RTP Payload Format for Video Codec 1 (VC-1)

IPR Notice

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Status of this Memo

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

Distribution of this document is unlimited.

Copyright Notice

Copyright (C) The Internet Society (2005).

Abstract

This memo specifies an RTP payload format for encapsulating Video Codec 1 (VC-1) compressed bit streams, as defined by the proposed Society of Motion Picture and Television Engineers (SMPTE) standard, SMPTE 421M. SMPTE is the main standardizing body in the motion imaging industry and the proposed SMPTE 421M standard defines a compressed video bit stream format and decoding process for television.
1. Introduction

The bit stream syntax for compressed video in Video Codec 1 (VC-1) format is defined by SMPTE 421M [1]. SMPTE 421M also specifies constraints that must be met by VC-1 conformant bit streams, and it specifies the complete process required to decode the bit stream. However, it does not specify the VC-1 compression algorithm, thus allowing for different ways to implement a VC-1 encoder.

The VC-1 bit stream syntax has three profiles. Each profile has specific bit stream syntax elements and algorithms associated with it. Depending on the application in which VC-1 is used, some profiles may be more suitable than others. For example, the Simple
profile is designed for low bit rate Internet streaming and for playback on devices that can only handle low complexity decoding. The Advanced profile is designed for broadcast applications, such as digital TV, HD DVD or HDTV. The Advanced profile is the only VC-1 profile that supports interlaced video frames.

RTP Payload Format for Video Codec 1 (VC-1) July 2005

Section 2 defines the abbreviations used in this document. Section 3 provides a more detailed overview of VC-1. Sections 4 and 5 define the RTP payload format for VC-1, and section 6 defines the MIME and SDP parameters for VC-1. See section 7 for security considerations.

1.1 Conventions used in this document

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 BCP 14, RFC 2119 [2].

2. Definitions and abbreviations

This document uses the definitions in SMPTE 421M [1]. For convenience, the following terms from SMPTE 421M are restated here:

B-picture: A picture that is coded using motion compensated prediction from past and/or future reference fields or frames. A B-picture cannot be used for predicting any other picture.

Bit-stream data unit (BDU): A unit of the compressed data which may be parsed (i.e., syntax decoded) independently of other information at the same hierarchical level. A BDU can be, for example, a sequence layer header, an entry-point segment header, a frame, or a slice.

Encapsulated BDU (EBDU): A BDU which has been encapsulated using the encapsulation mechanism described in Annex E of SMPTE 421M [1], to prevent emulation of the start code prefix in the bit stream.

Entry-point: A point in the bit stream that offers random access.

frame: A frame contains lines of spatial information of a video signal. For progressive video, these lines contain samples starting from one time instant and continuing through successive lines to the
bottom of the frame. For interlaced video, a frame consists of two fields, a top field and a bottom field. One of these fields will commence one field period later than the other.

interlace: The property of frames where alternating lines of the frame represent different instances in time. In an interlaced frame, one of the fields is meant to be displayed first.

I-picture: A picture coded using information only from itself.

level: A defined set of constraints on the values which may be taken by the parameters (such as bit rate and buffer size) within a particular profile. A profile may contain one or more levels.

P-picture: A picture that is coded using motion compensated prediction from past reference fields or frames.

picture: For progressive video, a picture is identical to a frame, while for interlaced video, a picture may refer to a frame, or the top field or the bottom field of the frame depending on the context.

profile: A defined subset of the syntax of VC-1, with a specific set of coding tools, algorithms, and syntax associated with it. There are three VC-1 profiles: Simple, Main and Advanced.

progressive: The property of frames where all the samples of the frame represent the same instance in time.

random access: A random access point in the bit stream is defined by the following guarantee: If decoding begins at this point, all frames needed for display after this point will have no decoding dependency on any data preceding this point, and are also present in the decoding sequence after this point. A random access point is also called an entry-point.

sequence: A coded representation of a series of one or more pictures. In VC-1 Advanced profile, a sequence consists of a series of one or more entry-point segments, where each entry-point segment consists of a series of one or more pictures, and where the first picture in each entry-point segment provides random access. In VC-1 Simple and Main profiles, the first picture in each sequence is an I-picture.
slice: A consecutive series of macroblock rows in a picture, which are encoded as a single unit.

start codes (SC): 32-bit codes embedded in that coded bit stream that are unique, and identify the beginning of a BDU. Start codes consist of a unique three-byte Start Code Prefix (SCP), and a one-byte Start Code Suffix (SCS).

3. Overview of VC-1

The VC-1 bit stream syntax consists of three profiles: Simple, Main, and Advanced. The Simple and Main profiles are designed for relatively low bit rate applications. For example, the maximum bit rate supported by the Simple profile is 384 kbps. To help achieve high compression efficiency, certain features such as non-square pixels and support for interlaced pictures, are only included in the Advanced profile.

The maximum bit rate supported by the Advanced profile is 135 Mbps, making it suitable for nearly lossless encoding of HDTV signals.

RTP Payload Format for Video Codec 1 (VC-1) July 2005

Only the Advanced profile supports carrying user-data (meta-data) in-band with the compressed bit stream. The user-data can be used for closed captioning support, for example.

Of the three profiles, only the Advanced profile allows codec configuration parameters, such as the picture aspect ratio, to be changed through in-band signaling in the compressed bit stream.

For each of the profiles, a certain number of "levels" have been defined. Unlike a "profile", which implies a certain set of features or syntax elements, a "level" is a set of constraints on the values of parameters in a profile, such as the bit rate or buffer size. The VC-1 Simple profile has two levels, the Main profile has three, and the Advanced profile has five levels. See Annex D of SMPTE 421M [1] for a detailed list of the profiles and levels.

3.1 VC-1 bit stream layering model

The VC-1 bit stream is defined as a hierarchy of layers. This is conceptually similar to the notion of a protocol stack of networking protocols. The outermost layer is called the sequence layer. The
other layers are entry-point, picture, slice, macroblock and block.

In the Simple and Main profiles, a sequence in the sequence layer consists of a series of one or more coded pictures. In the Advanced profile, a sequence consists of one or more entry-point segments, where each entry-point segment consists of a series of one or more pictures, and where the first picture in each entry-point segment provides random access. A picture is decomposed into macroblocks. A slice comprises one or more contiguous rows of macroblocks.

The entry-point and slice layers are only present in the Advanced profile. In the Advanced profile, the start of each entry-point layer segment indicates a random access point. In the Simple and Main profiles each I-picture is a random access point.

Each picture can be coded as an I-picture, P-picture, skipped picture, BI-picture, or as a B-picture. These terms are defined in section 2 of this document and in section 4.12 of SMPTE 421M [1].

3.2 Bit-stream Data Units in Advanced profile

In the Advanced profile only, each picture and slice is byte-aligned and is considered a Bit-stream Data Unit (BDU). A BDU is defined as a unit that can be parsed (i.e., syntax decoded) independently of other information in the same layer.

The beginning of a BDU is signaled by an identifier called Start Code (SC). Sequence layer headers and entry-point segment headers are also BDUs and thus can be easily identified by their Start Codes. See Annex E of SMPTE 421M [1] for a complete list of Start Codes. Note that blocks and macroblocks are not BDUs and thus do not have a Start Code and are not necessarily byte-aligned.

The Start Code consists of four bytes. The first three bytes are 0x00, 0x00 and 0x01. The fourth byte is called the Start Code Suffix (SCS) and it is used to indicate the type of BDU that follows the Start Code. For example, the SCS of a sequence layer header (0x0F) is different from the SCS of an entry-point segment header (0x0E). The Start Code is always byte-aligned and is transmitted in network byte order.

To prevent accidental emulation of the Start Code in the coded bit
stream, SMPTE 421M defines an encapsulation mechanism that uses byte stuffing. A BDU which has been encapsulated by this mechanism is referred to as an Encapsulated BDU, or EBDU.

3.3 Decoder initialization parameters

In the VC-1 Advanced profile, the sequence layer header contains parameters that are necessary to initialize the VC-1 decoder. These parameters apply to all entry-point segments until the next occurrence of a sequence layer header in the coded bit stream.

The parameters in the sequence layer header include, among other things, the Advanced profile level, the dimensions of the coded pictures, the aspect ratio, interlace information, the frame rate and up to 31 leaky bucket parameter sets for the Hypothetical Reference Decoder (HRD).

Section 6.1 of SMPTE 421M [1] provides the formal specification of the sequence layer header.

Each leaky bucket parameter set for the HRD specifies a peak transmission bit rate and a decoder buffer capacity. The coded bit stream is restricted by these parameters. The HRD model does not mandate buffering by the decoder. Its purpose is to limit the encoder's bit rate fluctuations according to a basic buffering model, so that the resources necessary to decode the bit stream are predictable. The HRD has a constant-delay mode and a variable-delay mode. The constant-delay mode is appropriate for broadcast and streaming applications, while the variable-delay mode is designed for video conferencing applications.

Annex C of SMPTE 421M [1] specifies the usage of the hypothetical reference decoder for VC-1 bit streams. A general description of the theory of the HRD can be found in [6].

The concept of an entry-point layer applies only to the VC-1 Advanced profile. The presence of an entry-point segment header indicates a random access point within the bit stream. The entry-point segment header specifies current buffer fullness values for the leaky buckets in the HRD. The header also specifies coding control parameters that are in effect until the occurrence of the next entry-point segment header in the bit stream. See Section 6.2 of SMPTE 421M [1] for the
formal specification of the entry-point segment header.

Neither a sequence layer header nor an entry-point segment header is defined for the VC-1 Simple and Main profiles. For these profiles, decoder initialization parameters MUST be conveyed out-of-band from the coded bit stream. Section 4.4 of this document specifies how the parameters are conveyed by this RTP payload format.

3.4 Ordering of frames

Frames are transmitted in the same order in which they are captured, except if the presence of B-pictures has been indicated in the decoder initialization parameters. In the latter case, the frames are reordered by the VC-1 encoder such that the frames that the B-pictures depend on are transmitted first. This is referred to as the coded order of the frames.

When the presence of B-pictures has been indicated, the decoder is required to buffer one picture. When an I-picture or a P-picture is received, the picture is not displayed until the next I- or P-picture is received. However, B-pictures are displayed immediately. These rules are stated in section 5.4 in SMPTE 421M [1].

Figure 1 illustrates the timing relationship between the capture of frames, their coded order, and the display order of the decoded frames. The figure shows that the display of frame P4 is delayed until frame P5 is received, while frames B2 and B3 are displayed immediately.

Capture: | I0 P1 B2 B3 P4 ...
Coded order: | I0 P1 P4 B2 B3 P5 ...
Display order: | I0 P1 B2 B3 P4 ...
| +----------+----------+----------+---------------------+------>
| 0 1 2 3 4 5 6 7 |

Figure 1. Frame reordering when B-pictures are indicated.
4. Encapsulation of VC-1 format bit streams in RTP

4.1 Access Units

Each RTP packet contains an integral number of application data units (ADUs). For VC-1 format bit streams, an ADU is equivalent to one Access Unit (AU), as defined in this section. Figure 2 shows the layout of an RTP packet with multiple AUs.

```
+-----------------+  ..  +-----------------+
| RTP             |  ..  | RTP             |
| AU(1)           |  ..  | AU(1)           |
| AU(2)           |  ..  | AU(2)           |
| AU(n)           |  ..  | AU(n)           |
+-----------------+  ..  +-----------------+
```

Figure 2. RTP packet structure.

Access Units MUST be byte-aligned. Each Access Unit MUST start with the AU header defined in section 5.2, and is followed by a variable length payload.

The AU payload MUST contain data belonging to exactly one VC-1 frame.

The following rules apply to the contents of each AU payload when the VC-1 Advanced profile is used:

- The AU payload MUST contain VC-1 bit stream data in EBDU format (i.e., the bit stream must use the byte-stuffing encapsulation mode defined in Annex E of SMPTE 421M [1].)

- The AU payload MAY contain multiple EBDUs, e.g., a sequence layer header, an entry-point segment header, a frame header and multiple slices and the associated user-data. (However, all slices and their corresponding macroblocks MUST belong to the same video frame.)

- The AU payload MUST start at an EBDU boundary, except when the AU payload contains a fragmented frame, in which case the rules in section 4.2 apply.

If the data in an AU (EBDUs in the case of Advanced profile and frame in the case of Simple and Main) does not end at an octet boundary, up to 7 zero-valued padding bits MUST be added to achieve octet-alignment.

4.2 Fragmentation of VC-1 frames

Each AU payload SHOULD contain a complete VC-1 frame. However, if this would cause the RTP packet to exceed the MTU size, the frame SHOULD be fragmented into multiple AUs to avoid IP-level
fragmentation. When an AU contains a fragmented frame, this MUST be indicated by setting the FRAG field in the AU header as defined in section 5.3.

AU payloads that do not contain a fragmented frame, or that contain the first fragment of a frame, MUST start at an EBDU boundary if Advanced profile is used. In this case, for Simple and Main profiles, the AU payload MUST begin with the start of a frame.

If Advanced profile is used, AU payloads that contain a fragment of a frame other than the first fragment, SHOULD start at an EBDU boundary, such as at the start of a slice.

However, slices are only defined for the Advanced profile, and are not always used. Blocks and macroblocks are not BDUs (have no Start Code) and are not byte-aligned. Therefore, it may not always be possible to continue a fragmented frame at an EBDU boundary.

If a RTP packet contains an AU with the last fragment of a frame, additional AUs SHOULD NOT be included in the RTP packet.

If the PTS Delta field in the AU header is used, each fragment of a frame MUST have the same presentation time. If the DTS Delta field in the AU header is used, each fragment of a frame MUST have the same decode time.

4.3 Time stamp considerations

Video frames MUST be transmitted in the coded order. Coded order implies that no frames are dependent on subsequent frames, as discussed in section 3.4. The RTP timestamp field MUST be set to the decode time of the video frame contained in the first AU in the RTP packet. The decode time is equivalent to the sampling instant of the frame, except when the codec initialization parameters indicate that the VC-1 bit stream contains B-pictures. When the presence of B-pictures has been indicated, the encoder may reorder frames, as explained in section 3.4 of this document and in section 5.4 of SMPTE 421M [1].

The VC-1 bit stream does not carry any time stamps other than an optional Temporal Frame Reference Counter field, which, if it is present, can be used to calculate the decode time of a frame. However, the RTP sender may have access to different externally
provided time stamps depending on the method used to ingest the VC-1 bit stream. For example, if VC-1 is encapsulated in MPEG-2 Transport Stream, each frame is assigned a presentation time (PTS) and optionally also a decode time (DTS). If a VC-1 bit stream is stored in an ASF file, only the decode time of each video frame is available.

If only presentation time information is available, the RTP sender can approximate the decode time of a frame by its presentation time, after taking frame reordering into account. Frame reordering can be handled by an algorithm similar to the one illustrated in Figure 1 in section 3.4. The algorithm requires buffering of only one frame.

If only decode time information is available, determining the presentation time of a P-frame requires buffering, or looking ahead, to the first frame that does not depend on the P-frame. Using the coded order sequence in Figure 1 as an example, the RTP sender cannot determine presentation time of frame P4 until it has seen frame P5. This would be a more complicated and costly procedure than to estimate a decode time from the presentation time. Hence, this RTP payload format defines that the RTP timestamp field must represent the decode time of the frame.

Knowing if the stream will contain B-pictures helps the decoder allocate resources more efficiently, as the encoder will not reorder any frames. In that case, the buffering of one frame as described in section 3.4 is not necessary. Avoiding this buffer reduces the end-to-end delay, which may be important for interactive applications. For Advanced profile, B-pictures are assumed to be present by default. If the coded bit stream never contains B-pictures, this can be indicated using the "bpic" MIME parameter defined in section 6.1.

For Simple and Main profiles, the presence of B-pictures is indicated by a non-zero value for the MAXBFRAMES field in STRUCT_C decoder initialization parameter. STRUCT_C conveyed in the MIME "config" parameter, which is defined in section 6.1.

4.4 Signaling of MIME format parameters

When this RTP payload format is used with SDP, the decoder initialization parameters described in section 3.3 MUST be signaled in SDP using the MIME parameters specified in section 6.1. Section
6.2 specifies how to map the MIME parameters to SDP.

When the Advanced profile is used, the decoder initialization parameters MAY be changed by inserting a new sequence layer header or an entry-point segment header in the coded bit stream.

Note that the sequence layer header specifies the encoding level, the maximum size of the coded pictures and possibly also the frame rate. Thus, if the sequence layer header changes, the new header supersedes the values of the MIME parameters "level", "width", "height" and "framerate".

To improve robustness against loss of RTP packets, it is RECOMMENDED that if the sequence layer header changes, it should be repeated frequently in the bit stream. Note that any data in the VC-1 bit stream, including the sequence header itself, must be accounted for when computing the leaky bucket parameters for the HRD. (See section 3.3 for a discussion about the HRD.)

The Seq Count field in the Access Unit header is used to track changes to the sequence layer header. A value of 0 is reserved for the case when the most recent sequence layer header of the bit stream is identical to the sequence layer header in the MIME "config" parameter (defined in section 6.1.)

If the RTP sender cannot determine the most recent sequence layer header, or if it is different from the sequence layer header in the MIME "config" parameter, a non-zero value MUST be used for the Seq Count field.

When the RTP sender transmits an AU containing a sequence layer header that is different from the previous sequence layer header, the value of the Seq Count field MUST be incremented. The Seq Count field of all subsequent AU headers MUST be set to this new value until the sequence layer header changes again.

In certain applications, the sequence layer header never changes. This MAY be signaled with the MIME parameter "mode=1" or "mode=3", as appropriate. (See the definition of the "mode" parameter in section 6.1.) If "mode=1" or "mode=3" is signaled and a sequence layer header is present in the coded bit stream, it MUST be identical to
the sequence layer header specified by the MIME "config" parameter.

The entry-point segment header contains information that is needed by the decoder to decode the frames in that segment. This means that in the event of lost RTP packets the decoder may be unable to decode frames until the next entry-point segment header is received. Access Units that contain an entry-point segment header MUST have the RA bit in AU header set to 1. (The RA bit is defined in section 5.3.)

In certain applications, the entry-point segment header never changes. This MUST be signaled with the MIME parameter "mode=2" or "mode=3", as appropriate. In this case, any entry-point segment headers that are present in the bit stream MAY be removed by the RTP sender. If "mode=2" or "mode=3" is signaled and an entry-point segment header is present in the coded bit stream, it MUST be identical to the entry-point segment header specified by the MIME "config" parameter.

5. RTP Payload Format syntax

5.1 RTP header usage

The format of the RTP header is specified in RFC 3550 [3] and is reprinted in Figure 3 for convenience.

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=2|P|X|  CC   |M|     PT      |       sequence number         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           timestamp                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           synchronization source (SSRC) identifier            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            contributing source (CSRC) identifiers             |
|                             ....                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3. RTP header according to RFC 3550
With the exception of the fields listed below, the RTP header fields are used as defined in RFC 3550 and by the RTP profile in use.

**Marker bit (M):** 1 bit

This bit is set to 1 if the RTP packet contains an Access Unit containing a complete VC-1 frame, or the last fragment of a VC-1 frame.

**Payload type (PT):** 7 bits

This document does not assign a RTP payload type for this RTP payload format. The assignment of a payload type has to be performed either through the RTP profile used or in a dynamic way.

**Timestamp:** 32 bits

The RTP timestamp is set to the decode time of the VC-1 frame in the first Access Unit. A 90 kHz clock rate MUST be used.

### 5.2 AU header syntax

The Access Unit header consists of a one-byte AU Control field, and 4 optional fields. The structure of the AU header is illustrated in Figure 4.

```
+----------------------------------+
|AU     | Seq   | PTS   | DTS   | AUP  |
|Control| Count | Delta | Delta | Len  |
+----------------------------------+
```

Figure 4. Structure of AU header.

**AU Control:** 8 bits

The usage of the AU Control field is defined in section 5.3.

**Seq Count:** 8 bits

Sequence Layer Counter. This field is a binary modulo 256 counter. The value of this field, if present, MUST be incremented by 1, each time an AU containing a new sequence layer header is transmitted. The value 0 is reserved for the
case when the RTP sender knows that the current sequence layer header is identical to the sequence layer header in the MIME "config" parameter (defined in section 6.1) and MUST NOT be used for any other purpose.

If this field is not present, a value of 0 MUST be assumed.

**PTS Delta: 32 bits**
Presentation time delta. Specifies the presentation time of the frame as a 2's complement offset (delta) from the timestamp in the RTP header of this RTP packet. The PTS Delta field MUST use the same clock rate as the timestamp field in the RTP header.

**DTS Delta: 32 bits**
Decode time delta. Specifies the decode time of the frame as a 2's complement offset (delta) from the timestamp in the RTP header of this RTP packet. The DTS Delta field MUST use the same clock rate as the timestamp field in the RTP header.
This field SHOULD NOT be included in the first AU header in the RTP packet, because the RTP timestamp field specifies the decode time of the frame in the first AU.

**AUP Len: 16 bits**
Access Unit Payload Length. Specifies the size, in bytes, of the payload of the Access Unit. The field does not include the size of the AU header itself. The field MUST be included in each AU header in an RTP packet, except for the last AU header in the packet.

### 5.3 AU Control field syntax

The structure of the 8-bit AU Control field is shown in Figure 5.

```
+----+----+----+----+----+----+----+----+
|  FRAG   | RA | SC | PT | DT | LP | R  |
+----------------------------------------+
```

**FRAG: 2 bits**

Fragmentation Information. This field indicates if the AU payload contains a complete frame or a fragment of a frame. It MUST be set as follows:
0: The AU payload contains a fragment of a frame other than the first or last fragment.
1: The AU payload contains the first fragment of a frame.
2: The AU payload contains the last fragment of a frame.
3: The AU payload contains a complete frame (not fragmented.)

SC: 1 bit
Sequence Layer Counter present. This bit MUST be set to 1 if the AU header includes the Seq Count field. The bit MUST be 0 for Simple and Main profile bit streams.

RA: 1 bit
Random Access Point indicator. This bit MUST be set to 1 if the AU contains a frame that is a random access point. In the case of Simple and Main profiles, any I-picture is a random access point. In the case of Advanced profile, the first frame after an entry-point segment header is a random access point. Note that if entry-point segment headers are not transmitted at every random access point, this MUST be indicated using the MIME parameter "mode=2" or "mode=3", as appropriate.

PT: 1 bit
PTS Delta Present. This bit MUST be set to 1 if the AU header includes the PTS Delta field.

DT: 1 bit
DTS Delta Present. This bit MUST be set to 1 if the AU header includes the DTS Delta field.

LP: 1 bit
Length Present. This bit MUST be set to 1 if the AU header includes the AUP Len field.

R: 1 bit
Reserved. This bit MUST be set to 0 and MUST be ignored by receivers.
6.1 Media Type Registration

The media subtype for VC-1 is allocated from the standards tree. The top-level media type under which this payload format is registered is 'video'.

The receiver MUST ignore any unrecognized parameter.

Media type: video

Media subtype: vc1

Required parameters:

profile:
The value is a decimal number indicating the VC-1 profile.
The following values are defined:
0: Simple profile.
1: Main profile.
3: Advanced profile.

config:
The value is a hexadecimal representation of an octet string that expresses the decoder initialization parameters. Decoder initialization parameters are mapped onto the hexadecimal octet string in an MSB-first basis. The first bit of the decoder initialization parameters MUST be located at the MSB of the first octet. If the decoder initialization parameters are not multiple of 8 bits, in the last octet up to 7 zero-valued padding bits MUST be added to achieve octet alignment.

For the Simple and Main profiles, the decoder initialization parameters are STRUCT_C, as defined in Annex J of SMPTE 421M [1].

For the Advanced profile, the decoder initialization parameters are a sequence layer header directly followed by an entry-point segment header. The two headers MUST be in EBDU format, meaning that they must include their Start Codes and must use the encapsulation method defined in Annex E of SMPTE 421M [1].

width:
The value is a decimal number specifying the maximum horizontal size of the coded picture in pixels.
Note: When Advanced profile is used, this parameter only applies while the sequence layer header specified in the config parameter is in use.

height:
The value is a decimal number specifying the maximum vertical size of the coded picture in pixels.

Note: When Advanced profile is used, this parameter only applies while the sequence layer header specified in the config parameter is in use.

bitrate:
The value is a decimal number specifying the peak transmission rate of the coded bit stream. The number does not include RTP overhead.

Note: When Advanced profile is used, this parameter only applies while the sequence layer header specified in the config parameter is in use.

buffer:
The value is a decimal number specifying the leaky bucket size, B, in milliseconds, required to contain a stream transmitted at the transmission rate specified by the bitrate parameter. This parameter is defined in the hypothetical reference decoder model for VC-1, in Annex C of SMPTE 421M [1].

Note: When Advanced profile is used, this parameter only applies while the sequence layer header specified in the config parameter is in use.

Optional parameters:

level:
The value is a decimal number specifying the level of the encoding profile.
For Advanced profile, valid values are 0 to 4, which correspond to levels L0 to L4, respectively. For Simple and Main profiles, the following values are defined:
1: Low Level
2: Medium Level
3: High Level (only valid for Main profile)

Note: When Advanced profile is used, this parameter only
applies while the sequence layer header specified in the config parameter is in use.

framerate:
The value is a decimal number specifying the number of frames per second, multiplied by 1000. For example, 29.97 frames per second is represented as 29970.

Note: When Advanced profile is used, this parameter only applies while the sequence layer header specified in the config parameter is in use.

bpic:
This parameter signals if B-pictures may be present when the Advanced profile is used. If this parameter is present, and B-pictures may be present in the coded bit stream, this parameter MUST be equal to 1. If B-pictures will never be present in the coded bit stream, even if the sequence layer header changes, this parameter SHOULD be present and its value SHOULD be equal to 0.

If this parameter is not specified, a value of 1 MUST be assumed.

mode:
The value is a decimal number specifying the use of the sequence layer header and the entry-point segment header. This parameter is only used for Advanced profile. The following values are defined:
0: Both the sequence layer header and the entry-point segment header may change, and changed headers will be included in the RTP packets.
1: The sequence layer header specified in the config parameter never changes.
2: The entry-point segment header specified in the config parameter never changes. Entry-point segment headers MAY not be included in the RTP packets. Each Access Unit that has the RA bit set to 1 contains a random access point even if an entry-point segment header is not included in the RTP packet.
3: Modes 1 and 2 combined.

If the mode parameter is not specified, a value of 0 MUST be
assumed. The mode parameter SHOULD be specified if any of the modes 1-3 apply to the VC-1 bit stream.

Encoding considerations:
This media type is framed and contains binary data. This media type depends on RTP framing, and hence is only defined for transfer via RTP [3].

Security considerations:
See Section 7 of this document.

Interoperability considerations:
None.

Published specification:
This payload format specification.

Applications which use this media type:
Multimedia streaming and conferencing tools.

Additional Information:
None.

Person & email address to contact for further information:
Anders Klemets <anderskl@microsoft.com>
IETF AVT working group.

Intended Usage:
COMMON

Restrictions on usage:
This media type depends on RTP framing, and hence is only defined for transfer via RTP [3].

Authors:
Anders Klemets

Change controller:
IETF Audio/Video Transport Working Group delegated from the IESG.

6.2 Mapping of MIME parameters to SDP
The information carried in the media type specification has a specific mapping to fields in the Session Description Protocol (SDP) [4]. If SDP is used to specify sessions using this payload format, the mapping is done as follows:

- The media name in the "m=" line of SDP MUST be video (the media type).
- The encoding name in the "a=rtpmap" line of SDP MUST be vc1 (the media subtype).
- The clock rate in the "a=rtpmap" line MUST be 90000.
- The REQUIRED parameters "profile", "config", "width", "height", "bitrate" and "buffer" MUST be included in the "a=fmtp" line of SDP.

These parameters are expressed as a MIME media type string, in the form of a semicolon separated list of parameter=value pairs.

- The OPTIONAL parameters "level", "framerate", "bpic" and "mode", when present, MUST be included in the "a=fmtp" line of SDP. These parameters are expressed as a MIME media type string, in the form of a semicolon separated list of parameter=value pairs:

  a=fmtp:<dynamic payload type> <parameter name>=<value>[,<value>][; <parameter name>=<value>]

- Any unknown parameters to the device that uses the SDP MUST be ignored. For example, parameters defined in later specifications MAY be copied into the SDP and MUST be ignored by receivers that do not understand them.

An example of media representation in SDP is as follows (Simple profile, Medium level):

m=video 49170 RTP/AVP 98
a=rtpmap:98 VC1/90000
a=fmtp:98 profile=0;level=2;width=352;height=288;framerate=15000;
bitrate=384000;buffer=2000;config=4e291800

7. Security Considerations
RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification [4], and in any appropriate RTP profile. This implies that confidentiality of the media streams is achieved by encryption; for example, through the application of SRTP [5].

A potential denial-of-service threat exists for data encodings using compression techniques that have non-uniform receiver-end computational load. The attacker can inject pathological RTP packets into the stream that are complex to decode and that cause the receiver to be overloaded. VC-1 is particularly vulnerable to such attacks, because it is possible for an attacker to generate RTP packets containing frames that affect the decoding process of many future frames. Therefore, the usage of data origin authentication and data integrity protection of at least the RTP packet is RECOMMENDED; for example, with SRTP [5].

Note that the appropriate mechanism to ensure confidentiality and integrity of RTP packets and their payloads is very dependent on the application and on the transport and signaling protocols employed. Thus, although SRTP is given as an example above, other possible choices exist.

8. IANA Considerations

IANA is requested to register the media subtype name "vc1" for the media type "video" as specified in section 6.1 of this document.

9. References

9.1 Normative references

9.2 Informative references


Author's Addresses

Anders Klemets
Microsoft Corp.
1 Microsoft Way
Redmond, WA 98052
USA
Email: anderskl@microsoft.com

Acknowledgements

Thanks to Shankar Regunathan for pointing out errors in the initial draft of this document.

IPR Notices

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.
The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Full Copyright Statement

Copyright (C) The Internet Society (2005).

This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.