session participants believes it "owns" the current bandwidth limitation. This is not critical for a number of reasons:

a) If a TMMBR message is lost in transmission, the media sender does not learn about the restrictions imposed on it. However, it also does not send a TMMBN message notifying reception of a request it has never received. Therefore, no new limit is established, the media receiver sending the more restrictive TMMBR is not the owner. Since this media receiver has not seen a notification corresponding to its request, it is free to re-send it.

b) Similarly, if a TMMBN message gets lost, the media receiver that has sent the corresponding TMMBR request does not receive acknowledgement. In that case, it is also not the "owner" of the restriction and is free to re-send the request.
c) If multiple competing TMMBR messages are sent by different session participants, then the resulting TMMBN indicates the lowest bandwidth requested; the owner is set to the sender of the TMMBR with the lowest requested bandwidth value.

TMMBR feedback SHOULD NOT be used if the underlying transport protocol is capable of providing similar feedback information from the receiver to the sender.

It also important to consider the security risks involved with faked TMMBRs. See security considerations in Section 6.

The feedback messages may be used in both multicast and unicast sessions of any of the specified topologies.

For sessions with a larger number of participants using the lowest common denominator, as required by this mechanism, may not be the most suitable course of action. Larger session may need to consider other ways to support adapted bit-rate to participants, such as partitioning the session in different quality tiers, or use some other method of achieving bit-rate scalability.

If the value set by a TMMBR message is expected to be permanent the TMMBR setting party is RECOMMENDED to renegotiate the session parameters to reflect that using the setup signalling.

4.2.1.2. Message Format

The Feedback control information (FCI) consists of one or more TMMBR FCI entries with the following syntax:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              SSRC                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Maximum bit-rate in units of 128 bits/s            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1 - Syntax for the TMMBR message

SSRC: The SSRC value of the target of this specific maximum bit-rate request.

Maximum bit-rate: The temporary maximum media bit-rate value in units of 128 bit/s. This provides range from 0 to 549755813888 bits/s (~550 Tbit/s) with a granularity of 128 bits/s.

The length of the FB message is be set to 2+2*N where N is the number of TMMBR FCI entries.

4.2.1.3. Timing Rules

The first transmission of the request message MAY use early or immediate feedback in cases when timeliness is desirable. Any repetition of a request message SHOULD use regular RTCP mode for its transmission timing.

4.2.2. Temporary Maximum Media Bit-rate Notification (TMMBN)

The FCI field of the TMMBN Feedback message SHALL contain one TMMBN FCI entry.

4.2.2.1. Semantics

This feedback message is used to notify the senders of any TMMBR message that one or more TMMBR messages have been received. It indicates to all participants the currently employed maximum bit-rate value and the "owner" of the current limitation. The "owner" of a limitation is the sender of the last (most restrictive) TMMBR message received by the media sender.

The "SSRC of the packet sender" field indicates the source of the notification. The "SSRC of media source" SHALL be set to the SSRC of the media receiver that currently owns the bit-rate limitation.

A TMMBN message SHALL be scheduled for transmission after the reception of a TMMBR message with a FCI including the session participant's SSRC. Only a single TMMBN SHALL be sent, even if more than one TMMBR messages are received between the scheduling of the transmission and the actual transmission of the TMMBN message. The TMMBN message indicates the limit and the owner of that limit at the time of transmitting the message. The limit SHALL be the lowest of all values received since the last TMMBN was transmitted. The one sending that request SHALL become the owner of the limit.

The reception of a TMMBR message with a transmission limit greater or equal than the current limit SHALL still result in the transmission of a TMMBN message. However the limit and owner is not changed, unless it was from the owner, and the current limit and owner is indicated in the TMMBN message. This procedure allows session participants that haven't seen the last TMMBN message to get a correct view of this media sender's state.

When a media sender determines an "owner" of a limitation has left the session, then the current limitation is removed, and the media sender SHALL send a TMMBN message indicating the maximum session bandwidth.

4.2.2.2. Message Format

The TMMBN Feedback control information (FCI) entry has the following syntax:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Maximum bit-rate in units of 128 bits/s            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2 - Syntax for the TMMBN message

Maximum bit-rate: The current temporary maximum media bit-rate value in units of 128 bit/s.

The length field value of the FB message SHALL be 3.

4.2.2.3. Timing Rules

The acknowledgement SHOULD be sent as soon as allowed by the applied timing rules for the session. Immediate or early feedback mode SHOULD be used for these messages.

4.3. Payload Specific Feedback Messages

Payload-Specific FB messages are identified by the value PT=PSFB (206) as RTCP packet type.

AVPF defines three payload-specific FB messages and one application layer FB message. This memo specifies four additional payload

specific feedback messages. All are identified by means of the FMT parameter as follows:

```
0:     unassigned
1:     Picture Loss Indication (PLI)
2:     Slice Lost Indication (SLI)
3:     Reference Picture Selection Indication (RPSI)
4:     Full Intra Request Command (FIR)
5:     Temporal-Spatial Trade-off Request (TSTR)
6:     Temporal-Spatial Trade-off Announcement (TSTA)
7:     Video Back Channel Message (VBCM)
8-14:  unassigned
15:    Application layer FB message
16-30: unassigned
31:    reserved for future expansion of the sequence number space
```

The following subsections define the new FCI formats for the payload-specific FB messages.

4.3.1. Full Intra Request (FIR) command

The FIR command FB message is identified by PT=PSFB and FMT=4.

There MUST be one or more FIR entry contained in the FCI field.

4.3.1.1. Semantics

Upon reception of a FIR message, an encoder MUST send a decoder refresh point (see Section 2.2) as soon as possible.

Note: Currently, video appears to be the only useful application for FIR, as it appears to be the only RTP payloads widely deployed that relies heavily on media prediction across RTP packet boundaries. However, use of FIR could also reasonably be envisioned for other media types that share essential properties with compressed video, namely cross-frame prediction (whatever a frame may be for that media type). One possible example may be the dynamic updates of MPEG-4 scene descriptions. It is suggested that payload formats for such media types refer to FIR and other message types defined in this specification and in AVPF, instead of creating similar mechanisms in the payload specifications. The payload specifications may have to explain how the payload specific terminologies map to the video-centric terminology used here.

Note: In environments where the sender has no control over the codec (e.g. when streaming pre-recorded and pre-coded content), the

reaction to this command cannot be specified. One suitable reaction of a sender would be to skip forward in the video bit stream to the next decoder refresh point. In other scenarios, it may be preferable not to react to the command at all, e.g. when streaming to a large multicast group. Other reactions may also be possible. When deciding on a strategy, a sender could take into account factors such as the size of the receiving multicast group, the "importance" of the sender of the FIR message (however "importance" may be defined in this specific application), the frequency of decoder refresh points in the content, and others. However the usage of FIR in a session which predominately handles pre-coded content shouldn't use the FIR at all.

The sender MUST consider congestion control as outlined in section 5, which MAY restrict its ability to send a decoder refresh point quickly.

Note: The relationship between the Picture Loss Indication and FIR is as follows. As discussed in section 6.3.1 of AVPF, a Picture Loss Indication informs the decoder about the loss of a picture and hence the likeliness of misalignment of the reference pictures in encoder and decoder. Such a scenario is normally related to losses in an ongoing connection. In point-to-point scenarios, and without the presence of advanced error resilience tools, one possible option an encoder has is to send a decoder refresh point. However, there are other options including ignoring the PLI, for example if only one receiver of many has sent a PLI or when the embedded stream redundancy is likely to clean up the reproduced picture within a reasonable amount of time.
The FIR, in contrast, leaves a real-time encoder no choice but to send a decoder refresh point. It disallows the encoder to take into account any considerations such as the ones mentioned above.

Note: Mandating a maximum delay for completing the sending of a decoder refresh point would be desirable from an application viewpoint, but may be problematic from a congestion control point of view. "As soon as possible" as mentioned above appears to be a reasonable compromise.

FIR SHALL NOT be sent as a reaction to picture losses - it is RECOMMENDED to use PLI instead. FIR SHOULD be used only in such situations where not sending a decoder refresh point would render the video unusable for the users.

Note: a typical example where sending FIR is adequate is when, in a multipoint conference, a new user joins the session and no regular decoder refresh point interval is established. Another example

would be a video switching MCU that changes streams. Here, normally, the MCU issues a freeze picture request (through protocol means outside this specification) to the receiver(s), switches the streams, and issues a FIR to the new sender so to force it to emit a decoder refresh point. The decoder refresh point includes normally a Freeze Picture Release (defined outside this specification), which re-starts the rendering process of the receivers. Both techniques mentioned are commonly used in MCU-based multipoint conferences.

Other RTP payload specifications such as RFC 2032 [RFC2032] already define a feedback mechanism for certain codecs. An application supporting both schemes MUST use the feedback mechanism defined in this specification when sending feedback. For backward compatibility reasons, such an application SHOULD also be capable to receive and react to the feedback scheme defined in the respective RTP payload format, if this is required by that payload format.

The "SSRC of the packet sender" field indicates the source of the request, and the "SSRC of media source" is not used and SHALL be set to 0. The SSRC of media sender to which the FIR command applies to is in the FCI.

4.3.1.2. Message Format

Full Intra Request uses one additional FCI field, the content of which is depicted in Figure 3 The length of the FB message MUST be set to 2+2*N, where N is the number of FCI entries.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              SSRC                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Seq. nr       |    Reserved                                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3 - Syntax for the FIR message

SSRC: The SSRC value of the media sender of this specific FIR command.

Seq. nr: Command sequence number. The sequence number space is unique for each tuple consisting of the SSRC of command

source and the SSRC of the command target. The sequence number SHALL be increased by 1 modulo 256 for each new command. A repetition SHALL NOT increase the sequence number. Initial value is arbitrary.

Reserved: All bits SHALL be set to 0 and SHALL be ignored on reception.

The semantics of this FB message is independent of the RTP payload type.

4.3.1.3. Timing Rules

The timing follows the rules outlined in section 3 of [AVPF]. FIR commands MAY be used with early or immediate feedback. The FIR feedback message MAY be repeated. If using immediate feedback mode the repetition SHOULD wait at least on RTT before being sent. In early or regular RTCP mode the repetition is sent in the next regular RTCP packet.

4.3.1.4. Remarks

FIR messages typically trigger the sending of full intra or IDR pictures. Both are several times larger then predicted (inter) pictures. Their size is independent of the time they are generated. In most environments, especially when employing bandwidth-limited links, the use of an intra picture implies an allowed delay that is a significant multitude of the typical frame duration. An example: If the sending frame rate is 10 fps, and an intra picture is assumed to be 10 times as big as an inter picture, then a full second of latency has to be accepted. In such an environment there is no need for a particular short delay in sending the FIR message. Hence waiting for the next possible time slot allowed by RTCP timing rules as per [AVPF] may not have an overly negative impact on the system performance.

4.3.2. Temporal-Spatial Trade-off Request (TSTR)

The TSTR FB message is identified by PT=PSFB and FMT=5.

There MUST be one or more TSTR entry contained in the FCI field.

4.3.2.1. Semantics

A decoder can suggest the use of a temporal-spatial trade-off by sending a TSTR message to an encoder. If the encoder is capable of adjusting its temporal-spatial trade-off, it SHOULD take into account the received TSTR message for future coding of pictures. A value of 0 suggests a high spatial quality and a value of 31 suggests a high frame rate. The values from 0 to 31 indicate monotonically a desire for higher frame rate. Actual values do not correspond to precise values of spatial quality or frame rate.

The reaction to the reception of more than one TSTR message by a media sender from different media receivers is left open to the implementation. The selected trade-off SHALL be communicated to the media receivers by the means of the TSTA message.

The "SSRC of the packet sender" field indicates the source of the request, and the "SSRC of media source" is not used and SHALL be set to 0. The SSRC of media sender to which the TSTR applies to is in the FCI entries.

A TSTR message may contain multiple requests to different media senders, using multiple FCI entries.

4.3.2.2. Message Format

The Temporal-Spatial Trade-off Request uses one FCI field, the content of which is depicted in Figure 4. The length of the FB message MUST be set to 2+2*N, where N is the number of FCI entries included.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              SSRC                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Seq nr.      |  Reserved                           | Index   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4 - Syntax of the TSTR

SSRC: The SSRC value of the target (or the media sender) of this specific TSTR request.

Seq. nr: Request sequence number. The sequence number space is unique for each tuple consisting of the SSRC of request source and the SSRC of the request target. The sequence number SHALL be increased by 1 modulo 256 for each new command. A repetition SHALL NOT increase the sequence number. Initial value is arbitrary.

Index: An integer value between 0 and 31 that indicates the relative trade off that is requested. An index value of 0 index highest possible spatial quality, while 31 indicates highest possible temporal resolution.

Reserved: All bits SHALL be set to 0 and SHALL be ignored on reception.

4.3.2.3. Timing Rules

The timing follows the rules outlined in section 3 of [AVPF]. This request message is not time critical and SHOULD be sent using regular RTCP timing. Only if it is known that the user interface requires a quick feedback, the message MAY be sent with early or immediate feedback timing.

4.3.2.4. Remarks

The term "spatial quality" does not necessarily refer to the resolution, measured by the number of pixels the reconstructed video is using. In fact, in most scenarios the video resolution stays constant during the lifetime of a session. However, all video compression standards have means to adjust the spatial quality at a given resolution, often influenced by the Quantizer Parameter or QP. A numerically low QP results in a good reconstructed picture quality, whereas a numerically high QP yields a coarse picture. The typical reaction of an encoder to this request is to change its rate control parameters to use a lower frame rate and a numerically lower (on average) QP, or vice versa. The precise mapping of Index, frame rate, and QP is intentionally left open here, as it depends on factors such as compression standard employed, spatial resolution, content, bit rate, and many more.

4.3.3. Temporal-Spatial Trade-off Announcement (TSTA)

The TSTA FB message is identified by PT=PSFB and FMT=6.

There SHALL be one or more TSTA contained in the FCI field.

4.3.3.1. Semantics

This feedback message is used to acknowledge the reception of a TSTR. A TSTA entry in a TSTA feedback message SHALL be sent for each TSTR entry targeted to this session participant, i.e. each TSTR received that in the SSRC field in the entry has the receiving entities SSRC. The acknowledgement SHALL be sent also for repetitions received. If the request receiver has received TSTR with several different sequence numbers from a single requestor it SHALL only respond to the request with the highest (modulo 256) sequence number.

The TSTA SHALL include the Temporal-Spatial Trade-off index that will be used as a result of the request. This is not necessarily the same index as requested, as media sender may need to aggregate requests from several requesting session participants. It may also have some other policies or rules that limit the selection.

A single TSTA message MAY acknowledge multiple requests using multiple FCI entries.

4.3.3.2. Message Format

The Temporal-Spatial Trade-off Announcement uses one additional FCI field, the content of which is depicted in Figure 5. The length of the FB message MUST be set to 2+2*N, where N is the number of FCI entries.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              SSRC                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Seq nr.      |  Reserved                           | Index   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5 - Syntax of the TSTA

SSRC: The SSRC of the source of the TSTA request that is acknowledged.

Seq. nr: The sequence number value from the TSTA request that is being acknowledged.

Index: The trade-off value the media sender is using henceforth.

Reserved: All bits SHALL be set to 0 and SHALL be ignored on reception.

Informative note: The returned trade-off value (Index) may differ from the requested one, for example in cases where a media encoder cannot tune its trade-off, or when pre-recorded content is used.

4.3.3.3. Timing Rules

The timing follows the rules outlined in section 3 of [AVPF]. This acknowledgement message is not extremely time critical and SHOULD be sent using regular RTCP timing.

4.3.3.4. Remarks

None

4.3.4. H.271 VideoBackChannelMessage (VBCM)

The VBCM FB message is identified by PT=PSFB and FMT=7.

There MUST be one or more VBCM entry contained in the FCI field.

Semantics

The "payload" of VBCM indication carries codec specific, different types of feedback information. The type of feedback information can be classified as "status report" such as receiving bit stream without errors, loss of partial or complete picture or block or "update requests" such as complete refresh of the bit stream.

Note: There are possible overlap between the VBCM sub-messages and CCM/AVPF feedback messages, such FIR. Please see section 3.5.3 for further discussions.

The different types of feedback sub-messages carried in the VBCM are indicated by the "payloadType" as defined in [VBCM]. The different sub-message types as defined in [VBCM] are re-produced below for convenience. "payloadType", in ITU-T Rec. H.271 terminology, refers to the sub-type of the H.271 message and should not be confused with an RTP payload type.

Payload Type   Message Content

```
0      One or more pictures without detected bitstream error mismatch
1      One or more pictures that are entirely or partially lost
2      A set of blocks of one picture that is entirely or partially
lost
3      CRC for one parameter set
4      CRC for all parameter sets of a certain type
5      A "reset" request indicating that the sender should completely
refresh the video bitstream as if no prior bitstream data had been
received
>      5    Reserved for future use by ITU-T
```

The bit string or the "payload" of VBCM message is of variable length and is self-contained and coded in a variable length, binary format. The media sender necessarily has to be able to parse this optimized binary format to make use of VBCM messages

Each of the different types of sub-messages (indicated by payloadType)e may have different semantic based on the codec used.

Message Format

The VBCM indication uses one FCI field and the syntax is depicted in Figure 6.

```
0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              SSRC                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Seq. nr       |0| Payload Type| Length                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    VBCM Bit String....             | Padding  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 6 - Syntax for VBCM Message

SSRC: The SSRC value of the media sender of this specific VBCM indication message.

Seq. nr : Command sequence number. The sequence number space is unique for each tuple consisting of the SSRC of command source and the SSRC of the command target. The sequence number SHALL be increased by 1 modulo 256 for each new command. A repetition SHALL NOT increase the sequence number. Initial value is arbitrary.

0: Must be set to 0 and should not be acted upon receiving.

Payload: The RTP payload type for which the VBCM bit stream must be interpreted.

NOTE : Stephan I think this payload type is redundant, since during session set up phase you do lock down the payload type that you are going to use for this session. But I am keeping it since it is there in RPSI of AVPF. Is there any reason for this ? From an implementation point of view I can see this being helpful (and fast) since you don't have to go back to your session information.

VBCM Bit String : This is the bit string generated by the decoder carrying a specific feedback sub-message. It is of variable length.

Padding: Bits set to 0 to make up a 32 bit boundry

Timing Rules

The timing follows the rules outlined in section 3 of [AVPF]

Remarks

Please see section 3.5.3 for the applicability of the VBCM message in relation to messages in both AVPF and this memo with similar functionality.

5. Congestion Control

The correct application of the AVPF timing rules prevents the network flooding by feedback messages. Hence, assuming a correct implementation, the RTCP channel cannot break its bit-rate commitment and introduce congestion.

The reception of some of the feedback messages modifies the behaviour of the media senders or, more specifically, the media encoders. All of these modifications MUST only be performed within the bandwidth limits the applied congestion control provides. For example, when reacting to a FIR, the unusually high number of packets that form the decoder refresh point have to be paced in compliance with the congestion control algorithm, even if the user experience suffers from a slowly transmitted decoder refresh point.

A change of the Temporary Maximum Media Bit-rate value can only mitigate congestion, but not cause congestion. An increase of the value by a request REQUIRES the media sender to use congestion control when increasing its transmission rate to that value. A reduction of the value results in a reduced transmission bit-rate thus reducing the risk for congestion.

6. Security Considerations

The defined messages have certain properties that have security implications. These must be addressed and taken into account by users of this protocol.

The defined setup signalling mechanism is sensitive to modification attacks that can result in session creation with sub-optimal configuration, and, in the worst case, session rejection. To prevent this type of attack, authentication and integrity protection of the setup signalling is required.

Spoofed or maliciously created feedback messages of the type defined in this specification can have the following implications:

a. Severely reduced media bit-rate due to false TMMBR messages that sets the maximum to a very low value.
b. The assignment of the ownership of a bit-rate limit with a TMMBN message to the wrong participant. Thus potentially freezing the mechanism until a correct TMMBN message reached the participants.
c. Sending TSTR that result in a video quality different from the user's desire, rendering the session less useful.
d. Frequent FIR commands will potentially reduce the frame-rate making the video jerky due to the frequent usage of decoder refresh points.

To prevent these attacks there is need to apply authentication and integrity protection of the feedback messages. This can be accomplished against group external threats using the RTP profile that combines SRTP [SRTP] and AVPF into SAVPF [SAVPF]. In the MCU cases, separate security contexts and filtering can be applied between the MCU and the participants thus protecting other MCU users from a misbehaving participant.

7. SDP Definitions

Section 4 of [AVPF] defines new SDP [RFC2327] attributes that are used for the capability exchange of the AVPF commands and indications, such as Reference Picture selection, Picture loss

indication etc. The defined SDP attribute is known as rtcp-fb and its ABNF is described in section 4.2 of [AVPF]. In this section we extend the rtcp-fb attribute to include the commands and indications that are described in this document for codec control protocol. We also discuss the Offer/Answer implications for the codec control commands and indications.

7.1. Extension of rtcp-fb attribute

As described in [AVPF], the rtcp-fb attribute is defined to indicate the capability of using RTCP feedback. As defined in AVPF the rtcp-fb attribute must only be used as a media level attribute and must not be provided at session level.
All the rules described in [AVPF] for rtcp-fb attribute relating to payload type, multiple rtcp-fb attributes in a session description hold for the new feedback messages for codec control defined in this document.

The ABNF for rtcp-fb attributed as defined in [AVPF] is

```
Rtcp-fb-syntax = "a=rtcp-fb: " rtcp-fb-pt SP rtcp-fb-val CRLF
```

Where rtcp-fb-pt is the payload type and rtcp-fb-val defines the type of the feedback message such as ack, nack, trr-int and rtcp-fb-id. For example to indicate the support of feedback of picture loss indication, the sender declares the following in SDP

```
v=0
o=alice 3203093520 3203093520 IN IP4 host.example.com
s=Media with feedback
t=0 0
c=IN IP4 host.example.com
m=audio 49170 RTP/AVPF 98
a=rtpmap:98 H263-1998/90000
a=rtcp-fb:98 nack pli
```

In this document we define a new feedback value type called "ccm" which indicates the support of codec control using RTCP feedback messages. The "ccm" feedback value should be used with parameters, which indicates the support of which codec commands the session may use. In this draft we define four parameters, which can be used with the ccm feedback value type.

- o "fir" indicates the support of Full Intra Request
- o "tmmbr" indicates the support of Temporal Maximum Media Bit-rate

- o "tstr" indicates the support of temporal spatial trade-off request.
- O "bbcm" indicates the support of H.271 video back channel messages.

In ABNF for rtcp-fb-val defined in [AVPF], there is a placeholder called rtcp-fb-id to define new feedback types. The ccm is defined as a new feedback type in this document and the ABNF for the parameters for ccm are defined here (please refer section 4.2 of [AVPF] for complete ABNF syntax).

```
Rtcp-fb-param = SP "app" [SP byte-string]
              / SP rtcp-fb-ccm-param
              /       ; empty

rtcp-fb-ccm-param = "ccm" SP ccm-param


ccm-param  = "fir"   ; Full Intra Request
            / "tmmbr" ; Temporary max media bit rate
            / "tstr"  ; Temporal Spatial Trade Off
            / "vbcm" 1*[SP subMessageType] ; H.271 VBCM messages
            / token [SP byte-string]
                      ; for future commands/indications
subMessageType = 1*[integer];
byte-string = <as defined in section 4.2 of [AVPF] >
```

7.2. Offer-Answer

The Offer/Answer [RFC3264] implications to codec control protocol feedback messages are similar to as described in [AVPF]. The offerer MAY indicate the capability to support selected codec commands and indications. The answerer MUST remove all ccm parameters, which it does not understand or does not wish to use in this particular media session. The answerer MUST NOT add new ccm parameters in addition to what has been offered. The answer is binding for the media session and both offerer and answerer MUST only use feedback messages negotiated in this way.

7.3. Examples

Example 1: The following SDP describes a point-to-point video call with H.263 with the originator of the call declaring its capability to support codec control messages - fir, tstr. The SDP is carried in a high level signalling protocol like SIP

```
v=0
o=alice 3203093520 3203093520 IN IP4 host.example.com
s=Point-to-Point call
c=IN IP4 172.11.1.124
m=audio 49170 RTP/AVP 0
a=rtpmap:0 PCMU/8000
m=video 51372 RTP/AVPF 98
a=rtpmap:98 H263-1998/90000
a=rtcp-fb:98 ccm tstr
a=rtcp-fb:98 ccm fir
```

In the above example the sender when it receives a TSTR message from the remote party can adjust the trade off as indicated in the RTCP TSTA feedback message.

Example 2: The following SDP describes a SIP end point joining a video MCU that is hosting a multiparty video conferencing session. The participant supports only the FIR (Full Intra Request) codec control command and it declares it in its session description. The video MCU can send an FIR RTCP feedback message to this end point when it needs to send this participants video to other participants of the conference.

```
v=0
o=alice 3203093520 3203093520 IN IP4 host.example.com
s=Multiparty Video Call
c=IN IP4 172.11.1.124
m=audio 49170 RTP/AVP 0
a=rtpmap:0 PCMU/8000
m=video 51372 RTP/AVPF 98
a=rtpmap:98 H263-1998/90000
a=rtcp-fb:98 ccm fir
```

When the video MCU decides to route the video of this participant it sends an RTCP FIR feedback message. Upon receiving this feedback message the end point is mandated to generate a full intra request.

Example 3: The following example describes the Offer/Answer implications for the codec control messages. The Offerer wishes to support all the commands and indications of codec control messages. The offered SDP is

```
-------------> Offer
      v=0
```

```
o=alice 3203093520 3203093520 IN IP4 host.example.com
s=Offer/Answer
c=IN IP4 172.11.1.124
m=audio 49170 RTP/AVP 0
a=rtpmap:0 PCMU/8000
m=video 51372 RTP/AVPF 98
a=rtpmap:98 H263-1998/90000
a=rtcp-fb:98 ccm tstr
a=rtcp-fb:98 ccm fir
a=rtcp-fb:98 ccm tmmbr
```

The answerer only wishes to support FIR and TSTR message as the codec control messages and the answerer SDP is

```
<---------------- Answer

      v=0
      o=alice 3203093520 3203093524 IN IP4 otherhost.example.com
      s=Offer/Answer
      c=IN IP4 189.13.1.37
      m=audio 47190 RTP/AVP 0
      a=rtpmap:0 PCMU/8000
      m=video 53273 RTP/AVPF 98
      a=rtpmap:98 H263-1998/90000
      a=rtcp-fb:98 ccm tstr
      a=rtcp-fb:98 ccm fir
```

Example 4: The following example describes the Offer/Answer implications for H.271 Video back channel messages (VBCM). The Offerer wishes to support VBCM and the submessages of payloadType 2(A set of blocks of one picture that is entirely or partially lost, 3 (CRC for one parameter set) and 4 (CRC for all parameter sets of a certain type).

```
-------------> Offer
      v=0
      o=alice 3203093520 3203093520 IN IP4 host.example.com
      s=Offer/Answer
      c=IN IP4 172.11.1.124
      m=audio 49170 RTP/AVP 0
      a=rtpmap:0 PCMU/8000
      m=video 51372 RTP/AVPF 98
      a=rtpmap:98 H263-1998/90000
      a=rtcp-fb:98 ccm vbcm 2 3 4
```

The answerer only wishes to support sub-messages 3 and 4 only

<---------------- Answer

```
v=0
o=alice 3203093520 3203093524 IN IP4 otherhost.example.com
s=Offer/Answer
c=IN IP4 189.13.1.37
m=audio 47190 RTP/AVP 0
a=rtpmap:0 PCMU/8000
m=video 53273 RTP/AVPF 98
a=rtpmap:98 H263-1998/90000
a=rtcp-fb:98 ccm vbcm 3 4
```

So in the above example only VBCM indication comprising of only "payloadType" 3 and 4 will be supported.

8. IANA Considerations

The new value of ccm for the rtcp-fb attribute needs to be registered with IANA.

```
Value name:       ccm
Long Name:        Codec Control Commands and Indications
Reference:        RFC XXXX
```

For use with "ccm" the following values also needs to be registered.

```
Value name:       fir
Long name:        Full Intra Request Command
Usable with:      ccm
Reference:        RFC XXXX

Value name:       tmmbr
Long name:        Temporary Maximum Media Bit-rate
Usable with:      ccm
Reference:        RFC XXXX

Value name:       tstr
Long name:        temporal Spatial Trade Off
Usable with:      ccm
Reference:        RFC XXXX

Value name:       vbcm
Long name:        H.271 video back channel messages
Usable with:      ccm
Reference:        RFC XXXX
```

9. Acknowledgements

The authors would like to thank Andrea Basso, Orit Levin, Nermeen Ismail for their work on the requirement and discussion draft [Basso].

10. References

10.1. Normative references

[AVPF] draft-ietf-avt-rtcp-feedback-11.txt

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

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.

[RFC2327] Handley, M. and V. Jacobson, "SDP: Session Description Protocol", RFC 2327, April 1998.

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

[Topologies] M. Westerlund, and S. Wenger, "Topologies", RFC xxxx, x

10.2. Informative references

[Basso] A. Basso, et. al., "Requirements for transport of video control commands", draft-basso-avt-videoconreq-02.txt, expired Internet Draft, October 2004.

[AVC] Joint Video Team of ITU-T and ISO/IEC JTC 1, Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC), Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-G050, March 2003.

[NEWPRED] S. Fukunaga, T. Nakai, and H. Inoue, "Error Resilient Video Coding by Dynamic Replacing of Reference Pictures," in Proc. Globcom'96, vol. 3, pp. 1503 - 1508, 1996.

[SRTP] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, March 2004.

[RFC2032] Turletti, T. and C. Huitema, "RTP Payload Format for H.261 Video Streams", RFC 2032, October 1996.

[SAVPF] J. Ott, E. Carrara, "Extended Secure RTP Profile for RTCP-based Feedback (RTP/SAVPF)," draft-ietf-avt-profile-savpf-02.txt, July, 2005.

[RFC3525] Groves, C., Pantaleo, M., Anderson, T., and T. Taylor, "Gateway Control Protocol Version 1", RFC 3525, June 2003.

[VBCM] ITU-T Rec. H.271, "Video Bach Channel Messages", pre-published, June 2006

Any 3GPP document can be downloaded from the 3GPP web server, "http://www.3gpp.org/", see specifications.

11. Authors' Addresses

Stephan Wenger
Nokia Corporation
P.O. Box 100
FIN-33721 Tampere
FINLAND

Phone: +358-50-486-0637
EMail: stewe@stewe.org

Umesh Chandra
Nokia Research Center
6000 Connection Drive
Irving, Texas 75063
USA

Phone: +1-972-894-6017
Email: Umesh.Chandra@nokia.com

Magnus Westerlund
Ericsson Research
Ericsson AB
SE-164 80 Stockholm, SWEDEN

Phone: +46 8 7190000
EMail: magnus.westerlund@ericsson.com

Bo Burman
Ericsson Research
Ericsson AB
SE-164 80 Stockholm, SWEDEN

Phone: +46 8 7190000
EMail: bo.burman@ericsson.com

12. List of Changes relative to previous drafts

The following changes since draft-wenger-avt-avpf-ccm-01 have been made:

- The topologies have been rewritten and clarified.
- The TMMBR mechanism has been completely revised to use notification and suppress messages in deployments with large common SSRC spaces.

The following changes since draft-wenger-avt-avpf-ccm-02 have been made:

- Update of section 4.2.2.1 (TMMBN) as per discussions between Harikishan Desineni and Magnus Westerlund on the AVT list around Feb 21, 2006
- Section 2.3.4 clarified as per email exchange between Colin Perkins and Magnus Westerlund around Feb 24
- Section 3.5.2 and other occurrences throughout the draft, Temporal/Spatial Acknowledgement renamed to Temporal/Spatial Annoucement

Changes relative to draft-wenger-avt-avpf-ccm-03

- Moved "topologies" out to another draft
- Editorial improvements
- Added new code point VBCM for H.271 Video back channel messages. Several sections - 3,4 and 7 were modified for this new CCM message.
- Removed Basso use case referring to forward Freeze command, added justification.

Full Copyright Statement

Intellectual Property Statement

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.

RFC Editor Considerations

The RFC editor is requested to replace all occurrences of XXXX with the RFC number this document receives.