INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 MPEG2016/N16144
February 2016, San Diego, CA, US

Source **Requirements**
Status **Approved**
Title **Requirements for the Common Media Application Format**

1 Introduction

1.1 Evolution Toward a Common Format

Several MPEG technologies have been adopted for the majority of video delivered over the internet and other IP networks (cellular, cable, broadcast, etc.). Various organizations have taken MPEG's core coding, file format and system standards, and combined them into their own specifications for their specific applications. While these specifications share major common parts, their differences result in both unnecessary duplication of engineering effort, and duplication of identical content in slightly different formats. The industry would benefit if application consortia could reference a single MPEG specification (a "common format") that would allow a single media encoding to work across many applications and devices.

1.2 The Internet Delivery Challenge

By most measures, audio-video content (typically called 'video') constitutes the majority of data delivered over the internet today, and the volume is growing at a rate that exceeds the growth of storage and transmission capacity of the internet. Content delivery networks can distribute media objects to thousands of different users, but their efficiency is impaired when users request the same media in many different formats.

1.3 The Internet Video Device Explosion

Most video, such as movies and TV shows, that has been delivered in the past only to TV sets by broadcast or disc, is now available over the internet. In the last decade, several billion new internet video devices such as phones, tablets, game consoles, and connected TVs have been bought and used by consumers to play video in a wide range of bitrates, resolutions, languages, etc. that reflect the variety of devices, networks, and the global reach of the internet. As a result, the industry has evolved an "adaptive" media format that allows each device to select and combine media objects, such as a particular audio and video tracks, and segments of video at a bitrate and resolution that is optimized for the device and the current network throughput. This object-oriented, late binding media format that has evolved allows one set of media objects to be

encoded that devices can combine into thousands of customized media presentations without having to encode and deliver thousands of different audio/video files or streams.

1.4 Standardizing a "Common Media Application Format"

The segmented media format that has been widely adopted for internet delivery using DASH, Web browsers, commercial services such as Netflix and YouTube, etc. is derived from the ISO Base Media File Format, using MPEG codecs, Common Encryption, etc. The same components have already been widely adopted and specified by many application consortia (ARIB, ATSC, DECE, DASH-IF, DVB, HbbTV, SCTE, 3GPP, DTG, DLNA, etc.), but the absence of a common media format, or minor differences in practice, mean that slightly different media files must often be prepared for the same content. The industry would further benefit from a common format, embodied in an MPEG standard, to improve interoperability and distribution efficiency.

2 Scope

The Common Media Application Format defines the media format only. MPEG technologies such as DASH and MMT may be used for the delivery of the Common Media Application Format Segment. Multiple delivery protocols may specify how the same Common Media Application Format Segments are delivered. The description and delivery of presentations are both considered to be in layers above the layer that defines the media format and the encoding and decoding of Media Segments, and therefore they are out of the proposed scope.

3 Timeline and Urgency of Standardization

With the wide deployment of the internet video in various market, and consortia defining their own media formats, we believe MPEG should start a project to develop and publish its first Common Media Application Format specification as soon as possible. Such a specification addressing the most common use-cases and defining a few Common Media Application Format profiles would help industry and consortia to reference this specification and avoid fragmentation of media formats.

4 MPEG Technologies in the Common Media Application Format

Some of the MPEG technologies that have been widely adopted for internet video and are expected to be referenced by the Common Media Application Format are:

- ISO Base Media File Format (MPEG-4 Part 12)
- MPEG codecs for audio and video
- MPEG-4 file specifications for the delivery of audio, video, and subtitles/captions derived from the Base format (MPEG-4 Part 15, Part 30, etc.)
- Common Encryption (ISO/IEC 23001-7)

5 Use Cases

1. Adaptive bitrate streaming of Media Segments using HTTP(S) and any Presentation Description, such as DASH MPD, Smooth Streaming Manifest, Apple HTTP Live Streaming Manifest (m3u8), etc. Note: Specific CMAF bindings to MPEG DASH and other presentation descriptions are expected to be defined separately.

2. Broadcast/multicast streaming of Media Segments over one-way networks such as terrestrial broadcast, satellite, or cellular network.
3. Hybrid network streaming of live content via broadcast/multicast or cellular, and unicast for time shifted or individualized content, and Segments lost during broadcast delivery.
4. Download of streaming files for local playback, or local playback of downloaded content combined with streamed Segments.
5. Server-side and Client-side ad insertion with media segment and manifest-level signaling of messages for ad insertion and other purposes, such as SCTE-35, VAST, VMAP, etc.

6 Requirements

1. The Common Media Application Format (CMAF) shall be based on existing MPEG standards and industry practices; and shall support the following requirements:
    a. Shall define specific profiles based on the ISO Base Media File Format (ISOBMFF), each defining a conformance point that provides interoperability between CMAF conformant devices (CMAF players) and CMAF presentations that support that profile. The highest priority is to specify the minimum requirements both CMAF presentations and CMAF players must support for interoperability, and to specify an easily adopted baseline profile.
    b. Shall be capable of carrying encoded audio, video, and subtitles, including signaling to support track selection and common accessibility use cases.

2. Shall be delivery mechanism agnostic, but support at least the following delivery mechanisms:
    a. Adaptive streaming with seamless adaptive switching of CMAF tracks encoded with different bitrates, frame rates, and video resolutions.
    b. Late binding of independently created and/or delivered tracks for combined playback.

        Note: In late binding, media components' segments are separately delivered and combined at the playback time, while in early binding, the media component segments are added together as part of delivery and fed combined into the player.

    c. Early binding of independently created and/or delivered tracks for combined delivery, i.e. playable delivery segments that contain media segments of differing media types but the same approximate time-range,
    d. Multicast and broadcast delivery,
    e. Hybrid network delivery (broadcast + unicast)
    f. Physical media delivery (memory sticks, discs, etc.)
    g. Low latency live streaming
    h. Multitrack file progressive download and playback,
    i. Efficient CDN delivery with uniform identification of each media segment by URI (and possible byte range) for efficient caching and reuse.
    j. Shall support random access of the content
    k. Shall support signaling of various random access points and their types
    l. Shall support delivery and playback of selected tracks of the content.

3. Each CMAF profile shall define:
    a. A video codec, its profile and level, associated color space, EOTF and other rendering constraints;

b. An audio codec, its profile, features, and channel configuration;
c. Optional closed caption and subtitle formats;
d. An optional encryption scheme and key management constraints for Common Encryption (CENC) of media samples;
e. Track and segment encryption metadata that enables the use of any DRM systems that conform to MPEG Common Encryption for key management and decryption;
f. A set of constraints for ISOBMFF files, tracks, movie fragments, samples, and elementary streams bindings;
g. CMAF profile identifier(s) for the conformance points supported by a presentation or device.