Network Working Group
Internet Draft
Intended status: Informational

A. Filippov
Huawei Technologies
A. Norkin
Netflix
J.R. Alvarez
Huawei Technologies

Expires: October 27, 2017

April 28, 2017

<Video Codec Requirements and Evaluation Methodology>

draft-ietf-netvc-requirements-06.txt

Status of this Memo

Copyright Notice

Abstract

This document provides requirements for a video codec designed mainly for use over the Internet. In addition, this document describes an evaluation methodology needed for measuring the compression efficiency to ensure whether the stated requirements are fulfilled or not.

Table of Contents

1. Introduction....3
2. Applications....3
  2.1. Internet Video Streaming....4
  2.2. Internet Protocol Television (IPTV)....6
  2.3. Video conferencing....7
  2.4. Video sharing....8
  2.5. Screencasting....9
  2.6. Game streaming....10
  2.7. Video monitoring / surveillance....11
3. Requirements....12
  3.1. General requirements....12
  3.2. Basic requirements....14
    3.2.1. Input source formats....14
    3.2.2. Coding delay....14
    3.2.3. Complexity....15
    3.2.4. Scalability....15
    3.2.5. Error resilience....15
  3.3. Optional requirements....16
    3.3.1. Input source formats....16
    3.3.2. Scalability....16
    3.3.3. Complexity....16
    3.3.4. Coding efficiency....16
4. Evaluation methodology....17
  4.1. Compression performance evaluation....17
  4.2. Reference software....20
5. Security Considerations....20
6. Conclusions....20
7. IANA Considerations....20
8. References....20
  8.1. Normative References....20
  8.2. Informative References....21
9. Acknowledgments....21

Appendix A. Abbreviations used in the text of this document.....23
Appendix B. Used terms.....24

1. Introduction

In this document, the requirements for a video codec designed mainly for use over the Internet are presented. The requirements encompass a wide range of applications that use data transmission over the Internet including Internet video streaming, IPTV, peer-to-peer video conferencing, video sharing, screencasting, game streaming and video monitoring / surveillance. For each application, typical resolutions, frame-rates and picture access modes are presented. Specific requirements related to data transmission over packet-loss networks are considered as well. In this document, when we discuss data protection techniques we only refer to methods designed and implemented to protect data inside the video codec since there are many existing techniques that protect generic data transmitted over networks with packet losses. From the theoretical point of view, both packet-loss and bit-error robustness can be beneficial for video codecs. In practice, packet losses are a more significant problem than bit corruption in IP networks. It is worth noting that there is an evident interdependence between possible amount of delay and the necessity of error robust video streams:

- If an amount of delay is not crucial for an application, then reliable transport protocols such as TCP that retransmits undelivered packets can be used to guarantee correct decoding of transmitted data.

- If the amount of delay must be kept low, then either data transmission should be error free (e.g., by using managed networks) or compressed video stream should be error resilient.

Thus, error resilience can be useful for delay-critical applications to provide low delay in packet-loss environment.

2. Applications

In this chapter, an overview of video codec applications that are currently available on the Internet market is presented. It is worth noting that there are different use cases for each application that define a target platform, and hence there are different types of communication channels involved (e.g., wired or wireless channels) that are characterized by different quality of service as well as bandwidth; for instance, wired channels are considerably more error-free than wireless channels and therefore require different QoS approaches. The target platform, the channel bandwidth and the

channel quality determine resolutions, frame-rates and quality or bit-rates for video streams to be encoded or decoded. By default, color format YCbCr 4:2:0 is assumed for the application scenarios listed below.

2.1. Internet Video Streaming

Typical content for this application is movies, TV-series and shows, and animation. Internet video streaming uses a variety of client devices and has to operate under changing network conditions. For this reason, an adaptive streaming model has been widely adopted. Video material is encoded at different quality levels and different resolutions, which are then chosen by a client depending on its capabilities and current network bandwidth. An example combination of resolutions and bitrates is shown in Table 1 below.

Resolution *	Frame-rate, fps	PAM
4K, 3840x2160	24/1.001, 24, 25,	RA
2K (1080p), 1920x1080	30/1.001, 30, 50,	RA
1080i, 1920x1080*	60/1.001, 60, 100,	RA
720p, 1280x720	120/1.001, 120	RA
576p (EDTV), 720x576	The set of frame-rates	RA
576i (SDTV), 720x576*	presented in this table	RA
480p (EDTV), 720x480	is taken from Table 2	RA
480i (SDTV), 720x480*	in [1]	RA
512x384		RA
QVGA, 320x240		RA

Table 1. Internet Video Streaming: typical values of resolutions, frame-rates, and RAPs

NB *: Interlaced content can be handled at the higher system level and not necessarily by using specialized video coding tools. It is included in this table only for the sake of completeness as most video content today is in the progressive format.

A video encoding pipeline in on-demand Internet video streaming typically operates as follows:

- Video is encoded in the cloud by software encoders.
- Source video is split into chunks, each of which is encoded separately, in parallel.
- Closed-GOP encoding with 2-5 second intra-picture intervals (or more) is used.
- Encoding is perceptually optimized. Perceptual quality is important and should be considered during the codec development.

Characteristics and requirements of this application scenario are as follows:

- High encoder complexity (up to 10x and more) can be tolerated since encoding happens once and in parallel for different segments.
- Decoding complexity should be kept at reasonable levels to enable efficient decoder implementation.
- Support and efficient encoding of a wide range of content types and formats is required:
  - High Dynamic Range (HDR), Wide Color Gamut (WCG), high resolution (currently, up to 4K), high frame-rate content are important use cases, the codec should be able to encode such content efficiently.
  - Coding efficiency improvement at both lower and higher resolutions is important since low resolutions are used when streaming in low bandwidth conditions.
  - Improvement on both "easy" and "difficult" content in terms of compression efficiency at the same quality level contributes to the overall bitrate/storage savings.
  - Film grain (and sometimes other types of noise) is often present in the streaming movie-type content and is usually a part of the creative intent.

- Significant improvements in compression efficiency between generations of video standards are desirable since this scenario typically assumes long-term support of legacy video codecs.

- Random access points are inserted frequently (one per 2-5 seconds) to enable switching between resolutions and fast-forward playback.

- Elementary stream should have a model that allows easy parsing and identification of the sample components.

- Middle QP values are normally used in streaming, this is also the range where compression efficiency is important for this scenario.

- Scalability or other forms of supporting multiple quality representations are beneficial if they do not incur significant bitrate overhead and if mandated in the first version.

2.2. Internet Protocol Television (IPTV)

This is a service for delivering television content over IP-based networks. IPTV may be classified into two main groups based on the type of delivery, as follows:

- unicast (e.g., for video on demand), where delay is not crucial;

- multicast/broadcast (e.g., for transmitting news) where zapping, i.e. stream changing, delay is important.

In the IPTV scenario, traffic is transmitted over managed (QoS-based) networks. Typical content used in this application is news, movies, cartoons, series, TV shows, etc. One important requirement for both groups is Random access to pictures, i.e. random access period (RAP) should be kept small enough (approximately, 1-5 seconds). Optional requirements are as follows:

- Temporal (frame-rate) scalability;

- Resolution and quality (SNR) scalability.

For this application, typical values of resolutions, frame-rates, and RAPs are presented in Table 2.

Resolution *	Frame-rate, fps	PAM
2160p (4K),3840x2160	24/1.001, 24, 25,	RA
1080p, 1920x1080	30/1.001, 30, 50,	RA
1080i, 1920x1080*	60/1.001, 60, 100,	RA
720p, 1280x720	120/1.001, 120	RA
576p (EDTV), 720x576	The set of frame-rates	RA
576i (SDTV), 720x576*	presented in this table	RA
480p (EDTV), 720x480	is taken from Table 2	RA
480i (SDTV), 720x480*	in [1]	RA

Table 2. IPTV: typical values of resolutions, frame-rates, and RAPs

NB *: Interlaced content can be handled at the higher system level and not necessarily by using specialized video coding tools. It is included in this table only for the sake of completeness as most video content today is in progressive format.

2.3. Video conferencing

This is a form of video connection over the Internet. This form allows users to establish connections to two or more people by two-way video and audio transmission for communication in real-time. For this application, both stationary and mobile devices can be used. The main requirements are as follows:

- Delay should be kept as low as possible (the preferable and maximum end-to-end delay values should be less than 100 ms [7] and 320 ms [2], respectively);

- Temporal (frame-rate) scalability;

- Error robustness.

Support of resolution and quality (SNR) scalability is highly desirable. For this application, typical values of resolutions, frame-rates, and RAPs are presented in Table 3.

Resolution	Frame-rate, fps	PAM
1080p, 1920x1080	15, 30	FIZD
720p, 1280x720	30, 60	FIZD
4CIF, 704x576	30, 60	FIZD
4SIF, 704x480	30, 60	FIZD
VGA, 640x480	30, 60	FIZD
360p, 640x360	30, 60	FIZD

Table 3. Video conferencing: typical values of resolutions, frame-rates, and RAPs

2.4. Video sharing

This is a service that allows people to upload and share video data (using live streaming or not) and to watch them. It is also known as video hosting. A typical User-generated Content (UGC) scenario for this application is to capture video using mobile cameras such as GoPro or cameras integrated into smartphones (amateur video). The main requirements are as follows:

- Random access to pictures for downloaded video data;
- Temporal (frame-rate) scalability;
- Error robustness.

Support of resolution and quality (SNR) scalability is highly desirable. For this application, typical values of resolutions, frame-rates, and RAPs are presented in Table 4.

Resolution	Frame-rate, fps	PAM
2160p (4K),3840x2160	24, 25, 30, 48, 50, 60	RA
1440p (2K),2560x1440	24, 25, 30, 48, 50, 60	RA

1080p, 1920x1080	24, 25, 30, 48, 50, 60	RA
720p, 1280x720	24, 25, 30, 48, 50, 60	RA
480p, 854x480	24, 25, 30, 48, 50, 60	RA
360p, 640x360	24, 25, 30, 48, 50, 60	RA

Table 4. Video sharing: typical values of resolutions, frame-rates [8], and RAPs

2.5. Screencasting

This is a service that allows users to record and distribute computer desktop screen output. This service requires efficient compression of computer-generated content with high visual quality up to visually and mathematically (numerically) lossless [9]. Currently, this application includes business presentations (powerpoint, word documents, email messages, etc.), animation (cartoons), gaming content, data visualization, i.e. such type of content that is characterized by fast motion, rotation, smooth shade, 3D effect, highly saturated colors with full resolution, clear textures and sharp edges with distinct colors [9]), virtual desktop infrastructure (VDI), screen/desktop sharing and collaboration, supervisory control and data acquisition (SCADA) display, automotive/navigation display, cloud gaming, factory automation display, wireless display, display wall, digital operating room (DiOR), etc. For this application, an important requirement is the support of low-delay configurations with zero structural delay, a wide range of video formats (e.g., RGB) in addition to YCbCr 4:2:0 and YCbCr 4:4:4 [9]. For this application, typical values of resolutions, frame-rates, and RAPs are presented in Table 5.

Resolution	Frame-rate, fps	PAM
Input color format: RGB 4:4:4		
5k, 5120x2880	15, 30, 60	AI, RA, FIZD
4k, 3840x2160	15, 30, 60	AI, RA, FIZD
WQXGA, 2560x1600	15, 30, 60	AI, RA, FIZD
WUXGA, 1920x1200	15, 30, 60	AI, RA, FIZD

WSXGA+, 1680x1050	15, 30, 60	AI, RA, FIZD
WXGA, 1280x800	15, 30, 60	AI, RA, FIZD
XGA, 1024x768	15, 30, 60	AI, RA, FIZD
SVGA, 800x600	15, 30, 60	AI, RA, FIZD
VGA, 640x480	15, 30, 60	AI, RA, FIZD
Input color format: YCbCr 4:4:4		
5k, 5120x2880	15, 30, 60	AI, RA, FIZD
4k, 3840x2160	15, 30, 60	AI, RA, FIZD
1440p (2K), 2560x1440	15, 30, 60	AI, RA, FIZD
1080p, 1920x1080	15, 30, 60	AI, RA, FIZD
720p, 1280x720	15, 30, 60	AI, RA, FIZD

Table 5. Screencasting for RGB and YCbCr 4:4:4 format: typical values of resolutions, frame-rates, and RAPs

2.6. Game streaming

This is a service that provides game content over the Internet to different local devices such as notebooks, gaming tablets, etc. In this category of applications, server renders 3D games in cloud server, and streams the game to any device with a wired or wireless broadband connection [10]. There are low latency requirements for transmitting user interactions and receiving game data in less than a turn-around delay of 100 ms. This allows anyone to play (or resume) full featured games from anywhere in the Internet [10]. An example of this application is Nvidia Grid [10]. Another category application is broadcast of video games played by people over the Internet in real time or for later viewing [10]. There are many companies such as Twitch, YY in China enable game broadcasting [10]. Games typically contain a lot of sharp edges and large motion [10]. The main requirements are as follows:

- Random access to pictures for game broadcasting;

- Temporal (frame-rate) scalability;

o Error robustness.

Support of resolution and quality (SNR) scalability is highly desirable. For this application, typical values of resolutions, frame-rates, and RAPs are similar to ones presented in Table 5.

2.7. Video monitoring / surveillance

This is a type of live broadcasting over IP-based networks. Video streams are sent to many receivers at the same time. A new receiver may connect to the stream at an arbitrary moment, so random access period should be kept small enough (approximately, ~1-5 seconds). Data are transmitted publicly in the case of video monitoring and privately in the case of video surveillance, respectively. For IP-cameras that have to capture, process and encode video data, complexity including computational and hardware complexity as well as memory bandwidth should be kept low to allow real-time processing. In addition, support of high dynamic range and a monochrome mode (e.g., for infrared cameras) as well as resolution and quality (SNR) scalability is an essential requirement for video surveillance. In some use-cases, high video signal fidelity is required even after lossy compression. Typical values of resolutions, frame-rates, and RAPs for video monitoring / surveillance applications are presented in Table 6.

Resolution	Frame-rate, fps	PAM
2160p (4K),3840x2160	12, 25, 30	RA, FIZD
5Mpixels, 2560x1920	12, 25, 30	RA, FIZD
1080p, 1920x1080	25, 30	RA, FIZD
1.3Mpixels, 1280x960	25, 30	RA, FIZD
720p, 1280x720	25, 30	RA, FIZD
SVGA, 800x600	25, 30	RA, FIZD

Table 6. Video monitoring / surveillance: typical values of resolutions, frame-rates, and RAPs

3. Requirements

Taking the requirements discussed above for specific video applications, this chapter proposes requirements for an internet video codec.

3.1. General requirements

3.1.1. The most basic requirement is coding efficiency, i.e. compression performance on both "easy" and "difficult" content for applications and use cases in Section 2. The codec should provide higher coding efficiency over state-of-the-art video codecs such as HEVC/H.265 and VP9, at least by 25% in accordance with the methodology described in Section 4.1 of this document. For higher resolutions, the coding efficiency improvements are expected to be higher than for lower resolutions.

3.1.2. Good quality specification and well-defined profiles and levels are required to enable device interoperability and facilitate decoder implementations. A profile consists of a subset of entire bitstream syntax elements and consequently it also defines the necessary tools for decoding a conforming bitstream of that profile. A level imposes a set of numerical limits to the values of some syntax elements. An example of codec levels to be supported is presented in Table 7. An actual level definition should include constraints on features that impact the decoder complexity. For example, these features might be as follows: maximum bit-rate, line buffer size, memory usage, etc.

Level	Example picture resolution at highest frame rate
1	128x96(12,288*)@30.0 176x144(25,344*)@15.0
2	352x288(101,376*)@30.0
3	352x288(101,376*)@60.0 640x360(230,400*)@30.0
4	640x360(230,400*)@60.0 960x540(518,400*)@30.0
5	720x576(414,720*)@75.0 960x540(518,400*)@60.0 1280x720(921,600*)@30.0

		1,280x720(921,600*)@68.0
	6	2,048x1,080(2,211,840*)@30.0
		1,280x720(921,600*)@120.0
	7	2,048x1,080(2,211,840*)@60.0
		1,920x1,080(2,073,600*)@120.0
	8	3,840x2,160(8,294,400*)@30.0
		4,096x2,160(8,847,360*)@30.0
		1,920x1,080(2,073,600*)@250.0
	9	4,096x2,160(8,847,360*)@60.0
		1,920x1,080(2,073,600*)@300.0
	10	4,096x2,160(8,847,360*)@120.0
		3,840x2,160(8,294,400*)@120.0
	11	8,192x4,320(35,389,440*)@30.0
		3,840x2,160(8,294,400*)@250.0
	12	8,192x4,320(35,389,440*)@60.0
		3,840x2,160(8,294,400*)@300.0
	13	8,192x4,320(35,389,440*)@120.0

Table 7. Codec levels

NB *: The quantities of pixels are presented for such applications where a picture can have an arbitrary size (e.g., screencasting)

3.1.3. Bitstream syntax should allow extensibility and backward compatibility. New features can be supported easily by using metadata (e.g., such as SEI messages, VUI, headers) without affecting the bitstream compatibility with legacy decoders. A newer version of the decoder shall be able to play bitstreams of an older version of the same or lower profile and level.

3.1.4. A bitstream should have a model that allows easy parsing and identification of the sample components (such as ISO/IEC14496-10, Annex B or ISO/IEC 14496-15). In particular, information needed for packet handling (e.g., frame type) should not require parsing anything below the header level.

3.1.5. Perceptual quality tools (such as adaptive QP and quantization matrices) should be supported by the codec bit-stream.

3.1.6. The codec specification shall define a buffer model such as hypothetical reference decoder (HRD).

3.1.7. Specifications providing integration with system and delivery layers should be developed.

3.2. Basic requirements

3.2.1. Input source formats:

- Bit depth: 8- and 10-bits (up to 12-bits for a high profile) per color component;
- Color sampling formats:
  - YCbCr 4:2:0;
  - YCbCr 4:4:4, YCbCr 4:2:2 and YCbCr 4:0:0 (preferably in different profile(s)).
- For profiles with bit depth of 10 bits per sample or higher, support of high dynamic range and wide color gamut.
- Support of arbitrary resolution according to the level constraints for such applications where a picture can have an arbitrary size (e.g., in screencasting).

3.2.2. Coding delay:

- Support of configurations with zero structural delay also referred to as "low-delay" configurations.
  - Note 1: end-to-end delay should be up to 320 ms [2] but its preferable value should be less than 100 ms [7]
- Support of efficient random access point encoding (such as intra coding and resetting of context variables) as well as efficient switching between multiple quality representations.
- Support of configurations with non-zero structural delay (such as out-of-order or multi-pass encoding) for applications without low-delay requirements if such configurations provide additional compression efficiency improvements.

3.2.3. Complexity:

- Feasible real-time implementation of both an encoder and a decoder supporting a chosen subset of tools for hardware and software implementation on a wide range of state-of-the-art platforms. The real-time encoder tools subset should provide meaningful improvement in compression efficiency at reasonable complexity of hardware and software encoder implementations as compared to current real-time implementations of state-of-the-art video compression technologies such as HEVC/H.265 and VP9.

- High-complexity software encoder implementations used by offline encoding applications can have 10x or more complexity increase compared to state-of-the-art video compression technologies such as HEVC/H.265 and VP9.

3.2.4. Scalability:

- Temporal (frame-rate) scalability should be supported.

3.2.5. Error resilience:

- Error resilience tools that are complementary to the error protection mechanisms implemented on transport level should be supported.

- The codec should support mechanisms that facilitate packetization of a bitstream for common network protocols.

- Packetization mechanisms should enable frame-level error recovery by means of retransmission or error concealment.

- The codec should support effective mechanisms for allowing decoding and reconstruction of significant parts of pictures in the event that parts of the picture data are lost in transmission.

- The bitstream specification shall support independently decodable sub-frame units similar to slices or independent tiles. It shall be possible for the encoder to restrict the bit-stream to allow parsing of the bit-stream after a packet-loss and to communicate it to the decoder.

3.3. Optional requirements

3.3.1. Input source formats

- Bit depth: up to 16-bits per color component.
- Color sampling formats: RGB 4:4:4.
- Auxiliary channel (e.g., alpha channel) support.

3.3.2. Scalability:

- Resolution and quality (SNR) scalability that provide low compression efficiency penalty (up to 5% of BD-rate [12] increase per layer with reasonable increase of both computational and hardware complexity) can be supported in the main profile of the codec being developed by the NETVC WG. Otherwise, a separate profile is needed to support these types of scalability.
- Computational complexity scalability(i.e. computational complexity is decreasing along with degrading picture quality) is desirable.

3.3.3. Complexity:

Tools that enable parallel processing (e.g., slices, tiles, wave front propagation processing) at both encoder and decoder sides are highly desirable for many applications.

- High-level multi-core parallelism: encoder and decoder operation, especially entropy encoding and decoding, should allow multiple frames or sub-frame regions (e.g. 1D slices, 2D tiles, or partitions) to be processed concurrently, either independently or with deterministic dependencies that can be efficiently pipelined
- Low-level instruction set parallelism: favor algorithms that are SIMD/GPU friendly over inherently serial algorithms

3.3.4. Coding efficiency

Compression efficiency on noisy content, content with film grain, computer generated content, and low resolution materials is desirable.

4. Evaluation methodology

4.1. Compression performance evaluation

As shown in Fig.1, compression performance testing is performed in 3 overlapped ranges that encompass 10 different bitrate values:

- Low bitrate range (LBR) is the range that contains the 4 lowest bitrates of the 10 specified bitrates (1 of the 4 bitrate values is shared with the neighboring range);

- Medium bitrate range (MBR) is the range that contains the 4 medium bitrates of the 10 specified bitrates (2 of the 4 bitrate values are shared with the neighboring ranges);

- High bitrate range (HBR) is the range that contains the 4 highest bitrates of the 10 specified bitrates (1 of the 4 bitrate values is shared with the neighboring range).

Initially, for the codec selected as a reference one (e.g., HEVC or VP9), a set of 10 QP (quantization parameter) values should be specified (in a separate document on Internet video codec testing) and corresponding quality values should be calculated. In Fig.1, QP and quality values are denoted as QP0, QP1, QP2,..., QP8, QP9 and Q0, Q1, Q2,..., Q8, Q9, respectively. To guarantee the overlaps of quality levels between the bitrate ranges of the reference and tested codecs, a quality alignment procedure should be performed for each range's outermost (left- and rightmost) quality levels Qk of the reference codec (i.e. for Q0, Q3, Q6, and Q9) and the quality levels Q'k (i.e. Q'0, Q'3, Q'6, and Q'9) of the tested codec. Thus, these quality levels Q'k and, hence, the corresponding QP value QP'k (i.e. QP'0, QP'3, QP'6, and QP'9) of the tested codec should be selected using the following formulas:

```
Q'k =   min { abs(Q'i - Qk) },
      i in R


QP'k = argmin { abs(Q'i(QP'i) - Qk(QPk)) },
       i in R
```

where R is the range of the QP indexes of the tested codec, i.e. the candidate Internet video codec. The inner quality levels (i.e. Q'1, Q'2, Q'4, Q'5, Q'7, and Q'8) as well as their corresponding QP values of each range (i.e. QP'1, QP'2, QP'4, QP'5, QP'7, and QP'8) should be as equidistantly spaced as possible between the left- and rightmost quality levels without explicitly mapping their values using the above described procedure.

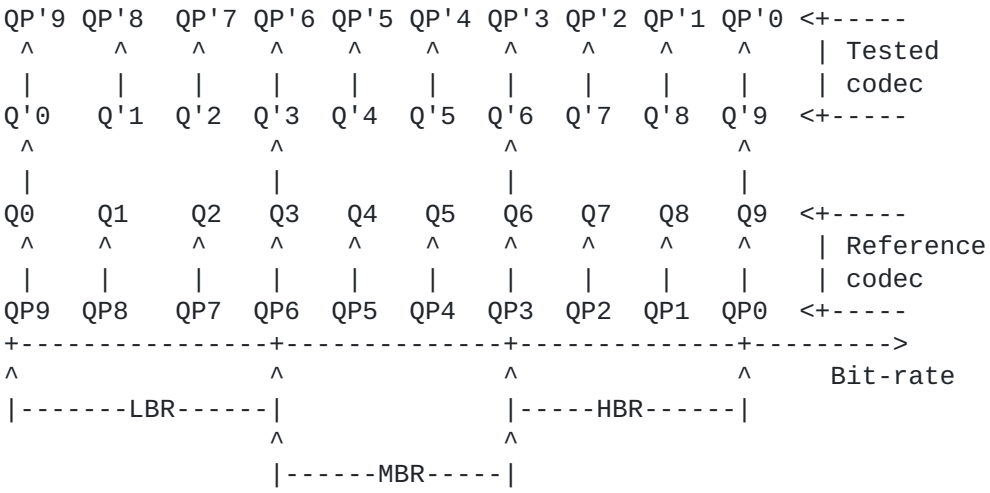

Figure 1 Quality/QP alignment for compression performance evaluation

Since the QP mapping results may vary for different sequences, eventually, this quality alignment procedure needs to be separately performed for each quality assessment index and each sequence used for codec performance evaluation to fulfill the above described requirements.

To assess the quality of output (decoded) sequences, two indexes, PSNR [3] and MS-SSIM [3,11] are separately computed. In the case of the YCbCr color format, PSNR should be calculated for each color plane whereas MS-SSIM is calculated for luma channel only. In the case of the RGB color format, both metrics are computed for R, G and B channels. Thus, for each sequence, 30 RD-points for PSNR (i.e. three RD-curves, one for each channel) and 10 RD-points for MS-SSIM (i.e. one RD-curve, for luma channel only) should be calculated in the case of YCbCr. If content is encoded as RGB, 60 RD-points (30 for PSNR and 30 for MS-SSIM) should be calculated, i.e. three RD-curves (one for each channel) are computed for PSNR as well as three RD-curves (one for each channel) for MS-SSIM.

Finally, to obtain an integral estimation, BD-rate savings [12] should be computed for each range and each quality index. In

addition, average values over all the 3 ranges should be provided for both PSNR and MS-SSIM. A list of video sequences that should be used for testing as well as the 10 QP values for the reference codec are defined in a separate document. Testing processes should use the information on the codec applications presented in this document. As the reference for evaluation, state-of-the-art video codecs such as HEVC/H.265 [4,5] or VP9 must be used. The reference source code of the HEVC/H.265 codec can be found at [6]. The HEVC/H.265 codec must be configured according to [13] and Table 8.

Intra-period, second	HEVC/H.265 encoding mode according to [13]
AI	Intra Main or Intra Main10
RA	Random access Main or Random access Main10
FIZD	Low delay Main or Low delay Main10

Table 8. Intra-periods for different HEVC/H.265 encoding modes according to [13]

According to the coding efficiency requirement described in Section 3.1.1, BD-rate savings calculated for each color plane and averaged for all the video sequences used to test the NETVC codec should be, at least,

- 25% if calculated over the whole bitrate range;

- 15% if calculated for each bitrate subrange (LBR, MBR, HBR).

Since values of the two objective metrics (PSNR and MS-SSIM) are available for some color planes, each value should meet these coding efficiency requirements, i.e. the final BD-rate saving denoted as S is calculated for a given color plane as follows:

S = min { S_psnr, S_ms-ssim },

where S_psnr and S_ms-ssim are BD-rate savings calculated for the given color plane using PSNR and MS-SSIM metrics, respectively.

In addition to the objective quality measures defined above, subjective evaluation must also be performed for the final NETVC codec adoption. For subjective tests, the MOS-based evaluation

procedure must be used as described in section 2.1 of [3]. For perception-oriented tools that primarily impact subjective quality, additional tests may also be individually assigned even for intermediate evaluation, subject to a decision of the NETVC WG.

4.2. Reference software

Reference software provided to the NETVC WG for candidate codecs should comprise a fully operational encoder supporting necessary rate controls, subjective quality optimization features and some degree of speed optimization and a "real-time" decoder.

5. Security Considerations

This document itself does not address any security considerations. However, it is worth noting that a codec implementation (for both an encoder and a decoder) should cover the worst case of computational complexity, memory bandwidth, and physical memory size (e.g., for decoded pictures used as references). Otherwise, it can be considered as a security vulnerability and lead to denial-of-service (DoS) in the case of attacks.

6. Conclusions

In this document, an overview of Internet video codec applications and typical use cases as well as a prioritized list of requirements for an Internet video codec are presented. An evaluation methodology for this codec is also proposed.

7. IANA Considerations

This document has no IANA actions.

8. References

8.1. Normative References

[1] Recommendation ITU-R BT.2020-2: Parameter values for ultra-high definition television systems for production and international programme exchange, 2015.

[2] Recommendation ITU-T G.1091: Quality of Experience requirements for telepresence services, 2014.

[3] ISO/IEC PDTR 29170-1: Information technology -- Advanced image coding and evaluation methodologies -- Part 1: Guidelines for codec evaluation.

[4] ISO/IEC 23008-2:2015. Information technology -- High efficiency coding and media delivery in heterogeneous environments -- Part 2: High efficiency video coding

[5] Recommendation ITU-T H.265: High efficiency video coding, 2013.

[6] https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/

8.2. Informative References

[7] S. Wenger, "The case for scalability support in version 1 of Future Video Coding," contribution COM 16-C 988 R1-E to ITU-T SG16/Q6, September 2015."Recommended upload encoding settings (Advanced)"

[8] "Recommended upload encoding settings (Advanced)" https://support.google.com/youtube/answer/1722171?hl=en

[9] H. Yu, K. McCann, R. Cohen, and P. Amon, "Requirements for future extensions of HEVC in coding screen content", ISO/IEC JTC1/SC29/WG11 MPEG2013/N14174, San Jose, USA, Jan. 2014

[10] Manindra Parhy, "Game streaming requirement for Future Video Coding," MPEG Contribution m36771, June 2015, Warsaw, Poland.

[11] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," Invited Paper, IEEE Asilomar Conference on Signals, Systems and Computers, Nov. 2003, Vol. 2, pp. 1398-1402.

[12] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves (VCEG-M33)," in VCEG Meeting (ITU-T SG16 Q.6), Austin, Texas, USA, Apr. 2-4 2001.

[13] F. Bossen, "Common test conditions and software reference configurations," JCTVC-L1100, Geneva, Switzerland, Jan. 2013.

[14] http://www.digitizationguidelines.gov/term.php?term=compressionvisuallylossless)

9. Acknowledgments

10. The authors would like to thank Mr. Jiantong Zhou, Mr. Paul Coverdale, Mr. Vasily Rufitskiy, and Dr. Haitao Yang for many useful discussions on this document and their help while preparing it as well as Mr. Mo Zanaty, Dr. Minhua Zhou, Dr. Ali Begen, Mr. Thomas

Daede, Dr. Thomas Davies, Mr. Jonathan Lennox, Dr. Timothy Terriberry, Mr. Peter Thatcher, Dr. Jean-Marc Valin, Mr. Jack Moffitt, Mr. Greg Coppa and Mr. Andrew Krupiczka for their valuable comments on different revisions of this document.

This document was prepared using 2-Word-v2.0.template.dot.

Appendix A. Abbreviations used in the text of this document

Abbreviation	Meaning
AI	All-Intra (each picture is intra-coded)
BD-Rate	Bjontegaard Delta Rate
FIZD	just the First picture is Intra-coded, Zero structural Delay
GOP	Group of Picture
HBR	High Bitrate Range
HDR	High Dynamic Range
HRD	Hypothetical Reference Decoder
IPTV	Internet Protocol Television
LBR	Low Bitrate Range
MBR	Medium Bitrate Range
MOS	Mean Opinion Score
MS-SSIM	Multi-Scale Structural Similarity quality index
PAM	Picture Access Mode
PSNR	Peak Signal-to-Noise Ratio
QoS	Quality of Service
QP	Quantization Parameter
RA	Random Access
RAP	Random Access Period
RD	Rate-Distortion
SEI	Supplemental Enhancement Information
UGC	User-Generated Content
VDI	Virtual Desktop Infrastructure
VUI	Video Usability Information
WCG	Wide Color Gamut

Appendix B. Used terms

Term	Meaning
High dynamic range imaging	is a set of techniques that allow a greater dynamic range of exposures or values (i.e., a wide range of values between light and dark areas) than normal digital imaging techniques. The intention is to accurately represent the wide range of intensity levels found in such examples as exterior scenes that include light-colored items struck by direct sunlight and areas of deep shadow [14].
Random access period	is the period of time between two closest independently decodable frames (pictures).
RD-point	A point in a 2 dimensional rate-distortion space where the values of bitrate and quality metric are used as x- and y-coordinates, respectively
Visually lossless compression	is a form or manner of lossy compression where the data that are lost after the file is compressed and decompressed is not detectable to the eye; the compressed data appearing identical to the uncompressed data [14].
Wide color gamut	is a certain complete color subset (e.g., considered in ITU-R BT.2020) that supports a wider range of colors (i.e., an extended range of colors that can be generated by a specific input or output device such as a video camera, monitor or printer and can be interpreted by a color model) than conventional color gamuts (e.g., considered in ITU-R BT.601 or BT.709).

Authors' Addresses

Alexey Filippov
Huawei Technologies

Email: alexey.filippov@huawei.com

Andrey Norkin
Netflix

Email: anorkin@netflix.com

Jose Roberto Alvarez
Huawei Technologies

Email: jose.roberto.alvarez@huawei.com