Network Working Group
Internet-Draft: draft-joetansey-alvc-codec-01
Intended status: Experimental
Expires: February 27, 2026

Joe Tansey
Cisco

August 28, 2025

Adaptive Layered Voice Codec (ALVC) for LPWAN Store-and-Forward

Joe Tansey
Cisco
Email: joetanse@cisco.com

draft: draft-joetansey-alvc-codec-01

Status of This Memo

Copyright Notice

Author: Joe Tansey
Cisco Systems
170 West Tasman Dr
San Jose, CA 95134
United States
Email: joetanse@cisco.com

Abstract

This document specifies the Adaptive Layered Voice Codec (ALVC), a scalable speech codec optimized for extremely constrained low-power

wide-area networks (LPWANs). ALVC enables intelligible base layer playback at sub-kilobit rates and progressive quality improvement via enhancement layers delivered asynchronously. The design supports store-and-forward operation, fragmentation, unequal error protection, and monotonic refinement from partial reception.

Table of Contents
---

1. Introduction

Low-power wide-area networks (LPWANs) such as LoRaWAN, Sigfox, and NB-IoT typically carry short telemetry messages and cannot sustain interactive voice. Nevertheless, many industrial, public safety, and remote operations scenarios benefit from delayed voice delivery, for example voice notes that can be forwarded opportunistically through gateways. This document introduces the Adaptive Layered Voice Codec (ALVC), which explicitly separates intelligibility from fidelity. A low-rate base layer provides immediate comprehension from sparse fragments, and one or more enhancement layers provide incremental quality upgrades when channel capacity permits. ALVC is designed for robustness to out-of-order arrival, long one-way latency, and loss. Receivers can start playback after decoding an initial window of base-layer fragments and then upgrade already buffered audio as enhancements arrive. The codec is independent of any particular transport; a companion document specifies a SCHC based profile suitable for LPWANs.

2. 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 (RFC 2119 and RFC 8174) when, and only when, they appear in all capitals, as shown here.

3. Use Cases and Deployment Scenarios

Industrial and Utilities: maintenance crews leave spoken annotations at remote assets; gateways collect fragments and forward them to control rooms. Public Safety and SAR: short updates where coverage is intermittent; base-layer-only reception provides meaningful information. Telemetry Plus Human Context: numeric telemetry with spoken rationale; base layer coexists with sensor payloads under strict duty-cycle limits.

4. Operational Constraints

Deployments share: 50-200 byte payload budgets; strict duty-cycle limits; minutes-scale latency and out-of-order delivery; MCU-class endpoints. Goals include intelligibility at or below 1.2 kbps,

progressive refinement without resending Layer-0, and graceful degradation under loss.

5. Signal Model and Analysis

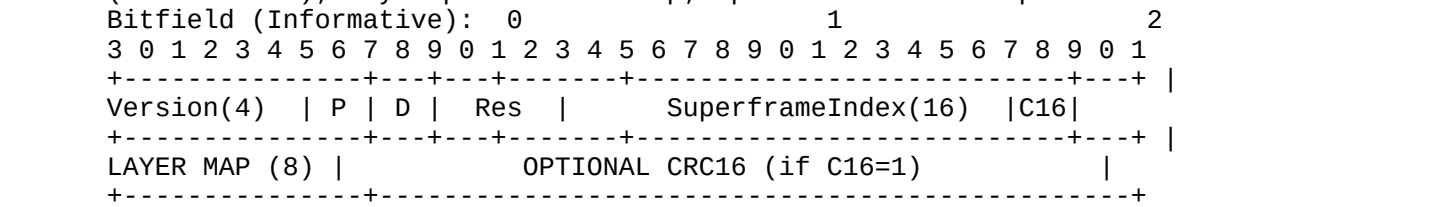
Narrowband speech at 8 kHz. Analysis on 20 ms frames (50% overlap), grouped into 40-80 ms superframes. Fixed-point pitch tracking and LPC analysis. Quantizers are chosen to support most-significant-bit-first emission for graceful truncation.

6. Layered Structure

Layer-0 (Base): independently decodable, ~0.8-1.2 kbps; voicing, pitch, gain, compact spectral envelope (e.g., LSF), sparse excitation indices. Layer-1 (Core): refines envelope, pitch, and gain; MSB-first emission recommended. Layer-2 (High-Band): adds high-band envelopes and sibilance cues; playback remains narrowband when absent. Layer-R (Residual): optional transform-coded residual in bit-planes; encoders may stop opportunistically.

7. Bitstream Syntax

Top-level fields: version and profile ID; superframe index and duration (40 or 80 ms); layer presence bitmap; optional CRC16. Top-Level Header Bitfield (Informative):



8. Framing and Bit Budgets

Default superframe duration: 40 ms (two frames) or 80 ms (four frames). Illustrative budgets: Layer-0 at 800 bps (32 bits per 40 ms); Layer-0 at 1200 bps (48 bits per 40 ms); Layer-1 adds 500-1200 bps; Layer-2 adds 300-800 bps; Layer-R opportunistic. Worked Example (Informative): Layer-0 at 0.8 kbps -> 32 bits per 40 ms; 20 s -> 500 superframes -> 16,000 bits ~2,000 bytes; with 5-byte headers, overhead ~2,500 bytes; total Layer-0 ~4.5 kB including parity; with 80 B MTU, ~57 payloads (+20-30% parity).

9. Packet Loss, FEC, and Concealment

Layer-0 MUST use FEC such as Reed-Solomon or fountain; enhancements SHOULD use lighter FEC or none; encoders MAY embed low-rate forward copies; enhancements MAY be MSB-first; when base fragments are unrecoverable, decoders MUST apply PLC using pitch-synchronous synthesis and noise fill.

10. Decoder Behavior from Partial Layers

Receivers MUST permit immediate playback from Layer-0 alone and MUST apply received enhancements idempotently to improve buffered audio without discontinuities. Implementations SHOULD support background re-synthesis and MAY cache undecoded enhancements.

11. Complexity and Memory

Targets fixed-point MCUs. Reference decoder < 40 MHz (Layer-0), < 80 MHz with one enhancement. Informative: Cortex-M4 @ 80 MHz decodes Layer-0 in real time with <25% CPU and <32 KiB RAM; +one enhancement ~50% CPU and +8 KiB RAM; M7/A-class provide headroom.

12. Interoperability and Profiles

Interoperability via profiles fixing Layer-0 bit allocation and frame durations; a LoRaWAN example is in the companion transport document.

13. Security Considerations

Bitstreams SHOULD be protected via authenticated encryption. With CoAP, use OSCORE. With SCHC, per-fragment AEAD is RECOMMENDED. Implementations MUST avoid variable-time decoding paths that leak content via timing.

14. Privacy Considerations

Voice content is sensitive; support redaction/transcript-only modes and retention limits; metadata timing can leak usage patterns; padding and batching MAY mitigate.

15. IANA Considerations

This document has no IANA actions.

16. Acknowledgments

Thanks to colleagues for feedback on layered speech models and LPWAN operation.

17. References

Normative:
  RFC 2119; RFC 8174; RFC 8613.
Informative:
  RFC 6716; RFC 8724; RFC 9363; RFC 7252.

Authors’ Addresses

Joe Tansey
Cisco Systems
170 West Tasman Dr
San Jose, CA 95134
United States

Email: joetanse@cisco.com