RTP Timestamps for Variable Frequency Codecs

Colin Perkins

UNIVERSITY
of
GLASGOW

Outline

- Definition of the RTP timestamp
- Conventions for RTP timestamp use with audio formats
- Use of the RTP timestamp in proposed new payload formats
  - VMR-WB
  - AMR-WB+
- Issues to consider
- Options going forward

RTP Timestamp Definition

- The timestamp reflects the sampling instant of first octet in the RTP data packet
- The sampling instant MUST be derived from a clock that increments monotonically and linearly in time to allow synchronization and jitter calculations

[RFC 3550 section 5.1]

⇒ *Cannot vary RTP timestamp rate within a session*

Conventions for Audio Timestamps

- The RTP clock rate used for generating the RTP timestamp is independent of the number of channels and the encoding; it usually equals the number of sampling periods per second.

[RFC 3551 section 4.1]

- Two exceptions:
  - G.722 uses an 8kHz RTP timestamp clock for a 16kHz sampling rate codec, for compatibility with a mistake in RFC 1890
  - For compatibility with other MPEG systems, MPEG Audio uses 90kHz RTP timestamp clock separate from the audio sampling clock

[RFC 3551 sections 4.5.2 and 4.5.13, RFC 3119]

⇒ *Using RTP clock rates other than the sampling rate is allowed, but must consider how it will affect the overall system design*

Conventions for Audio Timestamps

- The sampling frequency SHOULD be drawn from the set: 8,000, 11,025, 16,000, 22,050, 24,000, 32,000, 44,100 and 48,000 Hz. (Older Apple Macintosh computers had a native sample rate of 22,254.54 Hz, which can be converted to 22,050 with acceptable quality by dropping 4 samples in a 20 ms frame). However, most audio encodings are defined for a more restricted set of sampling frequencies.

[RFC 3551 section 4.1]

Use of Timestamp in New Audio Formats

- VMR-WB
  - Can accept 8 or 16 kHz of input sampling rates
  - By default produces 16kHz output, irrespective of input
  - RTP payload format desires to use a fixed 16kHz clock
- AMR-WB+
  - Can accepts a range of input sampling rates
  - Re-samples within the codec to one of 12 internal sampling frequencies
  - Can produce output at one of 8, 16, 24, 32 or 48 kHz
  - RTP payload format desires to use a fixed 72kHz clock
- Key points:
  - Codec input and output sampling rates are decoupled
  - May wish to switch input rates within a session; decoder is agnostic of input sampling rate

Issues With the New Formats

- The number of sampling periods per second varies in different parts of the system
  - The usual definition of the RTP clock rate for audio is not sufficient
  - Should the RTP clock rate match the input sampling rate, output sampling rate, or neither? How does this affect synchronisation?
- Is it necessary to signal input and/or output rates separately?
  - If so, how do we do this?
  - The “rate” MIME parameter is specified as “clock rate” with no definition of which clock

⇒ *Consistency across codecs is desirable, to simplify the protocol and implementations that support multiple codecs*

Options Going Forward

1. Only support codecs where input and output sampling rates are identical; use sampling rate as RTP timestamp rate
   - Simple but limits future development
   - Backwards compatible
2. Support codecs where input and output rates differ, but mandate a common RTP timestamp rate for all such codecs
   - See draft-ietf-avt-variable-rate-audio-00.txt and next presentation
3. Support codecs where input and output rates differ, allow each codec to specify its own definition for the RTP timestamp and signalling
   - Makes protocol, signalling and implementation more complex
   - Meaning of RTP timestamp codec dependent; rather than media dependent
4. Something else?

Options Going Forward

- Do we need to develop a set of guidelines for designers of new variable rate codecs?