

Payload Working Group	P. Westin
Internet-Draft	H.F. Lundin
Intended status: Standards Track	M. Glover
Expires: November 03, 2011	J. Uberti
	F. Galligan
	Google
	May 02, 2011

RTP Payload Format for VP8 Video
draft-ietf-payload-vp8-00

[Abstract](#)

This memo describes an RTP payload format for the VP8 video codec. The payload format has wide applicability, as it supports applications from low bit-rate peer-to-peer usage, to high bit-rate video conferences.

[Status of this Memo](#)

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on November 03, 2011.

[Copyright Notice](#)

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

[1. Introduction](#)

This memo describes an RTP payload specification applicable to the transmission of video streams encoded using the VP8 video codec [\[I-](#)

[D.bankoski-vp8-bitstream](#). The format described in this document can be used both in peer-to-peer and video conferencing applications.

The VP8 codec uses three different reference frames for interframe prediction: the previous frame, the golden frame, and the altref frame. The payload specification in this memo has elements that enable advanced use of the reference frames, e.g., for improved loss robustness.

Another property of the VP8 codec is that it applies data partitioning to the encoded data. Thus, an encoded VP8 frame can be divided into two or more partitions, as described in "VP8 Data Format and Decoding Guide" [\[I-D.bankoski-vp8-bitstream\]](#). The first partition (prediction or mode) contains prediction mode parameters and motion vectors for all macroblocks. The remaining partitions all contain the transform coefficients for the residuals. The first partition is decodable without the remaining residual partitions. The subsequent partitions may be useful even if some part of the frame is lost. This memo allows the partitions to be sent in the same RTP packet. Nevertheless, it may be beneficial for decoder error-concealment to use separate packets for the two partition types, even though it is not mandatory according to this specification.

The format specification is described in [Section 4](#). In [Section 5](#), a method to acknowledge receipt of reference frames using RTCP techniques is described.

The payload partitioning and the acknowledging method both serve as motivation for two of the fields included in the payload format: the "1st partition size" and "PictureID" fields.

[2. Conventions, Definitions and Acronyms](#)

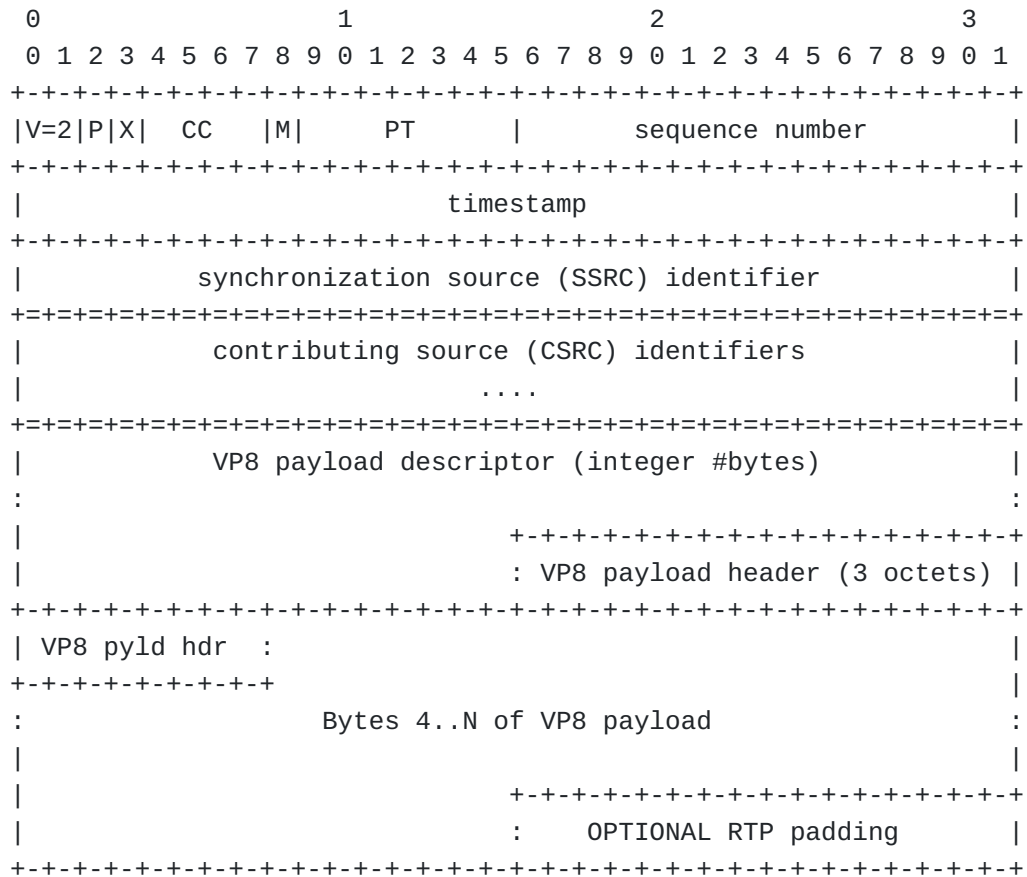
The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [\[RFC2119\]](#).

[3. Media Format Background](#)

VP8 is based on decomposition of frames into square sub-blocks of pixels, prediction of such sub-blocks using previously constructed blocks, and adjustment of such predictions (as well as synthesis of unpredicted blocks) using a discrete cosine transform (hereafter abbreviated as DCT). In one special case, however, VP8 uses a "Walsh-Hadamard" (hereafter abbreviated as WHT) transform instead of a DCT. An encoded VP8 frame is divided into two or more partitions, as described in [\[I-D.bankoski-vp8-bitstream\]](#). The first partition (prediction or mode) contains prediction mode parameters and motion vectors for all macroblocks. The remaining partitions all contain the quantized DCT/WHT coefficients for the residuals.

[4. Payload Format](#)

The general RTP payload format for VP8 is depicted below.



The VP8 payload descriptor and VP8 payload header will be described in the sequel. OPTIONAL RTP padding MUST NOT be included unless the P bit is set.

Marker bit: The marker bit indicates the last packet part of a frame. This enables a decoder to finish decoding the picture, where it otherwise may need to wait for the next packet to explicitly know that the frame is complete.

Timestamp: The RTP timestamp indicates the time when the frame was sampled at a clock rate of 90 kHz.

Sequence number: The sequence numbers are monotonically increasing and set as packets are sent.

The remaining RTP header fields are used as specified in [\[RFC3550\]](#).

4.1. VP8 Payload Descriptor

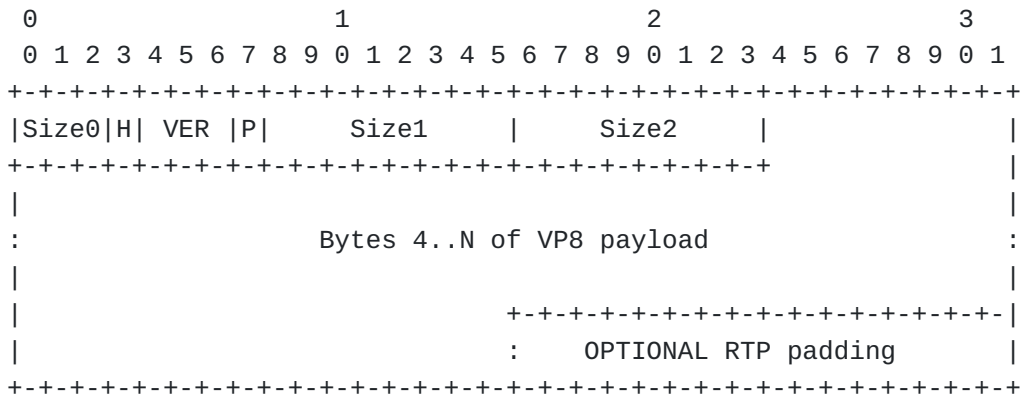
The first bytes after the RTP header are the VP8 payload descriptor, with the following structure.

PictureID:

8 or 16 bits. This is a running index of the frames. The field is present only if the I bit is equal to one. The most significant bit of the first octet is an extension flag. The 7 following bits carry (parts of) the PictureID. If the extension flag is one, the PictureID continues in the next octet forming a 15 bit index, where the 8 bits in the second octet are the least significant bits of the PictureID. If the extension flag is zero, there is no extension, and the PictureID is the 7 remaining bits of the first (and only) octet. The sender may choose 7 or 15 bits index. The PictureID SHOULD start on a random number, and MUST wrap after reaching the maximum ID.

4.2. VP8 Payload Header

The first three octets of an encoded VP8 frame are referred to as an "uncompressed data chunk" in [\[I-D.bankoski-vp8-bitstream\]](#), and co-serve as payload header in this RTP format. The codec bitstream format specifies two different variants of the uncompressed data chunk: a 3 octet version for interframes and a 10 octet version for key frames. The first 3 octets are common to both variants. In the case of a key frame the remaining 7 octets are considered to be part of the remaining payload in this RTP format. Note that the header is present only in packets which have the B bit equal to one in the payload descriptor. Subsequent packets for the same frame do not carry the payload header.



H: Show frame bit as defined in [\[I-D.bankoski-vp8-bitstream\]](#).

VER: A version number as defined in [\[I-D.bankoski-vp8-bitstream\]](#).

P: Inverse key frame flag. When set to 0 the current frame is a key frame. When set to 1 the current frame is an interframe. Defined in [\[I-D.bankoski-vp8-bitstream\]](#)

SizeN: The size of the first partition size in bytes is calculated from the 19 bits in Size0, Size1, and Size2 as 1stPartitionSize = Size0 + 8 * Size1 + 2048 * Size2. [\[I-D.bankoski-vp8-bitstream\]](#).

4.3. Aggregated and Fragmented Payloads

An encoded VP8 frame can be divided into two or more partitions, as described in [Section 1](#). The fragmentation information described in [Section 4.1](#) MUST be used to signal if any fragmentation is applied. Aggregation of encoded partitions is done without explicit signaling. Partitions MUST be aggregated in decoding order. Two fragments from different partitions MAY be aggregated into the same packet. An aggregation MUST have exactly one payload descriptor. Aggregated partitions MUST represent parts of one and the same video frame. Consequently, an aggregated packet will have one or no payload header, depending on whether the aggregate contains the first partition of a frame or not, respectively. Note that the length of the first partition can always be obtained from the first partition size parameter in the VP8 payload header.

The VP8 bitstream format [\[I-D.bankoski-vp8-bitstream\]](#) specifies that if multiple DCT partitions are produced, the location of each partition start is found at the end of the first (prediction/mode) partition. In this RTP payload specification, the location offsets are considered to be part of the first partition.

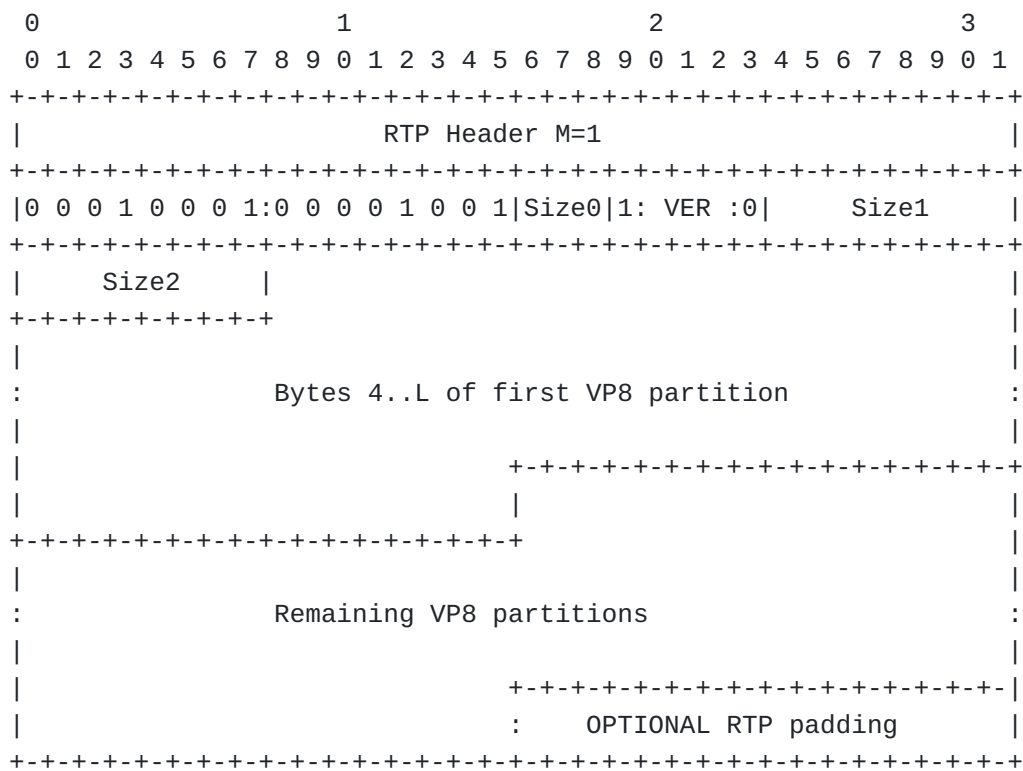
It is OPTIONAL for a packetizer implementing this RTP specification to pay attention to the partition boundaries within an encoded frame. If fragmentation of a frame is done without considering the partition boundaries, the FI field of the payload descriptors MUST be set as follows. The first packet of a frame MUST set FI to 01; the last packet of a frame MUST set FI to 10; all other packets MUST set FI to 11. If the frame is not fragmented over multiple RTP packets, the FI field MUST be set to 00.

4.4. Examples of VP8 RTP Stream

A few examples of how the VP8 RTP payload can be used are included below.

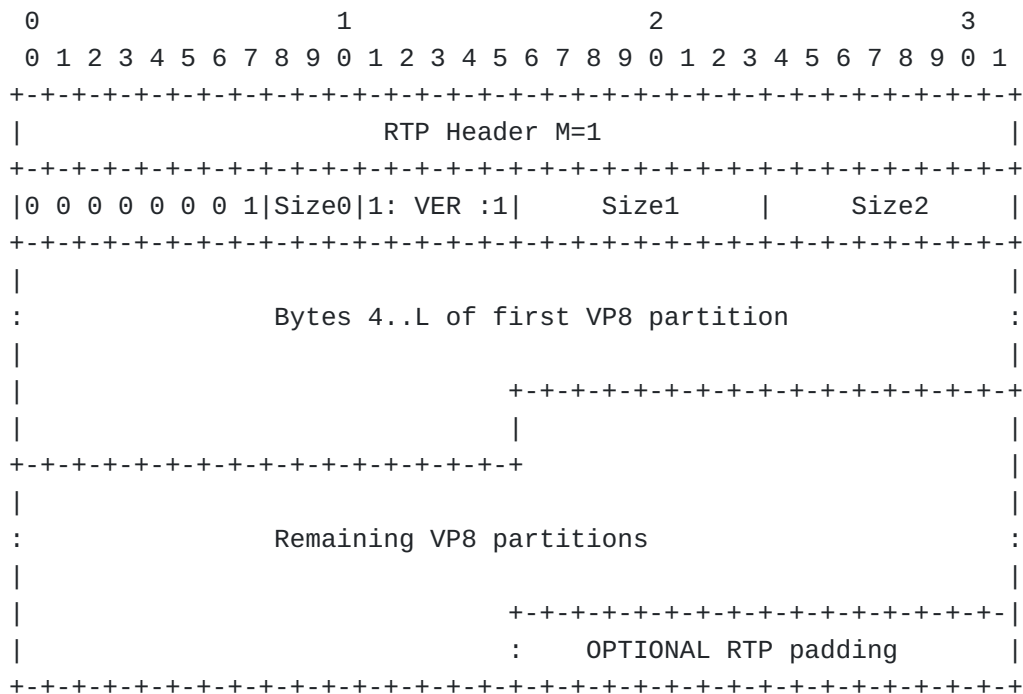
4.4.1. Key frame in a single RTP packet

Marker bit = 1. I = 1. B = 1. PictureID = 17 = 0001001 binary. P = 0.



4.4.2. VP8 interframe in a single RTP packet; no PictureID

Marker bit = 1. I = 0. B = 1. P = 1.



4.4.3. VP8 partitions in separate RTP packets

First RTP packet; marker bit = 0. I = 1. B = 1. PictureID = 17.

```

      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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                RTP Header M=0                                |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|0 0 0 1 0 0 0 1:0 0 0 0 1 0 0 1|Size0|1: VER :1|      Size1      |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|      Size2      |                                                    |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                                    |
:              Bytes 4..L of first VP8 partition              :
|                                                    |
|                                                    |
|                                                    |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

Second RTP packet; marker bit = 1. B = 0.

```

      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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                RTP Header M=1                                |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|0 0 0 1 0 0 0 0:0 0 0 0 1 0 0 1|                                                    |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                                    |
:              Remaining VP8 partitions              :
|                                                    |
|                                                    |
|              +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|              :   OPTIONAL RTP padding              |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

[4.4.4.](#) VP8 frame fragmented across RTP packets

First RTP packet; marker bit = 0. I = 1. FI = 00. B = 1.


```

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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     RTP Header M=0                                     |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|0 0 0 1 0 0 0 1:0 0 0 0 1 0 0 1|Size0|1: VER :1|          Size1          |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|      Size2      |                                                              |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                                              |
:              Bytes 4..L of first VP8 partition              :
|                                                              |
|                                                              |
|                                                              |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

Second RTP packet; marker bit = 0. FI = 01. B = 0.

```

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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     RTP Header M=0                                     |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|0 0 0 1 0 0 1 0:0 0 0 0 1 0 0 1|                                                              |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                                              |
:              First fragment of second VP8 partition              :
|                                                              |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

Third RTP packet; marker bit = 0. FI = 11. B = 0.

```

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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     RTP Header M=0                                     |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|0 0 0 1 0 1 1 0:0 0 0 0 1 0 0 1|                                                              |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                                              |
:              Middle fragment of second VP8 partition              :
|                                                              |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

Last RTP packet; marker bit = 1. FI = 10. B = 0.

```

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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     RTP Header M=1                                     |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|0 0 0 1 0 1 0 0:0 0 0 0 1 0 0 1|                                     |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     |                                     |
:                               Last fragment of second VP8 partition                               :
|                                     |                                     |
|                                     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     :   OPTIONAL RTP padding   :                                     |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

4.4.5. VP8 frame with long PictureID

PictureID = 4711 = 001001001100111 binary (first 7 bits: 0010010, last 8 bits: 01100111).

```

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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     RTP Header M=1                                     |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|0 0 0 1 0 0 0 1:1 0 0 1 0 0 1 0 0 1 1 0 0 1 1 1|Size0|1: VER :1|
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|      Size1      |      Size2      |                                     |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     |                                     |
:                               Bytes 4..N of first VP8 frame                               :
|                                     |                                     |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

5. Using VP8 with RPSI and SLI Feedback

The VP8 payload descriptor defined in [Section 4.1](#) above contains an optional PictureID parameter. This parameter is included mainly to enable use of reference picture selection index (RPSI) and slice loss indication (SLI), both defined in [\[RFC4585\]](#).

5.1. RPSI

The reference picture selection index is a payload-specific feedback message defined within the RTCP-based feedback format. The RPSI message is generated by a receiver and can be used in two ways. Either it can signal a preferred reference picture when a loss has been detected by the decoder -- preferably then a reference that the decoder knows is perfect -- or, it can be used as positive feedback information to

5.2. SLI

[illegible]

5.3. Example

The use of RPSI and SLI is best illustrated in an example. In this example, the encoder may not update the altref frame until the last sent golden frame has been acknowledged with an RPSI message. If an update is not received within some time, a new golden frame update is sent instead. Once the new golden frame is established and acknowledge, the same rule applies when updating the altref frame.

Event	Sender	Receiver	Established reference
1000	Send golden frame PictureID = 0		
		Receive and decode golden frame	
1001		Send RPSI(0)	
1002	Receive RPSI(0)		golden
...	(sending regular frames)		
1100	Send altref frame PictureID = 100		
		Altref corrupted or lost	golden
1101		Send SLI(100)	golden
1102	Receive SLI(100)		
1103	Send frame with reference to golden		
		Receive and decode frame (decoder state restored)	golden
...	(sending regular frames)		
1200	Send altref frame PictureID = 200		
		Receive and decode altref frame	golden
1201		Send RPSI(200)	
1202	Receive RPSI(200)		altref
...	(sending regular frames)		
1300	Send golden frame PictureID = 300		
		Receive and decode golden frame	altref
1301		Send RPSI(300)	altref
1302	RPSI lost		
1400	Send golden frame PictureID = 400		
		Receive and decode golden frame	altref

Event	Sender	Receiver	Established reference
1401		Send RPSI(400)	
1402	Receive RPSI(400)		golden

Exemple signaling between sender and receiver

Note that the scheme is robust to loss of the feedback messages. If the RPSI is lost, the sender will try to update the golden (or altref) again after a while, without releasing the established reference. Also, if an SLI is lost, the receiver can keep sending SLI messages at any interval, as long as the picture is corrupted.

6. Payload Format Parameters

This section specifies the parameters that MAY be used to select optional features of the payload format and certain features of the bitstream.

6.1. Restrictions on usage

This media type depends on RTP framing, and hence is only defined for transfer via RTP [RFC3550] [\[RFC3550\]](#). Transport within other framing protocols is not defined at this time.

6.2. Media Type Registration

This registration is done using the template defined in [\[RFC4288\]](#) and following [\[RFC4855\]](#).

Type name: video

Subtype name: VP8

Required parameters: none

Optional parameters: none

Encoding considerations:

This media type is framed and contains binary data; see Section 4.8 of [\[RFC4288\]](#).

Security considerations: See [Section 7](#) of RFC xxxx.

[RFC Editor: Upon publication as an RFC, please replace "XXXX" with the number assigned to this document and remove this note.]

Interoperability considerations: None.

Published specification: VP8 bitstream format [\[I-D.bankoski-vp8-bitstream\]](#) and RFC XXXX.

[RFC Editor: Upon publication as an RFC, please replace "XXXX" with

the number assigned to this document and remove this note.]

Applications which use this media type:

For example: Video over IP,
video conferencing.

Additional information: None.

Person & email address to contact for further information:

Patrik
Westin, patrik.westin@gmail.com

Intended usage: COMMON

Restrictions on usage:

This media type depends on RTP framing, and
hence is only defined for transfer via RTP [\[RFC3550\]](#).

Author: Patrik Westin, patrik.westin@gmail.com

Change controller:

IETF Payload Working Group delegated from the IESG.

[6.3.](#) SDP Parameters

The receiver MUST ignore any parameter unspecified in this memo.

[6.3.1.](#) Mapping of MIME Parameters to SDP

The MIME media type video/VP8 string is mapped to fields in the Session Description Protocol (SDP) [\[RFC2327\]](#) as follows:

*The media name in the "m=" line of SDP MUST be video.

*The encoding name in the "a=rtpmap" line of SDP MUST be VP8 (the MIME subtype).

*The clock rate in the "a=rtpmap" line MUST be 90000.

*The OPTIONAL parameter "version", if included, MUST be in the a=fmtp SDP field. This parameter matches the VP8 bitstream version.

[6.4.](#) Example

An example of media representation in SDP is as follows:

```
m=video 49170 RTP/AVPF 98
a=rtpmap:98 VP8/90000
a=fmtp:98 version=0
```

7. Security Considerations

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification [\[RFC3550\]](#), and in any applicable RTP profile. The main security considerations for the RTP packet carrying the RTP payload format defined within this memo are confidentiality, integrity and source authenticity. Confidentiality is achieved by encryption of the RTP payload. Integrity of the RTP packets through suitable cryptographic integrity protection mechanism. Cryptographic system may also allow the authentication of the source of the payload. A suitable security mechanism for this RTP payload format should provide confidentiality, integrity protection and at least source authentication capable of determining if an RTP packet is from a member of the RTP session or not. Note that the appropriate mechanism to provide security to RTP and payloads following this memo may vary. It is dependent on the application, the transport, and the signaling protocol employed. Therefore a single mechanism is not sufficient, although if suitable the usage of SRTP [\[RFC3711\]](#) is recommended. This RTP payload format and its media decoder do not exhibit any significant non-uniformity in the receiver-side computational complexity for packet processing, and thus are unlikely to pose a denial-of-service threat due to the receipt of pathological data. Nor does the RTP payload format contain any active content.

8. Congestion Control

Congestion control for RTP SHALL be used in accordance with RFC 3550 [\[RFC3550\]](#), and with any applicable RTP profile; e.g., RFC 3551 [\[RFC3551\]](#). The congestion control mechanism can, in a real-time encoding scenario, adapt the transmission rate by instructing the encoder to encode at a certain target rate. Media aware network elements MAY use the information in the VP8 payload descriptor in [Section 4.1](#) to identify non-reference frames and discard them in order to reduce network congestion.

9. IANA Considerations

The IANA is requested to register the following values:

- Media type registration as described in [Section 6.2](#).

10. References

[I-D.bankoski-vp8-bitstream]	Bankoski, J, Koleszar, J, Quillio, L, Salonen, J, Wilkins, P and Y Xu, " VP8 Data Format and Decoding Guide ", Internet-Draft draft-bankoski-vp8-bitstream-06, August 2011.
[RFC2119]	

	Bradner, S. , " Key words for use in RFCs to Indicate Requirement Levels ", BCP 14, RFC 2119, March 1997.
[RFC4585]	Ott, J., Wenger, S., Sato, N., Burmeister, C. and J. Rey, " Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF) ", RFC 4585, July 2006.
[RFC3550]	Schulzrinne, H., Casner, S., Frederick, R. and V. Jacobson, " RTP: A Transport Protocol for Real-Time Applications ", STD 64, RFC 3550, July 2003.
[RFC3711]	Baughner, M., McGrew, D., Naslund, M., Carrara, E. and K. Norrman, " The Secure Real-time Transport Protocol (SRTP) ", RFC 3711, March 2004.
[RFC2327]	Handley, M. and V. Jacobson , " SDP: Session Description Protocol ", RFC 2327, April 1998.
[RFC4288]	Freed, N. and J. Klensin, " Media Type Specifications and Registration Procedures ", BCP 13, RFC 4288, December 2005.
[RFC4855]	Casner, S., " Media Type Registration of RTP Payload Formats ", RFC 4855, February 2007.
[RFC3551]	Schulzrinne, H. and S. Casner, " RTP Profile for Audio and Video Conferences with Minimal Control ", STD 65, RFC 3551, July 2003.

[Authors' Addresses](#)

Patrik Westin Westin Google, Inc. Kungsbron 2 Stockholm, 11122 Sweden EMail: patrik.westin@gmail.com

Henrik F Lundin Lundin Google, Inc. Kungsbron 2 Stockholm, 11122 Sweden EMail: hlundin@google.com

Michael Glover Glover Google, Inc.

Justin Uberti Uberti Google, Inc.

Frank Galligan Galligan Google, Inc.

[Table of Contents](#)

- *1. [Introduction](#)
- *2. [Conventions, Definitions and Acronyms](#)
- *3. [Media Format Background](#)
- *4. [Payload Format](#)
 - *4.1. [VP8 Payload Descriptor](#)

- *4.2. [VP8 Payload Header](#)
- *4.3. [Aggregated and Fragmented Payloads](#)
- *4.4. [Examples of VP8 RTP Stream](#)
 - *4.4.1. [Key frame in a single RTP packet](#)
 - *4.4.2. [VP8 interframe in a single RTP packet; no PictureID](#)
 - *4.4.3. [VP8 partitions in separate RTP packets](#)
 - *4.4.4. [VP8 frame fragmented across RTP packets](#)
 - *4.4.5. [VP8 frame with long PictureID](#)
- *5. [Using VP8 with RPSI and SLI Feedback](#)
 - *5.1. [RPSI](#)
 - *5.2. [SLI](#)
 - *5.3. [Example](#)
- *6. [Payload Format Parameters](#)
 - *6.1. [Restrictions on usage](#)
 - *6.2. [Media Type Registration](#)
 - *6.3. [SDP Parameters](#)
 - *6.3.1. [Mapping of MIME Parameters to SDP](#)
 - *6.4. [Example](#)
- *7. [Security Considerations](#)
- *8. [Congestion Control](#)
- *9. [IANA Considerations](#)
- *10. [References](#)
- *[Authors' Addresses](#)