Unified Communication Specification for H.264/MPEG-4 Part 10 Scalable Video Coding RTP Transport Version 1.0

This document is now managed by IMTC. Please note that any change that affects backwards compatibility requires a vote of approval by not less than two-thirds of its members.

Version 1.0 – June 12, 2014
Development of this specification is performed through the mailing list:
ucif_twg264svcprofile@mail.ucif.org

Table of Contents

1. Introduction

This document contains a specification for the transport of H.264/MPEG-4 Part 10 Advanced Video Coding Annex G (Scalable Video Coding) [1] bitstreams in unified communication applications.

The goal of this specification is to support the use of AVC and SVC video in the entire gamut of unified communications applications that use video. Targeted application scenarios thus include low-end mobile phone video chat, all the way to high-end multi-monitor telepresence systems.

This specification assumes that the reader is familiar with the H.264/MPEG-4 Part 10 AVC standard specification and its Scalable Video Coding (SVC) extension specified in Annex G. In the following, the term 'AVC' refers to the H.264/MPEG-4 Part 10 specification excluding the scalability features of Annex G, whereas 'SVC' refers specifically to systems that use the scalability features.

2. Normative references

[1] ITU-T Rec. H.264 | ISO/IEC 14496-10 Advanced video coding for generic audiovisual services. The standard is available at http://www.itu.int/rec/T-REC-H.264. Unless otherwise specified, this document refers to the edition approved by ITU-T in February 2014 (posted at the ITU-T web site link above). Annex G of this specification contains the SVC extension.
[2] S. Bradner, "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119, March 1997.
[3] S. Wenger, Y.-K. Wang, T. Schierl, and A. Eleftheriadis, "RTP Payload Format for Scalable Video Coding," RFC 6190, May 2011.
[4] Y.-K. Wang, R. Even, T. Kristensen, and R. Jesup, "RTP Payload Format for H.264 Video," RFC 6184, May 2011.
[5] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," RFC 3550, July 2003.
[6] UCI Forum, "Unified Communication Specification for H.264/MPEG-4 Part 10 AVC and SVC Modes," Ver. 1.0, December 2011.

3. Terminology

3.1. Abbreviations

For the purposes of this specification, the following abbreviations apply (those with an asterisk '*' are copied from the H.264/MPEG-4 Part 10 specification):

IDR*	Instantaneous Decoding Refresh
NAL*	Network Abstraction Layer
PPS	Picture Parameter Set
SEI*	Supplemental Enhancement Information
SPS	Sequence Parameter Set
SVC*	Scalable Video Coding

UC	Unified Communications

3.2. Definition

For the purposes of this specification, the following definitions apply (those with an asterisk '*' are copied from the H.264/MPEG-4 Part 10 specification [1]):

Bitstream*	A sequence of bits comprising NAL units that forms the representation of coded pictures and associated data forming one or more coded video sequences.
NAL unit	the basic encapsulation structure in H.264/MPEG-4 Part 10; a syntax structure containing an indication of the type of data to follow, followed by that data in the form of a byte sequence interspersed as necessary with emulation prevention bytes

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 RFC 2119 [2].

4. General Properties

This specification defines the transport of AVC and SVC bitstreams over RTP. The bitstreams must comply to one of the UC Modes defined in [6] and summarized below. The transport mechanism is assumed to be RTP [5] and thus the corresponding RTP payload formats for H.264 AVC (RFC 6184 [4]) and SVC (RFC 6190 [3]) apply. This specification specifies a subset of the possible configurations available in these specifications for the purposes of ensuring and simplifying interoperable implementations.

4.1. Mode Structure

Five UC modes are defined in [6]. They include AVC single layer, SVC with temporal scalability, SVC with temporal and quality scalability[1], SVC with temporal and spatial scalability, and SVC with all of temporal, quality, and spatial scalability. The intention of including modes with incremental scalability capabilities is to allow encoder chip and device manufacturers to gradually incorporate the necessary support into their devices.

The UC Modes are as follows.

- UC Mode 0: Non-scalable single layer AVC bitstream.
- UC Mode 1: SVC with temporal scalability using hierarchical P pictures.
- UC Mode 2q: SVC with temporal scalability using hierarchical P pictures and quality scalability.

[1] Quality scalability is also known as SNR scalability.

- UC Mode 2s: SVC with temporal scalability using hierarchical P pictures and spatial scalability.
- UC Mode 3: SVC scalability with all of temporal scalability (using hierarchical P pictures), quality, and spatial scalability.

Encoders that conform to higher level modes include the capabilities of encoding bitstreams associated with lower level modes. For example, encoders that conform to UC Mode 2q must be able to generate a single layer AVC stream, i.e., UC Mode 0.

NOTE: Change of UC Mode may be requested through signaling means that are outside the scope of this specification.

Figure 1 depicts the hierarchical structure and relationship for all UC Modes. Details of each UC Mode are elaborated in [6].

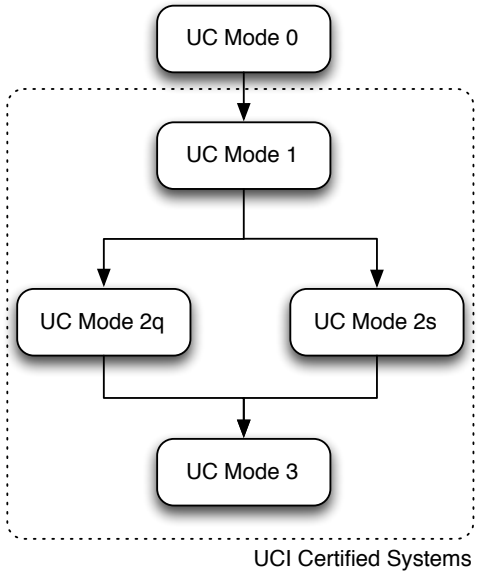

Figure 1. The hierarchical relationship of UC Modes

The UC Modes indicate properties of the bitstream that will be produced by an encoder or transmitted from a server without taking into account any adaptations that may happen dynamically at either the sender or during transport to the receiver. For example, a sender may elect to eliminate a layer to accommodate reduced available bit rate.

4.2. Bitstream Parameters

Bitstreams transported using this specification must conform to [6]. Consequently, for non-scalable bitstreams they must conform to the Constrained Baseline profile or Constrained High profile. Scalable bitstreams must conform to the Scalable Constrained Baseline or the Scalable Constrained High profile.

NOTE: Bitstreams conforming to UC Mode 0 may include NAL units of type 14 (referred to as prefix NAL units).

The UC Modes specification [6] specifies further minimum bitstream parameters, including minimum set of spatial resolutions, temporal resolutions, and levels, that systems conforming to that specification must support.

5. Transport

5.1. UC Mode 0: AVC

Mode 0 transport is outside the scope of this document.

5.2. UC Modes 1, 2Q, 2S, and 3: SVC

5.2.1. Transport Properties

The RTP payload format defined in RFC 6190 [3] must be used for the transport of video data in all the SVC Modes.

Single-Session Transmission (SST) as defined in RFC 6190, Section 1.2.2 must be used. In this transmission mode all SVC data are carried in a single RTP session using a single SSRC.

Non-interleaved packetization as defined in RFC 6190, Section 4.5.1, must be used.

Single-Time Aggregation Packet (STAP-A) NAL units, as defined in RFC 6184, Section 5.7, should be used to carry the SVC data. Single NAL units or Fragmentation Units (FU-A), defined in RFC 6184, Section 5.8, may be used to carry the SVC data as well. PACSI NAL units, as defined in RFC 6190, Section 4.9, must be used to indicate layer information, either in the same packet (in the case of STAP-A) or in a Single NAL unit packet that immediately precedes, in transmission order, the packet containing the SVC data it is associated with. In Single NAL unit mode or when FU-A NAL units are used, PACSI NAL units may be sent only when there is a change in the layer information of the NAL units that follow as compared with the NAL units that precede, in decoding order, the PACSI NAL unit.

NOTE: STAP-A NAL units are preferred over FU-A NAL units because the latter cannot carry PACSI NAL unit information, and thus offer fewer error resilience capabilities.

5.2.2. Bitstream Properties

Sequence Parameter Set (SPS) and Picture Parameter Set (PPS) shall be sent in-band in the access unit of every IDR picture either as an independent packet or aggregated with the first slice of the IDR picture.

Scalability Information SEI (SSEI) messages shall be sent in the access unit of every IDR picture.

The following SSEI parameters shall be included in the SSEI message:

- temporal_id_nesting_flag shall be set to 1.
- For each layer (num_layers_minus1)

- layer_id set to (temporal_id | quality_id <<3 | dependency_id <<7)
- dependency_id
- temporal_id
- quality_id
- constant_frm_rate_idc set to 1
- layer_dependency_info_present_flag set to 1
- frm_rate_info_present_flag set to 1
- avg_frame_rate set
- frm_size_info_present_flag set to 1
- parameter_sets_info_present_flag set to 1

Video usability information (VUI) parameters must be sent in the SPS sent with each access unit of every IDR picture

5.2.3. Dynamic Behavior

When a spatial or quality layer is removed or added back in by a sender a Layers Not Present SEI (LNPSEI) message shall be inserted in the bitstream at the start of the first picture from which the drop starts, and also following every SSEI in the bitstream.

* * *