Workgroup: Internet Engineering Task Force
Internet-Draft: draft-valin-opus-dred-00
Updates: 6716 (if approved)
Published: 8 March 2023
Intended Status: Standards Track
Expires: 9 September 2023
Authors: JM. Valin   J. Buethe
Amazon   Amazon

Deep Audio Redundancy (DRED) Extension for the Opus Codec

Abstract

This document proposes a mechanism for embedding very low bitrate deep audio redundancy (DRED) within the Opus codec (RFC6716) bitstream.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

This document proposes a mechanism for embedding very low bitrate deep audio redundancy (DRED) within the Opus codec [RFC6716] bitstream.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. DRED Extension Format

We use the Opus extension mechanism [opus-extension] to add deep redundancy within the padding of an Opus packet. We use the extension ID 32, which means that the L flag signals whether a length code is included. In this document, we define only the extension payload. [Note: until adoption by the IETF, experimental implementations of DRED MUST use experiment extension ID 127 to avoid causing interoperability problems]

The principles behind the DRED mechanism defined in this extension are explained in [dred-paper]. All the data in the extension payload is encoded using the Opus entropy coder defined in Section 4.1 of [RFC6716]. Since some of the fields at the beginning of the payload are encoded with flat binary probabilities, they can still be interpreted as bits.

The extension starts with an offset indicator, encoded as a signed 5-bit integer (two's complement) in units of 2.5 ms. The offset indicates the time of the last sample analysed for the transmitted features in the packet, measured from the time of the first sample in the Opus frame that contains the extension data.

The offset is followed by a 4-bit initial quantizer field (Q0) ranging from 0 to 15. That quantizer is used on the most recent frame encoded and is followed by the 3-bit quantizer slope dQ. The 3-bit dQ index selects from the following values: [0, 1/8, 3/16, 1/4, 3/8, 1/2, 3/4, 1] quantizer step per frame. The quantizer for frame k is thus given by: min(15, round(Q0 + dQ_table[dQ] * k)). For example, using Q0=5 and dQ=2 (3/16), frame k=20 would use a quantizer of round(5 + 3/16 * k) = 9.

The compressed redundancy information consists of an initial state coded with a pyramid vector quantizer (PVQ), followed by the entropy-coded latent representation. The number of 40-ms DRED blocks is not coded explicitly. Instead, the decoder MUST NOT decode blocks when fewer than 8 bits remain in the DRED payload.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Offset |  Q0   |  dQ |    PVQ                                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+                                     +
:                                                               :
|            ...                +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                               |  Latent coeffs                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
:                                                               :
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: Extension framing

3. IANA Considerations

This document assigns ID 32 to the "Opus Extension IDs" registry to implement the proposed DRED extension.

4. Security Considerations

As is the case for any media codec, the decoder must be robust against malicious payloads. Similarly, the encoder must also be robust to malicious audio input since the encoder input can often be controlled by an attacker. That can happen through browser JS, echo, or when the encoder is on a gateway.

DRED is designed to have a complexity that is independent of the signal characteristics. However, there exist implementation details that can cause signal-dependent complexity changes. One example is CPU treatement of denormals that can sometimes cause increased CPU load and could be triggered by malicious input. For that reason, it is important to minimize such impact to reduce the impact of DOS attacks. Similarly, since the encoding and decoding process can be cputationally costly, devices must manage the complexity to avoid attacks that could trigger too much DRED encoding or decoding to be performed.

The use of variable-bitrate (VBR) encoding in DRED poses a theoretical information leak threat [RFC6562], but that threat is believed to be significantly lower than that posed by VBR encoding in the main Opus payload. Since this document provides a way to dymanically vary the amount of redundancy transmitted, it is also possible to reduce the overall VBR risk of Opus by using DRED as a way of making the total Opus payload constant (CBR) or nearly constant.

5. References

5.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

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

[opus-extension] Valin, J.-M., "Extension Formatting for the Opus Codec (draft-valin-opus-extension)", March 2023.

5.2. Informative References

[RFC6562] Perkins, C. and JM. Valin, "Guidelines for the Use of Variable Bit Rate Audio with Secure RTP", RFC 6562, DOI 10.17487/RFC6562, March 2012, <https://www.rfc-editor.org/info/rfc6562>.

[dred-paper] Valin, J.-M., Buethe, J., and A. Mustafa, "Low-Bitrate Redundancy Coding of Speech Using a Rate-Distortion-Optimized Variational Autoencoder", 2023, <https://arxiv.org/abs/2212.04453>.

Authors' Addresses

Jean-Marc Valin
Amazon
Canada

Email: jmvalin@amazon.com

Jan Buethe
Amazon
Germany

Email: jbuethe@amazon.com