Network Working Group
Internet Draft
Intended status: Informational
Expires: September 29, 2016

A. Filippov
Huawei Technologies
March 29, 2016

<Video Codec Requirements and Evaluation Methodology>
draft-ietf-netvc-requirements-01.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, an evaluation methodology needed for measuring the parameters (compression efficiency, computational complexity, etc.) to ensure whether the stated requirements are fulfilled or not.

Table of Contents

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 IPTV (broadcasting over IP-based networks), peer-to-peer video conferencing, video sharing, screencasting, 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 packet-loss networks. 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 resends 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 YUV 4:2:0 is assumed for the application scenarios listed below.

2.1. Internet Protocol Television (IPTV) / IP-based over-the-top (OTT) video transmission

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 and, hence, error resilience is not needed;
- multicast/broadcast (e.g., for transmitting news) where zapping, i.e. stream changing, delay is important and, therefore, error resilience is required in the case of unmanaged networks like the Internet.

The main difference between IPTV and IP-based OTT video transmission is that traffic is transmitted over managed (QoS-based) and unmanaged networks in the above cases, respectively. 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, 1-15 seconds). For the second group, two further requirements should be met:

- Temporal (frame-rate) scalability;
- Error robustness (only for IP-based OTT video transmission).

For the first use case, the two above-mentioned requirements are optional. Support of resolution and quality (SNR) scalability is highly desirable for the both groups. For this application, typical values of resolutions, frame-rates, and RAPs are presented in Table 1.

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 1. 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.

2.2. 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 2.

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

2.3. 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 3.

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

2.4. 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 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 a wide range of video formats (e.g., RGB) in addition to YUV 4:2:0 and YUV 4:4:4 [9]. For this application, typical values of resolutions, frame-rates, and RAPs are presented in Table 4.

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

2.5. 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;
- 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 4.

2.6. 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 as well as resolution and quality (SNR) scalability is an essential requirement for video surveillance. For this application, typical values of resolutions, frame-rates, and RAPs are presented in Table 5.

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

Table 5. 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. The most basic requirement is coding efficiency, i.e. compression performance. It should be better than for state-of-the-art video codecs such as HEVC/H.265 and VP9. Levels to be supported by the new codec are presented in Table 6.

Level	Example picture resolution at highest frame rate
1	640x360(230,400*)@60.0
2	640x360(230,400*)@60.0 960x540(518,400*)@30.0
3	720x576(414,720*)@75.0 960x540(518,400*)@60.0 1280x720(921,600*)@30.0

	1,280x720(921,600*)@68.0
4	2,048x1,080(2,211,840*)@30.0
	1,280x720(921,600*)@120.0
5	2,048x1,080(2,211,840*)@60.0
	1,920x1,080(2,073,600*)@120.0
6	3,840x2,160(8,294,400*)@30.0
	4,096x2,160(8,847,360*)@30.0
	1,920x1,080(2,073,600*)@250.0
7	4,096x2,160(8,847,360*)@60.0
	1,920x1,080(2,073,600*)@300.0
8	4,096x2,160(8,847,360*)@120.0
	3,840x2,160(8,294,400*)@120.0
9	8,192x4,320(35,389,440*)@30.0
	3,840x2,160(8,294,400*)@250.0
10	8,192x4,320(35,389,440*)@60.0
	3,840x2,160(8,294,400*)@300.0
11	8,192x4,320(35,389,440*)@120.0

Table 6. 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. Basic requirements

3.1.1. Input source formats:

- Bit depth: 8- and 10-bits per color component;
- Color sampling formats: YUV 4:2:0, YUV 4:4:4
- Support of arbitrary resolution for such applications where a picture can have an arbitrary size (e.g., screencasting)

3.1.2. Coding delay:

- Support of configurations with zero structural delay also referred to as “low-delay” configurations (end-to-end delay should be up to 320 ms [2] but it’s preferable value should be less than 100 ms [7])

3.1.3. Complexity:

- Feasible real-time implementation of both an encoder and a decoder for hardware and software implementation based on a wide range of state-of-the-art platforms

3.1.4. Scalability:

- Temporal (frame-rate) scalability

3.1.5. Error resilience:

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

3.2. Optional requirements

3.2.1. Input source formats

- Bit depth: up to 16-bits per color component;
- Color sampling formats: RGB 4:4:4 and YUV 4:2:2;
- Auxiliary channel (e.g., alpha channel) support;
- Support of high dynamic range and wide color gamut

3.2.2. Scalability:

- Resolution and quality (SNR) scalability;
- Computational complexity scalability, i.e. computational complexity is decreasing along with degrading picture quality

3.2.3. 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

4. Evaluation methodology

4.1. Compression performance evaluation

As shown in Fig.1, compression performance testing is performed in 3 ranges that encompass 12 different bit-rate values:

- Low bit-rate range (LBR) is the range that contains the 4 lowest bit-rates of the 12 specified bit-rates;
- Medium bit-rate range (MBR) is the range that contains the 4 medium bit-rates of the 12 specified bit-rates;
- High bit-rate range (HBR) is the range that contains the 4 highest bit-rates of the 12 specified bit-rates.

To avoid any rate control mechanisms that can significantly impact evaluation results, just nominal values of bit-rates should be specified in a separate document on Internet video codec testing. The deviation between nominal and actual values of bit-rates obtained for both reference and tested codecs should be less than the threshold value defined in the above-mention document on Internet video codec testing. This deviation is calculated as follows:

$$D = abs((BRa - BRn)/ BRn)*100\%,$$

where BRn is a nominal value of bit-rate, BRa is an actual value of bit-rate obtained for either reference or tested codecs.

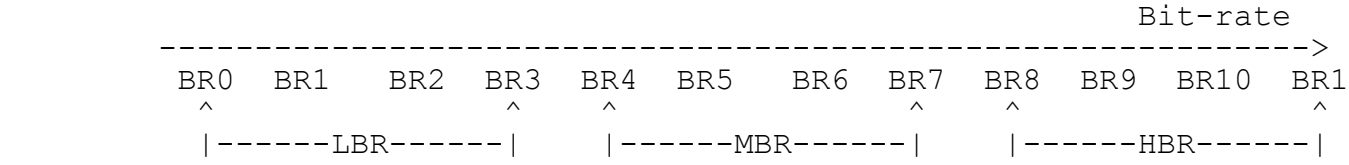

Figure 1 Bit-rate ranges for the CBR mode

To assess the quality of output (decoded) sequences, two indexes, PSNR [3] and MS-SSIM [3,11], should be separately calculated for each color plane. For obtaining an integral estimation, BD-rate [12] should be computed for each range and each quality index. Finally, 18 values should be obtained for a color format, which contains 3 color planes (e.g., for YUV or RGB). A list of video sequences that should be used for testing as well as the 6 values of bit-rates 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, the HEVC/H.265 codec [4,5] 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 9.

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
JIZD	Low delay Main or Low delay Main10

Table 9. Intra-periods for different HEVC/H.265 encoding modes according to [13]

In addition to the objective quality measures defined above, subjective evaluation must also be performed before adopting any new tool and a final codec standard. 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.

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. References

7.1. 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/

7.2. 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)

8. Acknowledgments

9. The author would like to thank Mr. Jiantong Zhou, Mr. Jose Alvarez, 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, 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
GOP	Group of Picture
HBR	High Bit-rate Range
PAM	Picture Access Mode
RA	Random Access
RAP	Random Access Period
IPTV	Internet Protocol Television
FIZD	just the First picture is Intra-coded, Zero structural Delay
LBR	Low Bit-rate Range
MBR	Medium Bit-rate Range
MOS	Mean Opinion Score
MS-SSIM	Multi-Scale Structural Similarity quality index
OTT	Over-The-Top
PSNR	Peak Signal-to-Noise Ratio
QoS	Quality of Service
UGC	User-Generated Content
VDI	Virtual Desktop Infrastructure

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).
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