Internet Engineering Task Force (IETF)
Request for Comments: 7875
Category: Informational
ISSN: 2070-1721

S. Proust, Ed.
Orange

May 2016

Additional WebRTC Audio Codecs for Interoperability

Abstract

To ensure a baseline of interoperability between WebRTC endpoints, a minimum set of required codecs is specified. However, to maximize the possibility of establishing the session without the need for audio transcoding, it is also recommended to include in the offer other suitable audio codecs that are available to the browser.

This document provides some guidelines on the suitable codecs to be considered for WebRTC endpoints to address the use cases most relevant to interoperability.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Engineering Task Force (IETF). It has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are a candidate for any level of Internet Standard; see Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc7875.

Copyright Notice

Table of Contents

1. Introduction

As indicated in [OVERVIEW], it has been anticipated that WebRTC will not remain an isolated island and that some WebRTC endpoints will need to communicate with devices used in other existing networks with the help of a gateway. Therefore, in order to maximize the possibility of establishing the session without the need for audio transcoding, it is recommended in [RFC7874] to include in the offer other suitable audio codecs beyond those that are mandatory to implement. This document provides some guidelines on the suitable codecs to be considered for WebRTC endpoints to address the use cases most relevant to interoperability.

The codecs considered in this document are recommended to be supported and included in the offer, only for WebRTC endpoints for which interoperability with other non-WebRTC endpoints and non-WebRTC-based services is relevant as described in Sections 4.1.2, 4.2.2, and 4.3.2. Other use cases may justify offering other additional codecs to avoid transcoding.

2. Definitions and Abbreviations

- Legacy networks: In this document, legacy networks encompass the conversational networks that are already deployed like the PSTN, the PLMN, the IP/IMS networks offering VoIP services, including 3GPP "4G" Evolved Packet System [TS23.002] supporting voice over LTE (VoLTE) radio access [IR.92].

- WebRTC endpoint: A WebRTC endpoint can be a WebRTC browser or a WebRTC non-browser (also called "WebRTC device" or "WebRTC native application") as defined in [OVERVIEW].

- AMR: Adaptive Multi-Rate

- AMR-WB: Adaptive Multi-Rate Wideband

- CAT-iq: Cordless Advanced Technology - internet and quality

- DECT: Digital Enhanced Cordless Telecommunications

- IMS: IP Multimedia Subsystems

- LTE: Long Term Evolution (3GPP "4G" wireless data transmission standard)

- MOS: Mean Opinion Score, defined in ITU-T Recommendation P.800 [P.800]

- PSTN: Public Switched Telephone Network
- PLMN: Public Land Mobile Network
- VoLTE: Voice over LTE

3. Rationale for Additional WebRTC Codecs

The mandatory implementation of Opus [RFC6716] in WebRTC endpoints can guarantee codec interoperability (without transcoding) at state-of-the-art voice quality (better than narrow-band "PSTN" quality) between WebRTC endpoints. The WebRTC technology is also expected to be used to communicate with other types of endpoints using other technologies. It can be used for instance as an access technology to VoLTE services (Voice over LTE as specified in [IR.92]) or to interoperate with fixed or mobile Circuit-Switched or VoIP services like mobile Circuit-Switched voice over 3GPP 2G/3G mobile networks [TS23.002] or DECT-based VoIP telephony [EN300175-1]. Consequently, a significant number of calls are likely to occur between terminals supporting WebRTC endpoints and other terminals like mobile handsets, fixed VoIP terminals, and DECT terminals that do not support WebRTC endpoints nor implement Opus. As a consequence, these calls are likely to be either of low narrow-band PSTN quality using G.711 [G.711] at both ends or affected by transcoding operations. The drawback of such transcoding operations are listed below:

- Degraded user experience with respect to voice quality: voice quality is significantly degraded by transcoding. For instance, the degradation is around 0.2 to 0.3 MOS for most of the transcoding use cases with AMR-WB codec (Section 4.1) at 12.65 kbit/s and in the same range for other wideband transcoding cases. It should be stressed that if G.711 is used as a fallback codec for interoperation, wideband voice quality will be lost. Such bandwidth reduction effect down to narrow band clearly degrades the user-perceived quality of service leading to shorter and less frequent calls. Such a switch to G.711 is a choice for customers. If transcoding is performed between Opus and any other wideband codec, wideband communication could be maintained but with degraded quality (MOSs of transcoding between AMR-WB at 12.65 kbit/s and Opus at 16 kbit/s in both directions are significantly lower than those of AMR-WB at 12.65 kbit/s or Opus at 16 kbit/s). Furthermore, in degraded conditions, the addition of defects, like (a) audio artifacts due to packet losses and (b) audio effects due to the cascading of different packet loss recovery algorithms, may result in a quality below the acceptable limit for the customers.

- Degraded user experience with respect to conversational interactivity: the degradation of conversational interactivity is due to the increase of end-to-end latency for both directions that is introduced by the transcoding operations. Transcoding requires full de-packetization for decoding of the media stream (including mechanisms of de-jitter buffering and packet loss recovery) then re-encoding, re-packetization, and resending. The delays produced by all these operations are additive and may increase the end-to-end delay up to 1 second, much beyond the acceptable limit.

- Additional cost in networks: transcoding places important additional cost on network gateways mainly related to codec implementation, codecs licenses, deployment, testing and validation cost. It must be noted that transcoding of wideband to wideband would require more CPU processing and be more costly than transcoding between narrowband codecs.

4. Additional Suitable Codecs for WebRTC

The following are considered relevant codecs with respect to the general purpose described in Section 3. This list reflects the current status of foreseen use cases for WebRTC. It is not limitative; it is open to inclusion of other codecs for which relevant use cases can be identified. It's recommended to include codecs (in addition to Opus and G.711) according to the foreseen interoperability cases to be addressed.

4.1. AMR-WB

4.1.1. AMR-WB General Description

The Adaptive Multi-Rate WideBand (AMR-WB) is a 3GPP-defined speech codec that is mandatory to implement in any 3GPP terminal that supports wideband speech communication. It is being used in circuit-switched mobile telephony services and new multimedia telephony services over IP/IMS. It is specially used for voice over LTE as specified by GSMA in [IR.92]. More detailed information on AMR-WB can be found in [IR.36]. References for AMR-WB-related specifications including the detailed codec description and source code are in [TS26.171], [TS26.173], [TS26.190], and [TS26.204].

4.1.2. WebRTC-Relevant Use Case for AMR-WB

The market of personal voice communication is driven by mobile terminals. AMR-WB is now very widely implemented in devices and networks offering "HD voice", where "HD" stands for "High Definition". Consequently, a high number of calls are likely to occur between WebRTC endpoints and mobile 3GPP terminals offering

AMR-WB. Thus, the use of AMR-WB by WebRTC endpoints would allow transcoding-free interoperation with all mobile 3GPP wideband terminals. Besides, WebRTC endpoints running on mobile terminals (smartphones) may reuse the AMR-WB codec already implemented on those devices.

4.1.3. Guidelines for AMR-WB Usage and Implementation with WebRTC

The payload format to be used for AMR-WB is described in [RFC4867] with a bandwidth-efficient format and one speech frame encapsulated in each RTP packet. Further guidelines for implementing and using AMR-WB and ensuring interoperability with 3GPP mobile services can be found in [TS26.114]. In order to ensure interoperability with 4G/VoLTE as specified by GSMA, the more specific IMS profile for voice derived from [TS26.114] should be considered in [IR.92]. In order to maximize the possibility of successful call establishment for WebRTC endpoints offering AMR-WB, it is important that the WebRTC endpoints:

- Offer AMR in addition to AMR-WB, with AMR-WB listed first (AMR-WB being a wideband codec) as the preferred payload type with respect to other narrow-band codecs (AMR, G.711, etc.) and with a bandwidth-efficient payload format preferred.

- Be capable of operating AMR-WB with any subset of the nine codec modes and source-controlled rate operation.

- Offer at least one AMR-WB configuration with parameter settings as defined in Table 6.1 of [TS26.114]. In order to maximize interoperability and quality, this offer does not restrict the codec modes offered. Restrictions on the use of codec modes may be included in the answer.

4.2. AMR

4.2.1. AMR General Description

Adaptive Multi-Rate (AMR) is a 3GPP-defined speech codec that is mandatory to implement in any 3GPP terminal that supports voice communication. This includes both mobile phone calls using GSM and 3G cellular systems as well as multimedia telephony services over IP/IMS and 4G/VoLTE, such as the GSMA voice IMS profile for VoLTE in [IR.92]. References for AMR-related specifications including detailed codec description and source code are [TS26.071], [TS26.073], [TS26.090], and [TS26.104].

4.2.2. WebRTC-Relevant Use Case for AMR

A user of a WebRTC endpoint on a device integrating an AMR module wants to communicate with another user that can only be reached on a mobile device that only supports AMR. Although more and more terminal devices are now "HD voice" and support AMR-WB; there are still a high number of legacy terminals supporting only AMR (terminals with no wideband / HD voice capabilities) that are still in use. The use of AMR by WebRTC endpoints would consequently allow transcoding free interoperation with all mobile 3GPP terminals. Besides, WebRTC endpoints running on mobile terminals (smartphones) may reuse the AMR codec already implemented on these devices.

4.2.3. Guidelines for AMR Usage and Implementation with WebRTC

The payload format to be used for AMR is described in [RFC4867] with bandwidth efficient format and one speech frame encapsulated in each RTP packet. Further guidelines for implementing and using AMR with purpose to ensure interoperability with 3GPP mobile services can be found in [TS26.114]. In order to ensure interoperability with 4G/VoLTE as specified by GSMA, the more specific IMS profile for voice derived from [TS26.114] should be considered in [IR.92]. In order to maximize the possibility of successful call establishment for WebRTC endpoints offering AMR, it is important that the WebRTC endpoints:

- Be capable of operating AMR with any subset of the eight codec modes and source-controlled rate operation.
- Offer at least one configuration with parameter settings as defined in Tables 6.1 and 6.2 of [TS26.114]. In order to maximize the interoperability and quality, this offer shall not restrict AMR codec modes offered. Restrictions on the use of codec modes may be included in the answer.

4.3. G.722

4.3.1. G.722 General Description

G.722 [G.722] is an ITU-T-defined wideband speech codec. G.722 was approved by the ITU-T in 1988. It is a royalty-free codec that is common in a wide range of terminals and endpoints supporting wideband speech and requiring low complexity. The complexity of G.722 is estimated to 10 MIPS [EN300175-8], which is 2.5 to 3 times lower than AMR-WB. In particular, G.722 has been chosen by ETSI DECT as the mandatory wideband codec for New Generation DECT in order to greatly increase the voice quality by extending the bandwidth from narrow band to wideband. G.722 is the wideband codec required for terminals that are certified as CAT-iq DECT, and version 2.0 of the CAT-iq

specifications have been approved by GSMA as the minimum requirements for the "HD voice" logo usage on "fixed" devices, i.e., broadband connections using the G.722 codec.

4.3.2. WebRTC-Relevant Use Case for G.722

G.722 is the wideband codec required for DECT CAT-iq terminals. DECT cordless phones are still widely used to offer short-range wireless connection to PSTN or VoIP services. G.722 has also been specified by ETSI in [TS181005] as the mandatory wideband codec for IMS multimedia telephony communication service and supplementary services using fixed broadband access. The support of G.722 would consequently allow transcoding-free IP interoperation between WebRTC endpoints and fixed VoIP terminals including DECT CAT-iq terminals supporting G.722. Besides, WebRTC endpoints running on fixed terminals that implement G.722 may reuse the G.722 codec already implemented on these devices.

4.3.3. Guidelines for G.722 Usage and Implementation with WebRTC

The payload format to be used for G.722 is defined in [RFC3551] with each octet of the stream of octets produced by the codec to be octet-aligned in an RTP packet. The sampling frequency for the G.722 codec is 16 kHz, but the RTP clock rate is set to 8000 Hz in SDP to stay backward compatible with an erroneous definition in the original version of the RTP audio/video profile. Further guidelines for implementing and using G.722 to ensure interoperability with multimedia telephony services over IMS can be found in Section 7 of [TS26.114]. Additional information about the G.722 implementation in DECT can be found in [EN300175-8], and the full codec description and C source code are in [G.722].

5. Security Considerations

Relevant security considerations can be found in [RFC7874], "WebRTC Audio Codec and Processing Requirements". Implementers making use of the additional codecs considered in this document are advised to also refer more specifically to the "Security Considerations" sections of [RFC4867] (for AMR and AMR-WB) and [RFC3551] (for the RTP audio/video profile).

6. References

6.1. Normative References

[G.722] ITU-T, "7 kHz audio-coding within 64 kbit/s", ITU-T Recommendation G.722, September 2012, <http://www.itu.int/rec/T-REC-G.722-201209-I/en>.

[IR.92] GSMA, "IMS Profile for Voice and SMS", IR.92, Version 9.0, April 2015, <http://www.gsma.com/newsroom/all-documents/ir-92-ims-profile-for-voice-and-sms/>.

[RFC3551] Schulzrinne, H. and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control", STD 65, RFC 3551, DOI 10.17487/RFC3551, July 2003, <http://www.rfc-editor.org/info/rfc3551>.

[RFC4867] Sjoberg, J., Westerlund, M., Lakaniemi, A., and Q. Xie, "RTP Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs", RFC 4867, DOI 10.17487/RFC4867, April 2007, <http://www.rfc-editor.org/info/rfc4867>.

[RFC7874] Valin, JM. and C. Bran, "WebRTC Audio Codec and Processing Requirements", RFC 7874, DOI 10.17487/RFC7874, May 2016, <http://www.rfc-editor.org/info/rfc7874>.

[TS26.071] 3GPP, "Mandatory Speech Codec speech processing functions; AMR Speech CODEC; General description", 3GPP TS 26.171 v13.0.0, December 2015, <http://www.3gpp.org/DynaReport/26071.htm>.

[TS26.073] 3GPP, "ANSI C code for the Adaptive Multi Rate (AMR) speech codec", 3GPP TS 26.073 v13.0.0, December 2015, <http://www.3gpp.org/DynaReport/26073.htm>.

[TS26.090] 3GPP, "Mandatory Speech Codec speech processing functions; Adaptive Multi-Rate (AMR) speech codec; Transcoding functions.", 3GPP TS 26.090 v13.0.0, December 2015, <http://www.3gpp.org/DynaReport/26090.htm>.

[TS26.104] 3GPP, "ANSI C code for the floating-point Adaptive Multi Rate (AMR) speech codec.", 3GPP TS 26.104 v13.0.0, December 2015, <http://www.3gpp.org/DynaReport/26090.htm>.

[TS26.114] 3GPP, "IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction", 3GPP TS 26.114 v13.3.0, March 2016, <http://www.3gpp.org/DynaReport/26114.htm>.

[TS26.171] 3GPP, "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; General description.", 3GPP TS 26.171 v13.0.0, December 2015, <http://www.3gpp.org/DynaReport/26171.htm>.

[TS26.173] 3GPP, "ANSI-C code for the Adaptive Multi-Rate - Wideband (AMR-WB) speech codec", 3GPP TS 26.173 v13.1.0, March 2016, <http://www.3gpp.org/DynaReport/26173.htm>.

[TS26.190] 3GPP, "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions", 3GPP TS 26.190 v13.0.0, December 2015, <http://www.3gpp.org/DynaReport/26190.htm>.

[TS26.204] 3GPP, "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; ANSI-C code.", 3GPP TS 26.204 v13.1.0, March 2016, <http://www.3gpp.org/DynaReport/26204.htm>.

6.2. Informative References

[EN300175-1] ETSI, "Digital Enhanced Cordless Telecommunications (DECT); Common Interface (CI); Part 1: Overview", ETSI EN 300 175-1, v2.6.1, 2015, <http://www.etsi.org/deliver/etsi_en/300100_300199/30017501/02.06.01_60/en_30017501v020601p.pdf>.

[EN300175-8] ETSI, "Digital Enhanced Cordless Telecommunications (DECT); Common Interface (CI); Part 8: Speech and audio coding and transmission.", ETSI EN 300 175-8, v2.6.1, 2015, <http://www.etsi.org/deliver/etsi_en/300100_300199/30017508/02.06.01_60/en_30017508v020601p.pdf>.

[G.711] ITU-T, "Pulse code modulation (PCM) of voice frequencies", ITU-T Recommendation G.711, November 1988, <http://www.itu.int/rec/T-REC-G.711-198811-I/en>.

[IR.36] GSMA, "Adaptive Multirate Wide Band", IR.36, Version 3.0, September 2014, <http://www.gsma.com/newsroom/all-documents/official-document-ir-36-adaptive-multirate-wide-band>.

[OVERVIEW] Alvestrand, H., "Overview: Real Time Protocols for Browser-based Applications", Work in Progress, draft-ietf-rtcweb-overview-15, January 2016.

[P.800] ITU-T, "Methods for subjective determination of transmission quality", ITU-T Recommendation P.800, August 1996, <https://www.itu.int/rec/T-REC-P.800-199608-I/en>.

[RFC6716] Valin, JM., Vos, K., and T. Terriberry, "Definition of the Opus Audio Codec", RFC 6716, DOI 10.17487/RFC6716, September 2012, <http://www.rfc-editor.org/info/rfc6716>.

[TS181005] ETSI, "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); Service and Capability Requirements V3.3.1 (2009-12)", ETSI TS 181005, 2009, <http://www.etsi.org/deliver/etsi_ts/181000_181099/181005/03.03.01_60/ts_181005v030301p.pdf>.

[TS23.002] 3GPP, "Network architecture", 3GPP TS23.002 v13.5.0, March 2016, <http://www.3gpp.org/dynareport/23002.htm>.

Acknowledgements

We would like to thank Magnus Westerlund, Barry Dingle, and Sanjay Mishra who carefully reviewed the document and helped to improve it.

Contributors

The following individuals contributed significantly to this document:

- Stephane Proust, Orange, stephane.proust@orange.com
- Espen Berger, Cisco, espeberg@cisco.com
- Bernhard Feiten, Deutsche Telekom, Bernhard.Feiten@telekom.de
- Bo Burman, Ericsson, bo.burman@ericsson.com
- Kalyani Bogineni, Verizon Wireless, Kalyani.Bogineni@VerizonWireless.com
- Mia Lei, Huawei, lei.miao@huawei.com
- Enrico Marocco, Telecom Italia, enrico.marocco@telecomitalia.it

though only the editor is listed on the front page.

Author's Address

Stephane Proust (editor)
Orange
2, avenue Pierre Marzin
Lannion 22307
France

Email: stephane.proust@orange.com