RTCWEB Working Group B. Burman
Internet-Draft Ericsson
Intended status: Standards Track M. Isomaki
Expires: August 29, 2013 Nokia
B. Aboba
Microsoft Corporation
G. Martin-Cocher
RIM
G. Mandyam
Qualcomm Innovation Center
X. Marjou
France Telecom Orange
February 25, 2013

H.264 as Mandatory to Implement Video Codec for WebRTC
draft-burman-rtcweb-h264-proposal-01

Abstract

This document proposes that, and motivates why, H.264 should be a Mandatory To Implement video codec for WebRTC.

Status of this Memo

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

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

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

This Internet-Draft will expire on August 29, 2013.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the
document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal
Provisions Relating to IETF Documents
(http://trustee.ietf.org/license-info) in effect on the date of

publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction

The selection of a Mandatory To Implement (MTI) video codec for WebRTC has been discussed for quite some time in the RTCWEB WG. This document proposes that the H.264 video codec should be mandatory to implement for WebRTC implementations and gives motivation to this proposal.

The core of the proposal is that H.264 Constrained Baseline Profile Level 1.2 MUST be supported as Mandatory To Implement video codec. To enable higher quality for devices capable of it, support for H.264 Constrained High Profile Level 1.3, extended to support 720p resolution at 30 Hz framerate is RECOMMENDED.

This draft discusses the advantages of H.264 as the authors of this draft see them; a richness of implementations and hardware support, well known licensing conditions, good performance, and well defined handling of varying device capabilities.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14, RFC 2119 [RFC2119].

3. H.264 Overview

The video coding standard Advanced Video Coding (ITU-T H.264 | ISO/IEC 14496-10 [H264]) has been around for almost ten years by now. Developed jointly by MPEG and ITU-T in the Joint Video Team, it was published in its first version in 2003 and amended with support for higher-fidelity video in 2004. Other significant updates include support for scalability (2007) and multiview (2009). The codec goes under the names H.264, AVC and MPEG-4 Part10. In this memo the term "H.264" will be used.

H.264 was from the start very successful and has become widely adopted for (video) content as well as (video) communication services worldwide.

4. Implementations

Arguably, hardware or DSP acceleration for video encoding/decoding would be mostly beneficial for devices that has relatively lower

capacity in terms of CPU and power (smaller batteries), and the most common devices in this category are phones and tablets. There is a long list of vendors offering hardware or DSP implementations of H.264. In particular all vendors of platforms for mobile high-range phones, smartphones, and tablets support H.264/AVC High Profile encoding and decoding at least 1080p30, but those platforms are currently in general not used for low- to mid-range devices. These vendors are ST-Ericsson, Qualcomm, TI, Nvidia, Renesas, Mediatek, Huawei Hisilicon, Intel, Broadcom, Samsung. Those platforms all support H.264/AVC codec with dedicated HW or DSP. For at least the ST-Ericsson and Qualcomm hardware it is verified that the implementation has low-delay real-time support, but it seems likely that this is the case for at least the majority of the others as well.

There are also other specifications that implement support for H.264, such as HDMI(TM).

Regarding software implementations there is a long list of available implementations. Wikipedia provides an illustration of this with their list [Implementations], and more implementations appear, e.g. [Woon]. Not only are there standalone implementations available, including open source, but in addition recent Windows and Mac OS X versions support H.264 encoding and decoding.

5. Licensing

H.264 is a mature codec with a mature and well-known licensing model.

MPEG-LA released their AVC Patent Portfolio License already in 2004 and in 2010 they announced that H.264 encoded Internet video is free to end users will never be charged royalties [MPEGLA]. Real-time generated content, the content most applicable to WebRTC, was free already from the establishment of the MPEG-LA license. License fees for products that decode and encode H.264 video remain though. Those fees are, and will very likely continue to be for the lifetime of MPEG-LA pool, $0.20 per codec or less. It can be noted that for MPEG-LA, since one license covers both an encoder and decoder, there is no additional cost of using an encoder to an implementation that supports decoding of H.264.

It is a well-established fact that not all H.264 right holders are MPEG-LA pool members. H.264 is however an ITU/ISO/IEC international standard, developed under their respective patent policies, and all contributors must license their patents under Reasonable And Non-Discriminatory (RAND) terms. In the field of video coding, most major research groups interested in patents do contribute to the ITU/ISO/IEC standards process and are therefore bound by those terms.

VP8 is a much younger codec than H.264 and it is fair to say that the licensing situation is less clear than for H.264. Google has provided their patent rights on VP8 under a open source friendly license with very restrictive reciprocity conditions. According to MPEG-LA's web page [MpegLaVp8], MPEG-LA is in the process of forming a royalty-bearing patent pool for VP8. Also, according to press reports [DoJ], at least the US Department of Justice investigate MPEG-LA for anticompetitive activity in conjunction with the VP8 pool formation. This indicates that the licensing situation for VP8 has not settled.

6. Performance

Comparing video quality is difficult. Practically no modern video encoding method includes any bit-exact encoding where a given (video) input produces a specified encoded output bitstream. Instead, the encoded bitstream syntax and semantics are specified such that a decoder can correctly interpret it and produce a known output. This is true both for H.264 and VP8. Significant freedom is left to the encoder implementation to choose how to represent the encoded video, for example given a specific targeted bitrate. Thus it cannot in general be expected that any encoded video bitstream represents the best possible or most efficient representation, given the defined bitstream syntax elements available to that codec. The actually achieved quality for a certain bitstream, how close it is to the optimally possible with available syntax, at any given bitrate rather depends on the performance of the individual encoder implementation.

Also, not only is the resulting experienced video quality subjective, but also depends on the source material, on the point of operation and a number of other considerations. In addition, performance can be measured vs. bitrate, but also vs. e.g. complexity - and here another can of worms can be opened because complexity depends on hardware used (some platforms have video codec accelerations), SW platform (and how efficient it can use the hardware) and so on. On top of this comes that different implementations can have different performance, and can be operated in different ways (e.g. tradeoffs between complexity and quality can be made). Regardless of how a performance evaluation is carried out it can always be said that it is not "fair". This section nevertheless attempts to shed some light on this subject, and specifically the performance (measured against bitrate) of H.264 compared to VP8.

A number of studies [H264perf1][H264perf2][H264perf3] have been made to compare the compression efficiency performance between H.264 and

VP8. These studies show that H.264 is in general performing better than VP8 but the studies are not specifically targeting video

conferencing. Therefore, Ericsson made a comparison where a number of video conferencing type sequences were encoded using both H.264 and VP8. Eight video conferencing type test sequences were used; three were taken from the MPEG/ITU test set (vidyo2-4) and five were recorded by Ericsson. The sequences were all 720p 25/30Hz.

The focus of that test was to evaluate the best compression efficiency that could be achieved with both codecs since it was believed to be harder to make a fair comparison trying to use complexity constraints. The results showed that H.264 High Profile provides an average bitrate compared to VP8 of -23% (minus here means that H.264 is better) using PSNR-based Bjontegaard Delta bitrate (BD-rate) [PSNRdiff]. H.264 Constrained High Profile provided -16% and Constrained Baseline Profile resulted in +16% (plus here means that VP8 is better).

For H.264, JM 18.3 in low-delay mode without reordering of B or P pictures was used. For VP8 encoding, v1.1.0 with the "best" preset was used.

Again, video quality is difficult to compare. The authors however believe that the data provided in this section shows that H.264 is at least on par with VP8. As a final note, the new H.265/HEVC standard clearly outperforms both of them, but the authors think it is premature to mandate HEVC for WebRTC.

7. Profile/level

H.264/AVC [H264] has a large number of encoding tools, grouped in functionally reasonable toolsets by codec profiles, and a wide range of possible implementation capability and complexity, specified by codec levels. It is typically not reasonable for H.264 encoders and decoders to implement maximum complexity capability for all of the available tools. Thus, any H.264 decoder implementation is typically not able to receive all possible H.264 streams. Which streams can be received is described by what profile and level the decoder conforms to. Any video stream produced by an H.264 encoder must keep within the limits defined by the intended receiving decoder's profile and level to ensure that the video stream can be correctly decoded.

Profiles can be "ranked" in terms of the amount of tools included, such that some profiles with few tools are "lower" than profiles with more tools. However, profiles are typically not strictly supersets or subsets of each other in terms of which tools are used, so a strict ranking cannot be defined. It is also in some cases possible

to express compliance to the common subset of tools between two different profiles. This is fairly well described in [RFC6184].

When choosing a Mandatory To Implement codec, it is desirable to use a profile and level that is as widely supported as possible. Therefore, H.264 Constrained Baseline Profile Level 1.2 MUST be supported as Mandatory To Implement video codec. This is possible to support with significant margin in hardware devices (Section 4) and should likely also not cause performance problems for software-only implementations. All Level definitions (Annex A of [H264]) include a maximum framesize in macroblocks (16*16 pixels) as well as a maximum processing requirement in macroblocks per second. That number of macroblocks per second can be almost freely distributed between framesize and framerate. The maximum framesize for Level 1.2 corresponds to 352*288 pixels (CIF). Examples of allowed framesize and framerate combinations for Level 1.2 are CIF (352*288 pixels) at 15 Hz, QVGA (320*240 pixels) at 20 Hz, and QCIF (176*144 pixels) at 60 Hz.

Recognizing that while the above profile and level will likely be possible to implement in any device, it is also likely not sufficient for applications that require higher quality. Therefore, it is RECOMMENDED that devices and implementations that can meet the additional requirements also implement at least H.264 Constrained High Profile Level 1.3, extended to support 720p resolution at 30 Hz framerate, but the extension MAY alternatively be made from any Level higher than 1.3.

Note that the lowest non-extended Level that support 720p30 is Level 3.1, but fully supporting Level 3.1 also requires fairly high bitrate, large buffers, and other encoding parameters included in that Level definition that are likely not reasonable for the targeted communication scenario. This method of extending a lower level in SDP (Section 8) with a smaller set of applicable parameters is fully in line with [RFC6184], and is already used by some video conferencing vendors.

When considering the main WebRTC use case, real-time communication, the lack of need to support interlaced image format in that context, the limited use of and added delay from bi-directionally predicted (B) pictures, and the added implementation and computation complexity that comes with interlace and B-picture handling suggests that Constrained High Profile should be preferred over High Profile as optional codec. Note also that while Constrained High Profile is currently less supported in devices than High Profile, any High Profile decoder will be capable of decoding a Constrained High Profile bitstream since it is a subset of High Profile. To make a High Profile encoder support Constrained High Profile encoding, it will have to turn off interlace encoding and turn off the use of bi-

directional prediction.

8. Negotiation

Given that there exist a fairly large set of defined profiles and levels (Section 7), the probability is rather low that randomly chosen H.264 encoder and decoder implementations have exactly matching capabilities. In any communication scenario, there is therefore a need for a decoder to be able to convey its maximum supported profile and level that the encoder must not exceed.

In addition and depending on the wanted use case and the conditions that apply at a certain communication instance, there may also be a need to describe the currently wanted profile and level at the start of the communication session, which may be lower than the maximum supported by the implementation. In this scenario it may also be of interest to communicate from the encoder to the decoder both which profile and level that will actually be used and what is the maximum supported profile and level. The reason to communicate not only the starting point but also the maximum assumes that communication conditions may change during the conditions, maybe multiple times, possibly making another profile and level be a more appropriate choice.

Communication of maximum supported profile and level is the only mandatory SDP [RFC4566] parameter in the H.264 payload format [RFC6184], which also includes a large set of optional parameters, describing available use (decoder) and intended use (encoder) of those parameters for a specific offered [RFC3264] stream.

If the above mentioned (Section 7) capability for 720p30 is supported as an extension to Constrained High Profile Level 1.3 (or higher), the level extension SHOULD be signaled in SDP using the following parameters as defined in section 8.1 of [RFC6184]:

- profile-level-id=640c0d (or corresponding to a higher Level of Constrained High profile)
- max-fs=3600 (or greater)
- max-mbps=108000 (or greater)
- max-br=768 (or greater, whatever the device implementation can support)

9. Summary

H.264 is widely adopted and used for a large set of video services. This in turn is because H.264 offers great performance, reasonable

licensing terms (and manageable risks). As a consequence of its adoption for many services, a multitude implementations in software and hardware are available. Another result of the widespread adoption is that all associated technologies, such as payload formats, negotiation mechanisms and so on are well defined and standardized. In addition, using H.264 enables interoperability with many other services without video transcoding.

We therefore propose to the WG that H.264 shall be mandatory to implement for all WebRTC endpoints that support video, according to the details described in Section 7 and Section 8.

10. IANA Considerations

This document makes no request of IANA.

Note to RFC Editor: this section may be removed on publication as an RFC.

11. Security Considerations

No specific considerations apply to the information in this document.

12. Acknowledgements

All that provided valuable descriptions, comments and insights about the H.264 codec on the IETF mailing lists.

13. References

13.1. Normative References

[H264] ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services", March 2010.

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

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

[RFC4566] Handley, M., Jacobson, V., and C. Perkins, "SDP: Session Description Protocol", RFC 4566, July 2006.

[RFC6184] Wang, Y., Even, R., Kristensen, T., and R. Jesup, "RTP Payload Format for H.264 Video", RFC 6184, May 2011.

13.2. Informative References

[DoJ] "Web Video Rivalry Sparks U.S. Probe", <http://online.wsj.com/article/SB10001424052748703752404576178833590548792.html>.

[H264perf1] Vatolin, D., "MPEG-4 AVC/H.264 Video Codecs Comparison 2010 - Appendixes", , May 2010, <http://compression.graphicon.ru/video/codec_comparison/h264_2010/appendixes.html#Appendix_8>.

[H264perf2] Shah, K., "Implementation, performance analysis and comparison of VP8 and H.264.", University of Texas at Arlington Department of Electrical Engineering, 2011, <http://www-ee.uta.edu/Dip/Courses/EE5359/2011SpringFinalReportPPT/Shah_EE5359Spring2011FinalPPT.pdf>.

[H264perf3] De Simone, F., Goldmann, L., Lee, J., and T. Ebrahimi, "Performance analysis of VP8 image and video compression based on subjective evaluations", Ecole Polytechnique F'd'rale de Lausanne (EPFL) , Aug 2011, <http://infoscience.epfl.ch/record/168259/files/article.pdf>.

[Implementations] Wikipedia, "H.264/MPEG-4 AVC products and implementations", October 2012, <http://en.wikipedia.org/wiki/H.264/MPEG-4_AVC_products_and_implementations>.

[MPEGLA] "MPEG LAs AVC License Will Not Charge Royalties for Internet Video that is Free to End Users through Life of License", MPEGLA News Release, August 2010, <www.mpegla.com/Lists/MPEG%20LA%20News%20List/Attachments/231/n-10-08-26.pdf>.

[MpegLaVp8] "MPEG LA Announces Call for Patents Essential to VP8 Video Codec", <http://www.mpeg-la.com/main/pid/vp8/default.aspx>.

[PSNRdiff] Bjontegaard, G., "Calculation of Average PSNR Differences

between RD-Curves", ITU-T SG16 Q.6 Document VCEG-M33, April 2001.

[Woon] "Polycom Delivers Open Standards-Based Scalable Video Coding (SVC) Technology, Royalty-Free to Industry", October 2012, <http://www.polycom.com/content/www/en/company/news/press-releases/2012/20121004.html>.

Authors' Addresses

Bo Burman
Ericsson
Farogatan 6
Stockholm 16480
Sweden

Email: bo.burman@ericsson.com

Markus Isomaki
Nokia
Keilalahdentie 2-4
Espoo FI-02150
Finland

Email: markus.isomaki@nokia.com

Bernard Aboba
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
US

Email: bernard_aboba@hotmail.com

Gaelle Martin-Cocher
RIM
1875 Buckhorn Gate
Mississauga, ON L4W 5P1
Canada

Email: gmartincocher@rim.com

Giri Mandyam
Qualcomm Innovation Center

Email: mandyam@quicinc.com

Xavier Marjou
France Telecom Orange
2, avenue Pierre Marzin
Lannion, 22307
France

Email: xavier.marjou@orange.com